I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14619

From Referrals to Suspensions: New Evidence on Racial Disparities in Exclusionary Discipline

Jing Liu
Michael S. Hayes
Seth Gershenson

JULY 2021

I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14619

From Referrals to Suspensions: New Evidence on Racial Disparities in Exclusionary Discipline

Jing Liu
University of Maryland

Michael S. Hayes
Rutgers University

Seth Gershenson
American University and IZA

JULY 2021

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

From Referrals to Suspensions: New Evidence on Racial Disparities in Exclusionary Discipline*

We use novel data on disciplinary referrals, including those that do not lead to suspensions, to better understand the origins of racial disparities in exclusionary discipline. We find significant differences between Black and white students in both referral rates and the rate at which referrals convert to suspensions. An infraction fixed-effects research design that compares the disciplinary outcomes of white and non-white students who were involved in the same multi-student incident identifies systematic racial biases in sentencing decisions. On both the intensive and extensive margins, minoritized students receive harsher sentences than their white co-conspirators. This result is driven by high school infractions and applies to all infraction types. Reducing racial disparities in exclusionary discipline will require addressing underlying gaps in disciplinary referrals and the systematic biases that appear in the adjudication process.

JEL Classification: I2, J7

Keywords: exclusionary discipline, intentional discrimination, office referrals

Corresponding author:
Jing Liu
University of Maryland
2203 Benjamin Building, College Park
MD 20742
USA
E-mail: liu28@umd.edu

* The authors thank participants in AEFP, APPAM, and SOLE, and seminars at Indiana University and New York University for helpful discussions. Wenjing Gao provides excellent research assistance. The data were provided via in-kind support from a local educational agency using a data sharing agreement that requires agency review of the findings of the research prior to publication. Opinions reflect those of the authors and not necessarily those of the funding agencies or the data-sharing partner.

1 Introduction

Racial disparities in exclusionary discipline (i.e., suspensions) in U.S. public schools are striking: for example, the 2013-14 Civil Rights Data Collection finds that Black students accounted for 40% of suspensions but only 16% of enrollments. Such disparities are the subject of much debate and concern for two broad reasons. First, suspensions likely affect important socioeconomic outcomes and are thus a precursor to analogous disparities in educational achievement, high school and college completion, employment, and involvement with the criminal justice system (Bacher-Hicks et al., 2019; Davison et al., 2021; Sorensen et al., Forthcoming; Weisburst, 2019). This motivates efforts to reduce the use of exclusionary discipline, which disproportionately harms students of color (Steinberg and Lacoe, 2017; Davison et al., 2021).

Second, racial disparities in exclusionary discipline may be artificial in the sense that they result from systematic biases in schools' handling of student indiscipline and not underlying racial differences in student behavior. A 2014 Dear Colleague Letter from the Obama Administration refers to these biases as "intentional discrimination," which occurs when students of color are penalized more harshly than white students who committed the same infraction; we use these terms interchangeably. Decisions of whether, and for how long, to suspend students are typically made by school principals who, like anyone else, are prone to having implicit and explicit biases that influence their decisions (Jarvis and Okonofua, 2020; Starck et al., 2020; Sorensen et al., Forthcoming). The prevalence of this type of intentional discrimination has implications for how schools and policy makers might go about reducing racial disparities in exclusionary discipline and for reducing its use more broadly.

Causal identification of systematic biases in sentencing decisions – what the Dear Colleague letter refers to as intentional discrimination – is challenging because no two infractions are identical and researchers typically do not observe the student behaviors that lead to student suspensions. Barrett et al. (2019) introduces a novel approach to addressing this empirical challenge by comparing the suspension lengths (in days) received by students of different races who were involved in the *same* incident.[1] Specifically, using administrative data on suspensions in Louisiana, the authors use an incident fixed effects (FE) strategy to look at the student-specific disciplinary outcomes following fights that involved both Black and white students. They find that Black students receive longer suspensions, on average, than their white counterparts. The difference is modest in size at about 0.05 days though it is statistically significant. This finding suggests that intentional discrimination in the adjudication of these fights contributes to the Black-white suspension gap.

Intuitively, since these fights involve students who attend the same school and are literally involved in the same incident, any systematic racial differences in sentencing suggest the presence of biased adjudication. An implicit identifying assumption is that there are no systematic racial differences in the intensity of participation (e.g., using a weapon or provoking or escalating the fight). Another is that the students involved in the fight had similar prior disciplinary records.

The current study extends this approach to testing for intentional discrimination and probes these identifying assumptions in a few important ways. We do so using rich administrative data from a large and diverse school district in California. Thus, our first contribution is to rigorously test for racial bias in exclusionary discipline in a new context, outside the American South, with sizable enrollments of white, Black, Hispanic, and Asian students. Understanding the pervasiveness of intentional discrimination against other demographic groups and in other parts of the country is important given the legacy of segregation and anti-Black racism in states such as Louisiana and North Carolina.

Our second contribution is constructing a sample of infractions using data on disciplinary referrals and not realized suspensions, as not all referrals lead to a suspension. This is an important point for a few reasons. First, if suspensions are the sole measure of misbehavior, prior referrals that did not lead to a suspension are an omitted variable that could influence

[1] Shi and Zhu (2021) adopt a similar strategy and replicate these findings in North Carolina.

the suspension assigned to subsequent incidents.[2] Second, by relying solely on suspensions, Barrett et al. (2019) omit students who were involved in the same fight but were not suspended. This form of sampling on the dependent variable is potentially problematic because there are consequences on the extensive margin (being suspended) over and above those of being suspended for an additional day, which would under-estimate the magnitude of intentional discrimination. Observing all referrals as well as the associated suspension outcomes allows us to avoid both problems.

Third, these data facilitate one of the first systematic, quantitative descriptions of the referral process, as nearly all existing research on racial disparities in exclusionary discipline focuses on suspensions (e.g., Anderson et al. 2017; Bacher-Hicks et al. 2019; Barrett et al. 2019; Holt and Gershenson 2019; Lindsay and Hart 2017; Kinsler 2011) and not the referral and reporting process that necessarily precedes the decision of whether, and for how long, to suspend a student.[3] Referrals merit the attention of researchers and policymakers independent of their connection to suspensions, because even when referrals do not result in exclusionary discipline, they are intermediate educational outcomes that can erode students' trust in teachers, the quality of student-teacher relationships, and students' engagement in school. In turn, strained student-teacher relationships and student disengagement can harm achievement and lead to future disciplinary infractions.

Finally, the universe of referrals allows us to test for intentional discrimination in all types of disciplinary infractions and not just fights. This is useful because fights are potentially

[2] A simple example illustrates the problem: 1) suppose two students, one Black and one white, are otherwise identical in terms of socioeconomic and academic background, and are in the same classes; 2) they get in a fight, and receive suspensions of 5 and 2 days, respectively 3) this was the first suspension of the school year for each student. This is the data available in previous research (e.g., Barret et al. 2019), and from this information it looks like a clear case of intentional discrimination, as the Black student received a harsher punishment than the white student, even though they had "identical" backgrounds and were involved in the exact same disciplinary incident. However, now consider some additional information: 1) the principal's leniency decreases with each incident (referral) 2) This was the Black student's third disciplinary referral; the white student's first. With this new information, the difference in suspension length seems less arbitrary, less biased, and more the result of underlying referral histories. Solely relying on suspension data without knowing each student's referral history can lead to misdiagnoses of intentional discrimination.

[3] An exception is Girvan et al. (2017), who conduct descriptive analyses of referral (but not suspension) data and conclude that implicit bias among teachers contribute to racial gaps in office referrals.

unique in terms of having an instigator or a "more violent" participant, which might lead to an omitted variables bias, and because principals' biases might vary by infraction type. Moreover, knowing whether intentional discrimination is more pronounced for certain types of infractions provides critical information for the design of interventions and policies that aim to reduce racial disparities in exclusionary discipline.

We begin our analyses by describing the distribution of disciplinary referrals and the rate at which referrals result in suspensions. Decompositions of the large, unconditional Black-white gaps in both suspensions and referrals show that these gaps are primarily driven by within-school variation. For example, Black students are about 4 percentage points more likely to have been suspended in a given year than their white peers in the same school. However, we go beyond past research on suspensions by conducting similar analyses of disciplinary referrals and find that Black students are 12 percentage points more likely to have received at least one disciplinary referral than their white peers in the same school. This suggests that part of the racial gap in suspensions is due to underlying differences in the frequency of office referrals. However, the racial gap in referral propensities is not the sole reason for the racial gap in suspensions, as we also find that the conversion rate of referrals into suspensions is significantly higher for Black than for white students.

Following Barrett et al. (2019), we then test for intentional discrimination by using an infraction-FE approach. These estimates show a clear and consistent pattern in which minoritized students, particularly Black students, are punished more severely than white students who were involved in the same incident and had the same prior disciplinary histories. Specifically, Black students were about 2 percentage points (67%) more likely to be suspended than white students involved in the exact same incident. This finding is robust to controlling for past achievement, referrals, and suspensions, suggesting that the reason is intentional discrimination. Interestingly, this type of intentional discrimination seems confined to high schools, though it appears in all types of incidents and not just arguably subjective "defiance" referrals. And while this result applies to all under-represented minority students, the effect is three times as large for males as for females.

2 Data

Administrative data come from a large and demographically diverse urban school district in California for the 2016-17 and 2017-18 school years. Panel A of Table 1 summarizes the student-by-year level analytic sample. The district served 68,686 unique students in grades K-12 (120,951 student-year observations) during this time, of which 13% are white, 8% are Black, 29% are Hispanic, and 34% are Asian. We use students' home addresses to identify their residential census tract, which serves as a proxy for socioeconomic status.

The distinguishing feature of the data is detailed information on disciplinary referrals, regardless of whether they lead to a suspension. Specifically, referral records include the individual who made the referral, the reason for the referral (i.e., type of incident), and the exact time, date, and location of the incident (e.g., 3pm, in the hallway, on Monday April 2nd). This precise information allows us to identify the multi-student incidents that are central to our main identification strategy. There were 40,431 unique incidents, of which 13.2% (4,821) involved multiple students. 40% of those involved students of different races, which provide identifying variation for the incident-FE identification strategy. The data also uniquely link referrals to suspension outcomes (measured in days).

Panel B of Table 1 summarizes disciplinary outcomes also at the student-year level. Column 1 shows that each year about 9% of students received at least one office referral. Among those who had at least one referral, the average student was referred about 4.6 times. These frequencies are higher than for suspensions, indicating that many referrals do not lead to a suspension: only 1% of students were suspended per year and among those suspended, the average student was suspended about 1.7 times for about 3.5 days. We measure the "conversion rate" as the ratio of suspensions to referrals, which is about 5% on average.

Columns 2-6 report these figures separately by the mutually exclusive race/ethnicity categories contained in the administrative data. The “Other” category contains multi-racial, American Indian, Arabic, Samoan, and other non-white students. Comparing across columns we see stark and statistically significant disparities on both the intensive and extensive margins in both referrals and suspensions. These gaps are largest when comparing Black students to white and Asian students: Black students are more than 5 times as likely to be referred and 7 times more likely to be suspended in a given year than white students, for example. There is a smaller but still sizable white-Hispanic gap as well. Appendix Table A1 reports referral rates by student race and school type. Referrals are most common in middle schools, in both absolute and relative terms, though they occur in all grade levels.

The data also provide the reason(s) for each referral. Many referrals are the result of multiple infractions, so for the purpose of heterogeneity analyses we follow Lindsay and Hart (2017) in making mutually exclusive, one-off categories based on the “most severe” reason listed for the referral: (a) violence; (b) drugs; (c) interpersonal offenses; (d) disruption or noncompliance; (e) class skipping or walkout; (f) other. For example, a referral where the student was charged with both class skipping and disruption would be coded as disruption. Appendix Table A2 summarizes the types of referrals by school type. Different types of referrals occur at different rates across school types, as might be expected. For example, drug-related offenses are rare overall, but predominantly occur in high school. Interpersonal offenses and offenses due to disruption, noncompliance, class skipping, or walkout are more prevalent in middle schools. Violence incidents are most common in elementary school.

3 Methods

We begin the descriptive analysis by decomposing Black-white and Hispanic-white referral and suspension gaps into within- and between-school components.[4] We then further drill down into within-school gaps by estimating linear regressions at the student-year level that

[4]This exercise follows Barrett et al. (2019); details are provided in Appendix B.

condition on school-by-year fixed effects (FE), a vector of race indicators, and a vector of other observed student characteristics including student's neighborhood poverty rate, prior achievement, and disciplinary outcomes. The main outcomes for these regressions are indicators for ever referred and ever suspended. To examine the intensive margin, we consider outcomes such as total referrals, total suspensions, and the likelihood that a referral results in a suspension. Regressions for these outcomes are estimated on the restricted sample of students who had at least one referral in a year.

Specifically, we estimate models of the form

$$Y_{ist} = \beta Race_i + \gamma X_{ist} + \theta_{st} + \epsilon_{ist}, \tag{1}$$

where Y_{ist} is a disciplinary outcome for student i in school s in year t. We estimate Equation (1) with and without covariates (X_{ist}), where X includes lagged academic achievement and discipline outcomes, gender, neighborhood FE, and special education status. This descriptive exercise provides novel, suggestive evidence that racial gaps in referrals *and* in the processing of referrals contribute to racial gaps in suspensions.

However, statistically significant estimates of β do not necessarily indicate the presence of racial bias, as these models have no way of controlling for the severity or frequency of the infractions that led to the referral. Following Barrett et al. (2019), we address this omitted variables concern using student-by-incident level data and controlling for infraction FE. Identifying variation in these regressions comes from incidents that involve students of different races. Importantly, this includes students who were not suspended at all, as incidents are defined by referrals and not suspensions.

Specifically, we estimate models of the form

$$S_{ijt} = \alpha Discipline_{i,j-1,t} + \beta Race_i + \gamma X_{it} + \theta_j + \epsilon_{ijt}, \tag{2}$$

where S is the suspension outcome (in days or an indicator for suspension) awarded to student i stemming from referral j. Building on Equation (1), in addition to prior year's test scores and disciplinary incidents, Equation (2) also controls for student i's disciplinary incidents in the same year that occurred prior to incident j to account for the possibility that principals consider the student's entire history of referrals before making a decision. Most importantly, we control for infraction FE (θ_j) to exploit within-infraction variation. These FE control for unobserved aspects of the severity and nature of the incident and make school and year FE redundant, as incidents can only involve students in the same school.

There are two threats to the validity of OLS estimates of β in Equation (2). The first is that, on average, students of different races did not participate "equally" in the incident in terms of instigation, showing remorse, or degree of misbehavior. For instance, if white students were more likely to be the instigator of fights, then comparing the disciplinary outcomes of white and non-white students who fought will conflate intentional discrimination with a harsher penalty for instigation. We cannot directly rule out this possibility, though it is unlikely to fully explain our results because we find similar point estimates across incident type. Intuitively, some incidents, like cutting class, are less likely to have an instigator or "heavier" participant, and thus the FE estimates provide an apples to apples comparison.

The second threat is to external validity, as identification comes from a selected sample of multi-student, multi-race incidents (Miller et al., 2019). Appendix Table A4 shows that the identifying sample differs systematically from the overall sample in terms of incident size, racial composition, and prior achievement. This means incident-FE estimates may not generalize to the full population. Moreover, if treatment effects are heterogeneous, the estimates can be biased. For example, if drug incidents are more likely to involve multiple students from different racial/ethnic backgrounds, and Black students are punished more harshly than white students in these incidents, we would have upward bias in our estimates. Following Miller et al. (2019), we conduct a weighting exercise based on the predicted likelihood of being in the identifying sample to verify that our findings are robust to this threat.

4 Main Results

4.1 Decomposing Racial Gaps in Referrals and Suspensions

Figure 1 decomposes Black-white, Hispanic-white, and Asian-white gaps in referral and suspension rates into between- and within-school components separately by grade. Panel A shows a sizable Black-white referral rate gap of 10 to 30 percentage points in each grade, even in kindergarten, which peaks in middle school. Two-thirds of the gap is due to within-school differences, suggesting that the gap is not due to racial sorting into schools. Similar patterns in Black-white suspension gaps, shown in panel B of Figure 1, are consistent with those for referrals and with what Barrett et al. (2019) find in Louisiana.

Panels C and D of Figure 1 report the same figures for Hispanic-white gaps in referral and suspension rates. Though smaller in size than the analogous Black-white gaps, the main features are the same. A key difference from the Black-white analog, however, is that the Hispanic-white suspension rate gap is much smaller and more closely resembles the Asian-white gap shown in panel F. The other difference is that overall, between-school differences tend to play a larger role in the Hispanic-white gap.

The Asian-white gaps summarized in panels E and F are close to zero in the early grades and slightly favor Asian students in high school. Interestingly, the only case of the between- and within-school gaps diverging is for the Asian-white referral gap shown in panel E, where the between-school gap favors white students and the within-school gap favors Asian students. This suggests that school sorting patterns for Asian students are different than for the other groups, a point to which we return when discussing the regression results.

4.2 Racial Gaps in Referrals, Suspensions, and Conversions

The decomposition exercises reported in Figure 1 make clear that racial differences in referral and suspension rates are not merely a product of sorting into schools, as nontrivial shares of the Black-white and Hispanic-white gaps are driven by within-school differences. However, it could be that even within schools there are racial differences in students' socioeconomic, academic, and disciplinary backgrounds that explain the differences. We investigate this question in Table 2, which reports estimates of Equation (1) that control for school-by-year fixed effects (FE) and a vector of time-varying student covariates that includes neighborhood characteristics and lagged achievement, referrals, and suspensions.[5]

Each column of Table 2 reports regression-adjusted racial gaps in a specific disciplinary outcome (relative to white students). Columns 1 and 2 report estimates for the extensive margin of receiving at least one suspension and at least one referral, respectively. Consistent with prior research, Black students are more likely to be suspended than white students. Specifically, the likelihood of receiving at least one suspension for a typical Black student is 3.8 percentage points higher than for a white student in the same school with the same observed academic and disciplinary history. Column 2 shows an even larger Black-white gap in the chances of receiving a referral of about 11.5 percentage points. This suggests that the disparity in referrals contributes to the gap in suspensions. These point estimates are six and two times larger than baseline (white student) suspension and referral rates of 0.6% and 4.9%, respectively. Hispanic-white gaps are smaller in both absolute and relative terms than the Black-white gap, although both are at least marginally statistically significant. Sizable and statistically significant Asian-white gaps favor Asian students.

Columns 3 and 4 of Table 2 report estimates of racial gaps on the intensive margin of total annual referrals and suspensions. These regressions are restricted to students who had at least one referral in year t. There is a large Black-white gap in annual suspensions. On average, a Black student received 0.17 more suspensions than a white student. In contrast, as shown in Column 4, the typical Black student received 2.44 more referrals to the principal's office each year than the typical white student. The Hispanic-white gaps for these

[5]Estimates of a parsimonious specification with FE but no student-level controls are reported in Appendix Table A3, which provides qualitatively similar results, and suggests that these racial disparities are not due to observable differences in students' academic or socioeconomic backgrounds.

two outcomes are not statistically significant after adjusting for student covariates, while a modest Asian-white gap in favor of Asian students remains.

The Black-white gap in referrals is an order of magnitude larger than the analogous gap in suspensions, which suggests that there are racial differences in the rate at which referrals convert to suspensions. Column 5 confirms this by estimating models in which the outcome is the ratio of each student's suspensions to referrals. Conditional on student demographics and prior discipline history, conversion rates are similar for white, Asian, and Hispanic students. However, the conversion rate for Black students is significantly higher (1.6 percentage points, or 25%) than for any other group. Together, the results in Table 2 suggest that Black-white gaps in suspensions are due to disparities in the frequency of disciplinary referrals *and* in the rate at which referrals convert to suspensions.

4.3 Intentional Discrimination

Table 3 reports baseline estimates of Equation (2), which compare the disciplinary outcomes of students involved in the same multi-student infraction. The unit of analysis is the student-incident, so students who were involved in multiple multi-student events appear in the data multiple times. To preserve power, in panel A of Table 3, we group Black, Hispanic, and "other" students into one "minoritized" category. The other category includes many mixed-race students and generally students in this category resemble Black and Hispanic students in terms of other observable characteristics. The outcome variable in columns 1 through 3 of Table 3 is an indicator for whether the student was suspended. The outcome in columns 4 through 6 of Table 3 is the length of the suspension (in days and including zero). Panel B of Table 3 re-estimates the model with a full set of race/ethnicity indicators.

Column 1 of Table 3 reports estimates of a simple model that only controls for infraction fixed effects. Subsequent columns add additional controls, including student characteristics, lagged test scores, lagged disciplinary incidents, and finally current-year incidents that occurred prior to the current incident. The estimated coefficients on the race/ethnicity indicators are qualitatively similar across model specifications, suggesting that infraction fixed effects do a decent job of controlling for possibly confounding factors. However, the estimates in columns 3 and 6 are slightly smaller and also more precisely estimated, which suggests that it is important to control for previous referrals and suspensions in the current year. This is intuitive, as principals likely factor prior behavior into their disciplinary decisions.[6] Accordingly, the fully specified regressions in columns 3 and 6 are our preferred estimates.

Column 3 of Panel A shows that on average, minoritized students were 2 percentage points more likely to be suspended than their white same-incident peers. Relative to the baseline white suspension rate in the analytic sample, this indicates a large (67%) increase in the likelihood of suspension and provides strong evidence of systematic bias in adjudications. Asian students were slightly more likely to be suspended than white same-incident peers, but this difference is not statistically significant at traditional confidence levels. The analogous estimates in column 3 of Panel B show that systematic bias against minoritized students is not unique to one group, but roughly similar across Black, Hispanic, and "other-race" students. The point estimate on the Hispanic coefficient is smaller and imprecisely estimated, though it is not significantly different from the Black coefficient.

The results for suspension length in columns 4-6 are qualitatively similar to those found in columns 1 through 3. Once again, the point estimates are roughly similar across model specifications, but slightly smaller in the fully specified model (column 6). Panel B shows that gaps are largest for Black students, but qualitatively similar to those of other minoritized groups. These basic results are robust to implementing the weighting procedure suggested by Miller et al. (2019), as shown in Appendix Table A6.[7]

Overall, the estimates in columns 4-6 of Table 3 are remarkably similar to the analogous

[6] Following Barrett et al. (2019), in Appendix Table A5 we restrict the analytic sample to incidents that were each students' first of the year and find qualitatively similar, yet less precise, estimates.

[7] For simplicity, these weighted regressions adopt a binary "treatment" where white and Asian students are compared to minoritized Black, Hispanic, and "other" students.

estimates in Barrett et al. (2019) and Shi and Zhu (2021) for intentional discrimination against Black students and suggest that systematic racial biases in disciplinary adjudications are not unique to Louisiana and North Carolina.[8] However, our results suggest that there is discrimination against Hispanic and other minoritized groups as well.

Having documented systematic racial bias in the district's disciplinary adjudications, we now test for heterogeneity along several dimensions to understand where these biases are most pronounced. The descriptive analysis in section 4.1 shows that raw racial gaps in referrals and suspensions peak in middle school and that the most common reasons for referrals vary by grade level. This suggests that racial biases might also vary by grade level. Accordingly, we re-estimate the preferred full-specification of Equation (2) separately by school type (i.e., elementary, middle, and high) in Appendix Table A7. Here, we see that racial biases in adjudications are almost entirely driven by decisions made for high school students. Interestingly, we also find similar levels of intentional discrimination against Black and Hispanic high school students. Finally, and somewhat surprisingly, we find substantial discrimination against Asian students in elementary school. This finding merits further consideration, though could be driven by outliers in the relatively small number of multi-student, multi-race incidents in elementary school that involve an Asian student.

One possible interpretation of the finding that intentional discrimination is most prevalent in high schools is that certain types of offenses, which predominantly occur in high schools, are more susceptible to subjective interpretations that lead to biased punishments (e.g., defiance). To investigate, in Table 4 we re-estimate the preferred full-specification (Equation 2) separately by referral reason. While the point estimates move around a bit and are imprecisely estimated for some of the rarer reasons (e.g., drugs), Table 4 generally shows that intentional discrimination is present in the adjudication of all referral types. This suggests that it is not the incidents themselves, but something about the disciplinary process in high school, and how minoritized high school students are viewed by administrators, that

[8] Barrett et al. (2019) only examines Black-white gaps, and Shi and Zhu (2021) does not find intentional discrimination against Hispanic students compared with white students.

underlies the intentional discrimination uncovered in this and related analyses.

Lastly, in Appendix Table A8 we test for heterogeneity in intentional discrimination by gender, special education status, and neighborhood poverty level. Previous theoretical and empirical evidence suggests that males, special education students, and students from low-income backgrounds experience exclusionary discipline at higher rates than other students (Mendez and Knoff, 2003; Steinberg and Lacoe, 2017). Here, we test whether systematic racial biases in adjudications might contribute to or exacerbate those gaps. We estimate each model for the full sample and for high school students, as this is where intentional discrimination is most pronounced. The only statistically significant interaction term is for high-school male students, which suggests that that minoritized male students are 6 percentage points more likely to be suspended than minoritized female students; however, minoritized female students are still 2 percentage points more likely to be suspended than their white peers.

5 Conclusion

This study investigates two potential reasons that racial disparities in exclusionary discipline (suspensions) might arise. First, the gap could be the natural result of analogous disparities in disciplinary referrals. Second, there could be systematic biases in the adjudication of such referrals. We find evidence that both explanations contribute to large and troubling racial disparities in exclusionary discipline. Specifically, using unusually detailed administrative data from a large and diverse urban school district in California, we show that Black-white disparities in exclusionary discipline are large, present in all grade levels, largest in middle school, and primarily due to within- rather than between-school differences. We expand on this descriptive result, which has been documented elsewhere, by showing that similar patterns exist in disciplinary referrals that do not always result in exclusionary discipline. Finally, we expand the incident FE strategy introduced by Barrett et al. (2019) to test for systematic racial biases in how referrals are adjudicated. Importantly, the referral data allow us to include students who do not get suspended, to fully control for student prior discipline history, and to test for the full range of incident types beyond just fights.

We find evidence of systematic racial bias in the district's disciplinary adjudications. Specifically, compared with white students who were involved in the same incident and had similar prior disciplinary histories, on average, minoritized students were 67% (2 percentage points) more likely to be suspended and were suspended for 0.045 days longer. We also find that these results do not just apply to Black students, but also Hispanic and other minoritized groups. These estimates are in line with results in Barrett et al. (2019) and Shi and Zhu (2021), but in a way particularly striking given the studied context, California, is at the forefront of efforts to address racial inequities in exclusionary discipline. Heterogeneity analyses shows that racial biases in adjudications mainly occur in high school and for all infraction types.

These results make clear that closing racial gaps in exclusionary discipline requires addressing both gaps in referrals and biases in the adjudication process. However, there are several issues the current study does not speak to. For example, are biases in disciplinary adjudications due to the adjudicator, the way in which the referral was made, or both? An implication for schools may be to leverage insights from social psychology regarding empathy interventions, which have been shown to change teachers' perceptions, reduce suspensions, and improve students' achievement (Okonofua et al., 2016). Future research should work to understand the types of teachers, school personnel, and schools that generate these disparities, and the conditions in which they do so.

References

Anderson, Kaitlin, Gary Ritter, and Gema Zamarro, "Understanding a vicious cycle: Do out-of-school suspensions impact student test scores?," 2017.

Bacher-Hicks, Andrew, Stephen B Billings, and David J Deming, "The school to prison pipeline: Long-run impacts of school suspensions on adult crime," 2019.

Barrett, Nathan, Andrew McEachin, Jonathan N Mills, and Jon Valant, "Disparities and discrimination in student discipline by race and family income," *Journal of Human Resources*, 2019, pp. 0118–9267R2.

Davison, Miles, Andrew Penner, Emily Penner, Nikolas Pharris-Ciurej, Sonya R. Porter, Evan Rose, Yotam Shem-Tov, and Paul Yoo, "School Discipline and Racial Disparities in Early Adulthood," 6 2021. CES Working Paper Series.

Girvan, Erik J, Cody Gion, Kent McIntosh, and Keith Smolkowski, "The relative contribution of subjective office referrals to racial disproportionality in school discipline.," *School Psychology Quarterly*, 2017, *32* (3), 392.

Holt, Stephen B and Seth Gershenson, "The impact of demographic representation on absences and suspensions," *Policy Studies Journal*, 2019, *47* (4), 1069–1099.

Jarvis, Shoshana N and Jason A Okonofua, "School deferred: When bias affects school leaders," *Social Psychological and Personality Science*, 2020, *11* (4), 492–498.

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

Lindsay, Constance A and Cassandra MD Hart, "Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina," *Educational Evaluation and Policy Analysis*, 2017, *39* (3), 485–510.

Mendez, Linda M Raffaele and Howard M Knoff, "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*, 2003, pp. 30–51.

Miller, Douglas L, Na'ama Shenhav, and Michel Z Grosz, "Selection into identification in fixed effects models, with application to Head Start," Technical Report, National Bureau of Economic Research 2019.

Okonofua, Jason A, David Paunesku, and Gregory M Walton, "Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents," *Proceedings of the National Academy of Sciences*, 2016, *113* (19), 5221–5226.

Shi, Ying and Maria Zhu, "Equal Time for Equal Crime? Racial Bias in School Discipline," 2021.

Sorensen, Lucy C, Shawn D Bushway, and Elizabeth J Gifford, "Getting tough? the effects of discretionary principal discipline on student outcomes," *Education Finance and Policy*, Forthcoming.

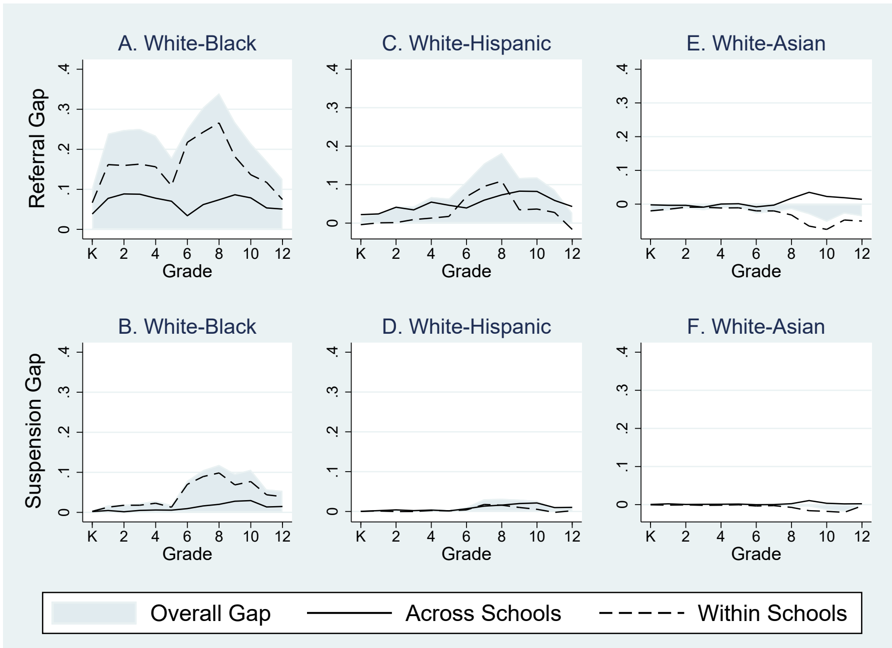
Starck, Jordan G, Travis Riddle, Stacey Sinclair, and Natasha Warikoo, "Teachers are people too: Examining the racial bias of teachers compared to other American adults," *Educational Researcher*, 2020, *49* (4), 273–284.

Steinberg, Matthew P and Johanna Lacoe, "What do we know about school discipline reform? Assessing the alternatives to suspensions and expulsions," *Education Next*, 2017, *17* (1), 44–53.

Weisburst, Emily K, "Patrolling public schools: The impact of funding for school police on student discipline and long-term education outcomes," *Journal of Policy Analysis and Management*, 2019, *38* (2), 338–365.

Figures and Tables

Figure 1: Racial Gaps in the Likelihood of Receiving a Referral and Suspension this Year



Notes: This figure shows the decomposition of the raw racial gaps in referrals and suspensions by grade. Data come from a large urban school district in California from school years 2016-17 to 2017-18. Technical details of the decomposition are documented in Appendix B.

Table 1: Student-by-Year Descriptive Statistics

	All Students	Race Comparison				
		White	Black	Hispanic	Asian	Other
Panel A: Student characteristics						
White	0.13	1.00				
Black	0.08		1.00			
Hispanic	0.29			1.00		
Asian	0.34				1.00	
Other Race	0.16					1.00
Female	0.48	0.49	0.49	0.46	0.48	0.47
Special Education	0.15	0.13	0.28	0.20	0.09	0.15
Elementary School	0.42	0.48	0.38	0.43	0.36	0.49
Middle School	0.21	0.13	0.24	0.20	0.24	0.12
High School	0.29	0.22	0.33	0.29	0.36	0.20
Missing Grade-Level	0.08	0.09	0.05	0.08	0.05	0.19
Resides in Poorest Neighborhood	0.12	0.07	0.23	0.14	0.10	0.10
Resides in Poor Neighborhood	0.09	0.08	0.06	0.12	0.09	0.08
Resides in Less Poor Neighborhood	0.11	0.14	0.07	0.07	0.13	0.11
Resides in Least Poor Neighborhood	0.08	0.12	0.05	0.07	0.09	0.08
Missing Poverty Data	0.60	0.60	0.59	0.60	0.59	0.64
Lagged Standardized Math Score	0.01	0.21	-0.35	-0.24	0.22	0.01
	[0.64]	[0.58]	[0.69]	[0.62]	[0.61]	[0.49]
Lagged Standardized Read Score	0.00	0.27	-0.32	-0.21	0.15	0.02
	[0.64]	[0.63]	[0.69]	[0.61]	[0.63]	[0.49]
Panel B: Disciplinary outcomes						
At least one referral	0.09	0.05	0.28	0.13	0.03	0.07
Total referrals	0.40	0.15	1.96	0.50	0.07	0.32
	[2.71]	[1.33]	[6.54]	[2.83]	[0.78]	[2.48]
Total referrals conditional on at least one referral	4.59	3.01	7.09	3.99	2.34	4.75
	[8.12]	[5.18]	[10.87]	[7.03]	[3.82]	[8.40]
At least one suspension	0.01	0.01	0.07	0.02	0.00	0.01
Total suspensions	0.02	0.01	0.11	0.03	0.01	0.02
	[0.26]	[0.12]	[0.61]	[0.27]	[0.14]	[0.10]
Total suspensions conditional on at least one suspension	1.65	1.35	1.83	1.56	1.52	1.59
	[1.49]	[0.90]	[1.69]	[1.33]	[1.64]	[1.33]
Total suspended days	3.51	2.93	3.92	3.37	2.81	3.56
	[3.51]	[2.52]	[3.84]	[3.24]	[3.98]	[3.15]
Ratio of Suspensions to Referrals	0.05	0.04	0.07	0.04	0.04	0.04
	[0.17]	[0.17]	[0.19]	[0.17]	[0.18]	[0.16]
Total Observations	120,951	15,813	9,525	35,468	41,234	18,911

Notes: Standard deviations are in brackets. Data come from a large urban school district in California from school year 2016-17 to 2017-18. The unit of analysis is at the student-by-year level. There are 120,951 student-by-year observations. The "other" race category includes multiracial students and student missing race data. All the statistics above are reported as proportions, except for the lagged standardized scores, the total referrals, total suspensions, total suspended days, and ratio of suspensions to referrals. Standard deviations are reported in brackets for all non-binary variables.

Table 2: Racial Gaps in Annual Student Discipline Outcomes

	At least one ...		Total Number of ...		Conversion
	Suspension	Referral	Suspensions	Referrals	Rate
	(1)	(2)	(3)	(4)	(5)
Black	0.038***	0.115***	0.168***	2.444***	0.016**
	(0.005)	(0.010)	(0.030)	(0.360)	(0.006)
Hispanic	0.002*	0.020***	-0.010	0.380	-0.007
	(0.001)	(0.004)	(0.025)	(0.306)	(0.006)
Asian	-0.009***	-0.035***	-0.032	-0.609**	-0.011
	(0.002)	(0.006)	(0.036)	(0.250)	(0.011)
Other Race	0.004**	0.003	0.062	0.530*	0.000
	(0.002)	(0.004)	(0.048)	(0.319)	(0.009)
Missing Race	0.000	0.007*	0.045	1.343***	-0.006
	(0.001)	(0.004)	(0.035)	(0.480)	(0.008)
White Student Mean	0.006	0.049	0.132	3.015	0.042
Controls for:					
School-Year FEs	✓	✓	✓	✓	✓
Time-varying controls	✓	✓	✓	✓	✓
Adjusted R-squared	0.065	0.155	0.089	0.149	0.006
Observations	120,951	120,951	10,393	10,393	10,393

Notes: Standard errors are in parentheses. The omitted race group is white students. The conversion rate is the ratio of total suspensions to total referrals. The time-varying controls include gender, special education status, grade-level, student's neighborhood poverty-rate, lagged math and reading standardized test scores, and lagged student discipline outcomes. Columns 3 through 5 include only student-year observations with at least one referral. $p<0.10$* $p<0.05$** $p<0.01$***.

Table 3: Within-Incident Racial Disparities in Disciplinary Outcomes

	Likelihood of Suspension			Suspension Days		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A – Race Categories Consolidated					
Minoritized	0.025**	0.025**	0.020**	0.057*	0.058*	0.045
	(0.012)	(0.012)	(0.010)	(0.033)	(0.033)	(0.029)
Asian	0.016	0.017	0.009	0.015	0.017	-0.003
	(0.015)	(0.015)	(0.012)	(0.043)	(0.042)	(0.037)
	Panel B – Detailed Race Categories					
Black	0.028**	0.027**	0.022**	0.066*	0.065*	0.054*
	(0.012)	(0.012)	(0.010)	(0.035)	(0.035)	(0.030)
Hispanic	0.021*	0.021*	0.016	0.051	0.053	0.039
	(0.012)	(0.013)	(0.011)	(0.034)	(0.034)	(0.029)
Other	0.031**	0.031**	0.025**	0.053	0.055	0.042
	(0.013)	(0.013)	(0.011)	(0.036)	(0.037)	(0.031)
Asian	0.016	0.017	0.009	0.016	0.017	-0.002
	(0.015)	(0.015)	(0.012)	(0.043)	(0.042)	(0.037)
White Student Mean		0.032			0.062	
Controls:						
Incident FEs	✓	✓	✓	✓	✓	✓
Student Characteristics		✓	✓		✓	✓
Prior Student Achievement		✓	✓		✓	✓
Prior Year's Discipline		✓	✓		✓	✓
Current Year's Discipline			✓			✓
Unique Multi-Race Referrals	6,011	6,011	6,011	6,011	6,011	6,011
Unique Multi-Race Incidents	2,544	2,544	2,544	2,544	2,544	2,544
Unique All Referrals	10,803	10,803	10,803	10,803	10,803	10,803
Unique All Incidents	4,766	4,766	4,766	4,766	4,766	4,766

Notes: Standard errors are in parentheses. Data come from a large urban school district in California from school year 2016-17 to 2017-18. The unit of analysis is at the incident level. The omitted group is white students. The "minoritized" category includes black, Hispanic, and "other" race students. The "other" race category includes multiracial, American Indian, Arabic, and Samoan students. The student characteristics includes gender, special education status, grade-level, student's neighborhood poverty-rate, lagged math and reading standardized test scores, and lagged student discipline outcomes. All model specifications include a race category called "missing race" for those students missing race data. $p<0.10$* $p<0.05$** $p<0.01$***.

Table 4: Within-Incident Racial Disparities in Disciplinary Outcomes by Incident Type

	All	Violence	Drugs	Interper	Defiance	Walkout	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A – Race Categories Consolidated						
Minoritized	0.020**	0.022	0.117	0.007	0.019	0.013**	-0.007
	(0.010)	(0.017)	(0.099)	(0.033)	(0.018)	(0.007)	(0.052)
Asian	0.009	0.031	-0.244	-0.023	0.001	-0.034	0.184
	(0.012)	(0.027)	(0.161)	(0.037)	(0.019)	(0.026)	(0.139)
	Panel B – Detailed Race Categories						
Black	0.022**	0.021	0.061	0.013	0.024	0.021*	0.026
	(0.010)	(0.016)	(0.156)	(0.035)	(0.018)	(0.011)	(0.065)
Hispanic	0.016	0.019	0.135	0.005	0.012	0.013	-0.002
	(0.011)	(0.019)	(0.103)	(0.033)	(0.020)	(0.011)	(0.061)
Other	0.025**	0.033**	0.105	-0.001	0.030	-0.002	-0.173
	(0.011)	(0.016)	(0.157)	(0.038)	(0.022)	(0.019)	(0.141)
Asian	0.009	0.030	-0.228	-0.022	0.001	-0.032	0.123
	(0.012)	(0.027)	(0.172)	(0.037)	(0.019)	(0.027)	(0.117)
White Student Mean	0.032	0.038	0.063	0.040	0.030	0.000	0.000
Multi-Race Referrals	6,011	1,732	66	1,193	2,144	815	61
Multi-Race Incidents	2,544	945	33	756	1,090	427	31
Referrals	10,803	3,114	159	2,134	3,803	1,473	120
Incidents	4,766	1,706	73	1,354	2,004	783	66

Notes: Standard errors are in parentheses. Data come from a large urban school district in California from school year 2016-17 to 2017-18. The unit of analysis is at the incident level. The omitted group is white students. The "minoritized" category includes black, Hispanic, and "other" race students. The "other" race category includes multiracial, American Indian, Arabic, and Samoan students students. All models include incident fixed effects, student characteristics, prior student achievement, and prior student discipline. The student characteristics includes gender, special education status, grade-level, student's neighborhood poverty-rate, lagged math and reading standardized test scores, and lagged student discipline outcomes. All model specifications include a race category called "missing race" for those students missing race data. $p<0.10^{*}$ $p<0.05^{**}$ $p<0.01^{***}$.

Appendix A

Table A1: Frequency of Referrals by Race and School Level

	White	Black	Hispanic	Asian	Other
Elementary	816	6,973	4,832	628	1,199
	1.70%	14.90%	10.30%	1.30%	2.60%
Middle	1,020	7,442	7.425	1,610	1,092
	2.20%	15.90%	15.90%	3.40%	2.30%
High School	581	5,098	6,085	875	1,135
	1.20%	10.90%	13.00%	1.90%	2.40%
All	2,417	19,513	18,342	3,113	3,426
	5.20%	41.70%	39.20%	6.70%	7.30%

Note: The unit of analysis is at the referral level.

Table A2: Frequency of Referrals by Reason and School Level

	Violence	Drugs	Interpersonal Offenses	Disruption/ Noncompliance	Class Skipping or Walkout	Other Reason	Total
Elem	7,447	18	2,836	3,421	573	153	14,448
	15.9%	0.1%	6.1%	7.3%	1.2%	0.3%	30.9%
Middle	4,101	80	5,178	6,880	2,169	181	18,589
	8.8%	0.2%	11.1%	14.7%	4.6%	0.4%	39.7%
High	1,537	457	4,036	5,405	2,178	161	13,774
	3.3%	0.9%	8.6%	11.6%	4.7%	0.3%	29.4%
All	13,085	555	12,050	15,706	4,920	495	46,811
	28.0%	1.2%	25.7%	33.6%	10.5%	1.1%	100.0%

Note: The unit of analysis is at the referral level.

Table A3: Racial Gaps in Annual Student Discipline Outcomes (Simple Model)

	At least one ...		Total Number of ...		Conversion
	Suspension	Referral	Suspensions	Referrals	Rate
	(1)	(2)	(3)	(4)	(5)
Black	0.060***	0.226***	0.248***	4.112***	0.025***
	(0.008)	(0.025)	(0.043)	(0.579)	(0.009)
Hispanic	0.013***	0.076***	0.048	0.982**	0.003
	(0.003)	(0.012)	(0.034)	(0.390)	(0.008)
Asian	-0.002*	-0.019***	0.007	-0.667**	0.001
	(0.001)	(0.006)	(0.041)	(0.299)	(0.011)
Other Race	0.012***	0.040***	0.116**	1.477***	0.011
	(0.003)	(0.008)	(0.052)	(0.454)	(0.011)
Missing Race	-0.001	0.001	-0.006	2.108***	-0.016*
	(0.001)	(0.006)	(0.037)	(0.650)	(0.009)
White Student Mean	0.006	0.049	0.132	3.015	0.042
Adjusted R-squared	0.019	0.059	0.013	0.039	0.004
Observations	120,951	120,951	10,393	10,393	10,393

Notes: The omitted race group is white students. The conversion rate is the ratio of total suspensions to total referrals. None of the regressions above include any fixed effects, or control variables. Columns 3 through 5 include only students with at least one referral. $p<0.10$* $p<0.05$** $p<0.01$***.

Table A4: Comparing Characteristics of Different Types of Incidents

	Type of Incidents				
Variables	(1) Single Student	(2) Multi-Student, Same-Race	(3) Multi-Student, Multi-Race	(4) P-value (1)=(3)	(5) P-value (2)=(3)
# of Students	1.00	2.27	2.82	0.00	0.00
White	0.05	0.02	0.06	0.00	0.00
Black	0.39	0.45	0.35	0.00	0.00
Hispanic	0.37	0.46	0.33	0.00	0.00
Asian	0.06	0.03	0.09	0.00	0.00
Other race	0.07	0.02	0.11	0.00	0.00
Female	0.25	0.33	0.30	0.00	0.00
Special Education	0.41	0.30	0.30	0.00	0.76
Lagged Standardized Math Score	-0.53	-0.73	-0.58	0.00	0.00
Lagged Standardized ELA Score	-0.58	-0.72	-0.60	0.04	0.00
Lagged # of Referrals	6.10	5.04	4.59	0.00	0.02
Lagged # of Suspensions	0.37	0.27	0.24	0.00	0.08
Elementary School	0.33	0.24	0.28	0.00	0.00
Middle School	0.39	0.50	0.47	0.00	0.00
High School	0.23	0.24	0.24	0.61	0.32
Violence	0.29	0.29	0.29	0.30	0.91
Drugs	0.01	0.02	0.01	0.86	0.00
Interpersonal Offenses	0.27	0.20	0.20	0.00	0.98
Disruption/Noncompliance	0.32	0.35	0.36	0.00	0.25
Class Skipping/Walkout	0.09	0.13	0.14	0.00	0.93
Other Reasons	0.01	0.01	0.01	1.00	0.25
# of Observations	36370	5061	6013		

Notes: This table compares characteristics of three types of incidents in our sample. Columns 4 and 5 provide p values for simple two-sample T tests comparing single-student and multi-student, same-race incidents to our identifying sample which is multi-student multi-race incidents.

Table A5: Regressions on Likelihood of Suspension and Suspension Days

	Likelihood of Suspension			Suspension Days		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A – Race Categories Consolidated					
Minoritized	0.024*	0.023	0.023	0.036	0.039	0.039
	(0.014)	(0.015)	(0.015)	(0.023)	(0.026)	(0.026)
Asian	0.020	0.021	0.021	0.001	0.006	0.006
	(0.189)	(0.19)	(0.019)	(0.035)	(0.036)	(0.036)
White Student Mean		0.015			0.012	
Controls:						
Incident FEs	✓	✓	✓	✓	✓	✓
Student Characteristics		✓	✓		✓	✓
Prior Student Achievement		✓	✓		✓	✓
Prior Year's Discipline		✓	✓		✓	✓
Current Year's Discipline			✓			✓
Unique Multi-Race Referrals	1,668	1,668	1,668	1,668	1,668	1,668
Unique Multi-Race Incidents	1,116	1,116	1,116	1,116	1,116	1,116
Unique All Referrals	2,966	2,966	2,966	2,966	2,966	2,966
Unique All Incidents	2,022	2,022	2,022	2,022	2,022	2,022

Notes: Cluster-robust standard errors at the incident level are in parentheses. Data come from a large urban school district in California from school year 2016-17 to 2017-18. The unit of analysis is at the incident level. The sample only includes observations where the student has no prior referrals this school year. The omitted group is white students. The "minority" category includes black, Hispanic, and "other" race students. The "other" race category includes multiracial students. The student characteristics include gender, special education status, grade-level, student's neighborhood poverty-rate, lagged math and reading standardized test scores, and lagged student discipline outcomes. All model specifications include a race category called "missing race" for those students missing race data. $p<0.10^{*}$ $p<0.05^{**}$ $p<0.01^{***}$.

Table A6: Weighted Regression Results Accounting for Selection into Identification

	Likelihood of Suspension		Suspension Days	
	(1)	(2)	(3)	(4)
	Unweighted	Weighted	Unweighted	Weighted
Minoritized	0.010*	0.008	0.028*	0.021
	(0.006)	(0.006)	(0.016)	(0.018)
R2	0.682	0.588	0.666	0.507
Observations	10808	2383	10808	2383

Notes: Different from our main specification, we combine Asian and white students as the referrence group so we only have one treatment group in order to implement the weighting strategy. Following Miller et al. (2019), we implement a one-step weighting strategy that uses the product of predicted likelihood of being in the identifying sample and inverse conditional variance as regression weights. p<0.10* p<0.05** p<0.01***.

Table A7: Regressions on Likelihood of Suspension by School Level

	All	Elem	Middle	High
	(1)	(2)	(3)	(4)
	Panel A – Race Categories Consolidated			
Minoritized	0.020**	0.011	0.000	0.057***
	(0.010)	(0.014)	(0.017)	(0.021)
Asian	0.009	0.035**	-0.014	0.020
	(0.012)	(0.018)	(0.021)	(0.027)
	Panel B – Detailed Race Categories			
Black	0.022**	0.012	0.005	0.058***
	(0.010)	(0.014)	(0.017)	(0.022)
Hispanic	0.016	0.010	-0.009	0.059***
	(0.011)	(0.016)	(0.018)	(0.022)
Other	0.025**	0.013	0.021	0.047**
	(0.011)	(0.014)	(0.022)	(0.022)
Asian	0.009	0.035**	-0.014	0.020
	(0.012)	(0.018)	(0.021)	(0.027)
White Student Mean	0.032	0.015	0.058	0.009
Controls:				
Incident FEs	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓
Prior Student Achievement	✓	✓	✓	✓
Prior Discipline	✓	✓	✓	✓
Unique Multi-Race Referrals	6,011	1,683	2,803	1,414
Unique Multi-Race Incidents	2,544	758	1,154	597
Unique Referrals	10,803	10,803	10,803	10,803
Unique Incidents	4,766	4,766	4,766	4,766

Notes: Standard errors are in parentheses. Data come from a large urban school district in California from school year 2016-17 to 2017-18. The unit of analysis is at the incident level. The omitted group is white students. The "minority" category includes black, Hispanic, and "other" race students. The "other" race category includes multiracial students. The student characteristics includes gender, special education status, grade-level, student's neighborhood poverty-rate, lagged math and reading standardized test scores, and lagged student discipline outcomes. All model specifications include a race category called "missing race" for those students missing race data. p<0.10* p<0.05** p<0.01***.

Table A8: Heterogeneity Results

	All	High School	All	High School	All	High School
	(1)	(2)	(3)	(4)	(5)	(6)
Minoritized	0.010	0.017*	0.021*	0.052***	0.002	0.015
	(0.021)	(0.010)	(0.011)	(0.019)	(0.019)	(0.017)
Asian	0.017	-0.009	0.012	0.008	0.015	-0.037
	(0.032)	(0.041)	(0.013)	(0.026)	(0.023)	(0.026)
Male	-0.017	-0.063**				
	(0.023)	(0.029)				
Minority × Male	0.013	0.059**				
	(0.023)	(0.030)				
Asian × Male	-0.010	0.044				
	(0.034)	(0.052)				
Special Education			0.007	-0.065		
			(0.019)	(0.090)		
Minority × Spec-Ed			-0.004	0.048		
			(0.019)	(0.089)		
Asian × Spec-Ed			-0.011	0.098		
			(0.026)	(0.091)		
Poor					-0.037	-0.040
					(0.024)	(0.029)
Minority × Poor					0.040	0.045
					(0.025)	(0.033)
Asian × Poor					0.027	0.081*
					(0.029)	(0.043)
White Student Mean	0.032	0.009	0.032	0.009	0.032	0.009
Unique Multi-Race Referrals	6,011	1,414	6,011.	1,414	6,011	1,414
Unique Multi-Race Incidents	2,544	597	2,544	597	2,544	597
Unique Referrals	10,803	2,548	10,803	2,548	10,803	2,548
Unique Incidents	4,766	1,115	4,766	1,115	4,766	1,115

Notes: Data come from a large urban school district in California from school year 2016-17 to 2017-18. The unit of analysis is at the incident level. The omitted group is white students. The "minority" category includes black, Hispanic, and "other" race students. The "other" race category includes multiracial students. All models include incident fixed effects (FEs), student characteristics, prior student achievement, and prior student discipline. The student characteristics includes gender, special education status, grade-level, student's neighborhood poverty-rate, lagged math and reading standardized test scores, and lagged student discipline outcomes. The "poor" category includes students residing in neighborhoods that have poverty rates below the 50th percentile. All model specifications include a race category called "missing race" for those students missing race data.

Appendix B

Decomposing Racial Gaps

We decompose racial gaps in referrals and suspensions into between-school and within-school components. We compare Black, Hispanic, and Asian students to their white peers by using both the likelihood of receiving a referral and the likelihood of having a suspension in a school year as our two outcomes. Following Barret et al. (2019), we define the raw average referral or suspension rate $\bar{D}_{is}$ for a given group of students in a given grade weighted across students and schools using equation (1) below:

$$\bar{D}_{is} = \frac{\sum_i \sum_s Group_{is} Y_{is}}{\sum_i \sum_s Group_{is}} \tag{1}$$

where i indicates students and s indicates schools. $Group_{is}$ indicates the student's racial or ethnic identity. Y_{is} takes the value of 1 if the student receives, for example, an office referral in the focal year, and 0 otherwise.

For simplicity, we use $\bar{D}_{is}$ to represent white students' referral or suspension rates and $\tilde{D}_{is}$ is to indicate the same measure for a non-white student group, which can be Black, Hispanic, or Asian students. Our goal is to decompose the raw gap $\bar{D}_{is}$-$\tilde{D}_{is}$ into between- and within-school components using equation (2) below:

$$\bar{D}_{is} - \tilde{D}_{is} = \bar{D}_s - \tilde{D}_s + ((\bar{D}_{is} - \tilde{D}_{is}) - (\bar{D}_s - \tilde{D}_s)) \tag{2}$$

$\bar{D}_s - \tilde{D}_s$ would be the measure on between-school gap and $(\bar{D}_{is} - \tilde{D}_{is}) - (\bar{D}_s - \tilde{D}_s)$ is the within-school gap. To plot Figure 1, we compute elements in Equation (2) for each grade (K-12) and each minoritized-white combinations for both referral and suspension rates using our analytic sample (school years 2016-2017 to 2017-2018).