



Troublemakers? The Role of Frequent Teacher Referrers in Expanding Racial Disciplinary Disproportionalities

Jing Liu
University of Maryland,
College Park

Emily K. Penner
University of California, Irvine

Wenjing Gao
University of Maryland,
College Park

Teachers' sense-making of student behavior determines whether students get in trouble and are formally disciplined. Status categories, such as race, can influence perceptions of student culpability, but the degree to which this contributes to racial disproportionality in discipline receipt is unknown. This study provides the first systematic documentation of teachers' use office discipline referrals (ODRs) in a large, diverse urban school district in California that specifies the identity of both the referred and referring individuals in all ODRs. We identify teachers exhibiting extensive referral behavior, or the top 5% referrers based on the number of ODRs they make in a given year and evaluate their contributions to disciplinary disparities. We find that "top referrers" effectively double the racial gaps in ODRs for both Black-White and Hispanic-White comparisons. These gaps are mainly driven by higher numbers of ODRs issued for Black and Hispanic students due to interpersonal offences and defiance, and also partially convert to racial gaps in suspensions. Both the level and racial compositions of the school sites where "top referrers" serve and their personal traits seem to explain some of their frequent referring behavior. Targeting supports and interventions to "top referrers" might afford an important opportunity to reduce racial disciplinary gaps.

VERSION: August 2022

Suggested citation: Liu, Jing, Emily Penner, and Wenjing Gao. (2022). Troublemakers? The Role of Frequent Teacher Referrers in Expanding Racial Disciplinary Disproportionalities. (EdWorkingPaper: 22-616). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/dlhx-sn90>

Troublemakers? The Role of Frequent Teacher Referrers in Expanding Racial Disciplinary Disproportionalities *

Jing Liu[†]

Emily K. Penner[‡]

Wenjing Gao[§]

ABSTRACT: Teachers' sense-making of student behavior determines whether students get in trouble and are formally disciplined. Status categories, such as race, can influence perceptions of student culpability, but the degree to which teachers' initial identification of student misbehavior exacerbate racial disproportionality in discipline receipt is unknown. This study provides the first systematic documentation of teachers' use of office discipline referrals (ODRs) in a large, diverse urban school district in California that specifies the identity of both the referred and referring individuals in all ODRs. We identify teachers exhibiting extensive referral behavior, or the top 5% referrers based on the number of ODRs they make in a given year and evaluate their contributions to disciplinary disparities. We find that "top referrers" effectively double the racial gaps in ODRs for both Black-White and Hispanic-White comparisons. These gaps are mainly driven by higher numbers of ODRs issued for Black and Hispanic students due to interpersonal offences and defiance, and also partially convert to racial gaps in suspensions. Both the level and racial compositions of the school sites where "top referrers" serve and their personal traits seem to explain some of their frequent referring behavior. Targeting supports and interventions to "top referrers" might afford an important opportunity to reduce racial disciplinary gaps.

KEYWORDS: Office Discipline Referrals, Exclusionary Discipline, Racial Disparities

*The authors thank David Blazar, Seth Gershenson, Betty Malen, Michael Hayes, and participants in the Annenberg Half-Baked Seminar for helpful discussions. 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.

[†][Corresponding author] jliu28@umd.edu. Assistant Professor, College of Education, University of Maryland, College Park; Research Associate, IZA Institute of Labor Economics.

[‡]emily.penner@uci.edu. Associate Professor, School of Education, University of California, Irvine.

[§]wgao12@umd.edu. Ph.D. student, College of Education, University of Maryland, College Park.

1 Introduction

Black students and students of color are substantially more likely to experience exclusionary discipline (e.g., suspensions and expulsions) than their White peers (U.S. Department of Education, 2016a). Racial disproportionalities have been identified in frequency of discipline referrals, receipt of discipline infractions, and severity of punishments for students involved in the same offences and the same discipline events (Barrett et al., 2021; Liu et al., 2022; Losen et al., 2015; Skiba et al., 2011). Exclusionary discipline has been associated with a host of negative academic impacts and school disengagement for disciplined students (Bell & Puckett, 2020; Chu & Ready, 2018; Noltemeyer et al., 2015; Hwang, 2018; Sorensen et al., 2022), and has widened racial gaps in absences and student achievement (e.g., Lacoë & Steinberg, 2019; Pearman et al., 2019). Racial inequities in receipt of school discipline have also been linked to disparities in young adulthood, including exposure to the criminal justice system, SNAP receipt, and college completion (Bacher-Hicks et al., 2019; Davison et al., 2021; Mittleman, 2018; Sorensen et al., 2022). The pervasiveness and lasting consequences of racial disparities school discipline underscore the need to identify the root causes of such disparities.

Despite longstanding interest in identifying the sources of unequal discipline receipt (Children’s Defense Fund, 1975; Kim, 2011; McCarthy & Hoge, 1987; Skiba et al., 2002; U.S. Department of Education, 2014), successful reductions in absolute levels of discipline following state and national policy efforts, and the adoption of popular discipline reform efforts such as restorative justice, racial disproportionality in school discipline persists (Anderson & Ritter, 2016; Hwang et al., 2022; Noltemeyer & McLoughlin, 2010; Executive Office of the President, 2016; U.S. Department of Education, 2014, 2016b; Davison et al., 2019; Gregory et al., 2018; Hashim et al., 2018). We thus argue that further reductions in racial disparities in exclusionary discipline hinge not just on formal programmatic and policy changes, but on also understanding the specific actions of educators. Research examining educators’

role in creating discipline disparities focuses primarily on *differential processing* of students through unequal consequences for similar behaviors (Barrett et al., 2021; Girvan et al., 2017; Skiba et al., 2002). A much smaller body of research is beginning to document how teachers also influence *differential selection* into formalized discipline through unequal use of Office Discipline Referrals (ODRs) (Girvan et al., 2017; Santiago-Rosario et al., 2021; Skiba et al., 2002, 2011). Recent evidence indicates variability in the use of ODRs by teacher (Holt et al., 2022), but it is unclear how much of the racial disparities in the receipt of ODRs and formal consequences can be attributed to differences in referral behavior.

We extend this literature by examining variability teachers’ use of ODRs and considering whether extensive referral behavior by a small number of teachers widens racial gaps in discipline receipt. We use four years of teacher-referral linked data from an urban school district to identify the top five percent of referrers, hereafter “top-referrers” and their contributions to racial referral and discipline gaps. We also test whether ODRs for more subjective offences (e.g., defiance versus weapons) contribute more to racial referral and discipline gaps. Understanding whether discipline is heavily concentrated among a small number of individuals, and identifying the characteristics and locations of teachers who produce a substantial amount of ODRs can help school districts to concentrate their intervention and support efforts to hopefully mitigate racial disparities in exclusionary discipline.

2 Literature Review

School discipline is a transactional process in which educators make a series of determinations about whether and how to react to undesired student behaviors (Girvan et al., 2017; Okonofua et al., 2016). These “moment-by-moment” interactions in which discipline decisions are made are instances where status categories, such as race, influence perceptions about students’ esteem, competence, and culpability, often in unconscious ways (Shedd, 2015; Vavrus & Cole, 2002). Lewis & Diamond (2015) build on a typology developed by Skiba et al.

(2002) and Piquero (2008) highlighting how the construction and enactment of school rules can infuse racial dynamics into decisions about who is selected for punishment (because a behavior they commit is perceived as needing discipline) and what severity of punishment (if any) is judged appropriate to the behavior. This combination of *differential selection* for discipline and *differential processing* of discipline by race are hypothesized to exacerbate the observed differential outcomes in discipline between racial and ethnic groups.

The existing literature has foregrounded investigations of racial disparities that emerge from *differential processing* in part because exclusionary punishment frequency and severity are routinely documented in district data systems and officially collected by the Office of Civil Rights (Bradshaw et al., 2010; Skiba et al., 2002, 2011; U.S. Department of Education, 2014). Research using these records of exclusionary discipline indicates that Black and Latinx students are more likely to be criminalized and punished more harshly by educators and school personnel, even for minor infractions (Allen, 2017; Blake et al., 2011; Lewis & Diamond, 2015; Rios, 2011; Sorensen et al., 2022). Differences in teachers' use of discipline for these minor and subjectively defined behaviors uniquely perpetuate racial discipline gaps (Barrett et al., 2021; Girvan et al., 2017; Skiba et al., 2002).¹

School context also contributes to *differential processing* (Kinsler, 2011; Mendez et al., 2002). Schools with higher proportions of Black students, other students of color, and low-income students have higher rates of exclusionary discipline (Gottfredson et al., 2005; Mendez et al., 2002; Welch & Payne, 2010). Likewise, schools' authoritative climate and disciplinary structure distinguish high and low-suspending schools and are associated with different magnitudes of Black-White suspension rate gaps (Heilbrun et al., 2018).

Yet to understand how disparities in these outcomes emerge, it is also important to attend to whether and how they emerge through *differential selection*, including both the

¹We follow a long line of literature to distinguish subjectively vs. objectively defined student behaviors (e.g., Lindsay & Hart, 2017; Girvan et al., 2017; Smolkowski et al., 2016). The main difference between these two types of behaviors is that behaviors such as defiance and disrespect are more ambiguously defined and allow more staff discretion than more objectively defined behaviors such as violence. See Smolkowski et al. (2016) for a more detailed discussion.

perception of a student action as misbehavior, as well as the formal reporting of that misbehavior through an ODR. Research provides several decades of evidence indicating teachers' differential perceptions of similar behaviors by Black and White students (Gilliam et al., 2016; Neal et al., 2003; Okonofua & Eberhardt, 2015; Owens & McLanahan, 2020; Skiba et al., 2011; Zimmerman et al., 1995). For example, educators are more likely to perceive behaviors committed by Black and other students of color as more problematic and deserving of punishment (Gregory & Roberts, 2017; Okonofua & Eberhardt, 2015; Owens, 2020; Santiago-Rosario et al., 2021). Yet student self-reports and other research suggest few racial differences in misbehavior (American Psychological Association, 2008; Gregory et al., 2010; Losen et al., 2015; McCarthy & Hoge, 1987; Wu et al., 1982).

A much smaller literature has documented *differential selection* resulting from the formalization of misbehavior through ODRs, finding evidence of disproportionalities in the likelihood of receiving ODRs based on student and school characteristics (Girvan et al., 2017; Santiago-Rosario et al., 2021; Skiba et al., 2002, 2011; Smolkowski et al., 2016). For example, Liu et al. (2022) find that Black students are more likely to receive ODRs and to have their ODRs converted to suspensions than White students even if they are involved in the same incident. There is also growing evidence that teachers' use of referrals, particularly in response to more subjective incidents, exacerbates gaps in exclusionary discipline and can also expand Black-White disparities in absenteeism and achievement (Girvan et al., 2021; Holt et al., 2022; Lindsay & Hart, 2017; Liu et al., 2022).

Collectively, these findings reinforce the need to examine and potentially intervene in teachers' use of ODRs as a means for potentially reducing racial disproportionality in referrals and other discipline outcomes. However, the lack of data linking specific teachers to office referral records has made it difficult to examine the contributions of specific teachers to racial disproportionalities in these early stages of the discipline process. One exception is Holt et al. (2022) who identify variability in the degree of late-elementary educators' punitiveness in the use of ODRs in North Carolina, developing a teacher-specific estimate of their contribution

to their average students' number of ODRs. More punitive teachers contribute to adverse academic and behavioral outcomes for Black students, particularly from referrals for issues like defiance. This highlights the idea that teachers vary in their use of ODRs and that this has negative consequences for students. It also suggests that examining the prevalence and distribution of the very punitive teachers and providing them with additional supports may be a promising path toward reducing racial disproportionality in ODRs.

Our work builds on this study to examine the distribution of ODRs across educators of all school levels and the degree to which extensive referral behavior contributes to racial disparities in referrals and exclusionary discipline. We exploit unusually rich administrative data that track all ODRs and the identity of both the referred and referring individuals from a large, diverse urban school district in California to provide the first systematic documentation of teachers' referring behavior. We identify how the ODRs of teachers who are at the top of the distribution of annual referring frequency translate into racial disproportionalities in both ODR and suspension receipt among students. We further probe how school context and personal traits might explain their frequent referring behavior. Specifically, this paper answers the following three research questions:

Research Question 1: Distribution. What is the prevalence of ODR use by unique teacher referrers in the focal district? Are there teachers who engage in extensive referring?

Research Question 2: Disciplinary Disproportionalities. How much do “top referrers” contribute to referral and suspension disproportionalities by race/ethnicity and by incident type?

Research Question 3: Predictors. How do school context and individual characteristics associate with a teacher's likelihood of being a referrer or a “top referrer”?

3 Data

Our analyses draw on unusually rich administrative data from a large, diverse urban school district in California for the 2016-17 through 2019-20 school years. During this period, this district served over 79,000 unique students in grades K-12 in 114 unique schools, totaling 227,922 student-year observations. We identify students' race/ethnicity using district records of students' self-reports of racial/ethnic identity following district classification categories. About 33 percent of the students are Asian. Hispanic students account for 30 percent. Black and White students account for 7 and 12 percent, respectively. The remaining 19 percent of students are identified as multi-racial or are missing race/ethnicity information.

One unique feature of the data is that it provides exceptionally detailed ODR records, regardless of whether an ODR leads to a suspension. Specifically, aside from the information on which students are involved in an ODR and the reason for the referral (i.e., type of incident), the data also contain a unique identifier for the educator who issued the referral, incident time, date, and location.² In addition, the data allow us to link ODR data to suspension data so that we can identify the discipline consequences of each ODR record. Such detailed information allows us to, for the first time, systematically describe educators' referring behavior and map out how such behavior might contribute to racial disproportionalities in both ODR and suspension. Further, the district also provides us with rich data on educator demographics, their appointment type, credential area, and years of experience, which we use to characterize educators with differential referring behavior.

Based on our conversation with district leaders, we exclude ODRs issued by administrators for two reasons. First, some teachers, especially substitute teachers and university interns who may not have access to the ODR reporting system, may rely on responsible administrative staff to issue an ODR. This may obscure the source of ODRs from administrators, causing some data accuracy issues. Second, by focusing on classroom teachers as

²In Appendix Figures A1, A2, and A3, we show the distribution of ODRs along those dimensions.

referrers, we examine the major source of ODRs that affords more opportunities for potential intervention, given teachers are the primary adults interacting with students in school. As a result, our analytic sample comprises 75,229 ODRs that were issued by 2,928 unique teachers (5,855 teacher-year observations). The demographics of teacher referrers are similar to all referrers, with 48 percent White, 5 percent Black, 16 percent Hispanic, 19 percent Asian and 12 percent identified as multi-racial or are missing race/ethnicity information.

The data also provide reason(s) each ODR was issued. A significant number of ODRs are the result of multiple infraction categories. We follow Lindsay and Hart (2017) in making mutually exclusive categories based on the “most severe” reason listed for an ODR: (a) violence; (b) drugs; (c) interpersonal offenses; (d) disruption or non-compliance; (e) class skipping or walkout; (f) other. For example, an ODR for which the student was disruptive and used violence would be coded as violence. We further categorize these reasons into three incident types: (a) interpersonal/defiance; (b) violence; (c) truancy/drug. To illustrate, “interpersonal offenses” and “disruption or non-compliance” are the most prevalent subjective reasons (57 percent of total ODRs); “violence” is the most common objective reason for an ODR (28 percent of total ODRs); whereas “drug”, “class skipping”, and “other” reasons together consists 15 percent of total ODRs.

4 Method

To answer RQ 1, we plot the distribution of ODRs at the teacher-by-year level and use the sample of teachers who make at least one ODR in a given year. As the distribution is positively skewed with a very long tail, we choose to group teachers into bins with unequal intervals for enhanced data visualization and ease of interpretation. Specifically, the first bins include teacher-year observations for teachers issued one to two and three to five ODRs, as these are the most prevalent annual quantities of ODRs issued. We then increase the interval by five ODRs incrementally until reaching up to 50 annual ODRs. We follow the same logic

and gradually expand the interval, and eventually top-code all teacher-year observations that have more than 200 ODRs into the last bin. We then identify individuals who issued the top 5% of unique ODRs within each of the four years to identify a group we refer to as the “top referrers”.

For RQ 2, we evaluate the contribution of “top referrers” to disciplinary disparities in both ODRs and suspensions through examining how the cumulative racial gaps in a particular disciplinary outcome evolve when considering ODRs issued by teachers with different referring frequencies.³ To do this, we first define the average ODRs for a given subgroup of students using all students who received at least one ODR from teachers who issued fewer than n ODRs in a given year by using the following equation:

$$\bar{D}_{i,j(N \leq n)} = \frac{\sum_i \sum_j Group_{ij} Y_{ij}}{\sum_i \sum_j Group_{ij}} \quad (1)$$

where i indicates students and j indicates teachers. $Group_{ij}$ indicates the student’s racial or ethnic identity. Y_{ij} indicated the number of ODRs the student received from teacher j in the focal school year.

For simplicity, we use $\bar{D}_{i,j(N \leq n)}$ to represent White students’ ODR rate and $\tilde{D}_{i,j(N \leq n)}$ to indicate the same measure for a non-White student group, which can be Asian, Black, Hispanic, or “other” students who self-identify as multi-racial or are missing race/ethnicity information. Our goal is to calculate how raw racial gaps evolve as we gradually include teachers who issue a higher volume of ODRs using equation (2) below:

$$\text{Cumulative racial gap} = \bar{D}_{i,j(N \leq n)} - \tilde{D}_{i,j(N \leq n)} \quad (2)$$

We then follow the same method but change Y_{ij} to indicate the number of ODRs the

³As our method is based on a teacher’s annual referring volume, some might worry that the Covid-19 pandemic which started in March 2020 might affect the data for school year 2019-2020 and thus our results. We show in Appendix Figure A4 that excluding year 2020 results in almost identical distribution of ODRs and cumulative racial gaps. We are also able to replicate our findings for RQ1 and RQ4 in Figure A5 and Table A1 by only using the first three years of data.

student received from teacher j that converts to a suspension.

RQ 3 asks how school context and individual characteristics associate with a teacher’s likelihood of being a referrer or a “top referrer”. To answer this question, we first compare the proportion of different racial/ethnic groups in four categories: (1) all students in the district; (2) all students who are referred at least once; (3) all students who are from schools that have at least one “top referrer”; (4) all students who are referred at least once by a “top referrer”. Through comparing how the racial composition of students referred by “top referrers” differ from demographics of all referred students, we will provide additional evidence on whether “top referrers” increase or decrease the overall racial gaps. The comparisons to the racial composition of all students in the district and the school sites where “top referrers” serve can provide some clues on the working environments of the “top referrers” and how such contexts might explain their referring behavior.

To further probe into the influence of individual characteristics, we then conduct regression analyses using all classroom teachers who served the district in the time span we examine. We use two indicator variables as outcomes—one indicates whether a teacher made at least one ODR in a given year and the other indicates whether a teacher is a “top referrer”. We regress these two outcomes on teachers’ gender, race/ethnicity, credentials, and years of experience, controlling for school and year fixed effects to account for factors that are constant at the school and year levels, respectively. We first estimate the regressions for the entire sample. We then estimate the same models separately for elementary, middle, and high schools to examine whether such correlations vary by school level.

5 Results

RQ1. Distribution. Among all classroom teachers who served the district during the four years we study, 34.2% of them ever made an ODR. We thus exclude teachers who made zero referrals in a given year and describe the distribution of individual teachers’ referring

frequency for referrers. This distribution is visualized in Figure 1. The first Y axis indicates the proportion of unique referring teachers in a given referring frequency among all referrers. The distribution is skewed toward the right with a long tail. More specifically, over half of referring teachers (50.4%) have an annual count of ODRs below five.

At the same time, some individual teachers’ referring frequency is extensive. Teachers with an annual referring frequency in the top 5% of referrers (1.7% of all teachers), which we label as “top referrers”, issued over 45 referrals per year (meaning roughly one referral every four school days), several magnitudes greater than their average-referring colleagues (who issue less than one referral for every two months of school). While only about 80 teachers are *ever* “top referrers” in our sample, they account for 36.9% of all ODRs made in this four-year period.⁴ Motivated by understanding the implications of this extensive referring behavior by a small group of teachers, we turn to RQ2 to investigate their potential contribution to racial and ethnic disproportionalities in both ODRs and suspensions.

RQ2. Disciplinary Disproportionalities. As discussed in Section 4, we answer this question by calculating cumulative referral gaps based on the annual counts of ODRs at the teacher-by-year level. The results are visualized in Figure 1, overlaid on the bar graph that we use to answer RQ1. Specifically, the second Y axis indicates cumulative referral gaps we define in equation 2, while the four lines indicate how such gaps for specific racial group comparisons change as we increasingly include teachers whose annual volume of ODRs corresponds to the bins on the X axis. For example, racial gaps in ODRs are close to zero for all comparisons when we consider ODRs made by teachers who only referred one or two students in a given year. However, if we consider all ODRs by teachers who made at most 25 ODRs per year (the top 15 percentile number of referrals), the Black-White gap increases to one referral, and the Hispanic-White and other-White gaps also increase somewhat, although the magnitudes are not as large.

⁴We also use the top 10% of referring teachers as the definition for “top referrers”, and replicate the main analyses in Appendix Figures A6,A7, and A8 This alternate threshold does not affect our main findings.

Overall, Figure 1 presents several distinct patterns. First, the Black-White gap in ODRs increases steadily as we gradually include more frequent referrers. When we consider all referrers, the Black-White referral gap reaches slightly above three ODRs. This pattern is similar for Hispanic-White and other-White gaps, but the growth for these two gaps are not nearly as fast as the Black-White gap. In contrast, the Asian-White gap stays relatively stable and is actually negative.

Second, if we combine the two sets of information presented in Figure 1, a less readily visible but critical finding is that the top 5% referrers, as identified in RQ1, effectively double the Black-White, Hispanic-White, and other-White referral gaps. For example, the Black-White gap in ODRs is about 1.6 when we consider all referrers with less than 46 ODRs. Once we include “top referrers”, this number jumps to around 3.4. In other words, while the “top referrers” represent a small group of teachers by definition, the sheer volume of ODRs they made has an outsized influence on overall racial gaps in ODRs.

We further unpack Figure 1 to explore how different types of ODRs might contribute to the patterns we have seen so far. Figure 2 provides the same plot but by race and ODR type. We group all ODRs into three buckets based on the most severe reasons for issuing a given ODR—interpersonal/defiance, violence, and truancy/drugs. Overall, what primarily drives the increase of racial gaps in ODRs, especially those by “top referrers”, are ODRs issued for interpersonal/defiance reasons, which are arguable more likely to be subject to bias (Lindsay & Hart, 2017). For example, “top referrers” double the Black-White referral gap in the interpersonal/defiance category (2.0 ODRs), which also account for more than half of the total Black-White referral gap (3.4 ODRs).

While ODRs, regardless of whether they lead to a suspension, are consequential on their own, the negative and long-lasting impact of suspensions are more well documented (e.g., Bacher-Hicks et al., 2019). We now turn to examine how the patterns we observe in ODRs convert to racial gaps in suspensions. In Figure 3, we present the parallels of Figure 1 but change the second Y axis to cumulative suspension gaps. As documented in Liu et al. (2022)

which uses identical data to this study, only 4% of all ODRs convert to suspensions. Thus, it is expected that the overall magnitudes of suspension gaps are much smaller compared with ODR gaps.

The patterns of suspension gaps largely align with those of ODR gaps, but show some distinct features. The most obvious difference is that the ODR gaps map onto suspension gaps with a much larger conversion rate for the Black-White gap compared with other racial group comparisons. Specifically, the total suspension gap for the Black-White comparison is 5% (.17/3.4) of the referral gap, but this rate is less than 3% for Hispanic-White and other-White comparisons. This is consistent with findings by Liu et al. that there is also a sizable referral conversion gap for Black students compared with their White peers.

Another distinctive pattern is that the Black-White suspension gap grows most quickly when we gradually include referrers who made up to 50 referrals per year (from zero to around .15 suspensions). Top 5% referrers continue to widen this gap until reaching a gap of .17 suspensions, but at a slower growth rate. We also observe that the Black-White and other-White gaps initially increase quickly but slow down when including “top referrers”. In sum, this suggests that “top referrers” contribute to racial suspension gaps. While their contribution is not commensurate with the sheer volume of the referrals they made, which is partially explained by the fact that they frequently referred students of color for minor reasons like interpersonal offences and defiance, they exacerbate the already large suspensions gaps.

RQ3. Predictors. Thus far, we have documented the contributions of “top referrers” on racial disproportionalities in both ODRs and suspensions. In RQ4, we start to probe into how working environments and personal traits might explain their referring behavior. We first explore the role of school context by comparing the racial compositions of several groups of referred and non-referred students. The results are shown in Figure 4.

We find that Black and Hispanic students are overrepresented among students who were referred by “top referrers” compared to the other three groups. First, not surprisingly, there

are clear disproportionalities in the racial composition of students who received ODRs from “top referrers” considering the overall racial composition of the district. More importantly, “top referrers” also referred higher proportions of Black and Hispanic students compared to all referred students and relative to their representation among students in their schools. For example, Black students only account for 7% of all students enrolled in the district but represent close to 22% of all referred students and 27% of students referred by “top referrers”. This confirms our finding that “top referrers” widen racial gaps in ODRs. While we find that “top referrers” indeed teach at schools with slightly higher proportions of Black and Hispanic students, the differences between the racial compositions of their schools and the district seem to explain only some, but not all of the differences between the racial composition of the students who they referred and all referred students. We now turn to examining how referrers’ individual characteristics might explain their referring frequency.

In Table 1, Columns (1) and (2) present results from regressions using the overall sample, and the remaining columns are from school-level specific samples. The first row reports the means of the outcome variables for each sample. There are notable school-level differences in ODR activity; the majority of teachers (70%) made at least one ODR at the middle school level, while fewer than half of teachers do so in elementary and high school.

Holding everything else equal, we do not find any statistically significant differences on teachers’ referring behavior by gender, but we find clear evidence that teachers of color are much less likely to be a referrer or a “top referrer” compared with their White colleagues. For example, across all school levels, a Black teacher’s likelihood of being a referrer is 4.6 percentage points lower than a White teacher’s. This number is 2.1% when comparing likelihoods of being a “top referrer”. Hispanic teachers are also significantly less likely to refer students or be a “top referrer” than White teachers—these coefficients are 1/3 to 1/2 of the magnitudes compared with the coefficients for Black teachers. While Asian teachers have even a lower probability of referring students compared with Black and Hispanic teachers, they are equally as likely to be a “top referrer” as their White colleagues when they do make

ODRs, suggesting varied referral behavior among Asian teachers. Teachers' credentials seem to also play a role. In particular, teachers who hold single subject credentials in English and math are both more likely to be a referrer and a "top referrer" compared with teachers with other credential types. The patterns of the results are not consistent in school-level specific regressions, suggesting school levels might confound the results on credentials.

Years of experience seems to be a particularly salient predictor that holds across samples. Specifically, as teachers accumulate more years of teaching experience, especially after three years when they receive tenure, their likelihood of being a referrer or a "top referrer" quickly drops. Notably, at the middle school level where "top referrers" are the most prevalent, extra years of experience does not reduce a teacher's likelihood of being a referrer until they reach at least 11 years of experience. However, any additional experience does greatly reduce the likelihood of being a "top referrer". Together, these results suggest that teachers who are White, early-career, and who serve middle schools are the ones who engage more in extensive referring.

6 Conclusion and Discussion

This study contributes to the literature on racial disparities in school discipline exposure by considering the degree to which a small group of teachers drives *differential selection* into school discipline by extensively using ODRs following an act they identify as misbehavior. Using exceptionally detailed administrative data from a large and diverse urban school district in California, we find that "top referrers", defined as the top 5% of teachers based on their annual ODR numbers, effectively double the racial gaps in ODRs for both Black-White and Hispanic-White comparisons. These gaps are mainly driven by higher numbers of ODRs issued for Black and Hispanic students due to more subjective reasons like interpersonal offenses and defiance. These ODRs also partially, although not entirely, convert to racial gaps in suspensions. Both the level and racial compositions of the school sites where "top refer-

ers” serve and their personal and professional traits seem to explain some of their extensive referring behavior.

As one of the first studies that examines the variability in teachers’ referral behavior and its contributions to racial disproportionalities in ODRs and exclusionary discipline, this study serves as a first step to understand the contributions of “top referrers” to these dynamics and how such information can be leveraged to provide them with targeted supports and interventions. In a companion study, we extend this line of inquiry by causally estimating how teachers’ qualifications and personal traits affect their referring frequency and more importantly, which students they refer (Hayes et al., 2022). Together, our hope is this strand of work can help further the knowledge on educators’ role in disciplinary disparities and identify effective strategies to combat them.

The approach focused on examining extensive referral behavior provides a tangible and transparent policy tool for school leaders in any district. In our district partner’s context, this group of educators contribute a meaningful proportion of the overall racial disproportionality in ODRs. At the same time, it also represents a relatively manageable number of educators to receive focused supports in ways that may adjust referral behavior. In other contexts, different proportions of individuals that the district feels it can effectively support may be identified through a similar approach. This approach may also help to identify school contexts or position types where individuals engaging in extensive referral behavior may be concentrated, highlighting where structural supports at the school level, subject area, or career stage may also be warranted.

References

- Allen, Q. (2017). “They Write Me Off and Don’t Give Me a Chance to Learn Anything”: Positioning, Discipline, and Black Masculinities in School. *Anthropology & Education Quarterly*, 48(3), 269–283. (.eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/aeq.12199>) doi: 10.1111/aeq.12199
- American Psychological Association. (2008, December). Are zero tolerance policies effective in the schools?: an evidentiary review and recommendations. *American Psychologist*, 63(9), 852–862. doi: 10.1037/0003-066X.63.9.852
- Anderson, K., & Ritter, G. (2016). *Disparate Use of Exclusionary Discipline: Evidence on Inequities in School Discipline from a U.S. State* (SSRN Scholarly Paper No. 2838464). Rochester, NY: Social Science Research Network. Retrieved 2022-06-14, from <https://papers.ssrn.com/abstract=2838464>
- Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). *The school to prison pipeline: Long-run impacts of school suspensions on adult crime* (Tech. Rep.). National Bureau of Economic Research.
- Barrett, N., McEachin, A., Mills, J. N., & Valant, J. (2021). Disparities and discrimination in student discipline by race and family income. *Journal of Human Resources*, 56(3), 711–748. (ISBN: 0022-166X publisher: University of Wisconsin Press)
- Bell, C., & Puckett, T. (2020). I want to learn but they won’t let me: Exploring the impact of school discipline on academic achievement. *Urban Education*, 0042085920968629. (publisher: SAGE Publications Sage CA: Los Angeles, CA)
- Blake, J. J., Butler, B. R., Lewis, C. W., & Darenbourg, A. (2011). Unmasking the Inequitable Discipline Experiences of Urban Black Girls: Implications for Urban Educational Stakeholders. *Urban Rev*, 43(1), 90–106. doi: 10.1007/s11256-009-0148-8

- Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology, 102*(2), 508–520. (publisher: American Psychological Association) doi: 10.1037/a0018450
- Children's Defense Fund. (1975). School suspensions: Are they helping children. *Cambridge, MA: Washington Research Project.*
- Chu, E. M., & Ready, D. D. (2018). Exclusion and urban public high schools: Short-and long-term consequences of school suspensions. *American Journal of Education, 124*(4), 479–509. (ISBN: 0195-6744 publisher: University of Chicago Press Chicago, IL)
- Davison, M., Penner, A. M., & Penner, E. K. (2019). Restorative for all? Racial disproportionality and school discipline under restorative justice. *American Educational Research Journal, 00028312211062613*. (ISBN: 0002-8312 publisher: SAGE Publications Sage CA: Los Angeles, CA)
- Davison, M., Penner, A. M., Penner, E. K., Pharris-Ciurej, N., Porter, S. R., Rose, E. K., . . . Yoo, P. (2021). School Discipline and Racial Disparities in Early Adulthood. *Educational Researcher, 0013189X211061732*. (ISBN: 0013-189X publisher: SAGE Publications Sage CA: Los Angeles, CA)
- Executive Office of the President. (2016). *The continuing need to rethink discipline* (Tech. Rep.). Retrieved 2022-06-26, from <https://eric.ed.gov/?q=source%3A%22Executive+Office+of+the+President%22&id=ED575980>
- Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions. *Yale University Child Study Center, 9*(28), 1–16.

- Girvan, E. J., Gion, C., McIntosh, K., & Smolkowski, K. (2017). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School psychology quarterly*, *32*(3), 392. (ISBN: 1939-1560 publisher: Educational Publishing Foundation)
- Girvan, E. J., McIntosh, K., & Santiago-Rosario, M. R. (2021). Associations Between Community-Level Racial Biases, Office Discipline Referrals, and Out-of-School Suspensions. *School Psychology Review*, *50*(2-3), 288–302. (publisher: Routledge eprint: <https://doi.org/10.1080/2372966X.2020.1838232>) doi: 10.1080/2372966X.2020.1838232
- Gottfredson, G. D., Gottfredson, D. C., Payne, A. A., & Gottfredson, N. C. (2005). School Climate Predictors of School Disorder: Results from a National Study of Delinquency Prevention in Schools. *Journal of Research in Crime and Delinquency*, *42*(4), 412–444. (publisher: SAGE Publications Inc) doi: 10.1177/0022427804271931
- Gregory, A., Huang, F. L., Anyon, Y., Greer, E., & Downing, B. (2018). An examination of restorative interventions and racial equity in out-of-school suspensions. *School Psychology Review*, *47*(2), 167–182. (ISBN: 2372-966X publisher: National Association of School Psychologists 4340 East West Highway, Suite ...)
- Gregory, A., & Roberts, G. (2017). Teacher Beliefs and the Overrepresentation of Black Students in Classroom Discipline. *Theory Into Practice*, *56*(3), 187–194. (publisher: Routledge eprint: <https://doi.org/10.1080/00405841.2017.1336035>) doi: 10.1080/00405841.2017.1336035
- 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. (publisher: American Educational Research Association) doi: 10.3102/0013189X09357621
- Hashim, A. K., Strunk, K. O., & Dhaliwal, T. K. (2018). Justice for all? Suspension bans and restorative justice programs in the Los Angeles Unified School District. *Peabody Journal of Education*, *93*(2), 174–189. (ISBN: 0161-956X publisher: Taylor & Francis)

- Hayes, M., Liu, J., & Gershenson, S. (2022). Who refers whom? Variation in teachers' proclivities to make disciplinary referrals. *Working Paper*.
- Heilbrun, A., Cornell, D., & Konold, T. (2018). Authoritative school climate and suspension rates in middle schools: Implications for reducing the racial disparity in school discipline. *Journal of School Violence, 17*(3), 324–338. (publisher: Routledge eprint: <https://doi.org/10.1080/15388220.2017.1368395>) doi: 10.1080/15388220.2017.1368395
- Holt, S., Vinopal, K., Choi, H., & Sorensen, L. (2022). *Strictly speaking: Examining teacher use of punishment and student outcomes* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2022-06-20, from <https://papers.ssrn.com/abstract=4093118> doi: 10.2139/ssrn.4093118
- Hwang, N. (2018). Suspensions and Achievement: Varying Links by Type, Frequency, and Subgroup. *Educational Researcher, 47*(6), 363–374. (publisher: American Educational Research Association) doi: 10.3102/0013189X18779579
- Hwang, N., Emily K. Penner, Miles Davison, Tanya Sanabria, Paul Hanselman, Thurston Domina, & Andrew M. Penner. (2022). Reining in Punitive Discipline: Recent Trends in Exclusionary School Discipline Disparities. *Socius: Sociological Research for a Dynamic World*. doi: doi.org/10.1177/23780231221103044
- Kim, C. Y. (2011). Policing School Discipline. *Brook. L. Rev., 77*(3), 861–904.
- Kinsler, J. (2011). Understanding the black–white school discipline gap. *Economics of Education Review, 30*(6), 1370–1383. doi: 10.1016/j.econedurev.2011.07.004
- Lacoe, J., & Steinberg, M. P. (2019). Do suspensions affect student outcomes? *Educational Evaluation and Policy Analysis, 41*(1), 34–62. (ISBN: 0162-3737 publisher: SAGE Publications Sage CA: Los Angeles, CA)

- Lewis, A. E., & Diamond, J. B. (2015). *Despite the best intentions: How racial inequality thrives in good schools*. Oxford University Press.
- Lindsay, C. A., & Hart, C. M. D. (2017). Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina. *Educational Evaluation and Policy Analysis*, *39*(3), 485–510. (publisher: American Educational Research Association) doi: 10.3102/0162373717693109
- Liu, J., Hayes, M. S., & Gershenson, S. (2022). JUE insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline. *Journal of Urban Economics*, 103453. (ISBN: 0094-1190 publisher: Elsevier)
- Losen, D. J., Hodson, C. L., Keith, I. I., & MA, M. (2015). *Are we closing the school discipline gap* (Tech. Rep.). Los Angeles: UCLA: The Civil Rights Project/Proyecto Derechos Civiles.
- McCarthy, J. D., & Hoge, D. R. (1987). The social construction of school punishment: Racial disadvantage out of universalistic process. *Social Forces*, *65*(4), 1101–1120. (ISBN: 1534-7605 publisher: The University of North Carolina Press)
- 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. (eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pits.10020>) doi: 10.1002/pits.10020
- Mittleman, J. (2018). A downward spiral? Childhood suspension and the path to juvenile arrest. *Sociology of Education*, *91*(3), 183–204. (ISBN: 0038-0407 publisher: Sage Publications Sage CA: Los Angeles, CA)
- Neal, L. V. I., McCray, A. D., Webb-Johnson, G., & Bridgest, S. T. (2003). The Effects of African American Movement Styles on Teachers' Perceptions and Reactions. *J Spec Educ*, *37*(1), 49–57. (publisher: SAGE Publications Inc) doi: 10.1177/00224669030370010501

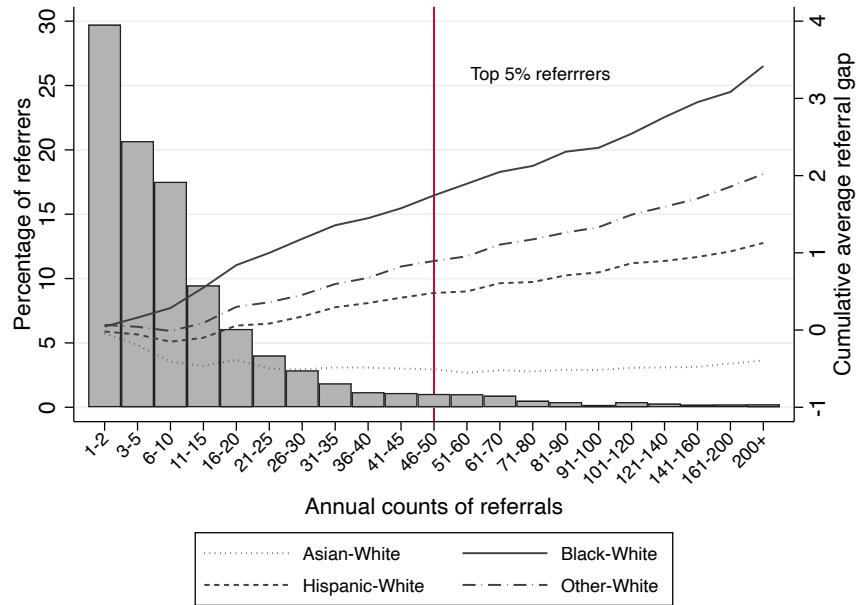
- Noltemeyer, A. L., & Mcloughlin, C. S. (2010). Patterns of Exclusionary Discipline by School Typology, Ethnicity, and Their Interaction. *Penn GSE Perspectives on Urban Education*, 7(1), 27–40. (publisher: University of Pennsylvania, Graduate School of Education)
- Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, 44(2), 224–240. (ISBN: 2372-966X publisher: National Association of School Psychologists Bethesda, MD)
- 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., Walton, G. M., & Eberhardt, J. L. (2016). A Vicious Cycle: A Social–Psychological Account of Extreme Racial Disparities in School Discipline. *Perspectives on Psychological Science*, 11(3), 381–398.
- Owens, J. (2020). *Exclusionary Discipline: Racial Disparities in How Educators Evaluate and Sanction Misbehavior*. San Francisco, CA.
- Owens, J., & McLanahan, S. S. (2020). Unpacking the Drivers of Racial Disparities in School Suspension and Expulsion. *Social Forces*, 98(4), 1548–1577. doi: 10.1093/sf/soz095
- Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are achievement gaps related to discipline gaps? Evidence from national data. *Aera Open*, 5(4), 2332858419875440. (ISBN: 2332-8584 publisher: SAGE Publications Sage CA: Los Angeles, CA)
- Piquero, A. R. (2008). Disproportionate Minority Contact. *The Future of Children*, 18(2), 59–79. (publisher: Princeton University)
- Rios, V. M. (2011). *Punished: Policing the Lives of Black and Latino Boys*. NYU Press.

- Santiago-Rosario, M. R., Whitcomb, S. A., Pearlman, J., & McIntosh, K. (2021). Associations between teacher expectations and racial disproportionality in discipline referrals. *Journal of School Psychology, 85*, 80–93. doi: 10.1016/j.jsp.2021.02.004
- Shedd, C. (2015). *Unequal City: Race, Schools, and Perceptions of Injustice*. Russell Sage Foundation.
- 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. (ISBN: 2372-966X publisher: Taylor & Francis)
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The urban review, 34*(4), 317–342. (ISBN: 1573-1960 publisher: Springer)
- Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N. T., & Horner, R. H. (2016). Vulnerable Decision Points for Disproportionate Office Discipline Referrals: Comparisons of Discipline for African American and White Elementary School Students. *Behavioral Disorders, 41*(4), 178–195. Retrieved 2022-06-21, from <https://doi.org/10.17988/bedi-41-04-178-195.1> (Publisher: SAGE Publications Inc) doi: 10.17988/bedi-41-04-178-195.1
- Sorensen, L. C., Bushway, S. D., & Gifford, E. J. (2022). Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes. *Education Finance and Policy, 17*(2), 255–284. doi: 10.1162/edfp_a.00341
- U.S. Department of Education. (2014). *Dear colleague letter on the nondiscriminatory administration of school discipline* [Letters (Correspondence)]. Retrieved 2022-06-14, from <https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html> (U.S. Department of Education (ED))

- U.S. Department of Education. (2016a). securing equal educational opportunity: Report to the president and secretary of education under section 203(b)(1) of the department of education organization act. , 44.
- U.S. Department of Education. (2016b). *Dear Colleague*. Retrieved 2022-06-14, from <https://www2.ed.gov/policy/gen/guid/school-discipline/files/dcl-on-pbis-in-ieps--08-01-2016.pdf>
- Vavrus, F., & Cole, K. (2002). “I didn’t do nothin’”: The discursive construction of school suspension. *The Urban Review*, *34*(2), 87–111. (ISBN: 1573-1960 publisher: Springer)
- Welch, K., & Payne, A. A. (2010). Racial Threat and Punitive School Discipline. *Social Problems*, *57*(1), 25–48. doi: 10.1525/sp.2010.57.1.25
- Wu, S.-C., Pink, W., Crain, R., & Moles, O. (1982). Student suspension: A critical reappraisal. *Urban Rev*, *14*(4), 245–303. doi: 10.1007/BF02171974
- Zimmerman, R. S., Khoury, E. L., Vega, W. A., Gil, A. G., & Warheit, G. J. (1995). Teacher and parent perceptions of behavior problems among a sample of African American, Hispanic, and non-Hispanic white students. *Am J Commun Psychol*, *23*(2), 181–197. doi: 10.1007/BF02506935

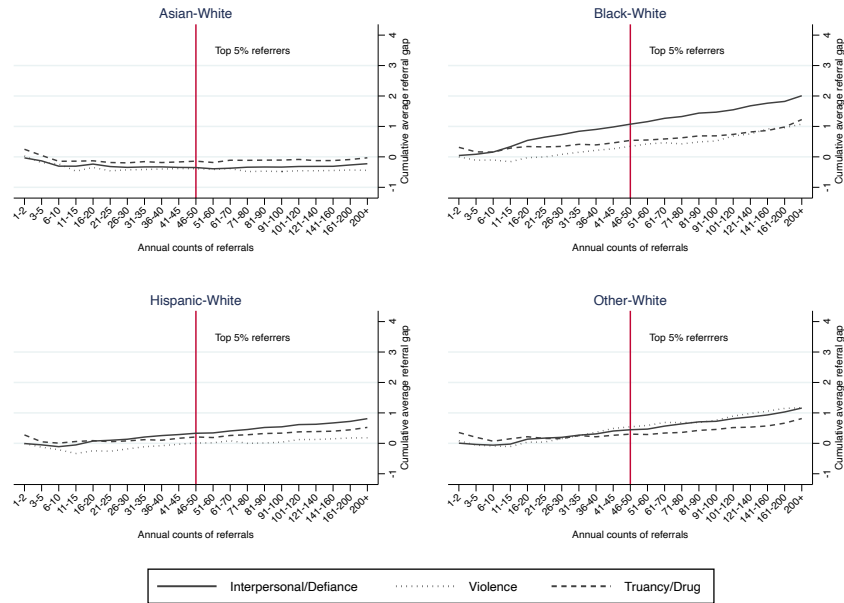
Figures and Tables

Figure 1: Distribution of Referring Frequency & The Corresponding Cumulative Referral Gaps



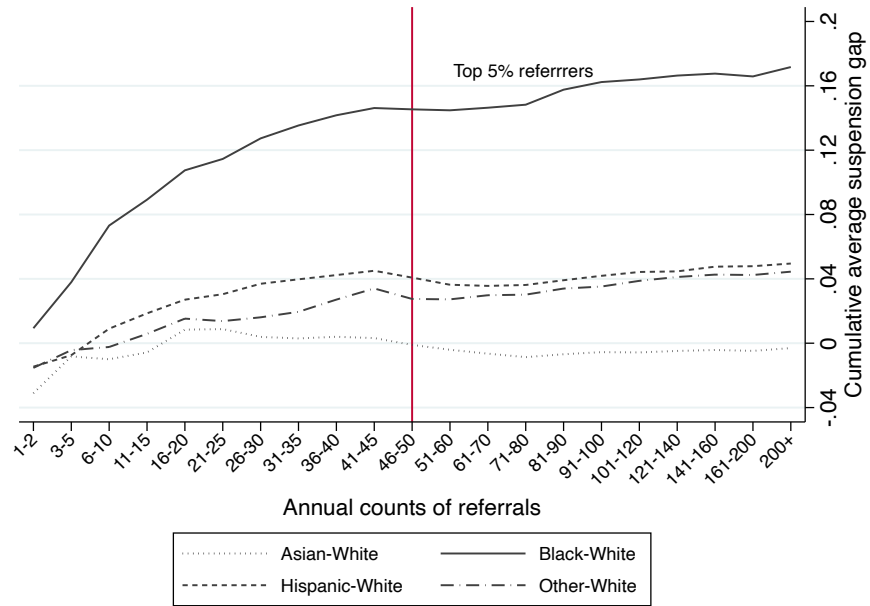
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The percentage of referrers represents the four-year average of the percentages of referrers in a given referral group. The cumulative average referral gap is the four-year average of the percentage differences between students of other races and White students (the reference group) who received at least one referral in a given referral group. Our referral group is formed by the counts of referrals at teacher-by-year level. The average number of referrals issued by top 5% referrers is 48.25 across years, falling into the 46-50 referral group.

Figure 2: Distribution of Referring Frequency & The Corresponding Cumulative Referral Gaps: by Race and Referral Type



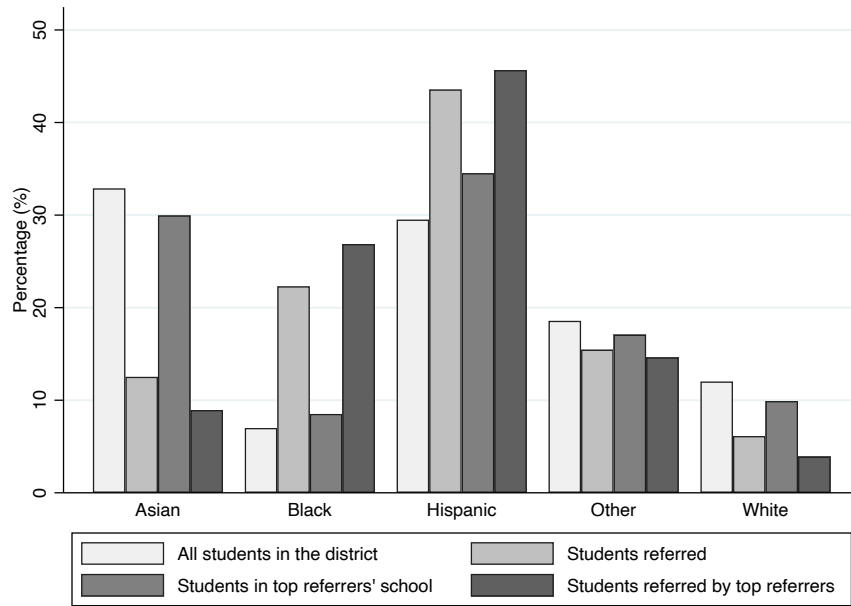
Notes: This figure decomposes the cumulative average referral gap by race and referral type, respectively. Data come from a large school district in California between 2016-17 and 2019-20 school years. For each race/ethnic group, the cumulative average referral gap is the four-year average of the percentage differences between students of a specific race/ethnic group and White students (the reference group) who received at least one referral in a given referral group by referral type. Our referral group is formed by the counts of referrals at teacher-by-year level. Referral types are grouped by their most severe reasons. Specifically, interpersonal offense, class disruption and non-compliance are grouped as Interpersonal/Defiance; violence is a stand-alone group of Violence; drug use, class skipping and other reasons are grouped as Truancy/Drug. The average number of referrals issued by top 5% referrers is 48.25 across years, falling into the 46-50 referral group.

Figure 3: Referring Frequency & The Corresponding Cumulative Suspension Gaps



Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The cumulative average suspension gap is the four-year average of percentage differences between students of other races and White students who received at least one suspension in a given referral group. Our referral group is formed by the counts of referrals at teacher-by-year level. The average number of referrals issued by top 5% referrers is 48.25 across years, falling into the 46-50 referral group.

Figure 4: Comparing Student Racial Representations



Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. Students are counted at student-by-year level. For each racial/ethnic group, this figure shows their percentages among all students in the district, their percentages among all students who are referred at least once, their percentages among all students who were in the top 5% referrers' school, and their percentages among all students who were referred at least once by any of the top 5% referrers.

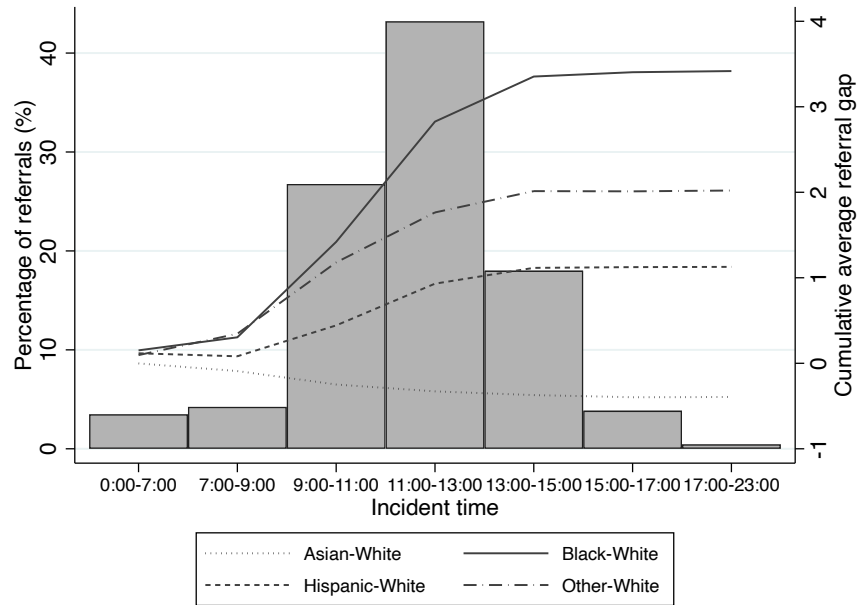
Table 1: Predictors of Being A Referrer or A “Top Referrer”

	Overall		Elementary		Middle		High	
	All	Top	All	Top	All	Top	All	Top
	Referrers	Referrers	Referrers	Referrers	Referrers	Referrers	Referrers	Referrers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Mean	0.342	0.017	0.278	0.013	0.707	0.057	0.425	0.010
Female	0.008	-0.005	-0.002	0.005	-0.030	-0.031**	0.042**	0.001
	(0.010)	(0.003)	(0.014)	(0.003)	(0.023)	(0.012)	(0.019)	(0.004)
Black	-0.046***	-0.021**	-0.048**	-0.012***	-0.128**	-0.074*	0.004	-0.010**
	(0.016)	(0.008)	(0.024)	(0.004)	(0.047)	(0.034)	(0.028)	(0.005)
Hispanic	-0.022**	-0.008**	-0.019	-0.002	-0.039	-0.035*	-0.024	-0.009*
	(0.011)	(0.004)	(0.016)	(0.005)	(0.039)	(0.019)	(0.024)	(0.005)
Asian	-0.055***	0.001	-0.047***	0.001	-0.107***	-0.005	-0.064*	-0.001
	(0.014)	(0.003)	(0.016)	(0.005)	(0.033)	(0.011)	(0.034)	(0.003)
Missing Race	-0.019*	-0.006*	-0.002	-0.002	-0.061	-0.040***	-0.010	0.002
	(0.011)	(0.003)	(0.020)	(0.004)	(0.035)	(0.013)	(0.019)	(0.004)
Other Race	-0.014	-0.003	0.004	0.009	-0.055	-0.027	-0.016	-0.001
	(0.019)	(0.006)	(0.034)	(0.010)	(0.044)	(0.016)	(0.045)	(0.011)
Cred in ELL	0.019**	0.001	-0.007	0.002	0.067***	-0.000	0.045**	0.002
	(0.009)	(0.003)	(0.015)	(0.005)	(0.015)	(0.008)	(0.021)	(0.005)
Cred in SPECED	-0.008	-0.006*	-0.042**	-0.004	-0.043	-0.027**	0.039	-0.000
	(0.018)	(0.004)	(0.019)	(0.006)	(0.026)	(0.009)	(0.036)	(0.002)
Cred in English	0.053***	0.013**	-0.038	-0.002	0.056	0.044**	0.102***	0.007
	(0.017)	(0.005)	(0.027)	(0.006)	(0.042)	(0.018)	(0.030)	(0.004)
Cred in Math	0.058***	0.012*	-0.003	-0.013***	0.084**	0.023	0.093***	0.016
	(0.017)	(0.007)	(0.025)	(0.005)	(0.033)	(0.013)	(0.029)	(0.013)
Cred in Science	0.083***	0.012	0.066	0.013	0.125***	0.019	0.116***	0.013
	(0.024)	(0.008)	(0.053)	(0.014)	(0.018)	(0.024)	(0.035)	(0.009)
Temporary	-0.075***	-0.020***	-0.091***	-0.020**	-0.017	-0.037*	-0.095***	-0.020
	(0.014)	(0.006)	(0.024)	(0.008)	(0.040)	(0.021)	(0.027)	(0.012)
Years 2-3	-0.006	-0.011*	-0.011	-0.002	0.018	-0.054**	0.028	0.006
	(0.016)	(0.006)	(0.024)	(0.010)	(0.022)	(0.018)	(0.032)	(0.007)
Years 4-5	-0.041***	-0.013	-0.042*	-0.006	-0.000	-0.052*	-0.043	-0.009
	(0.014)	(0.009)	(0.025)	(0.013)	(0.030)	(0.025)	(0.034)	(0.019)
Years 6-10	-0.077***	-0.029***	-0.077***	-0.024*	-0.022	-0.069***	-0.086**	-0.025
	(0.018)	(0.009)	(0.027)	(0.013)	(0.027)	(0.022)	(0.039)	(0.015)
Years 11-30	-0.093***	-0.030***	-0.083***	-0.023*	-0.066*	-0.073**	-0.081***	-0.024
	(0.019)	(0.008)	(0.026)	(0.012)	(0.035)	(0.026)	(0.029)	(0.015)
Controls for:								
School FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.359	0.083	0.295	0.046	0.137	0.107	0.313	0.070
Observations	17026	17026	7430	7430	2407	2407	4179	4179

Notes: Clustered-robust standard errors at the school level are in parentheses. Data come from a large school district in California between 2016-17 and 2019-20 school years. The unit of analysis is at teacher-by-year level. The “other” race category includes multiracial individuals and teachers missing race data. All the statistics above are reported as proportions. The omitted group is male, White, tenured or nontenured referrers with one year of teaching experience. Columns 1, 3, 5, 7 include referrers making at least one referral. $p < 0.10^*$ $p < 0.05^{**}$ $p < 0.01^{***}$.

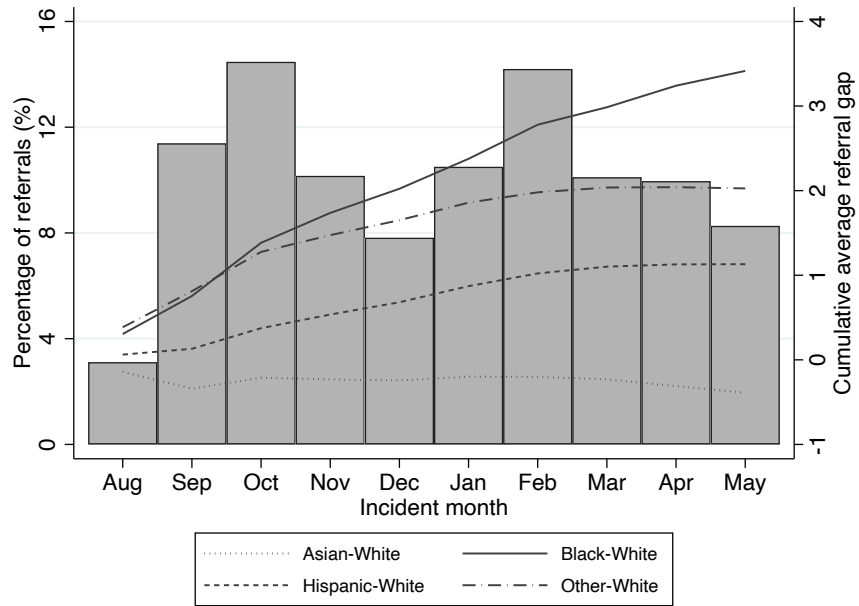
Appendix A

Figure A1: Distribution of referrals by incident time



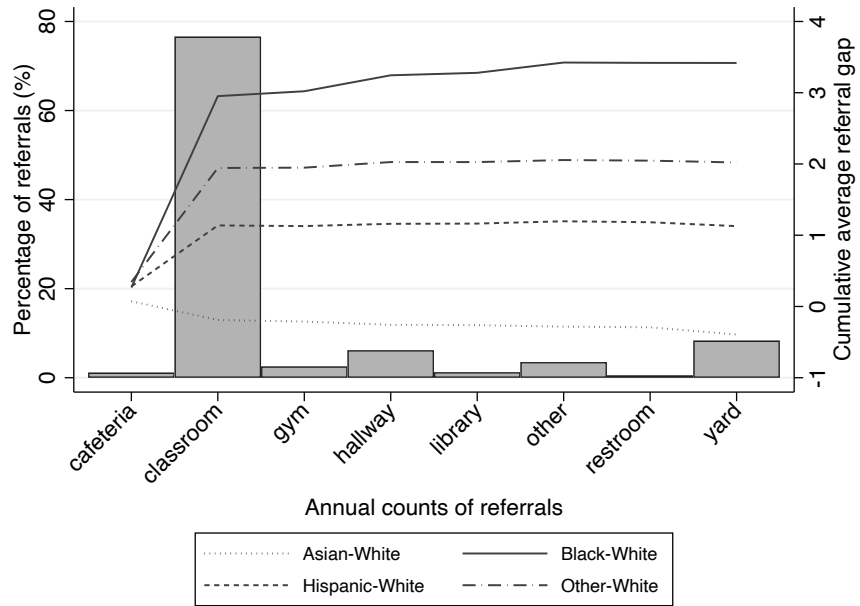
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. This figure shows the four-year average of the percentages of referrals by incident time.

Figure A2: Distribution of referrals by month



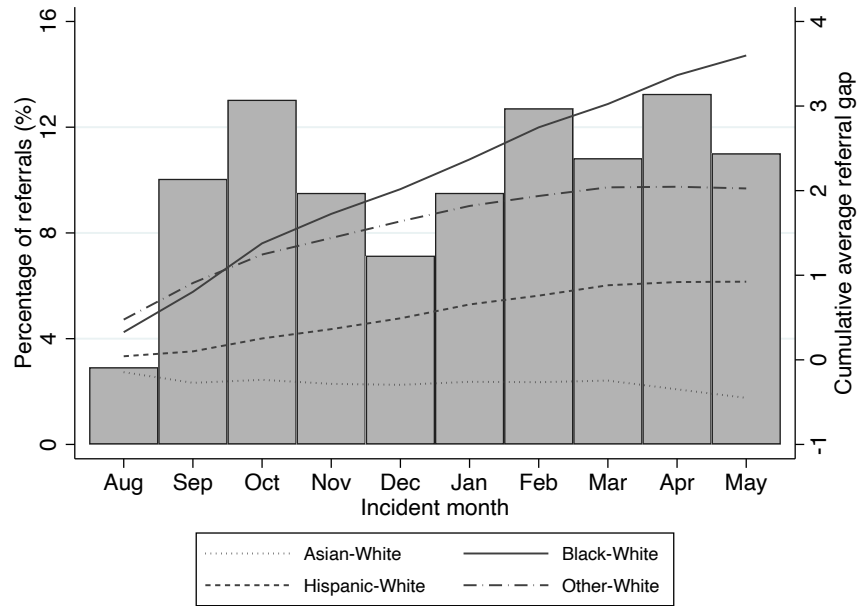
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. This figure shows the four-year average of the percentages of referrals by month. Due to summer holidays, June and July are excluded when producing this figure.

Figure A3: Distribution of referrals by location



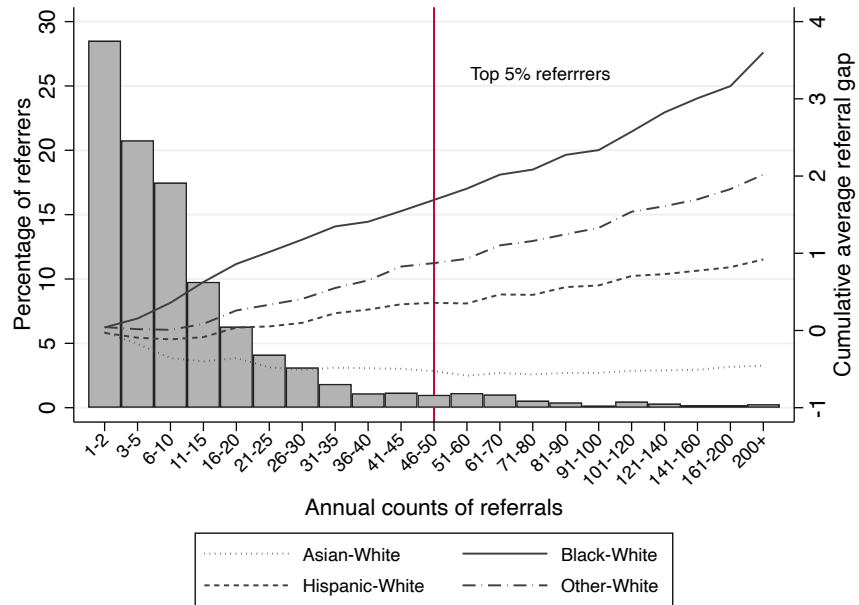
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. This figure shows the four-year average of the percentages of referrals by location.

Figure A4: Distribution of referrals by month (Year 2020 Excluded)



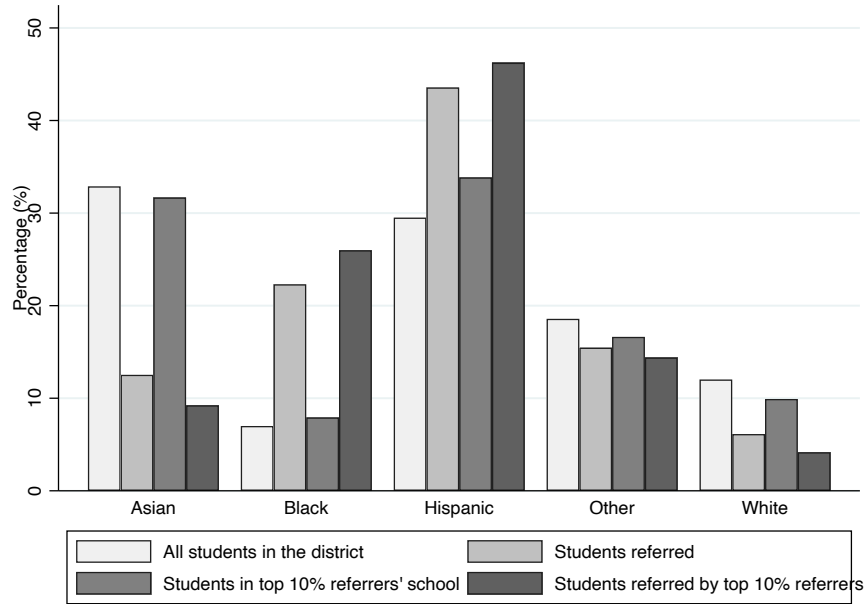
Notes: Data come from a large school district in California between 2016-17 and 2018-19 school years. This figure shows the three-year average of the percentages of referrals by month. Due to summer holidays, June and July are excluded when producing this figure.

Figure A5: Distribution of Referring Frequency & The Corresponding Cumulative Referral Gap (Year 2020 excluded)



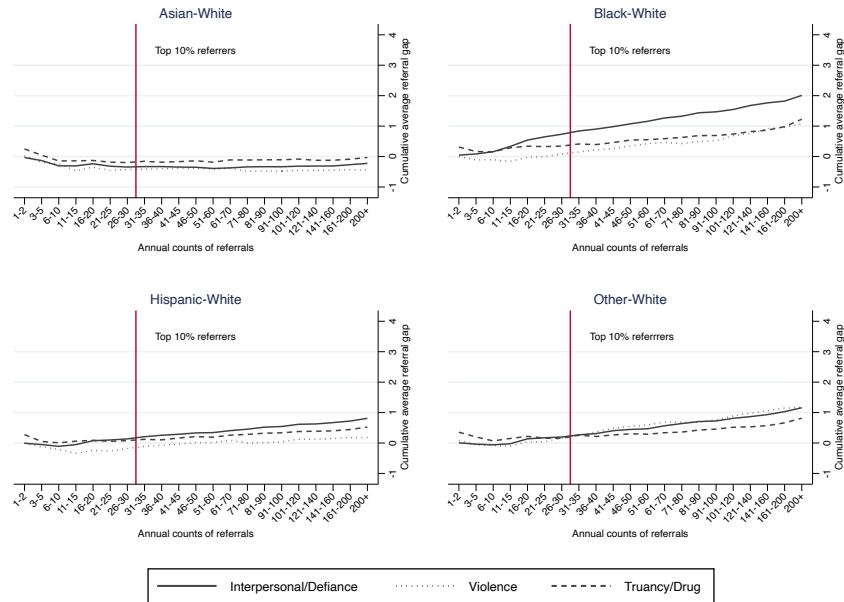
Notes: Data come from a large school district in California between 2016-17 and 2018-19 school years. The percentage of referrers represents the three-year average of the percentages of referrers in a given referral group. The cumulative average referral gap is the three-year average of the percentage differences between students of other races and White students (the reference group) who received at least one referral in a given referral group. Our referral group is formed by the counts of referrals at teacher-by-year level. The average number of referrals issued by top 5% referrers is 48.25 across years, falling into the 46-50 referral group.

Figure A6: Comparing Student Racial Representations



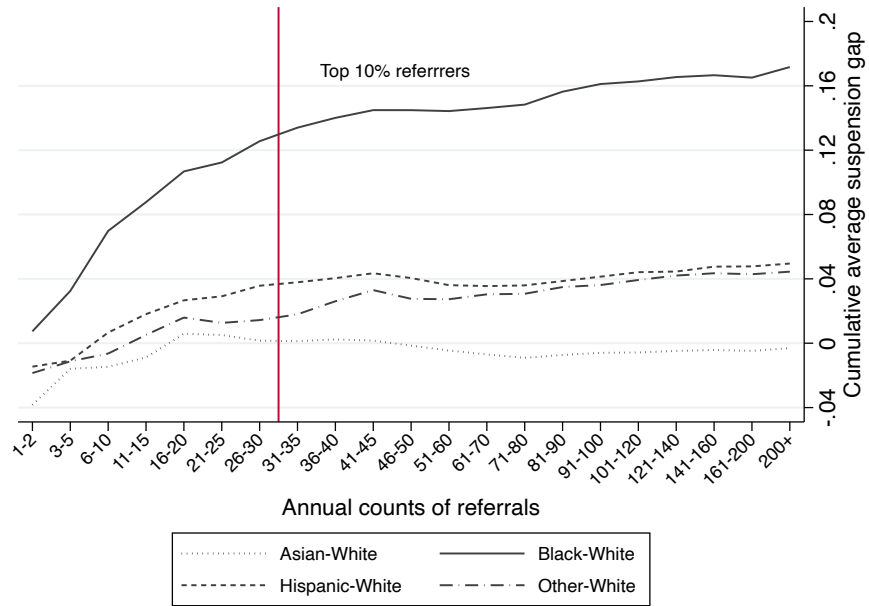
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. Students are counted at student-by-year level. For each racial/ethnic group, this figure shows their percentages among all students in the district, their percentages among all students who are referred at least once, their percentages among all students who were in the top 10% referrers' school, and their percentages among all students who were referred at least once by any of the top 10% referrers.

Figure A7: Distribution of Referring Frequency & The Corresponding Cumulative Referral Gap: by Race and Referral Type



Notes: This figure decomposes the cumulative average referral gap by race and referral type, respectively. Data come from a large school district in California between 2016-17 and 2019-20 school years. For each race/ethnic group, the cumulative average referral gap is the four-year average of the percentage differences between students of a specific race/ethnic group and White students (the reference group) who received at least one referral in a given referral group by referral type. Our referral group is formed by the counts of referrals at teacher-by-year level. Referral types are grouped by their most severe reasons. Specifically, interpersonal offense, class disruption and non-compliance are grouped as Interpersonal/Defiance; violence is a stand-alone group of Violence; drug use, class skipping and other reasons are grouped as Truancy/Drug. The average number of referrals issued by top 10% referrers is 30.50 across years, falling in-between the 26-30 and 31-35 referral group.

Figure A8: Referring Frequency & The Corresponding Cumulative Suspension Gap



Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The cumulative average suspension gap is the four-year average of percentage differences between students of other races and White students who received at least one suspension in a given referral group. Our referral group is formed by the counts of referrals at teacher-by-year level. The average number of referrals issued by top 10% referrers is 30.50 across years, falling in-between the 26-30 and 31-35 referral group.

Table A1: Predictors of Being A Referrer or A Top Referrer (Year 2020 Excluded)

	Overall		Elementary		Middle		High	
	All	Top	All	Top	All	Top	All	Top
	Referrers	Referrers	Referrers	Referrers	Referrers	Referrers	Referrers	Referrers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Mean	0.341	0.017	0.261	0.014	0.732	0.056	0.436	0.010
Female	0.013	-0.005	0.006	0.005	-0.018	-0.031*	0.042**	-0.001
	(0.010)	(0.004)	(0.015)	(0.004)	(0.023)	(0.015)	(0.020)	(0.004)
Black	-0.052***	-0.022**	-0.047*	-0.014**	-0.096*	-0.063	-0.043	-0.012*
	(0.017)	(0.010)	(0.028)	(0.005)	(0.051)	(0.044)	(0.032)	(0.007)
Hispanic	-0.017	-0.004	-0.016	0.000	-0.043	-0.021	-0.030	-0.005
	(0.013)	(0.004)	(0.017)	(0.005)	(0.034)	(0.023)	(0.028)	(0.006)
Asian	-0.052***	0.002	-0.048**	0.000	-0.111***	0.001	-0.067	0.000
	(0.016)	(0.004)	(0.018)	(0.005)	(0.033)	(0.017)	(0.040)	(0.004)
Missing Race	-0.016	-0.003	-0.017	-0.000	-0.047	-0.030**	0.002	0.005
	(0.012)	(0.003)	(0.021)	(0.005)	(0.038)	(0.010)	(0.021)	(0.006)
Other Race	-0.020	-0.001	0.014	0.003	-0.076	-0.002	-0.011	-0.002
	(0.021)	(0.006)	(0.037)	(0.011)	(0.044)	(0.016)	(0.051)	(0.013)
Cred in ELL	0.017	-0.001	-0.003	-0.001	0.048***	-0.006	0.051*	0.001
	(0.010)	(0.003)	(0.016)	(0.005)	(0.011)	(0.008)	(0.026)	(0.006)
Cred in SPECED	-0.024	-0.007**	-0.055***	-0.004	-0.075**	-0.023***	0.005	-0.003
	(0.021)	(0.003)	(0.021)	(0.005)	(0.029)	(0.006)	(0.037)	(0.004)
Cred in English	0.050***	0.012**	-0.033	-0.004	0.037	0.042**	0.091***	0.006*
	(0.019)	(0.005)	(0.029)	(0.006)	(0.050)	(0.018)	(0.029)	(0.003)
Cred in Math	0.050***	0.011	-0.016	-0.014***	0.055	0.028	0.089***	0.012
	(0.017)	(0.007)	(0.026)	(0.005)	(0.034)	(0.018)	(0.030)	(0.011)
Cred in Science	0.078***	0.016*	0.070	0.023	0.125***	0.025	0.101**	0.016
	(0.024)	(0.010)	(0.055)	(0.021)	(0.019)	(0.032)	(0.036)	(0.011)
Temporary	-0.072***	-0.021***	-0.088***	-0.021**	-0.009	-0.037	-0.093**	-0.022
	(0.016)	(0.008)	(0.027)	(0.010)	(0.040)	(0.031)	(0.034)	(0.013)
Years 2-3	-0.008	-0.014*	-0.010	-0.003	0.010	-0.073***	0.015	0.007
	(0.015)	(0.008)	(0.027)	(0.011)	(0.021)	(0.023)	(0.033)	(0.011)
Years 4-5	-0.038**	-0.017	-0.026	-0.006	0.004	-0.059	-0.069*	-0.020
	(0.015)	(0.011)	(0.029)	(0.014)	(0.023)	(0.034)	(0.036)	(0.023)
Years 6-10	-0.072***	-0.033***	-0.069**	-0.027*	0.000	-0.079**	-0.092**	-0.029
	(0.019)	(0.010)	(0.030)	(0.014)	(0.031)	(0.031)	(0.034)	(0.017)
Years 11-30	-0.091***	-0.034***	-0.079***	-0.024*	-0.056	-0.082**	-0.085***	-0.031
	(0.020)	(0.010)	(0.028)	(0.014)	(0.034)	(0.033)	(0.027)	(0.020)
Controls for:								
School FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.383	0.094	0.315	0.049	0.139	0.133	0.316	0.061
Observations	12749	12749	5566	5566	1801	1801	3120	3120

Notes: Clustered-robust standard errors at the school level are in parentheses. Data come from a large school district in California between 2016-17 and 2018-19 school years. The unit of analysis is at teacher-by-year level. The “other” race category includes multiracial individuals and teachers missing race data. All the statistics above are reported as proportions. The omitted group is male, White, tenured or nontenured referrers with one year of teaching experience. Columns 1, 3, 5, 7 include referrers making at least one referral. $p < 0.10^*$ $p < 0.05^{**}$ $p < 0.01^{***}$.