ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 22-616

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 teachers' initial identification of student misbehavior exacerbates 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 referring behavior, or the top 5 percent 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: February 2023

Suggested citation: Liu, Jing, Emily Penner, and Wenjing Gao. (2023). 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/d1bx-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 exacerbates 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 referring behavior, or the top 5 percent 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 and the Association for Public Policy Analysis and Management Conference 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 the frequency of office discipline referrals (ODRs), 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., 2011a). Exclusionary discipline has been associated with a host of negative academic impacts and school disengagement for disciplined students (Bell & Puckett, 2020; Chu & Ready, 2018; Hwang, 2018; Noltemeyer et al., 2015; Sorensen et al., 2022) and has widened racial gaps in absences and student achievement (e.g., Lacoe & Steinberg, 2019; Pearman et al., 2019). Racial inequities in the 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 in school discipline underscore the need to identify their root causes.

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; Davison et al., 2022; Executive Office of the President, 2016; Gregory et al., 2018; Hashim et al., 2018; Hwang et al., 2022; Nishioka et al., 2021; Noltemeyer & Mcloughlin, 2010; U.S. Department of Education, 2014, 2016b; Zakszeski et al., 2021). We thus argue that further reductions in racial disparities in exclusionary discipline hinge not just on formal programmatic and policy changes but also on understanding the specific actions

of educators. Research examining educators' roles 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 ODRs (Girvan et al., 2017; Santiago-Rosario et al., 2021; Skiba et al., 2002, 2011a). Recent evidence indicates variability in the use of ODRs by teachers (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 referring behavior.

We extend this literature by examining the variability in teachers' use of ODRs and considering whether the extensive referring behavior by a small number of teachers widens racial gaps in discipline receipt. We use four years of teacher-ODR linked data from an urban school district to identify the top 5 percent of referrers (hereafter "top-referrers") and their contributions to racial ODR and suspension gaps. We also test whether ODRs for more subjective offences (e.g., defiance versus weapons) contribute more to racial ODR and suspension 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 interventions 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, Walton, & Eberhardt, 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

(Okonofua & Eberhardt, 2015; Shedd, 2015; Vavrus & Cole, 2002). The construction and enactment of school rules can infuse racialized dynamics into decisions about who is selected for punishment (because a behavior they commit is perceived as needing discipline) and who is formally processed (i.e., punished) and how severely (Lewis & Diamond, 2015; Piquero, 2008; Skiba et al., 2002). [1] This combination of *differential selection* for discipline and *differential processing* of discipline by race is hypothesized to exacerbate the observed differential outcomes in discipline among racial and ethnic groups.

The existing literature has foregrounded investigations of racial disparities that emerge from *differential processing* in part because the frequency and severity of exclusionary punishment are routinely documented in district data systems and officially collected by the Office for Civil Rights (Bradshaw et al., 2010; Skiba et al., 2002, 2011a; U.S. Department of Education, 2014). Research examining exclusionary discipline records 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, and even within the same incidents (Allen, 2017; Barrett et al., 2021; Blake et al., 2011; Lewis & Diamond, 2015; Rios, 2011; Sorensen et al., 2022). Differences in teachers' use of exclusionary discipline for minor and subjectively defined behaviors uniquely perpetuate racial discipline gaps (Barrett et al., 2021; Girvan et al., 2017; Skiba et al., 2002).[2]

Yet to understand how disparities in these outcomes emerge, it is also important to attend to how they emerge through *differential selection*, including the observation of student behavior, the perception of a student action as misbehavior, as well as the formal reporting of that misbehavior through an ODR. Black students' disporportionate treatment is evident throughout. Educators are more likely to monitor and surveil the behaviors of Black students

[1] See Rodriguez & Welsh (2022) for framework detailing these processes and Curran (2020); Girvan et al. (2019); Nishioka (2017) for discussions of the various metrics used to examine racial disproportionalities throughout the discipline process.

[2] We follow a long line of literature to distinguish subjectively vs. objectively defined student behaviors (e.g., Girvan et al., 2017; Lindsay & Hart, 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.

than White students, even in early childhood (Zimmermann, 2018; Gansen, 2021). Decades of evidence highlights 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., 2011a; 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), despite evidence documenting few racial differences in misbehavior (American Psychological Association, 2008; Gregory et al., 2010; Losen et al., 2015; McCarthy & Hoge, 1987; Wu et al., 1982). National data also indicate that Black students are two to three times more likely to receive an ODR for problem behavior than their White peers, which then extend to an increased likelihood of receiving exclusionary discipline as a consequence (Skiba et al., 2011b). Others also find evidence that teachers' use of ODRs in response to more subjective incidents expands racial disproportionalities in ODRs and exclusionary discipline(Cook et al., 2018; Gion et al., 2022; Girvan et al., 2017, 2021; Holt et al., 2022; Lindsay & Hart, 2017; Liu et al., 2022; Santiago-Rosario et al., 2021; Skiba et al., 2002, 2011a; Smolkowski et al., 2016).

As teachers are often the initial source of ODRs following a perceived misbehavior in the classroom, they play an active role in *differential selection*. For example, teacher behavior, including classroom management (Cho et al., 2020; Gregory & Weinstein, 2008; Gion et al., 2022; Skiba et al., 2014), and their beliefs, biases, and expectations of students, relate to their likelihood of giving students ODRs and having those ODRs result in exclusionary discipline (Atiles et al., 2017; Gregory & Mosely, 2004; Okonofua & Eberhardt, 2015; Okonofua, Walton, & Eberhardt, 2016; Santiago-Rosario et al., 2021; Skiba et al., 2011a). Teacher background also relates to frequency of ODR use. While both Black and White teachers show evidence of racial bias in ODR use, Black teachers are slightly less likely to give out ODRs for subjective incidents than White teachers (Holt et al., 2022). Having a racially mismatched teacher significantly increases both the number and likelihood of suspensions

students receive (Gershenson et al., 2022; Holt & Gershenson, 2019; Lindsay & Hart, 2017; Shirrell et al., 2021). Likewise, teacher professional characteristics relate to ODR use and impact. Gershenson et al. (2022) find that students are less likely to be referred by more experienced teachers. Holt et al. (2022) reach slightly different conclusions about the relationship between teacher experience and referrals, finding that while teachers differ in their relative use of ODRs, they are fairly consistent in their use of ODRs as they gain experience. Students are also less likely to be referred by teachers who hold either an English language learner authorization or special education credential (Gershenson et al., 2022). The negative consequences of the harshness of ODRs are also slightly larger for students of more effective teachers (as defined by value-added to achievement) (Holt et al., 2022). Thus, while some features of teachers' background and professional qualifications relate to their referral behavior, our understanding of which teachers contribute most to racial disproportionalities in ODR use is incomplete.

School context also contributes to *differential selection* and *differential processing* of discipline incidents (Cruz et al., 2021; Gopalan & Nelson, 2019; Kinsler, 2011; Mendez et al., 2002). For example, the stress of emotionally taxing environments results in what McIntosh et al. (2014) have called Vulnerable Decision Points (VDPs). VDPs increase teachers' susceptibility to implicit biases, increase their likelihood of disproporitonately giving ODRs to students of color, and contribute to overall discipline disproportionalities (Gion et al., 2018; McIntosh et al., 2014; Smolkowski et al., 2016). Scholars have identified several examples of specific types of VDPs that contribute to the differential use of discipline, including incidents that occur in teachers' classrooms, incidents that occur at particular times of day, more subjective incidents, and incidents classified as more severe (Gion et al., 2018; McIntosh et al., 2014; Smolkowski et al., 2016).

Aspects of school procedure and routine also influence discipline disparities. Discipline protocols are enacted inconsistently within and between schools (Griffith & Tyner, 2019). Variability in school leaders' philosophies around school discipline and their relationships

with teachers in addressing discipline incidents influence the severity of their responses (Sorensen et al., 2022; Welsh, 2023). Administrators' biases contribute to their adherence to specific response guidelines and are associated with racial disparities in the severity of the consequences students receive, even for the same behaviors (Jarvis & Okonofua, 2020). Schools' differing logics around behavior for Black and White students, their authoritative climate, and their disciplinary structure cultivate differential responses to student behavior and are associated with different magnitudes of Black-White suspension rate gaps (Heilbrun et al., 2018; Wiley, 2021).

The student demographics of schools also relate to racial disproportionalities in *differential processing*, which ultimately stem from differences in *differential selection*. While Latinx and Native American students are disproportionately given ODRs relative to their White peers, disparities are the most pronounced for African American students (Skiba et al., 2011a; Smolkowski et al., 2016). Schools with higher proportions of Black students, other students of color, and low-income students have higher rates of exclusionary discipline use (Gottfredson et al., 2005; Mendez et al., 2002; Welch & Payne, 2010). At the same time, racialized logics in majority-White schools also contribute to the disproportionate monitoring and punishment of Black students (Wiley, 2021).

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 ODRs and other discipline outcomes. However, data limitations have made it difficult to examine specific teachers' contributions to racial disproportionalities in these early stages of the discipline process. One exception is Holt et al. (2022), who identify the variability in the degree of late-elementary educators' punitiveness and racial bias in the use of ODRs in North Carolina. Specifically, they measure teachers' punitiveness by estimating teachers' contributions to their average students' number of subjective ODRs[3], and teachers' racial bias

[3]The subjective ODRs considered in Holt et al. (2022) include behaviors like tardiness, talking back to teachers, being disrespectful, disruptive behavior, and other infractions that involved subjective interpretation on whether a behavior rises to the level of ODR.

by estimating how teachers' use of ODRs differs by student race/ethnicity. More punitive teachers contribute to adverse academic and behavioral outcomes for both referred students and non-referred students. Importantly, teachers with a racial bias in using ODRs produce strong negative impacts on achievement and attendance for Black students but not White students. These findings suggest that examining the prevalence and distribution of the very punitive teachers and providing them with additional support may be a promising path toward reducing racial disproportionality in ODRs.

Our work builds on Holt et al. (2022) to examine the distribution of ODRs across educators of all school levels and the degree to which extensive referring behavior contributes to racial disparities in ODRs 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 from 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 into 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" compared to all referrers contribute to racial/ethnic disparities in ODR and suspension receipt overall and by incident type?

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

3 Data

We examine how referring behavior contributes to racial disparities in discipline receipt in a large, diverse urban-intensive[4] school district in California. More than half of its residents are people of color from a diverse set of racial and ethnic backgrounds. It is also a city that had a once-thriving Black population that has declined after decades of under-investment in Black communities, rising costs of living, and gentrification.

Our analyses draw on unusually rich administrative data from the 2016-17 through 2019-20 school years. During this period, this district served over 79,000 unique students in grades K-12 in 191 unique schools, totaling 227,922 student-year observations. We identify students' race/ethnicity using district records of students' self-reported racial/ethnic identity following district classification categories. About 33 percent of the students are Asian. Hispanic students account for about 30 percent. Black and White students account for 7 and 12 percent, respectively. 10 percent of students are identified as multi-racial or other races. The remaining 9 percent are missing race/ethnicity information.

One unique feature of the data is that it provides exceptionally detailed ODR records, regardless of the consequences. Specifically, aside from the information on which students are involved in an ODR and the reason for the ODR (i.e., type of incident), the data also contain a unique identifier of the educator who issued the ODR, incident time, date, and location.[5] In addition, the data allow us to identify the disciplinary consequences of each ODR, if administered. Such detailed information allows us to systematically describe educators' referring behavior and map out how such behavior might contribute to racial disproportionalities in both ODRs and suspensions. Further, the district provides us with rich data on educator demographics, job positions, credential area, and years of experience, which we use to characterize educators with differential referring behavior.

[4]Urban Intensive schools are those where "outside of school factors such as housing, poverty, and transportation are directly connected to what happens inside of the school"(Milner, 2012; Welsh & Swain, 2020).

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

During the time span we examine, about 91.4 percent of ODRs were issued by classroom teachers and the rest were from school administrators. 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 referrers, we examine the primary source of ODRs which afford the most opportunities for potential intervention. As a result, our analytic sample comprises 75,229 ODRs that were issued by 2,928 unique teachers (5,855 teacher-year observations).[6] The demographics of "top referrers" are similar to all referrers, with 49 percent White, 5 percent Black, 16 percent Hispanic and 18 percent Asian. 5 percent of teachers are identified as either multi-racial or other races and the remaining 7 percent are missing race/ethnicity information.[7]

The data also provide the reason(s) for which each ODR was issued. A significant number of ODRs are the result of multiple infraction categories. We follow Lindsay & Hart (2017) in making mutually exclusive categories based on the "most severe" reason listed for an ODR, stated here in decreasing order of severity: (a) violence; (b) drugs; (c) interpersonal offenses; (d) disruption or non-compliance; (e) class skipping or walkout; and (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; and (c) truancy/drug. "Interpersonal offenses" and "disruption or non-compliance" are the most prevalent reasons for ODRs (57 percent of total ODRs); "violence" is the second most prevalent reason (28 percent of total ODRs); whereas "drugs", "class skipping", and

[6]We do not include teachers who made zero ODRs in most of our analyses and focus on teachers who made at least one ODR to precisely map teachers' referring behavior to racial gaps in ODRs, and consider only students who were referred. In this way, we can more concretely evaluate the contribution of teachers with different referring frequencies. However, we acknowledge that teachers who made no ODRs can also contribute to racial gaps in ODRs at an extensive margin.

[7]The demographics of the overall teacher population are 45 percent White, 5 percent Black, 15 percent Hispanic and 23 percent Asian. 4 percent of teachers identified themselves as either multi-racial or other races. We are missing race/ethnicity information for the remaining 8 percent.

"other" reasons together comprise the remaining 15 percent.

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 issued at least one ODR in a given year. As the distribution is positively skewed with a long tail, we group teachers into bins with unequal intervals for enhanced data visualization and ease of interpretation. Specifically, the first two bins include teacher-year observations for teachers who 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 rank among the top five percent based on total annual ODRs issued in each of the four years to identify a group of teachers whom 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. 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:

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

where i indicates students and j indicates teachers. $Group_{ij}$ indicates the student's racial or ethnic identity. Y_{ij} indicates 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 "multiracial/other" students who self-identify as multi-racial or other races.[8] 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:

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

We then follow the same method but change Y_{ij} to indicate the number of ODRs the student received from teacher j that converts to a suspension. One caveat here is that given we only consider students who received at least one ODR in calculating the ODR gaps, our approach explicitly focuses on the intensive margin or the raw differences in ODRs received from different racial groups among those who were referred. This is different from other popular measures for discipline gaps such as risk ratio, risk differences, and the raw differential representation (Curran, 2020; Girvan et al., 2019), which require accounting for *all students* who were enrolled in a school or in a district, yielding quite different findings about the magnitudes of discipline gaps. While our measure does not capture certain information about discipline gaps present in these alternative measures, it serves our purpose of precisely mapping each teacher referrer to the corresponding ODR gaps and evaluating their individual contributions.

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 five categories: (1) all students in the district; (2) all students who were referred at least once; (3) all students who were from schools that have at least one "top referrer"; (4) all students who were from "top referrers"' classrooms;

[8]We also use Asian students as the reference group and replicate the main analysis in Appendix Figure A4. For students for whom we are missing racial information, we do not consider them in the main analysis but report their ODR and suspension gaps in Figures A5, A6, A7, and A8.

and (5) all students who were 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 provide additional evidence on whether "top referrers" increased or decreased the overall racial discipline gaps. Comparisons to the racial composition of all students in the district, the school sites, and the classrooms where "top referrers" serve can provide some clues about the working environments of the "top referrers" and how such contexts might explain their referring behavior.

To further probe into the influence of teachers' characteristics to referral behavior, we then conduct regression analyses including all classroom teachers who served the district in the time span we examine in our analysis sample. We use two indicator variables as outcomes–whether a teacher made at least one ODR and whether a teacher is identified as a "top referrer" in a given year. 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 including all school levels. We then estimate the same models separately for elementary, middle, and high schools to examine whether such correlations vary by school level. To test the robustness of this model, we 1) estimate the same model but exclude data from the school year 2019-2020 (Table A2); and 2) account for the impact of principals by controlling for their race, gender, and years of experience (Table A3). The results are similar to our main model.

5 Results

RQ1. Distribution. Among all classroom teachers who served the district during the four years we study, 34.1 percent of them ever made an ODR in a given year. We subsequently exclude teachers who made zero ODRs 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. Specifically, over half of referring teachers (50.4 percent) 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 percent of referrers (1.7 percent of all teachers), which we label as "top referrers," issued over 48 ODRs per year (meaning roughly one ODR every four school days), several times greater than their average-referring colleagues (who issue less than one ODR for every two months of school). While fewer than 80 teachers are *ever* "top referrers" in our sample, they account for 34.8 percent of all ODRs made in this four-year period.[9] Motivated by understanding the implications of this extensive referring behavior by a small group of teachers, we turn to RQ2 to investigate their potential contributions 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 ODR 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 ODR 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 count 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 issued one or two ODRs 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 ODRs), the Black-White gap increases to one ODR, and the multiracial/other-White and Hispanic-White gaps also increase somewhat, although the magnitudes are not as large.

[9]We also use the top 10 percent of referring teachers as the definition for "top referrers," and replicate the main analyses in Appendix Figures A9, A10, and A11 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 ODR gap reaches slightly above three ODRs. This pattern is similar for Hispanic-White and multiracial/other-White gaps, but the growth of these two gaps is not nearly as fast as that of the Black-White gap. In contrast, the Asian-White gap stays relatively stable and is actually negative.

Second, if we combine the information presented in Figure 1, a less readily visible but critical finding is that the top 5 percent of referrers, as identified in RQ1, effectively double the Black-White, Hispanic-White, and multiracial/other-White ODR gaps. For example, the Black-White gap in ODRs is about 1.6 when we consider all referrers with fewer 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. [10]

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 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 arguably more likely to be subject to bias (Lindsay & Hart, 2017). For example, "top referrers" double the Black-White ODR gap in the interpersonal/defiance category (2.0 ODRs), which also account for more than half of the total Black-White ODR gap (3.4 ODRs).

While ODRs, regardless of whether they lead to a suspension, are consequential on their

[10]We consider several alternatives to evaluate the robustness of Figure 1. First, we include administrators in our sample (Appendix Figure A12). As our method is based on a teacher's annual referring count, 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 A13 that excluding year 2020 results in almost identical distribution of ODRs by months. We are also able to replicate our findings for RQ1 and RQ3 in Figure A14 and Table A1 by only using the first three years of data.

own, the negative and long-lasting impact of suspensions are better documented. 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), who use identical data to this study, only 4 percent 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 percent (.17/3.4) of the ODR gap, but this rate is about 4 percent for Hispanic-White and multiracial/other-White comparisons. This is consistent with findings by Liu et al. that there is also a sizable ODR 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 ODRs per year (from zero to around .15 suspensions). Top 5 percent 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 multiracial/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 ODRs 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 still exacerbate the already large suspensions gaps.

RQ3. Predictors. Thus far, we have documented the contributions of "top referrers" to racial disproportionalities in both ODRs and suspensions. In RQ3, we probe into how working environments and personal traits might explain their referring behavior. We first

explore the role of school and classroom contexts 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. However, "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 and classrooms. For example, Black students only accounted for 7 percent of all students enrolled in the district and 12 percent of students in "top referrers"' classrooms, but they represented close to 22 percent of all referred students and 27 percent 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 and even more so in their classrooms, the differences between the racial compositions of their classrooms, 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 percent) 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 find mostly statistically insignificant 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 4.8

percentage points lower than a White teacher's. This number is 2.2 percentage points when comparing relative 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 of 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 referring 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, but this makes sense given the differing requirements for single- versus multiple-subject credentials at different school levels.

Years of experience seems to be a particularly salient predictor that holds across school levels. 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.

We also replicate this exercise by replacing the outcomes using whether a teacher issued at least one referral for specific reasons. While results are mostly consistent with findings from Table 1, Table A3 reveals some interesting patterns, especially on teacher race. Compared with their White colleagues, Black teachers are less likely to issue ODRs for interpersonal, defiance, and violence reasons, but not for drugs, class skipping, or other reasons. Hispanic teachers are less likely to issue violence referrals but not for other reasons. In contrast, Asian

teachers seem to not favor the use of ODRs for any reasons.

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 offences 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 referrers" 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' referring 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 identifying how same-race teachers and other teacher qualifications and personal traits affect their overall referring frequency and, more importantly, which students they refer (Hayes et al., 2022). Only 7% of students in our district identify as Black, which is slightly higher than the state average, yet they are substantially over-represented among ODR recipients. While this disporportionality is mirrored elsewhere, it is important to replicate these studies and evaluate whether the specific teachers play an outsize role in exacerbating these disparities

in other contexts. Our hope is this collective body of work can contribute to our understanding of educators' disparate roles in disciplinary disparities, and ultimately identify effective strategies to reduce their prevalence and impact.

Our approach focused on examining extensive referring behavior provides a tangible and transparent policy tool for school leaders in any district. In our district partner's context, a small group of educators contribute a meaningful proportion of the overall racial disproportionality in ODRs. At the same time, this group represents a relatively manageable number of educators to receive focused supports. In other contexts, different proportions of individuals that the district feels it can effectively support may be identified through a similar approach. We can well imagine a tiered system to provide differential support to teachers who engage in different levels ODR use. Recent literature has surfaced on some promising interventions that may help adjust teachers' referring behavior. For example, an experiment in some California middle schools found that prompts about the utility of empathic (punitive) mindsets in the classroom caused teachers to change their stated response to hypothetical situations in the classroom to be less (more) punitive (Okonofua, Paunesku, & Walton, 2016). Given that "top referrers" tend to be teachers early in their career, targeting professional development supports on classroom management skills for this group of teachers might also be a viable approach to reducing their referring frequency.

Focusing on extensive referring behavior may also help to identify school contexts or position types where individuals engaging in extensive referring may be concentrated. Our analyses highlight that structural supports at certain school levels and subject areas are warranted. Moreover, detailed data similar to what is used in this study can help reveal the many VDPs that contribute to racial disciplinary dissproportionalities, such as time of the day, month of the year, and location. Concerted efforts such as multi-tiered systems of support (Stoiber & Gettinger, 2016) or restorative justice (Darling-Hammond & Fronius, 2022) can take into account these dimensions to better prevent student behavioral challenges.

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`

Atiles, J. T., Gresham, T. M., & Washburn, I. (2017). Values and beliefs regarding discipline practices: How school culture impacts teacher responses to student misbehavior. *Educational Research Quarterly*, *40*(3), 3–24.

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., & Darensbourg, 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*.

Cho, V., Mansfield, K. C., & Claughton, J. (2020). The past and future technology in classroom management and school discipline: A systematic review. *Teaching and Teacher Education*, *90*, 103037.

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)

Cook, C. R., Duong, M. T., McIntosh, K., Fiat, A. E., Larson, M., Pullmann, M. D., & McGinnis, J. (2018). Addressing discipline disparities for black male students: Linking malleable root causes to feasible and effective practices. *School Psychology Review*, *47*(2), 135–152.

Cruz, R. A., Kulkarni, S. S., & Firestone, A. R. (2021). A quantcrit analysis of context, discipline, special education, and disproportionality. *AERA Open*, *7*, 23328584211041354.

Curran, F. C. (2020). A matter of measurement: How different ways of measuring racial gaps in school discipline can yield drastically different conclusions about racial disparities in discipline. *Educational Researcher*, *49*(5), 382–387.

Darling-Hammond, S., & Fronius, T. (2022). Restorative practices in schools. In *Handbook of classroom management* (pp. 54–73). Routledge.

Davison, M., Penner, A. M., & Penner, E. K. (2022). Restorative for all? racial disproportionality and school discipline under restorative justice. *American educational research journal*, *59*(4), 687–718.

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`

Gansen, H. M. (2021). Disciplining difference (s): Reproducing inequalities through disciplinary interactions in preschool. *Social Problems*, *68*(3), 740–760.

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C. A., & Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, *14*(4), 300–342.

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.

Gion, C., McIntosh, K., & Falcon, S. (2022). Effects of a multifaceted classroom intervention on racial disproportionality. *School psychology review*, *51*(1), 67–83.

Gion, C., McIntosh, K., & Smolkowski, K. (2018). Examination of american indian/alaska native school discipline disproportionality using the vulnerable decision points approach. *Behavioral Disorders*, *44*(1), 40–52.

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

Girvan, E. J., McIntosh, K., & Smolkowski, K. (2019). Tail, tusk, and trunk: What different metrics reveal about racial disproportionality in school discipline. *Educational Psychologist*, *54*(1), 40–59.

Gopalan, M., & Nelson, A. A. (2019). Understanding the racial discipline gap in schools. *Aera Open*, *5*(2), 2332858419844613.

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., & Mosely, P. M. (2004). The discipline gap: Teachers' views on the overrepresentation of african american students in the discipline system. *Equity & Excellence in Education*, *37*(1), 18–30.

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

Gregory, A., & Weinstein, R. S. (2008). The discipline gap and african americans: Defiance or cooperation in the high school classroom. *Journal of school psychology*, *46*(4), 455–475.

Griffith, D., & Tyner, A. (2019). Discipline reform through the eyes of teachers. *Thomas B. Fordham Institute*.

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., & Gershenson, S. (2019). The impact of demographic representation on absences and suspensions. *Policy Studies Journal*, *47*(4), 1069–1099.

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. *Socious: Sociological Research for a Dynamic World*. doi: doi.org/10.1177/23780231221103044

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

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)

McIntosh, K., Girvan, E. J., Horner, R., & Smolkowski, K. (2014). Education not incarceration: A conceptual model for reducing racial and ethnic disproportionality in school discipline. *Kent McIntosh, Erik J. Girvan, Robert H. Horner, & Keith Smolkowski, Education not Incarceration: A Conceptual Model For Reducing Racial and Ethnic Disproportionality in School Discipline*, *5*.

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

Milner, H. R. (2012). *But what is urban education?* (Vol. 47) (No. 3). Sage Publications Sage CA: Los Angeles, CA.

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

Nishioka, V. (2017). School discipline data indicators: A guide for districts and schools. rel 2017-240. *Regional Educational Laboratory Northwest*.

Nishioka, V., Merrill, B., & Hanson, H. (2021). Changes in exclusionary and nonexclusionary discipline in grades k-5 following state policy reform in oregon. rel 2021-061. *Regional Educational Laboratory Northwest*.

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

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.

Rodriguez, L. A., & Welsh, R. O. (2022). The dimensions of school discipline: Toward a comprehensive framework for measuring discipline patterns and outcomes in schools. *AERA Open*, *8*, 23328584221083669.

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.

Shirrell, M., Bristol, T. J., & Britton, T. A. (2021). The effects of student-teacher ethnoracial matching on exclusionary discipline for asian american, black, and latinx students:

Evidence from new york city. edworkingpaper no. 21-475. *Annenberg Institute for School Reform at Brown University*.

Skiba, R. J., Chung, C.-G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion. *American Educational Research Journal*, *51*(4), 640–670. (publisher: [American Educational Research Association, Sage Publications, Inc.])

Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011a). 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., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011b). Race is not neutral: A national investigation of african american and latino disproportionality in school discipline. *School psychology review*, *40*(1), 85–107.

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

Stoiber, K. C., & Gettinger, M. (2016). Multi-tiered systems of support and evidence-based practices. In *Handbook of response to intervention* (pp. 121–141). Springer.

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

Welsh, R. O. (2023). Navigating tensions in school discipline: Examining school leaders, teachers, and the conversion of referrals into suspensions. *American Journal of Education*, *129*(2), 000–000.

Welsh, R. O., & Swain, W. A. (2020). (re) defining urban education: A conceptual review and empirical exploration of the definition of urban education. *Educational Researcher*, *49*(2), 90–100.

Wiley, K. E. (2021). A tale of two logics: School discipline and racial disparities in a "mostly white" middle school. *American Journal of Education*, *127*(2), 163–192.

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

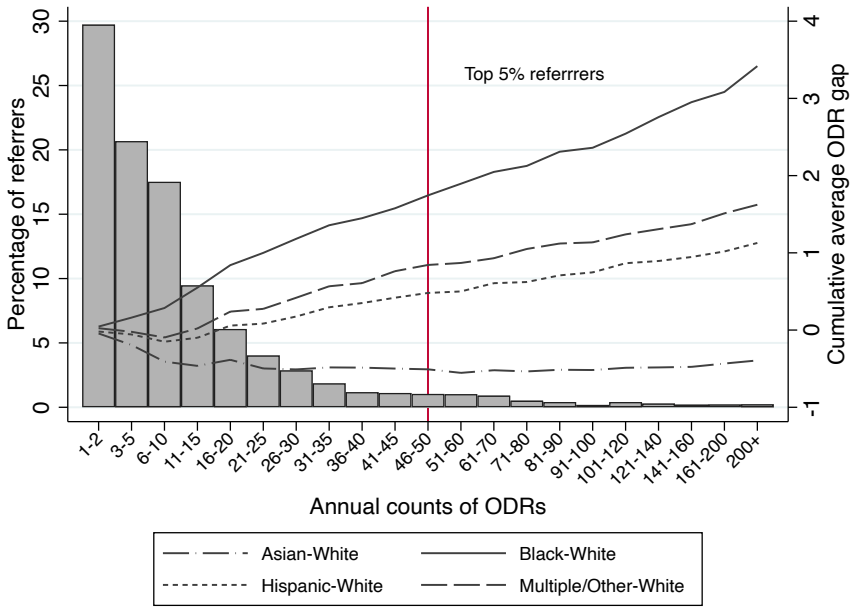
Zakszeski, B., Rutherford, L., Heidelburg, K., & Thomas, L. (2021). In pursuit of equity: Discipline disproportionality and swpbis implementation in urban schools. *School Psychology*, *36*(2), 122.

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

Zimmermann, C. R. (2018). The penalty of being a young black girl: Kindergarten teachers' perceptions of children's problem behaviors and student–teacher conflict by the intersection of race and gender. *The Journal of Negro Education*, *87*(2), 154–168.

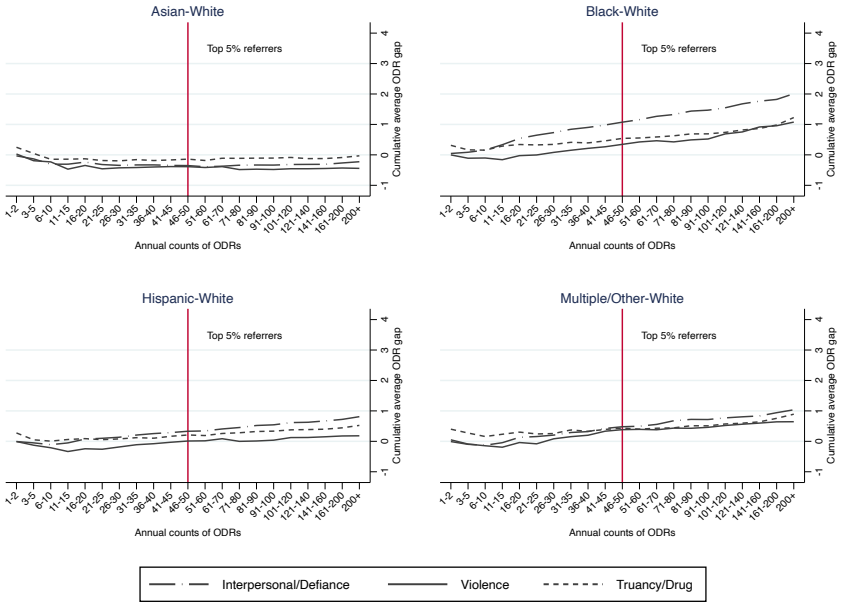
Figures and Tables

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



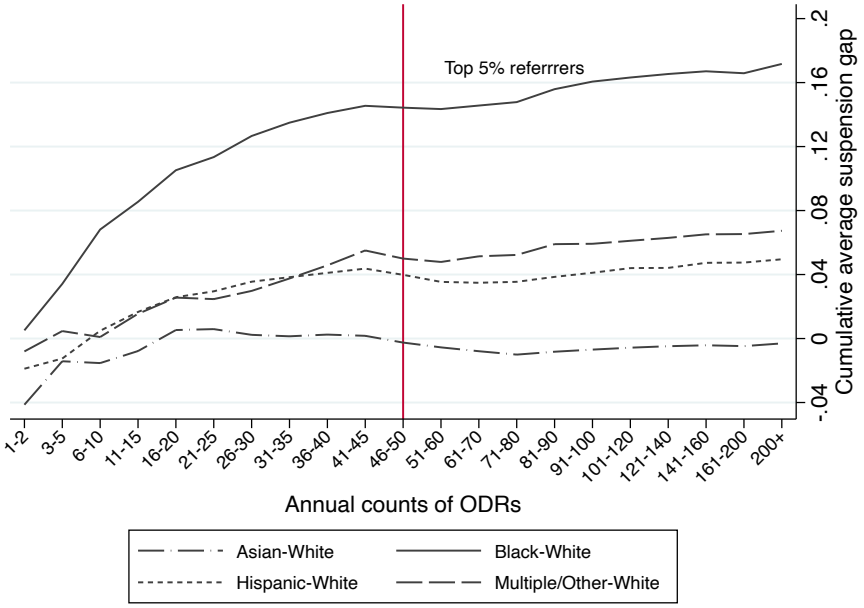
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The percentage of referrers represents the four-year average of the percentages of referrers in a given ODR group. The cumulative average ODR gap is the four-year average of the percentage differences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

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



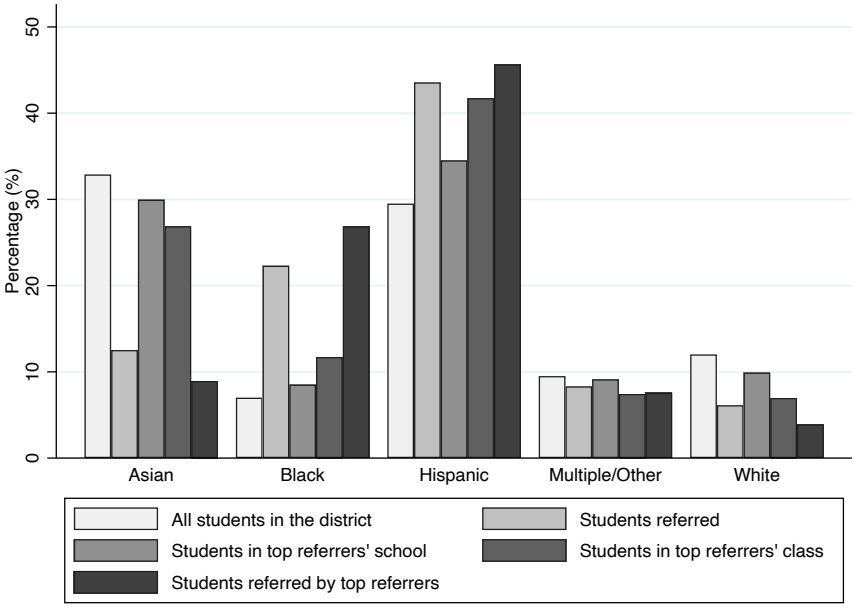
Notes: This figure decomposes the cumulative average ODR gap by race and ODR type, respectively. Data come from a large school district in California between 2016-17 and 2019-20 school years. For each race/ethnic group, the cumulative average ODR gap is the four-year average of the percentage differences between students of a specific race/ethnic group and White students (the reference group) who received at least one ODR in a given ODR group by ODR type. Our ODR group is formed by the counts of ODRs at the teacher-by-year level. ODR 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 ODRs issued by top 5% referrers is 48.25 across years, falling into the 46-50 ODR group.

Figure 3: Distribution of 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 ODR group. Our ODR group is formed by the counts of ODRs at the teacher-by-year level. The average number of ODRs issued by top 5% referrers is 48.25 across years, falling into the 46-50 ODR 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, their percentages among all students who were in the top 5% referrers' class, and their percentages among all students who were referred at least once by any of the top 5% referrers.

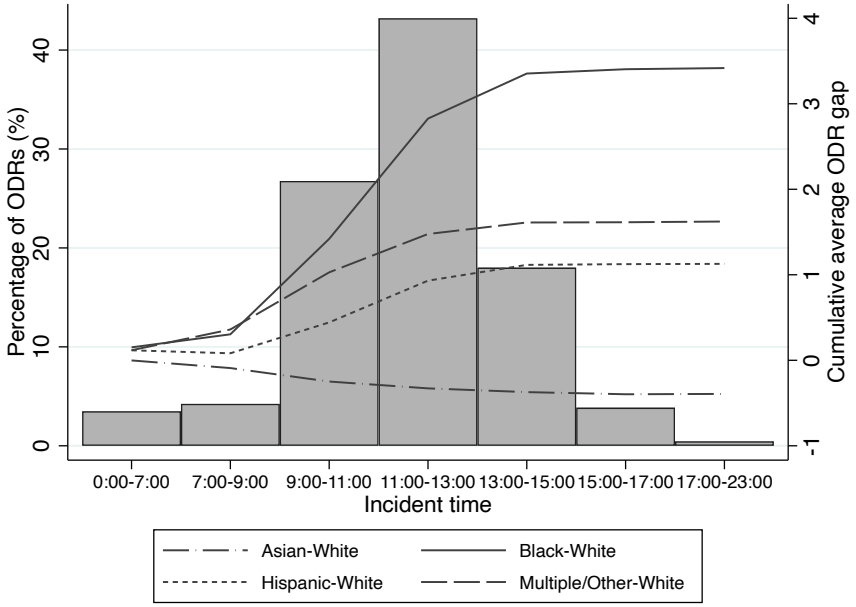
Table 1: Predictors of Being A Referrer or A "Top Referrer"

	Overall		Elementary		Middle		High	
	All Referrers	Top Referrers	All Referrers	Top Referrers	All Referrers	Top Referrers	All Referrers	Top Referrers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Mean	0.342	0.017	0.277	0.014	0.707	0.057	0.424	0.010
Female	0.009	-0.005	-0.002	0.005	-0.029	-0.030**	0.040**	0.001
	(0.010)	(0.004)	(0.014)	(0.003)	(0.022)	(0.012)	(0.019)	(0.004)
Black	-0.048***	-0.022**	-0.047*	-0.013***	-0.132**	-0.079**	-0.000	-0.009**
	(0.017)	(0.009)	(0.024)	(0.005)	(0.050)	(0.033)	(0.029)	(0.004)
Hispanic	-0.024**	-0.008**	-0.016	-0.003	-0.040	-0.035*	-0.022	-0.009**
	(0.011)	(0.003)	(0.017)	(0.005)	(0.040)	(0.018)	(0.023)	(0.005)
Asian	-0.058***	0.001	-0.047***	0.001	-0.106***	-0.005	-0.065*	-0.001
	(0.014)	(0.003)	(0.016)	(0.005)	(0.033)	(0.011)	(0.033)	(0.003)
Missing Race	-0.017	-0.006**	0.001	-0.002	-0.060	-0.040***	-0.010	0.000
	(0.011)	(0.003)	(0.020)	(0.004)	(0.035)	(0.013)	(0.020)	(0.003)
multiracial/other Race	-0.013	-0.003	0.005	0.008	-0.055	-0.026	-0.012	-0.001
	(0.018)	(0.005)	(0.034)	(0.010)	(0.045)	(0.016)	(0.045)	(0.010)
Cred in ELL	0.018*	0.002	-0.006	0.002	0.066***	-0.002	0.047**	0.002
	(0.009)	(0.002)	(0.016)	(0.005)	(0.015)	(0.007)	(0.020)	(0.005)
Cred in SPECED	-0.012	-0.007**	-0.040**	-0.005	-0.043	-0.028**	0.039	-0.001
	(0.013)	(0.003)	(0.019)	(0.006)	(0.027)	(0.009)	(0.035)	(0.002)
Cred in English	0.046**	0.013**	-0.037	-0.002	0.056	0.043**	0.104***	0.006
	(0.018)	(0.005)	(0.027)	(0.006)	(0.042)	(0.017)	(0.029)	(0.004)
Cred in Math	0.055***	0.011*	-0.005	-0.013***	0.083**	0.020	0.093***	0.014
	(0.016)	(0.007)	(0.026)	(0.005)	(0.033)	(0.013)	(0.029)	(0.012)
Cred in Science	0.086***	0.013*	0.067	0.013	0.124***	0.019	0.117***	0.012
	(0.022)	(0.007)	(0.052)	(0.015)	(0.018)	(0.024)	(0.036)	(0.008)
Temporary	-0.071***	-0.019***	-0.090***	-0.020**	-0.016	-0.034	-0.100***	-0.020*
	(0.016)	(0.006)	(0.024)	(0.008)	(0.039)	(0.021)	(0.026)	(0.012)
Years 2-3	-0.003	-0.009	-0.007	-0.001	0.020	-0.050**	0.024	0.007
	(0.014)	(0.006)	(0.024)	(0.010)	(0.022)	(0.020)	(0.032)	(0.007)
Years 4-5	-0.035**	-0.013	-0.037	-0.006	0.001	-0.046*	-0.047	-0.008
	(0.015)	(0.008)	(0.025)	(0.013)	(0.030)	(0.025)	(0.033)	(0.017)
Years 6-10	-0.062***	-0.028***	-0.070**	-0.024*	-0.021	-0.063**	-0.086**	-0.024*
	(0.017)	(0.008)	(0.028)	(0.013)	(0.026)	(0.023)	(0.038)	(0.013)
Years 11-30	-0.075***	-0.029***	-0.076***	-0.023*	-0.064*	-0.068**	-0.083***	-0.023*
	(0.016)	(0.008)	(0.026)	(0.012)	(0.034)	(0.027)	(0.027)	(0.013)
Controls for:								
School FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R2	0.413	0.084	0.302	0.046	0.137	0.108	0.322	0.076
Observations	17026	17026	7425	7425	2403	2403	4172	4172

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 the teacher-by-year level. 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 ODR. $p<0.10^{*}$ $p<0.05^{**}$ $p<0.01^{***}$.

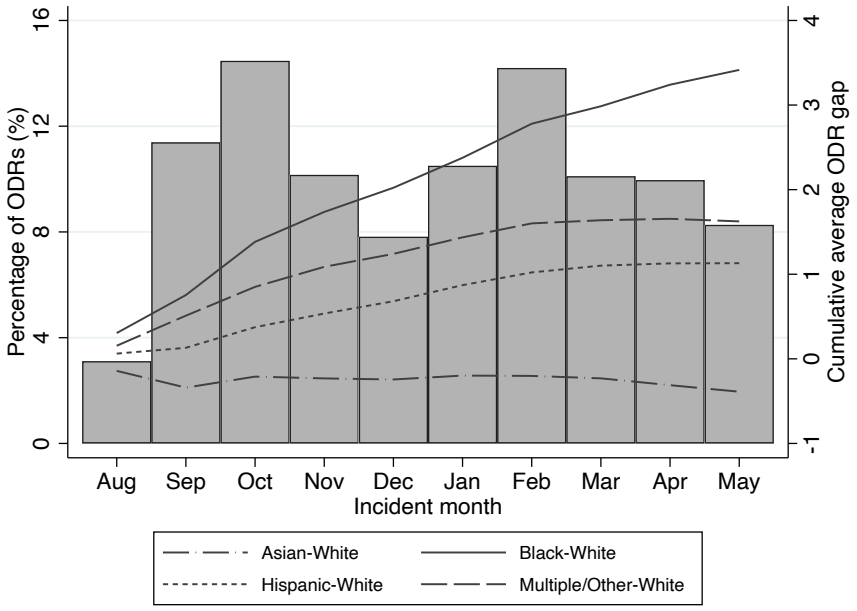
Appendix A

Figure A1: Distribution of ODRs by Incident Time



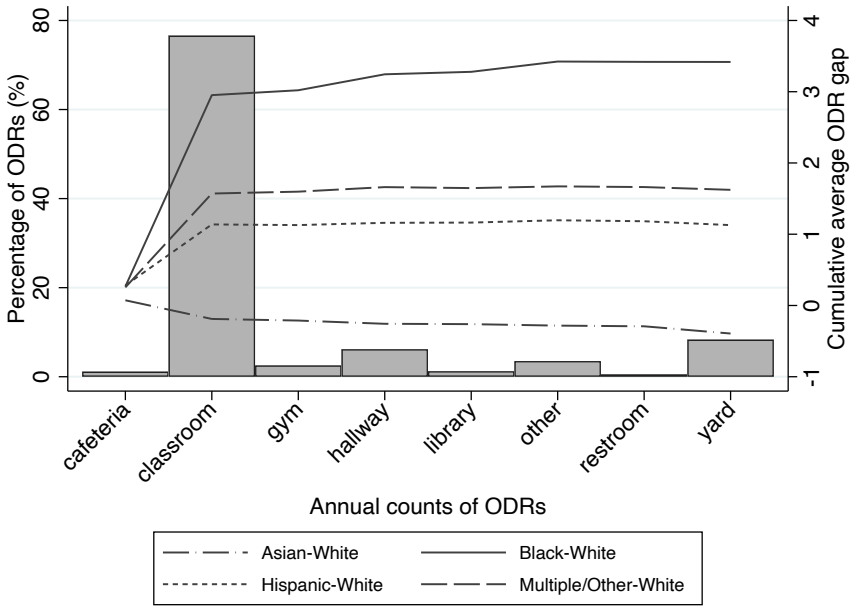
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. This figure shows the four-year average of the percentages of ODRs by incident time.

Figure A2: Distribution of ODRs by month



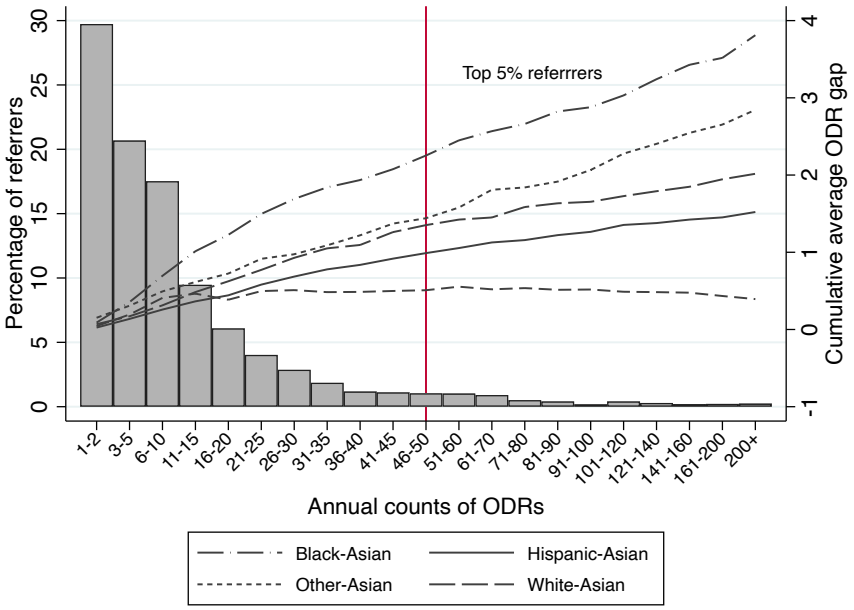
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. This figure shows the four-year average of the percentages of ODRs by month. Due to summer holidays, June and July are excluded when producing this figure.

Figure A3: Distribution of ODRs by location



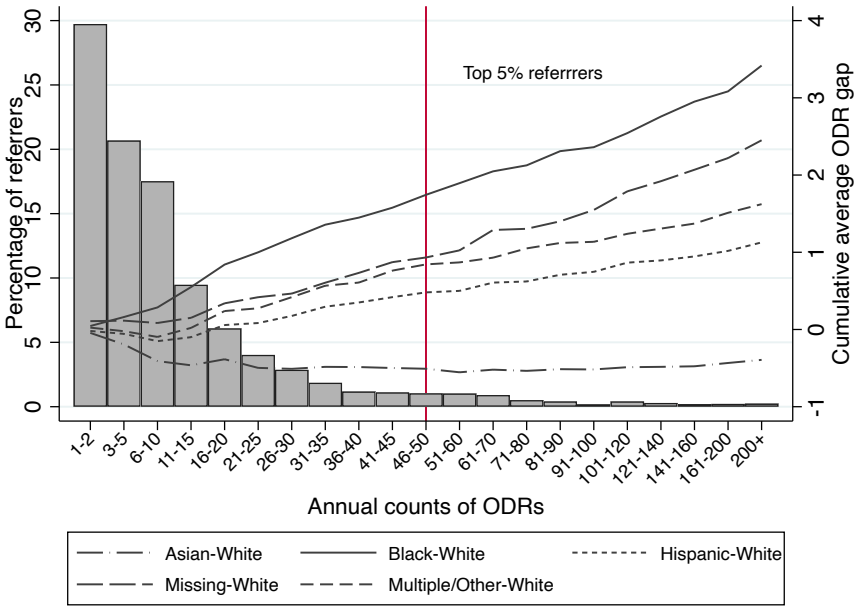
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. This figure shows the four-year average of the percentages of ODRs by location.

Figure A4: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps (with Asian Students as Referrence Group)



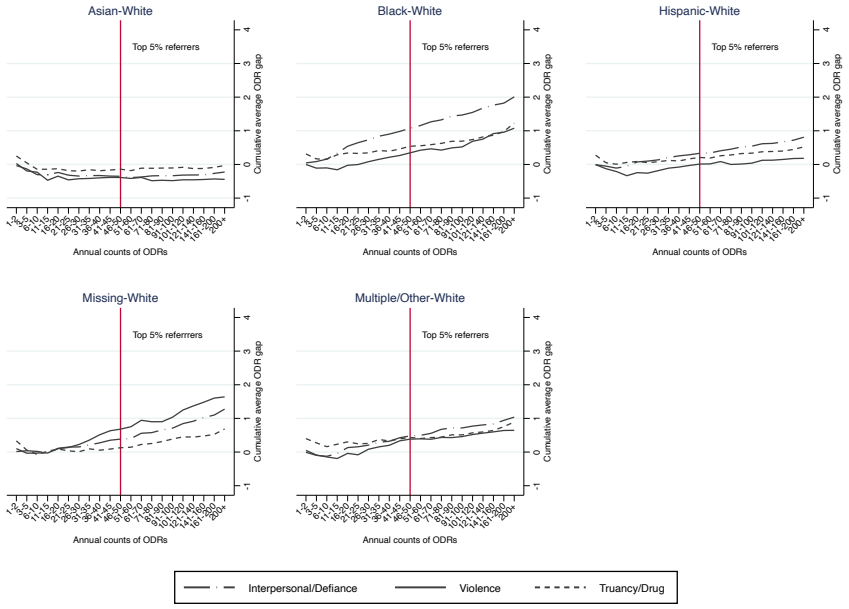
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The percentage of referrers represents the four-year average of the percentages of referrers in a given ODR group. The cumulative average ODR gap is the four-year average of the percentage differences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

Figure A5: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps (with Missing Racial Information)



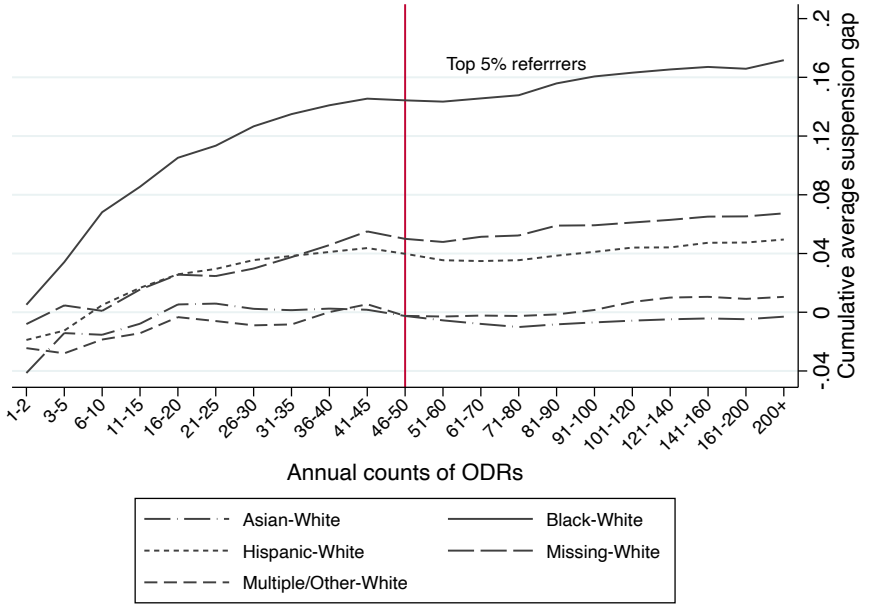
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The percentage of referrers represents the four-year average of the percentages of referrers in a given ODR group. The cumulative average ODR gap is the four-year average of the percentage differences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

Figure A6: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps: by Race and ODR Type (with Missing Racial Information)



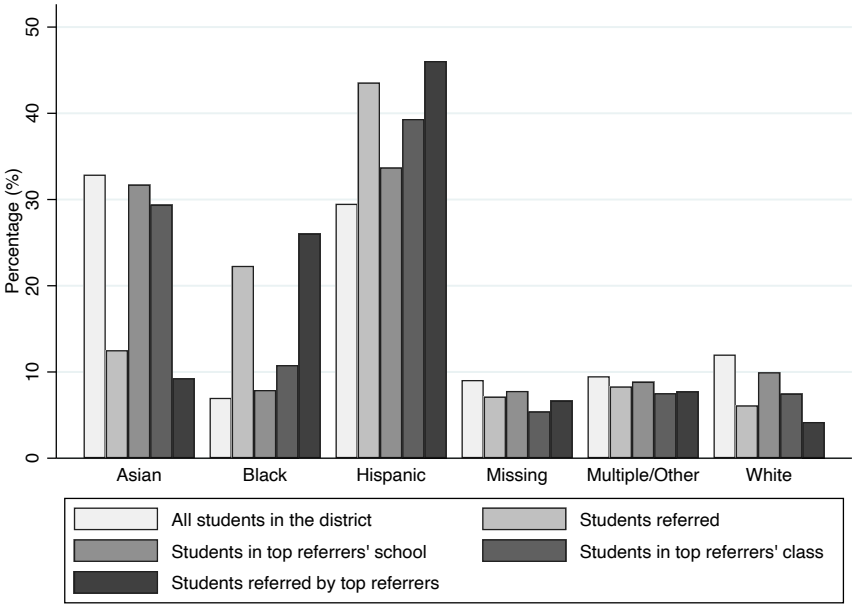
Notes: This figure decomposes the cumulative average ODR gap by race and ODR type, respectively. Data come from a large school district in California between 2016-17 and 2019-20 school years. For each race/ethnic group, the cumulative average ODR gap is the four-year average of the percentage differences between students of a specific race/ethnic group and White students (the reference group) who received at least one ODR in a given ODR group by ODR type. Our ODR group is formed by the counts of ODRs at the teacher-by-year level. ODR 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 ODRs issued by top 5% referrers is 48.25 across years, falling into the 46-50 ODR group.

Figure A7: Referring Frequency & The Corresponding Cumulative Suspension Gaps (with Missing Racial Information)



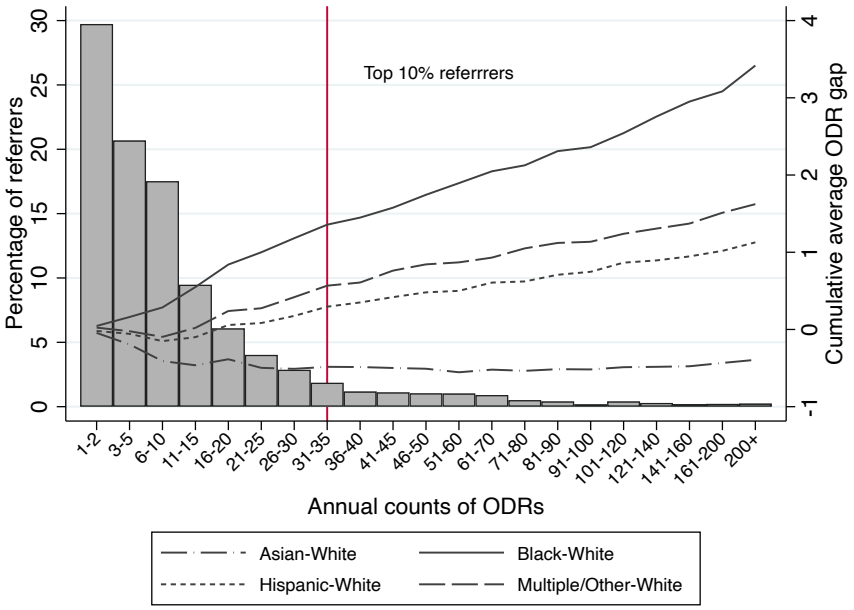
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The cumulative average suspension gap is the four-year average of percentage differences between students of other races and White students who received at least one suspension in a given ODR group.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

Figure A8: Comparing Student Racial Representations (with Missing Racial Information)



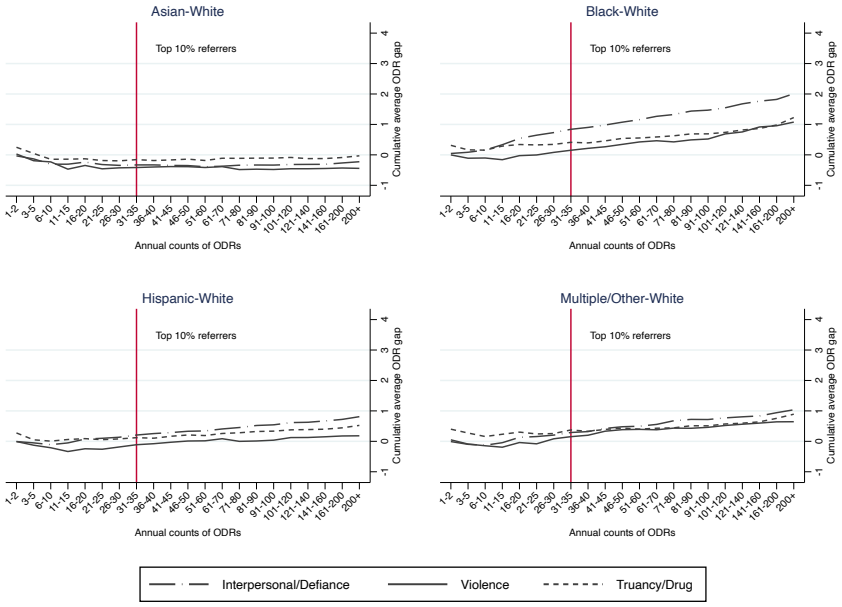
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. Students are counted at student-by-year level. For each racial/ethnic group, this figure shows their percentages among all students in the district, their percentages among all students who are referred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

Figure A9: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps (Top 10% Referrers)



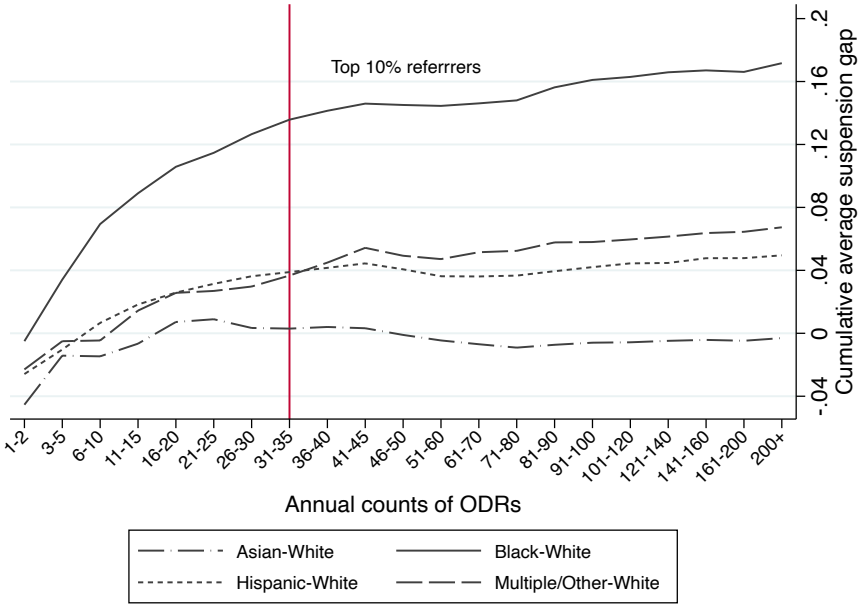
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The cumulative average ODR gap is the four-year average of the percentage differences between students of other races and White students (the reference group) who received at least one ODR in a given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

Figure A10: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps: by Race and ODR Type (Top 10% Referrers)



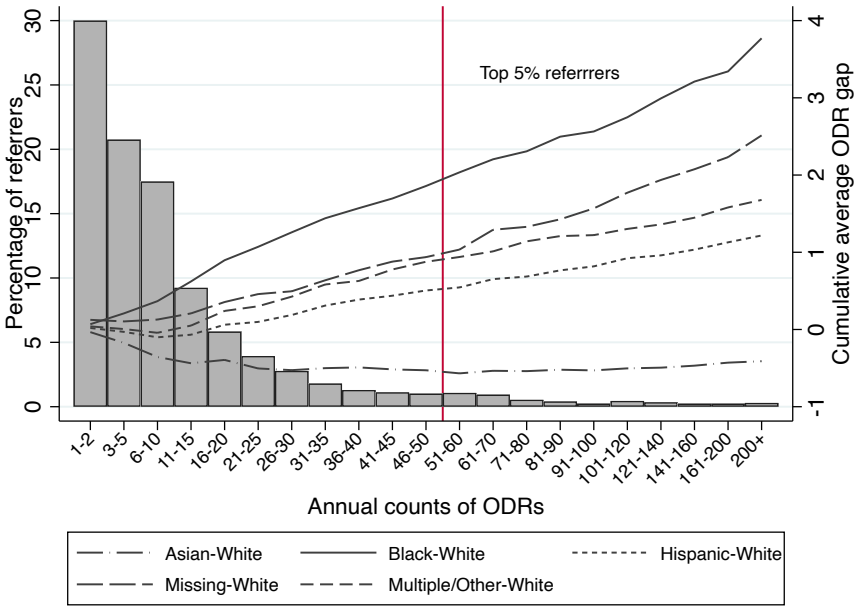
Notes: This figure decomposes the cumulative average ODR gap by race and ODR type, respectively. Data come from a large school district in California between 2016-17 and 2019-20 school years. For each race/ethnic group, the cumulative average ODR gap is the four-year average of the percentage differences between students of a specific race/ethnic group and White students (the reference group) who received at least one ODR in a given ODR group by ODR type. Our ODR group is formed by the counts of ODRs at the teacher-by-year level. ODR 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 ODRs issued by top 10% referrers is 31 across years, falling into 31-35 ODR group.

Figure A11: Referring Frequency & The Corresponding Cumulative Suspension Gaps (Top 10% Referrers)



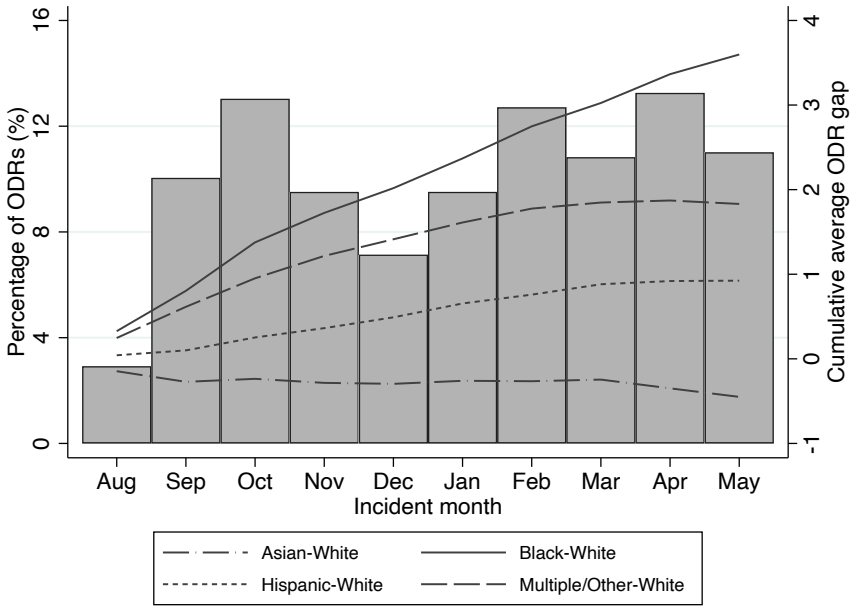
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The cumulative average suspension gap is the four-year average of percentage differences between students of other races and White students who received at least one suspension in a given ODR group.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

Figure A12: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps (Administrators Included)



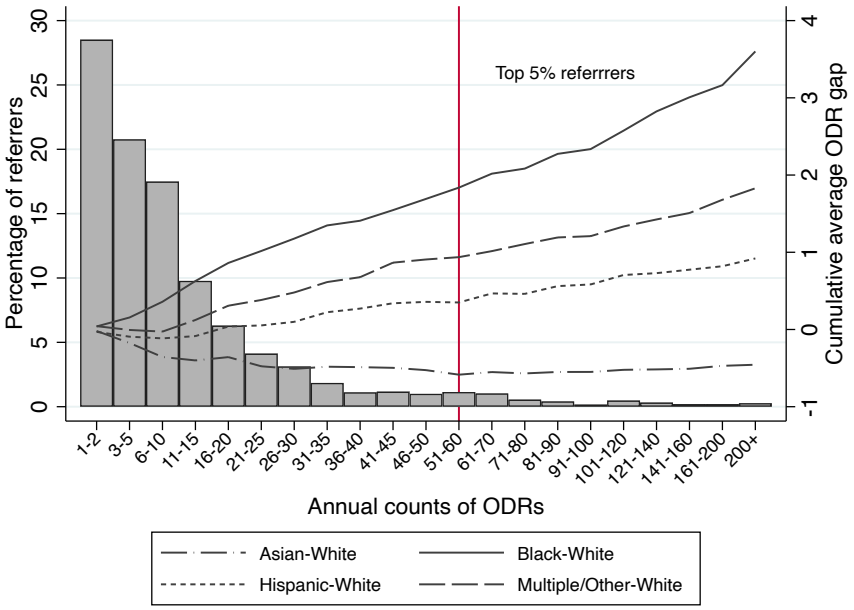
Notes: Data come from a large school district in California between 2016-17 and 2019-20 school years. The percentage of referrers represents the four-year average of the percentages of referrers in a given ODR group. The cumulative average ODR gap is the four-year average of the percentage differences between students of other races and White students (the reference group) who received at least one ODR in a given ODR group. Our ODR group is formed by the counts of ODRs at educator-by-year level. The average number of ODRs issued by top 5% referrers is 50.05 across years, falling in between the 46-50 and 51-60 ODR groups.

Figure A13: Distribution of ODRs by Month (Year 2020 Excluded)



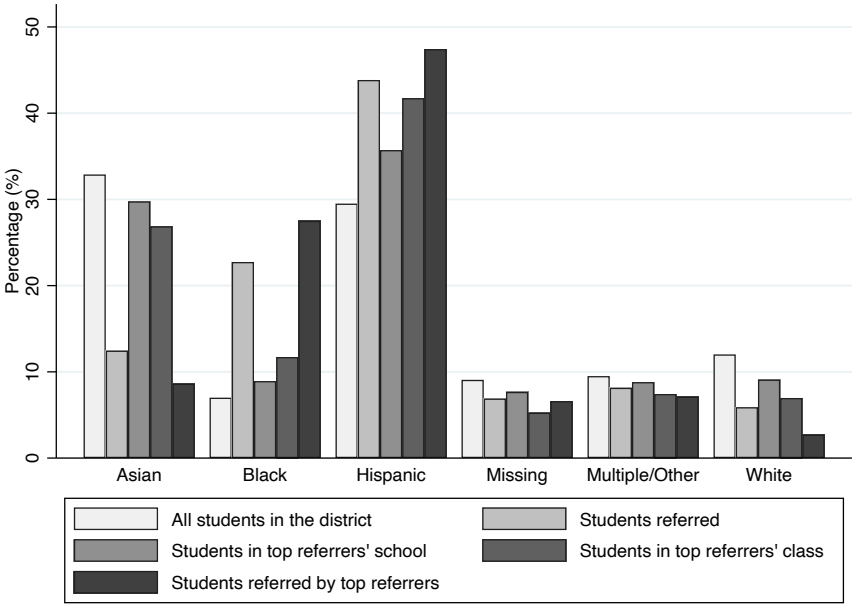
Notes: Data come from a large school district in California between 2016-17 and 2018-19 school years. This figure shows the three-year average of the percentages of ODRs by month. Due to summer holidays, June and July are excluded when producing this figure.

Figure A14: Distribution of Referring Frequency & The Corresponding Cumulative ODR Gaps (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 ODR group. The cumulative average ODR 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 ODR in a given ODR group. Our ODR group is formed by the counts of ODRs at the teacher-by-year level. The average number of ODRs issued by top 5% referrers is 53 across years, falling into the 51-60 ODR group.

Figure A15: 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, their percentages among all students who were in the top 5% referrers' class, and their percentages among all students who were referred at least once by any of the top 5% referrers.

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

	Overall		Elementary		Middle		High	
	All Referrers	Top Referrers	All Referrers	Top Referrers	All Referrers	Top Referrers	All Referrers	Top Referrers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Mean	0.341	0.017	0.260	0.014	0.731	0.055	0.434	0.010
Female	0.013	-0.005	0.004	0.006	-0.016	-0.030*	0.040*	-0.000
	(0.010)	(0.004)	(0.015)	(0.004)	(0.022)	(0.015)	(0.020)	(0.004)
Black	-0.051***	-0.023**	-0.046	-0.014**	-0.102*	-0.071	-0.039	-0.012*
	(0.018)	(0.010)	(0.028)	(0.006)	(0.054)	(0.042)	(0.032)	(0.006)
Hispanic	-0.021*	-0.003	-0.012	-0.000	-0.045	-0.021	-0.028	-0.005
	(0.011)	(0.004)	(0.018)	(0.006)	(0.035)	(0.023)	(0.026)	(0.006)
Asian	-0.057***	0.001	-0.048***	0.000	-0.110***	0.001	-0.068*	0.000
	(0.015)	(0.004)	(0.018)	(0.005)	(0.032)	(0.017)	(0.040)	(0.004)
Missing Race	-0.014	-0.004	-0.014	-0.000	-0.046	-0.029***	0.003	0.002
	(0.012)	(0.003)	(0.021)	(0.005)	(0.038)	(0.010)	(0.023)	(0.004)
multiracial/other Race	-0.016	-0.001	0.014	0.003	-0.075	0.001	-0.009	-0.002
	(0.020)	(0.006)	(0.037)	(0.011)	(0.046)	(0.016)	(0.047)	(0.013)
Cred in ELL	0.017*	-0.001	-0.002	-0.001	0.047***	-0.007	0.053**	0.001
	(0.010)	(0.003)	(0.017)	(0.005)	(0.012)	(0.008)	(0.026)	(0.005)
Cred in SPECED	-0.028**	-0.007**	-0.051**	-0.005	-0.075**	-0.024***	0.006	-0.003
	(0.014)	(0.003)	(0.020)	(0.005)	(0.029)	(0.006)	(0.035)	(0.004)
Cred in English	0.041**	0.012**	-0.032	-0.004	0.036	0.042**	0.095***	0.006
	(0.019)	(0.005)	(0.029)	(0.006)	(0.050)	(0.018)	(0.028)	(0.003)
Cred in Math	0.047***	0.010	-0.017	-0.014***	0.053	0.024	0.090***	0.010
	(0.017)	(0.007)	(0.026)	(0.005)	(0.034)	(0.018)	(0.030)	(0.009)
Cred in Science	0.081***	0.016*	0.070	0.023	0.125***	0.025	0.103**	0.014
	(0.023)	(0.009)	(0.054)	(0.021)	(0.019)	(0.032)	(0.037)	(0.009)
Temporary	-0.067***	-0.020***	-0.086***	-0.022**	-0.007	-0.033	-0.096***	-0.021*
	(0.018)	(0.007)	(0.027)	(0.010)	(0.040)	(0.031)	(0.034)	(0.012)
Years 2-3	-0.006	-0.012	-0.007	-0.002	0.013	-0.066**	0.015	0.008
	(0.015)	(0.008)	(0.026)	(0.011)	(0.018)	(0.024)	(0.033)	(0.010)
Years 4-5	-0.033**	-0.017*	-0.022	-0.006	0.006	-0.051	-0.069*	-0.019
	(0.016)	(0.010)	(0.029)	(0.015)	(0.022)	(0.035)	(0.036)	(0.020)
Years 6-10	-0.057***	-0.032***	-0.061**	-0.028*	0.002	-0.072**	-0.089**	-0.027*
	(0.017)	(0.010)	(0.030)	(0.015)	(0.030)	(0.033)	(0.034)	(0.014)
Years 11-30	-0.072***	-0.033***	-0.071**	-0.025*	-0.053	-0.075**	-0.084***	-0.030*
	(0.016)	(0.010)	(0.028)	(0.014)	(0.032)	(0.034)	(0.026)	(0.017)
Controls for								
School FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.442	0.093	0.326	0.049	0.140	0.135	0.326	0.069
Observations	12749	12749	5560	5560	1797	1797	3112	3112

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 the teacher-by-year level. 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 ODR. p<0.10* p<0.05** p<0.01***.

Table A2: Predictors of Being A Referrer or A Top Referrer (With Controls for Principals' Traits)

	Overall		Elementary		Middle		High	
	All Referrers	Top Referrers	All Referrers	Top Referrers	All Referrers	Top Referrers	All Referrers	Top Referrers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Mean	0.342	0.017	0.277	0.014	0.707	0.057	0.424	0.010
Female	0.009	-0.005	-0.003	0.005	-0.027	-0.028**	0.040**	0.001
	(0.010)	(0.003)	(0.014)	(0.003)	(0.023)	(0.012)	(0.019)	(0.004)
Black	-0.048***	-0.022***	-0.047*	-0.013***	-0.130**	-0.076**	0.001	-0.009**
	(0.017)	(0.009)	(0.024)	(0.005)	(0.049)	(0.033)	(0.028)	(0.004)
Hispanic	-0.024**	-0.007**	-0.017	-0.003	-0.039	-0.033*	-0.022	-0.009**
	(0.011)	(0.003)	(0.017)	(0.005)	(0.039)	(0.018)	(0.023)	(0.005)
Asian	-0.058***	0.001	-0.046***	0.001	-0.107***	-0.007	-0.065*	-0.001
	(0.014)	(0.003)	(0.016)	(0.005)	(0.033)	(0.010)	(0.034)	(0.003)
Missing Race	-0.016	-0.006**	0.002	-0.002	-0.057	-0.037**	-0.010	0.000
	(0.011)	(0.003)	(0.020)	(0.004)	(0.035)	(0.013)	(0.019)	(0.003)
multiracial/other Race	-0.013	-0.002	0.004	0.008	-0.053	-0.024	-0.013	-0.001
	(0.018)	(0.005)	(0.035)	(0.010)	(0.045)	(0.015)	(0.045)	(0.010)
Cred in ELL	0.018*	0.002	-0.006	0.002	0.065***	-0.002	0.047**	0.002
	(0.009)	(0.002)	(0.016)	(0.005)	(0.015)	(0.007)	(0.020)	(0.005)
Cred in SPECED	-0.012	-0.007**	-0.039**	-0.005	-0.043	-0.028**	0.039	-0.001
	(0.013)	(0.003)	(0.019)	(0.006)	(0.027)	(0.010)	(0.035)	(0.002)
Cred in English	0.046**	0.012**	-0.038	-0.003	0.055	0.041**	0.104***	0.006
	(0.018)	(0.005)	(0.026)	(0.006)	(0.042)	(0.017)	(0.029)	(0.004)
Cred in Math	0.055***	0.011*	-0.008	-0.013***	0.083**	0.021	0.092***	0.014
	(0.016)	(0.007)	(0.025)	(0.005)	(0.034)	(0.013)	(0.029)	(0.012)
Cred in Science	0.086***	0.013*	0.068	0.012	0.124***	0.020	0.117***	0.012
	(0.022)	(0.007)	(0.052)	(0.015)	(0.018)	(0.024)	(0.036)	(0.008)
Temporary	-0.070***	-0.020***	-0.088***	-0.020**	-0.015	-0.033	-0.100***	-0.020*
	(0.016)	(0.006)	(0.024)	(0.008)	(0.040)	(0.021)	(0.026)	(0.012)
Years 2-3	-0.002	-0.009	-0.002	-0.001	0.027	-0.045**	0.023	0.007
	(0.014)	(0.006)	(0.024)	(0.010)	(0.024)	(0.021)	(0.032)	(0.007)
Years 4-5	-0.035**	-0.012	-0.035	-0.006	0.004	-0.043	-0.047	-0.007
	(0.015)	(0.008)	(0.025)	(0.014)	(0.031)	(0.026)	(0.033)	(0.017)
Years 6-10	-0.062***	-0.027***	-0.069**	-0.024*	-0.014	-0.058**	-0.087**	-0.024*
	(0.017)	(0.008)	(0.028)	(0.013)	(0.028)	(0.024)	(0.038)	(0.013)
Years 11-30	-0.075***	-0.028***	-0.075***	-0.023*	-0.059	-0.065**	-0.083***	-0.023*
	(0.016)	(0.008)	(0.026)	(0.012)	(0.035)	(0.028)	(0.028)	(0.013)
Controls for								
School FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.414	0.092	0.309	0.047	0.141	0.144	0.322	0.076
Observations	17026	17026	7425	7425	2403	2403	4172	4172

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 the teacher-by-year level. 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 ODR. Principals' gender, race, and years of experience are also included as covariates. p<0.10* p<0.05** p<0.01***.

Table A3: Predictors of Referring Type

	Interpersonal Offenses/Defiance	Violence	Drug/Class Skipping/Other
	(1)	(2)	(3)
Outcome Mean	0.290	0.214	0.135
Female	0.013	-0.001	-0.002
	(0.009)	(0.007)	(0.008)
Black	-0.042**	-0.064***	-0.028
	(0.019)	(0.021)	(0.017)
Hispanic	-0.017	-0.026**	-0.007
	(0.011)	(0.011)	(0.009)
Asian	-0.053***	-0.042***	-0.023***
	(0.013)	(0.012)	(0.009)
Missing Race	-0.018	-0.015	-0.007
	(0.012)	(0.013)	(0.009)
multiracial/other Race	-0.011	-0.014	-0.031*
	(0.017)	(0.020)	(0.018)
Cred in ELL	0.017**	0.004	0.012
	(0.008)	(0.008)	(0.008)
Cred in SPECED	-0.008	-0.003	-0.008
	(0.013)	(0.014)	(0.012)
Cred in English	0.051**	0.002	0.035***
	(0.020)	(0.011)	(0.013)
Cred in Math	0.059***	0.014	0.032**
	(0.016)	(0.015)	(0.016)
Cred in Science	0.092***	0.067***	0.071***
	(0.025)	(0.022)	(0.019)
Temporary	-0.067***	-0.085***	-0.055***
	(0.016)	(0.017)	(0.015)
Years 2-3	-0.008	-0.007	0.004
	(0.013)	(0.014)	(0.015)
Years 4-5	-0.049***	-0.040**	-0.021
	(0.017)	(0.018)	(0.015)
Years 6-10	-0.071***	-0.072***	-0.053***
	(0.018)	(0.020)	(0.014)
Years 11-30	-0.077***	-0.089***	-0.060***
	(0.017)	(0.020)	(0.016)
Controls for:			
School FE	✓	✓	✓
Year FE	✓	✓	✓
Adjusted R^2	0.366	0.275	0.242
Observations	17026	17026	17026

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 the teacher-by-year level. All the statistics above are reported as proportions. The omitted group is male, White, tenured or nontenured referrers with one year of teaching experience. p<0.10* p<0.05** p<0.01***.