I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15663

Who Refers Whom? The Effects of Teacher Characteristics on Disciplinary Office Referrals

Michael S. Hayes
Jing Liu
Seth Gershenson

OCTOBER 2022

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15663

Who Refers Whom? The Effects of Teacher Characteristics on Disciplinary Office Referrals

Michael S. Hayes
Rutgers University

Jing Liu
University of Maryland College Park and IZA

Seth Gershenson
American University and IZA

OCTOBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Who Refers Whom? The Effects of Teacher Characteristics on Disciplinary Office Referrals*

Teachers affect a wide range of students' educational and social outcomes, but how they contribute to students' involvement in school discipline is less understood. We estimate the impact of teacher demographics and other observed qualifications on students' likelihood of receiving a disciplinary referral. Using data that track all disciplinary referrals and the identity of both the referred and referring individuals from a large and diverse urban school district in California, we find students are about 0.2 to 0.5 percentage points (7% to 18%) less likely to receive a disciplinary referral from teachers of the same race or gender than from teachers of different demographic backgrounds. Students are also less likely to be referred by more experienced teachers and by teachers who hold either an English language learners or special education credential. These results are mostly driven by referrals for defiance and violence infractions, Black and Hispanic male students, and middle school students. While it is unclear whether these findings are due to variation in teachers' effects on actual student behavior, variation in teachers' proclivities to make disciplinary referrals, or a combination of the two, these results nonetheless suggest that teachers play a central role in the prevalence of, and inequities in, office referrals and subsequent student discipline.

JEL Classification: I2, J7

Keywords: exclusionary discipline, teacher effectiveness, office referrals

Corresponding author:
Jing Liu
University of Maryland
2203 Benjamin Building, College Park
MD 20742
USA

E-mail: liu28@umd.edu

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

1 Introduction

Racial disparities in exclusionary discipline (i.e., suspensions) exist both between and within U.S. public schools (Barrett et al., 2019; Chin, 2021; Kinsler, 2011; Liu, Hayes, & Gershenson, 2022). Specifically, Black students are suspended from school at significantly higher rates than any other demographic group. These disparities are troubling for two broad and related reasons. First, suspensions are harmful in the sense that they likely hinder economic mobility and related long-run outcomes (Bacher-Hicks et al., 2019; Sorensen et al., in press; Weisburst, 2019). Second, these racial disparities in exclusionary discipline are at least partly due to systematic biases, or "intentional discrimination," in schools' handling of student indiscipline (Barrett et al., 2019; Liu, Hayes, & Gershenson, 2022; Shi & Zhu, 2022).

Accordingly, closing racial gaps in suspensions and reducing the use of suspensions in general are growing priorities for policymakers and education practitioners (Steinberg & Lacoe, 2017; Davison et al., 2022). Achieving these goals requires a clear understanding of the production of suspensions and the determinants of racial gaps in suspensions. (Sorensen et al., in press) study the role of principals, the final arbiters of a disciplinary decisions, in shaping racial disparities in exclusionary discipline, but less attention has been paid to office (disciplinary) referrals and the role of teachers in initiating that process. Indeed, office referrals necessarily precede suspensions and the majority (84% in our data) of referrals are made by classroom teachers. However, little is known about the types of teachers who make the most referrals. This is in stark contrast to a large literature on teachers' effects on a host of academic, behavioral, and non-cognitive outcomes including test scores, educational attainment, attendance, and earnings (Chetty et al., 2014; Gershenson, 2016; Jackson, 2018; Kraft, 2019; Ladd & Sorensen, 2017; Liu & Loeb, 2021).[1]

The current study extends the large literature on teacher effectiveness by examining teachers' impacts on office referrals. There are two reasons that teachers might vary in their

[1] A notable exception is that having a same-race teacher significantly reduces both the number and likelihood of suspensions Holt & Gershenson (2019); Lindsay & Hart (2017); Shirrell et al. (2021).

ability to cause students to receive office referrals. First, teachers may vary in their proclivity to make referrals, because they vary in their interpretation of classroom behavior (Girvan et al., 2017; Okonofua et al., 2016). Thus, all else equal, being assigned to a "frequent referrer" will increase the number of referrals a student receives. Second, teachers may affect student behavior, either directly via teaching socio-emotional skills or indirectly by changing the classroom climate, which in turn leads to changes in referral frequency (Kraft, 2019).

There is suggestive evidence that teachers do vary in how they perceive student (mis)behavior, though there is little research that uses actual referral data, perhaps because it is rarely available. 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 et al., 2016). A descriptive study that does utilize referral data is Skiba et al. (2002), who show that Black students are referred more often for arguably more subjective infractions, such as "disrespect" and "excessive noise." Taken together, it is easy to see how variation in teachers' perceptions of behavior can manifest in different referral rates across teachers and across student subgroups.

That said, we are aware of only two studies that explicitly examine teachers' referring behavior. First, Holt et al. (2022) analyze longitudinal data from North Carolina to identify the variability of elementary school teachers' punitiveness in the use of referrals. However, the authors do not observe the individuals who make the referrals; rather, they assume all referrals were made by the self-contained classroom teacher. Using a value-added model in which referrals are the outcome, they then identify more and less punitive teachers. More punitive teachers contribute to adverse academic and behavioral outcomes for Black students. Second, Liu, Penner, & Gao (2022) use the same data analyzed in the current study to describe the distribution of teachers' annual referral frequencies with an explicit focus on "chronic referrers." The top 5% of teachers who make the most referrals per year effectively double the racial gaps in referrals between Black and white, and between Hispanic and white, students.

While the two studies discussed in the prior paragraph suggest that there is in fact some variation in teachers' use of disciplinary referrals, neither addresses our titular question of who refers whom. The current study fills this gap in the literature by providing systematic quantitative evidence on how teacher traits, such as their demographic background, qualifications, and experience, affect the frequency with which their students receive office referrals, and perhaps more importantly, how this varies by student subgroup. Our study is thus closely related to a line of research on teacher effectiveness that associates observed teacher characteristics and qualifications with achievement gains and other educational outcomes. For example, having a same-race teacher can improve student short-run academic achievement, reduce receipts of suspensions and absences, and boost educational attainment (Dee, 2004a; Lindsay & Hart, 2017; Holt & Gershenson, 2019; Gershenson et al., 2018). Qualifications such as experience and undergraduate performance and coursework matter as well for student achievement (Clotfelter et al., 2007; Kukla-Acevedo, 2009). Thus, it is plausible that observable teacher characteristics and qualifications explain some of the variation in teachers' effects on student office referrals.

We conduct this research using unusually rich administrative data from a large and diverse urban district in California that tracks all disciplinary referrals and the identity of school personnel who issued them. We construct a novel panel data set that links student outcomes over time, including office referrals that do not result in a suspension, to the precise classroom and teacher that initiated the referral. These data allow us to estimate standard value-added models of the education production function as well as stacked regressions that compare students across subjects (classrooms) within a given year. These student-by-year fixed effects specifications control for unobserved time-varying student characteristics that might otherwise bias value-added estimates of teacher effects.

Our findings suggest that students are less likely to receive a disciplinary referral from teachers who share the same race/ethnicity and/or gender, who have more experience in the current school, or who hold a credential in teaching English language learners (ELL) and special education. The results are mainly driven by reduced likelihood of being referred due to defiance or violence reasons, Black and Hispanic male students, and middle schools. Our results add additional evidence to the large literature on student-teacher demographic match and teacher effectiveness more generally from the novel angle of disciplinary referrals. They also contribute to our growing knowledge of the disciplinary referral process that results in unequal rates of exclusionary discipline (Liu, Hayes, & Gershenson, 2022). It is here that our findings have rich policy implications: for example, to reduce the overall use of punitive strategies and ameliorate racial disparities in exclusionary discipline, providing targeted support for certain groups of teachers, such as novice teachers, might prove fruitful. Similarly, the classroom management techniques incorporated in ELL and special education certification programs might be adopted more broadly in teacher training programs. We revisit these implications in the conclusion.

2 Data

We use rich administrative data from a large and demographically diverse urban school district in California for the 2016-17 through 2019-20 school years. These data are ideal for the current study because they contain detailed information on all disciplinary referrals, regardless of whether or not they ultimately led to a suspension, as well as the individual who made and received the referral, the reason for the referral (i.e., type of incident), and the exact time, date, and location of the incident (e.g., 3 PM, in the library, on Monday April 2nd). We also observe student and teacher demographics and characteristics commonly found in administrative data systems. For students, we know their race/ethnicity, gender, special education status, test scores (for tested grades), grade point averages (GPA), and residential addresses which we use to match on to census data to identify neighborhood characteristics. For teachers, besides basic demographics, we also observe their credentials and total years of experience as a teacher as well as their experience at the current school.

We focus our analysis at the middle and high school level for two reasons. First, secondary school students have multiple teachers in different class subjects, while elementary students mostly are in self-contained classrooms with one primary teacher. Matching secondary students to all their teachers through course rosters, we can exploit within-student variation for a given year to identify how teacher characteristics affect a student's likelihood of receiving a referral, an identification strategy we detail in Section 3. Second, disciplinary incidents are far more common in secondary than elementary school; this is the more policy relevant context and provides adequate identifying variation. For example, during the 2017-18 school year, the average middle school student received 0.05 office referrals per year, compared to 0.02 for the average elementary student.

We merge the various administrative data sets on office referrals, suspensions, student and teacher demographics, and student course enrollment to create our main analytic sample, which is at the student-teacher-year level. Table 1 presents summary statistics on student demographics and their outcomes for our entire analytic sample and also separately by race. Panel A reports statistics at the student-by-year level. The district is racially diverse: Asian (43%) and Hispanic (27%) students are the two largest student subgroups and account for the majority of the student body, with the rest remainder being 11% white, 7% Black, and 12% who self identify as multi-racial or for whom we have missing race/ethnicity information. About 14% of students receive special education. Based on neighborhood poverty rates, we classify students' neighborhoods into quartiles and label them as poorest, poor, less poor, and least poor. It is evident that students of color, especially Black students, are more likely to receive special education, reside in the poorest neighborhoods, and have low math and reading test scores. About 11% of students received at least one referral during a given year, though this rate varies dramatically by race as well: Black students were almost six times more likely to receive a referral than white students and almost two times more likely than Hispanic students.

Panel B of Table 1 summarizes the analytic sample at the student-teacher-year level. The focus here is on the teacher characteristics that are the main educational inputs (independent variables) of interest and the student outcomes specific to individual teachers. We also summarize teacher characteristics at the teacher-by-year level in Appendix Table A1. One characteristic of interest is the demographic representation of the teaching force, as prior research finds significant, arguably causal effects of same-race teachers on a variety of student outcomes, including achievement (Dee, 2004a), suspensions (Holt & Gershenson, 2019; Lindsay & Hart, 2017), attendance (Tran & Gershenson, 2021), and educational attainment (Gershenson et al., 2018). In our sample, about 11% of student-teacher pairs are of the same gender *and* same race each year, 10% the same race but different gender, 40% same gender but different race, and the rest (39%) different gender and race. Similar to the overall composition of the K-12 teaching force in the U.S., teachers in the focal district are disproportionately white (48%), meaning that white students are significantly more likely than students of color to have a same-race teacher. Indeed, only 9% of Black students have a same-race teacher (5% same race and gender, and 4% same race only), a rate far lower than other racial/ethnic student groups.

Another easily observed teacher characteristic known to improve student performance and attendance is teaching experience (Gershenson, 2016; Ladd & Sorensen, 2017; Papay & Kraft, 2015; Wiswall, 2013). We consider two variables that capture teaching experience, each of which may be relevant in the context of classroom discipline: total teaching experience and experience in the current school.[2] Overall, 19% of teachers in our sample are new to their schools (17% at the student-teacher-year level), but this number varies significantly by student race: about 20% of Black and Hispanic students are in classrooms with a teacher who is new to the school compared to about 15% of white and Asian students. These differences are seen on the intensive margin as well: the average teacher has been in the school for about 7.2 years (7.7 at the student-teacher-year level) but is slightly higher for white (7.8) and

[2]Both measures yield similar results, so we focus on experience in the current school in the main text, while replicating the analysis using total experience in Appendix Table A3.

Asian (8.5) students than for Black (6.4) and Hispanic (6.7) students. Analogous patterns are observed in the total teaching experience variable. These differences are consistent with evidence that teacher turnover rates are higher in schools that serve higher shares of Black and Hispanic students (Hanushek et al., 2004; Lankford et al., 2002).

Other traditional teacher qualifications such as degrees and certificates tend to be only modestly associated with student outcomes (Clotfelter et al., 2007). Overall, about 11% of the analytic sample had a teacher with a masters degree, 49% had a teacher with a credential in English Language Learning (ELL), and 13% had a credential in special education. Most of these credentials, with the exception of special-education, are roughly evenly distributed across students. Black students were more than twice as likely as white students to have teachers with special-education credentials, which is consistent with the higher rates of special education classifications observed among Black students we report above.

In addition to the teacher characteristics and qualifications discussed thus far, we also consider a few other classroom characteristics known to affect achievement, attendance, and attainment, including class size (Cho et al., 2012; Dynarski et al., 2013; Tran & Gershenson, 2021) and course subjects (Whitney & Liu, 2017). We classify course subjects into seven categories, including ELA (15%), math (15%), science (13%), social studies (12%), foreign language classes (5%), physical education (12%), and "other (28%)."[3]

3 Methods

We seek to estimate the impact of observable teacher characteristics on a student's likelihood of receiving an office referral. As in studies of how teacher credentials affect academic achievement, the biggest threat to identification is the non-random sorting of students and teachers into classrooms (Clotfelter et al., 2007, 2010). Indeed, this type of within-school non-random sorting is well documented (Kalogrides & Loeb, 2013; Kalogrides et al., 2013). For

[3]The category "other" includes elective courses such as art, computer science, and vocational courses.

example, if early-career teachers are more likely to teach students with a higher likelihood of misbehaving, we might wrongly conclude that novice teachers cause misbehavior. Leveraging longitudinal data that span four school years and contain multiple teacher-subject pairings for each student in a given year, our primary identification strategy exploits within student-year variation in exposure to different teachers. This strategy helps control for the unobserved student characteristics that are constant across class periods in a given year that affect their propensity to misbehave.

Specifically, we estimate models of the form

$$R_{ijt} = \beta Match_{ijt} + \gamma X_{jt} + \theta_{it} + \epsilon_{ijt}, \tag{1}$$

where i, j, and t index students, teachers, and years, respectively. R is a binary indicator equal to one if the student was referred by a specific teacher in a specific year, and zero otherwise. $Match$ is a set of mutually exclusive indicators for the demographic (race and gender) match between student and teacher, where different race *and* different gender is the omitted group; this allows for intersectionality between race and gender (Gershenson et al., 2016). X is a set of teacher and classroom characteristics including experience, the various credentials summarized in Table A1, class size, class composition, class-period indicators, and subject indicators. θ is a student-year fixed effect (FE), so OLS estimates of β and γ are identified from within student-year (i.e., between subject) variation in teacher characteristics. Standard errors are two-way clustered at the teacher level and student level (Cameron et al., 2011). The student-year FE make student, school, and year FE redundant, as well as other student and school controls that that are constant within a given school-year or student-year.

The validity of OLS estimates of Equation (1) requires that sorting into classrooms is random conditional on the student-year FE. However, the student-year FE do not account for sorting within or across school *days*. There are two ways this may occur. First, there might exist time-of-day sorting of students and teachers into specific class *periods*. The class-period indicators included in X adjust for general time-of-day effects on student behavior (Williams & Shapiro, 2018), but it could still be the case that within a given period, some assignments are made in a strategic, non-random way. To address this concern, in some models we replace X_{jt} with classroom FE (ω_{jt}). An additional benefit of this approach is that it controls away any unobserved teacher qualities that may be correlated with teacher demographics (Dee, 2004b). The cost of this exercise, of course, is that because it is identified from within-classroom variation only β is identified because the other teacher qualifications do not vary across the students within a classroom.

Second, prior research suggests that most sorting of students into classrooms is due to prior academic performance (Dieterle et al., 2015), which suggests students might select into different classrooms based on their prior performance in a given subject. We address the potential subject-year specific sorting by adding a subject-specific lagged score to Equation (1). The lagged scores are not co-linear with the student-by-year FE because there are two lagged scores per student year: one in math and one in English language arts (ELA). The trade-off here is that this specification can only be estimated on a smaller analytic sample that contains two observations per student year in the two regularly tested subjects.

4 Results

4.1 Main Results

Table 2 reports estimates of different specifications of Equation (1).[4] Column (1) estimates a lagged dependent variable model that controls for observed student, teacher, and classroom characteristics and an indicator for whether the student received a referral in the prior year. Column (2) adds lagged math and ELA test scores to the model estimated in column (1) to further control for potential sorting based on prior academic performance. Column

[4] As shown in Appendix Table A2, the main results are robust to using a more restrictive sample of classrooms with at least 7 to 32 students. These numbers are the 10th and 90th percentiles of the class size distribution, respectively.

(3) presents estimates of our preferred student-year FE model that exploits between-class, within-year variation for a given student.

All three specifications yield similar results. The demographic match variables in column (3) of Table 2 show that students who share a gender and/or race with their teacher are about 0.2 to 0.5 percentage points (or 7% to 18%) less likely to receive a disciplinary referral than in classes with a demographically mismatched teacher during the same school year. This effect is strongest when students share the same race *and* same gender as their teacher, and this difference is statistically significant. These findings are consistent with prior research on suspensions (Holt & Gershenson, 2019; Lindsay & Hart, 2017).[5]

The teacher experience coefficients are also remarkably consistent across columns (1)-(3) in Table 2 and statistically significant. Following Wiswall (2013), our baseline models include a linear term of experience (years in current school), offset by an indicator for new to school that allows us to identify whether novice teachers are particularly likely to engage in referring students.[6] The preferred estimates in column (3) suggest that students taught by new-to-school teachers are 0.4 percentage points (or 14%) more likely to receive a disciplinary referral than when they are taught by more experienced teachers during the same school year. The coefficient on the linear years-of-experience term is small, but negative and statistically significant, suggesting that students are less likely to receive referrals as their teachers progress in their careers.

To compare the linear effects of teacher experience with alternative specifications, we visualize their differences in Figure 1. Specifically, in addition to the linear effect, we also plot estimates from quadratic and non-parametric specifications of teaching experience. The figure suggests that regardless of the specification, the effects of experience on disciplinary referrals concentrate in the first few years of teaching, especially the first year. Also, such

[5]To examine the intensive margin, we use total referrals from the teacher as the dependent variable and find similar results; see Appendix Table A3.

[6]Appendix Table A4 mimics Table 2, but instead measures experience as total years teaching. The results are quite similar regardless of how experience is measured.

effects attenuate over time and approach zero at about three to five years of experience based on the quadratic and non-parametric models. There are a few reasons why new-to-school teachers may be more likely to issue disciplinary referrals than their more senior colleagues, such as lacking classroom management experience. These results indicate that these teachers might need targeted support such as coaching or peer mentoring to assist in their handling of students' behavioral issues.

Finally, three types of teacher certifications appear to affect disciplinary referrals: compared to teachers with none of the five credentials (ELL, special education, English, math, and science), having a teacher with an ELL or special education credential reduces the chances a student receives a referral. This finding is intuitive, as communication and classroom management skills are often a particular focus of these programs. English credentials have a marginally significant, modest positive effect, though it is unclear why this is the case. There is no evidence that certifications in math or science affect student referrals.

As described in the methods section, we test the validity of the estimates of Equation (1) in two ways. First, we adjust for possibly endogenous unobserved classroom or teacher characteristics by adding classroom FE to the model, which subsume the class period indicators, subject FE, and observed teacher and classroom characteristics. These estimates are reported in Column (4) of Table 2. The three demographic-match indicators remain strongly jointly significant and the same-gender and same-gender and race point estimates remain similarly sized and individually significant. However, the same-race indicator shrinks and loses statistical significance, though remains negative. The reason is likely that there is too little variation within classrooms, particularly for non-white students, to separately identify all three demographic-match effects.

Second, we adjust for dynamic sorting into classrooms by adding a subject-specific lagged score to Equation (1). This requires restricting the sample to math and ELA classrooms, as these are the only tested subjects. As reported in Appendix Table A5, we find qualitatively similar results using this stacked lag score specification. This provides further support to our assumption that subject-year specific sorting does not bias our baseline estimates.

4.2 Heterogeneity

We allow for heterogeneity along several dimensions by estimating the baseline model on separate subsamples. We start by considering specific reasons for the referral (i.e., infraction type). Many referrals are the result of multiple infractions, so we follow Lindsay & Hart (2017) in coding five mutually exclusive categories based on the "most severe" reason listed for the referral: violence; drugs; interpersonal offenses; defiance; class skipping or walkout. For example, a referral of a student who skipped class and was disruptive would be coded as disruption. The impact of teacher traits might vary by referral reason because teachers may vary in either their ability to de-escalate certain types of situations or vary in how they perceive the severity of more subjective infractions, such as defiance.

The analysis by referral reason reported in Table 3 yields a few interesting findings. First, the demographic-match, experience, and special-education certificate effects primarily load on referrals for defiance. For example, about half of the demographic-match effect, 75% of the new-to-school teacher effect, and 80% of the special-education credential effect are driven by defiance referrals. This is consistent with extant evidence that defiance referrals are arguably the most subjective and most prone to racial biases on the part of the referrer (Barrett et al., 2019; Girvan et al., 2017; Skiba et al., 2002). Indeed, the California State Legislature introduced a statewide ban on willful defiance suspensions in grades K–3 in response to the prevalence and disproportionate occurrence of such suspensions among Black and Hispanic students (ACLUNorCal, 2014). Shortly thereafter, several large districts in the state banned willful defiance suspensions in all grades (Wang, 2022). Our results suggest that teachers who share the same demographic background as their students and who are more experienced and/or trained in dealing with student behavioral issues might be better able to mitigate their own biases and cultivate trusting relationships with students.

Referrals for violence are also more strongly influenced by teacher traits, particularly the demographic match variables. Same-race teachers reduce the likelihood of a student receiving a referral due to violence by 0.3 percentage points. Compared to behaviors like defiance, violence infractions are arguably subject to less subjective judgement and are more severe in nature, which in certain circumstances require a disciplinary referral. This suggests that if the results for defiance referrals are mostly explained by teachers' biases, the results for violence referrals might work through actual reductions in violent behavior.

We also estimate the preferred student-year FE specification given in equation (1) separately for different student demographic groups. The rationale is that students of color may be particularly affected by having a teacher of the same demographic group and, more generally, there are pronounced differences by race and sex in referral rates. Overall, we find that same-race and same-gender teachers have larger effects on Black, Hispanic, and male students than on their white, Asian, and female counterparts.[7] This is consistent with prior research on teachers' effects on student suspensions (Holt & Gershenson, 2019).

Additionally, we allow for intersectional heterogeneity by estimating the preferred student-year FE specification given in equation (1) separately by both race *and* gender. As shown in Table 4, the male, Black, and Hispanic demographic-match effects observed in Appendix Table A6 are driven by Black males, Hispanic males, and to a lesser extent Asian males. The increased referral probability due to new-to-school teachers, however, is driven by Black female as opposed to Black male students. This effect is roughly similar for male and female Hispanic students.

Finally, Appendix Table A7 investigates possible heterogeneity by school type. Specifically, we estimate the preferred student-year FE specification given in equation (1) separately for middle and high schools, higher and lower performing schools, and schools serving more and less advantaged students. A comparison of columns 1 and 2 shows that many of the teacher effects discussed to this point mostly occur in middle schools. This is consistent with

[7] See Appendix Table A6 for the full set of estimates.

the results reported in Table 3, as violence and defiance infractions are more common in middle schools than in high schools (Liu, Hayes, & Gershenson, 2022).

The remaining columns of Appendix Table A7 estimate the baseline model separately for different quartiles of schools. Columns 3 and 4 use the bottom and top quartile schools of the achievement distribution, where achievement is measured using math standardized test scores.[8] With the exception of ELL certification, the teacher effects are concentrated in the lowest performing schools. Similarly, in columns 5 and 6 we split the sample into the least and most economically advantaged schools based on neighborhood poverty rates among enrolled students. Once again, with the exception of ELL certification, the effects of teacher characteristics on student referrals are strongest in the most disadvantaged schools. That these effects tend to be largest in schools serving the lowest performing and most disadvantaged students suggests that many disciplinary referrals are marginal in the sense that they (and the associated negative consequences) might be driven by implicit biases and avoided entirely if those students had access to more effective teachers.

4.3 Linking to Suspensions

Prior research finds that student-teacher demographic match reduces the likelihood of student suspensions (e.g., (Holt & Gershenson, 2019; Lindsay & Hart, 2017; Shirrell et al., 2021). One of the main findings of the current study is that student-teacher demographic match similarly reduces the likelihood of receiving office referrals. This is consistent with extant evidence on suspensions, of course, and we now extend our analysis of the effects of teachers' observed qualifications on referrals to see how they directly affect the likelihood of being suspended. This exercise is useful for at least two reasons. First, replicating the established result that students receive fewer suspensions when assigned to a same-race teacher in a new context reinforces the general importance of teacher diversity for behavioral as well as academic outcomes; it also cross validates our dataset and our referral results. Second,

[8] We find similar results if achievement is instead measured using ELA standardized test scores.

if we observe effects of these teacher characteristics on actual suspensions, this implies that the additional referrals associated with said characteristics matter in the sense that many of them ultimately converted into suspensions.

Student-teacher demographic match does indeed affect the likelihood of a referral converting into a suspension, as reported in Appendix Table A8. Overall, students of the same gender and race as their teacher are about 0.04 of a percentage point (17%) less likely to receive a disciplinary referral that converts to a suspension than when they have a demographically mismatched teacher in a different class during the same school year. This effect is strongest for defiance referrals and for Black male students. This effect size is an order of magnitude smaller than the most comparable estimate in Holt & Gershenson (2019), who find the effect to be about 0.5 of a percentage point. However, Holt & Gershenson (2019) study self-contained elementary school classrooms, while the current study is set in middle and high schools in which students change classrooms throughout the day. Given students in our sample on average have seven periods per school day, we scale the current estimates up by a factor of seven, which yields a more similar, and more comparable, estimate of about 0.3 of a percentage point.[9]

Overall, these findings add to the existing evidence base that access to demographically matched teachers reduces the occurrence of exclusionary discipline, particularly among students of color, at the middle and high school level. Coupled with our earlier results for referrals, this implies that effects on suspensions are in large part driven by higher rates of office referrals. Importantly, then, providing all students with access to a representative and diverse set of teachers who have the skills and training to manage classroom discipline can reduce both the frequency and disproportionate occurrence of office referrals and ultimately, suspensions.

[9] Appendix Table A9 reports similar results on the intensive margin (i.e., the count of suspensions).

5 Conclusion

This study estimates the impact of teacher characteristics on the likelihood that students receive a disciplinary referral from said teacher. Using detailed administrative data from a large urban school district, we investigate the impact of student-teacher demographic match, teacher experience, teacher degree level, and teacher credentials on the likelihood of receiving a referral, the total number of referrals, and having at least one disciplinary referral convert into a suspension. Students who share a gender or race with their teacher are about 7% to 18% less likely to receive a disciplinary referral than from other teachers during the same school year. Heterogeneity analyses show that this effect is largest for defiance referrals, middle schools, schools with more economically disadvantaged students, and Black and Hispanic male students. Similarly, student-teacher demographic match also reduces the total number of referrals received by a student and the likelihood that students are ultimately suspended. These findings are consistent with extant evidence that student-teacher demographic match reduces the likelihood of student suspensions (Holt & Gershenson, 2019; Lindsay & Hart, 2017; Shirrell et al., 2021).

Teacher experience also affects the likelihood that students receive office referrals: specifically, the likelihood of receiving a disciplinary referral is 14% higher when students are taught by a novice teacher. Once again this effect is largest for defiance referrals and for Black and Hispanic students. Interestingly, the effect of experience fades out fairly quickly and approaches zero after about five years of teaching experience.

Finally, we find limited evidence that certain teacher certifications affect the likelihood and frequency of disciplinary referrals. Specifically, having teachers with ELL and special education credentials reduces the chances that students receive disciplinary referrals. This is likely due to the classroom management and communication skills taught in these programs. Interestingly, teachers with an English credential are more likely to refer a student to the office, and this effect is largest for Hispanic male students.

A limitation of these analyses is that we cannot identify the precise mechanisms through which teacher characteristics and qualifications affect student referrals. It is likely that two non-mutually exclusive channels are in play. First, teachers likely vary in their use of office referrals as a disciplinary tool both because they vary in their interpretation of classroom behavior and in their sense of how productive referrals, which carry the risk of exclusionary discipline, will be (Girvan et al., 2017; Okonofua et al., 2016). Second, teachers likely affect actual student behavior, making referrals more or less necessary in certain classrooms, both by teaching social emotional skills and by changing the classroom climate (Jackson, 2018; Kraft, 2019). It would be fruitful for future research to examine whether, and how much, each channel contributes to the effects on office referrals because each provides different policy implications. Another useful area for future research is to investigate the curricular aspects of ELL and special education certification programs that may be associated with the observed effect of these certifications on student referrals.

The question of exact mechanisms notwithstanding, the reduced-form findings of the current study do offer some guidance for policy and practice. At a basic level, our results provide concrete evidence that teachers play a pivotal role in the production of suspensions, as referrals necessarily precede suspensions. The heterogeneous effects and differential access to teachers with different qualifications documented here indicate that socio-demographic disparities in suspensions are at least partly due to teachers and not solely biases in the adjudication process (Liu, Hayes, & Gershenson, 2022). This further bolsters the importance of recruiting and retaining a diverse teaching force that is representative of the student body in its charge (Gershenson et al., 2021). The findings on teaching experiences highlight the importance of mentoring, coaching, and discussing school disciplinary protocols and practices with teachers who are both new to teaching and new to the school.

References

ACLUNorCal. (2014). *California enacts first-in-the-nation law to eliminate student suspensions for minor misbehavior.* Retrieved from https://www.aclunc.org/news/california-enacts-first-nation-law-eliminate-student-suspensions-minor-misbehavior (Last accessed on 3/20/2022)

Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime.

Barrett, N., McEachin, A., Mills, J. N., & Valant, J. (2019). Disparities and discrimination in student discipline by race and family income. *Journal of Human Resources*, 0118–9267R2.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, *104*(9), 2633–79.

Chin, M. J. (2021). Desegregated but still separated? The impact of school integration on student suspensions and special education classification. *Journal of Urban Economics*.

Cho, H., Glewwe, P., & Whitler, M. (2012). Do reductions in class size raise students' test scores? Evidence from population variation in minnesota's elementary schools. *Economics of Education Review*, *31*(3), 77–95.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of education review*, *26*(6), 673–682.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655–681.

Davison, M., Penner, A. M., Penner, E. K., Pharris-Ciurej, N., Porter, S. R., Rose, E. K., ... Yoo, P. (2022). School discipline and racial disparities in early adulthood. *Educational Researcher*, *51*(3), 231–234.

Dee, T. S. (2004a). Are there civic returns to education? *Journal of public economics*, *88*(9-10), 1697–1720.

Dee, T. S. (2004b). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics*, *86*(1), 195–210.

Dieterle, S., Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management*, *34*(1), 32–58.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of policy Analysis and management*, *32*(4), 692–717.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125–149.

Gershenson, S., Hansen, M., & Lindsay, C. A. (2021). *Teacher diversity and student success: Why racial representation matters in the classroom.* Harvard Education Press.

Gershenson, S., Hart, C., Hyman, J., Lindsay, C., & Papageorge, N. W. (2018). The long-run impacts of same-race teachers (no. w25254). *National Bureau of Economic Research. Retrieved January*, *15*, 2019.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of education review*, *52*, 209–224.

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.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of human resources*, *39*(2), 326–354.

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. *Available at SSRN 4093118*.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107.

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, *42*(6), 304–316.

Kalogrides, D., Loeb, S., & Béteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, *86*(2), 103–123.

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

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, *54*(1), 1–36.

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review*, *28*(1), 49–57.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241–279.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62.

Lindsay, C. A., & Hart, C. M. (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.

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.

Liu, J., & Loeb, S. (2021). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, *56*(2), 343–379.

Liu, J., Penner, E. K., & Gao, W. (2022). Troublemakers? The role of frequent teacher referrers in expanding racial disciplinary disproportionalities (edworkingpaper: 22-616).

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.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105–119.

Shi, Y., & Zhu, M. (2022). Equal time for equal crime? Racial bias in school discipline. *Economics of Education Review*, *88*, 102256.

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., 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)

Sorensen, L. C., Bushway, S. D., & Gifford, E. J. (in press). Getting tough? The effects of discretionary principal discipline on student outcomes. *Education Finance and Policy*.

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

Tran, L., & Gershenson, S. (2021). Experimental estimates of the student attendance production function. *Educational Evaluation and Policy Analysis*, *43*(2), 183–199.

Wang, R. (2022). The impact of suspension reforms on discipline outcomes: Evidence from california high schools. *AERA Open*, *8*, 23328584211068067.

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

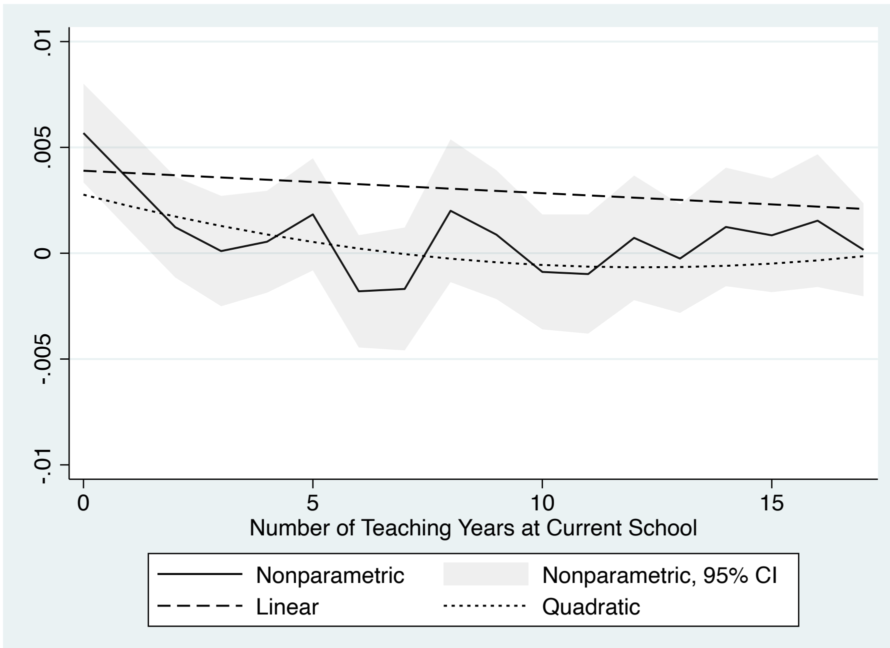
Whitney, C. R., & Liu, J. (2017). What we're missing: A descriptive analysis of part-day absenteeism in secondary school. *AERA Open*, *3*(2), 2332858417703660.

Williams, K. M., & Shapiro, T. M. (2018). Academic achievement across the day: Evidence from randomized class schedules. *Economics of Education Review*, *67*, 158–170.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, *100*, 61–78.

Figures and Tables

Figure 1: The Effects of Teaching Experiences on Disciplinary Office Referrals



Notes: Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is at the student-teacher-year level. Omitted group are teachers with more than 17 years of experience at current school. The p-values for the zero teaching years at current school indicator are less than 0.00 and 0.02 for the linear and quadratic models, respectively.

Table 1: Student Characteristics at the Student-Year Level

	All Students	Student Race Comparison				
		White	Black	Hispanic	Asian	Other
Panel A – Student-year level						
Female	0.48	0.49	0.50	0.46	0.49	0.48
White	0.11	1.00				
Black	0.07		1.00			
Hispanic	0.27			1.00		
Asian	0.43				1.00	
Other race	0.12					1.00
Special education	0.14	0.13	0.33	0.20	0.08	0.11
Middle school	0.41	0.48	0.41	0.42	0.39	0.40
High school	0.59	0.52	0.59	0.58	0.61	0.60
Resides in poorest neighborhood	0.15	0.05	0.34	0.18	0.13	0.14
Resides in poor neighborhood	0.18	0.15	0.13	0.21	0.18	0.16
Resides in less poor neighborhood	0.16	0.18	0.09	0.13	0.18	0.17
Resides in least poor neighborhood	0.16	0.27	0.08	0.13	0.17	0.19
Missing poverty data	0.35	0.34	0.36	0.36	0.35	0.34
Lagged Math Score	0.02	0.34	-0.50	-0.39	0.28	0.05
	[0.78]	[0.69]	[0.78]	[0.74]	[0.67]	[0.75]
Lagged Reading Score	0.01	0.42	-0.47	-0.33	0.20	0.06
	[0.78]	[0.72]	[0.79]	[0.75]	[0.69]	[0.76]
Missing Test Score	0.39	0.34	0.44	0.42	0.35	0.42
Lagged Non-Cumulative GPA	3.14	3.40	2.56	2.76	3.41	3.19
	[0.80]	[0.60]	[0.92]	[0.86]	[0.60]	[0.74]
Missing Non-Cumulative GPA	0.21	0.24	0.24	0.24	0.16	0.24
At least one referral this year	0.11	0.06	0.34	0.18	0.03	0.11
At least one referral last year	0.06	0.03	0.25	0.11	0.01	0.06
Student-Year Observations	107,361	11,751	7,774	29,500	45,783	12,553
Panel B – Student-teacher-year level						
Same student-teacher race and gender	0.11	0.26	0.05	0.09	0.12	0.00
Same student-teacher race only	0.10	0.25	0.04	0.09	0.12	0.00
Same student-teacher gender only	0.40	0.25	0.46	0.42	0.38	0.51
Zero years at current school	0.17	0.16	0.22	0.20	0.14	0.17
Years of experience at current school	7.68	7.81	6.39	6.65	8.51	7.73
	[6.48]	[6.46]	[6.03]	[6.08]	[6.67]	[6.50]
Total years of teaching experience	11.90	12.18	10.46	10.76	12.78	11.99
	[8.70]	[8.65]	[8.42]	[8.41]	[8.82]	[8.71]
At least one referral from teacher this year	0.03	0.01	0.11	0.05	0.01	0.03
Total referrals from teacher this year	0.06	0.02	0.26	0.09	0.01	0.05
	[0.50]	[0.27]	[1.11]	[0.64]	[0.18]	[0.47]
Student-teacher-year observations	719,096	77,679	52,162	196,924	307,668	84,663

Notes: Standard deviations are reported in brackets for all non-binary variables. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The "other" race category includes multiracial individuals and students missing race data. All the statistics above are reported as proportions, except for the non-binary variables.

Table 2: Regressions on the Likelihood of Referring

	(1)	(2)	(3)	(4)
Lagged Referral	0.195***	0.188***		
	(0.005)	(0.005)		
Same Race and Gender	-0.004***	-0.004***	-0.005***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Same Race Only	-0.003***	-0.003***	-0.004***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Same Gender Only	-0.002***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Zero Experience at Current School	0.004***	0.004***	0.004***	
	(0.001)	(0.001)	(0.001)	
Experience at Current School	-0.000	-0.000*	-0.000**	
	(0.000)	(0.000)	(0.000)	
Master's Degree	0.000	0.000	0.001	
	(0.001)	(0.001)	(0.001)	
Credential in ELL	-0.001*	-0.001*	-0.002**	
	(0.001)	(0.001)	(0.001)	
Credential in Special Education	-0.004**	-0.003	-0.005**	
	(0.002)	(0.002)	(0.002)	
Credential in English	0.002*	0.002*	0.002**	
	(0.001)	(0.001)	(0.001)	
Credential in Math	-0.000	-0.000	-0.000	
	(0.001)	(0.001)	(0.001)	
Credential in Science	0.001	0.001	0.000	
	(0.001)	(0.001)	(0.001)	
Lagged Student Test Score		-0.003***		
		(0.001)		
Class Size	-0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Class Average Lagged GPA	0.016**	0.025***	0.019***	
	(0.001)	(0.001)	(0.001)	
Class Average Lagged Referral rate	0.944***	0.950***	1.003***	
	(0.011)	(0.011)	(0.012)	
Average Referral Rate		0.028		
Controls for:				
Time-Varying Controls	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	
Lagged Student Achievement		✓		
Student by Year FEs			✓	✓
Class Period Indicators	✓	✓	✓	
Classroom FEs				✓
Joint significance tests (p-values)				
Student-Teacher Match Indicators	0.000***	0.000***	0.000***	0.001***
Adjusted R-squared	0.245	0.247	0.317	0.313
Observations	719,096	719,096	719,096	719,096

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.

Table 3: Regressions on the Likelihood of Referring by Referral Reason

	All	Violence	Drugs	Interper	Defiance	Walkout
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race and Gender	-0.005***	-0.002***	-0.000	-0.001	-0.002***	-0.001**
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Same Race Only	-0.004***	-0.001**	-0.000*	-0.000	-0.002***	-0.001
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Same Gender Only	-0.002***	0.000	-0.000	-0.001***	-0.001***	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Zero Experience at Current School	0.004***	0.001**	-0.000***	0.001**	0.003***	-0.001
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Experience at Current School	-0.000**	-0.000*	-0.000	-0.000	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Master's Degree	0.001	-0.000	0.000	0.001	0.001	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Credential in ELL	-0.002**	-0.001***	-0.000	-0.001**	0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Credential in Special Education	-0.005**	0.002*	-0.000**	0.002	-0.004**	-0.003***
	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Credential in English	0.002**	0.001*	-0.000	0.000	0.001	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Credential in Math	-0.000	0.001	-0.000**	-0.001	-0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Credential in Science	0.000	0.000	-0.000	0.001	-0.001	0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Class Size	0.000	0.000	-0.000	-0.000	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Class Average Lagged GPA	0.019***	0.003***	0.000**	0.004***	0.008***	0.005***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Class Average Lagged Referral rate	1.003***	0.125***	0.015***	0.258***	0.395***	0.197***
	(0.012)	(0.007)	(0.003)	(0.010)	(0.013)	(0.011)
Average Referral Rate	0.028	0.004	0.001	0.007	0.011	0.005
Controls for:						
School by Year FEs	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓
Joint significance tests (p-values)						
Student-Teacher Match Indicators	0.000***	0.001***	0.253	0.001***	0.001***	0.170
Adjusted R-squared	0.317	0.068	0.027	0.102	0.119	0.082
Observations	719,096	719,096	719,096	719,096	719,096	719,096

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. Walkout includes both walkouts and skipping class. We find no practically or statistically significant results when the model is run separately for "other" referral reasons, and therefore, those results are not included in the table for brevity reasons. p<0.10* p<0.05** p<0.01***.

Table 4: Regressions on the Likelihood of Referring by Both Race and Gender

	Black		Hispanic		Asian	
	Female	Male	Female	Male	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race and Gender	-0.028*	-0.060***	-0.001	-0.013***	0.001	-0.001
	(0.014)	(0.012)	(0.003)	(0.003)	(0.001)	(0.001)
Same Race Only	-0.012	-0.046***	-0.008***	-0.003	0.001*	0.001
	(0.012)	(0.013)	(0.003)	(0.003)	(0.000)	(0.001)
Same Gender Only	0.004	-0.011**	-0.001	-0.006***	0.001**	-0.002**
	(0.004)	(0.005)	(0.002)	(0.002)	(0.000)	(0.001)
Zero Experience at Current School	0.010*	-0.002	0.009***	0.007**	0.000	0.001
	(0.006)	(0.006)	(0.002)	(0.003)	(0.001)	(0.001)
Experience at Current School	-0.000	-0.001*	-0.000	-0.000**	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Master's Degree	0.001	-0.003	0.002	0.001	0.000	0.001
	(0.007)	(0.007)	(0.002)	(0.003)	(0.000)	(0.001)
Credential in ELL	-0.003	-0.011**	-0.001	-0.004**	-0.001***	-0.002***
	(0.004)	(0.005)	(0.002)	(0.002)	(0.000)	(0.001)
Credential in Special Education	-0.004	-0.007	-0.008**	-0.010**	-0.001	0.003
	(0.009)	(0.009)	(0.004)	(0.005)	(0.002)	(0.003)
Credential in English	0.020***	0.001	0.002	0.008**	0.001	0.001
	(0.005)	(0.007)	(0.002)	(0.003)	(0.000)	(0.001)
Credential in Math	0.007	0.008	-0.003	-0.002	-0.000	-0.001
	(0.008)	(0.007)	(0.003)	(0.004)	(0.001)	(0.001)
Credential in Science	0.002	0.001	0.002	0.006	-0.002***	0.001
	(0.008)	(0.008)	(0.003)	(0.004)	(0.001)	(0.001)
Class Size	0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Class Average Lagged GPA	0.036***	0.052***	0.017***	0.022***	0.005***	0.008***
	(0.006)	(0.006)	(0.002)	(0.002)	(0.001)	(0.001)
Class Average Lagged Referral rate	1.278***	1.275***	0.872***	1.152***	0.286***	0.536***
	(0.038)	(0.030)	(0.030)	(0.023)	(0.035)	(0.034)
Average Referral Rate	0.087	0.131	0.031	0.059	0.003	0.010
Controls for:						
Time-varying Controls	✓	✓	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓
Joint significance tests (p-values)						
Student-Teacher Match Indicators	0.083*	0.000***	0.036**	0.001***	0.060	0.022**
Adjusted R-squared	0.379	0.398	0.251	0.307	0.123	0.161
Observations	25,906	25,256	90,191	106,733	151,512	156,156

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.

Appendix A

Table A1: Teacher Characteristics at the Teacher-Year Level

	Mean
Female	0.54
White	0.48
Black	0.06
Hispanic	0.14
Asian	0.19
Other Race	0.13
0 years of exper at current school	0.19
# years of exper at current school	7.15
	[6.33]
Master's degree	0.11
Missing data on teacher education	0.06
Credential in ELL	0.49
Credential in special education	0.13
Credential in English	0.22
Credential in math	0.17
Credential in science	0.14
Missing data on teacher credential	0.01
Teacher-Year Observations	6,397

Notes: Standard deviations are reported in brackets for all non-binary variables. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the teacher-year level. The "other" race category includes multiracial individuals and individuals missing race data, and for this reason, "other" race students are coded a 0 for the same race as teacher indicators. All the statistics above are reported as proportions, except for the non-binary variables.

Table A2: Regressions on the Likelihood of Referring Using Classrooms with 7 to 32 Students

	(1)	(2)	(3)	(6)
Lagged Referral	0.214***	0.207***		
	(0.006)	(0.005)		
Same Race and Gender	-0.004***	-0.004***	-0.005***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Same Race Only	-0.002**	-0.002**	-0.004***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Same Gender Only	-0.002***	-0.002***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Zero Total Teaching Years	0.005***	0.005***	0.005***	
	(0.001)	(0.001)	(0.001)	
Total Teaching Experience	-0.000	-0.000	-0.000	
	(0.000)	(0.000)	(0.000)	
Master's Degree	0.000	0.000	0.001	
	(0.001)	(0.001)	(0.001)	
Credential in ELL	-0.001	-0.001	-0.002**	
	(0.001)	(0.001)	(0.001)	
Credential in Special Education	-0.007**	-0.006*	-0.003	
	(0.003)	(0.003)	(0.003)	
Credential in English	0.002	0.002	0.002**	
	(0.001)	(0.001)	(0.001)	
Credential in Math	-0.001	-0.001	-0.000	
	(0.001)	(0.001)	(0.001)	
Credential in Science	0.001	0.001	0.000	
	(0.002)	(0.002)	(0.001)	
Lagged Student Test Score		-0.003***		
		(0.001)		
Class Size	-0.000***	-0.000***	-0.000***	
	(0.000)	(0.000)	(0.000)	
Class Average Lagged GPA	0.016***	0.025***	0.021***	
	(0.001)	(0.001)	(0.001)	
Class Average Lagged Referral rate	1.038***	1.044***	1.079***	
	(0.019)	(0.019)	(0.020)	
Average Referral Rate		0.027		
Controls for:				
Time-Varying Controls	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	
Lagged Student Achievement		✓		
Student by Year FEs			✓	✓
Class Period Indicators	✓	✓	✓	
Classroom FEs				✓
Joint significance tests (p-values)				
Student-Teacher Match Indicators	0.000***	0.000***	0.000***	0.001***
Adjusted R-squared	0.214	0.217	0.286	0.296
Observations	590,795	590,795	590,795	590,795

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. The sample is restricted to classrooms with a class size between the 10th percentile (7 students) and the 90th percentile (32 students). All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.

Table A3: Regressions on the Total Referrals from Teacher

	(1)	(2)	(3)	(6)
Lagged Total Referrals	0.044***	0.043***		
	(0.003)	(0.003)		
Same Race and Gender	-0.012***	-0.012***	-0.015***	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
Same Race Only	-0.009***	-0.009***	-0.013***	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)
Same Gender Only	-0.004**	-0.004**	-0.004***	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Zero Experience at Current School	0.014***	0.013***	0.012***	
	(0.004)	(0.004)	(0.004)	
Experience at Current School	0.000	0.000	-0.000	
	(0.000)	(0.000)	(0.000)	
Master's Degree	-0.002	-0.002	0.000	
	(0.003)	(0.003)	(0.002)	
Credential in ELL	-0.001	-0.001	-0.003	
	(0.003)	(0.003)	(0.003)	
Credential in Special Education	0.001	0.006	0.007	
	(0.010)	(0.010)	(0.009)	
Credential in English	0.006	0.005	0.006*	
	(0.004)	(0.004)	(0.003)	
Credential in Math	-0.000	-0.000	-0.000	
	(0.004)	(0.004)	(0.004)	
Credential in Science	0.001	0.001	0.001	
	(0.005)	(0.005)	(0.005)	
Lagged Student Test Score		-0.008***		
		(0.002)		
Class Size	-0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Class Average Lagged GPA	0.034***	0.062***	0.044***	
	(0.004)	(0.004)	(0.003)	
Class Average Lagged Referral rate	2.443***	2.454***	2.387***	
	(0.083)	(0.083)	(0.077)	
Average Total Referrals	0.057			
Controls for:				
Time-Varying Controls	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	
Lagged Student Achievement		✓		
Student by Year FEs			✓	✓
Class Period Indicators	✓	✓	✓	
Classroom FEs				✓
Joint significance tests (p-values)				
Student-Teacher Match Indicators	0.002***	0.001***	0.000***	0.078*
Adjusted R-squared	0.170	0.172	0.293	0.342
Observations	719,096	719,096	719,096	719,096

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.

Table A4: Regressions on the Likelihood of Referring Using Total Teaching Years

	(1)	(2)	(3)	(4)
Lagged Referral	0.195***	0.188***		
	(0.005)	(0.005)		
Same Race and Gender	-0.004***	-0.005***	-0.006***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
Same Race Only	-0.003***	-0.003***	-0.004***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Same Gender Only	-0.002***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Zero Total Teaching Years	0.003**	0.003**	0.003**	
	(0.001)	(0.001)	(0.001)	
Total Teaching Experience	-0.000**	-0.000**	-0.000***	
	(0.000)	(0.000)	(0.000)	
Master's Degree	0.001	0.001	0.001	
	(0.001)	(0.001)	(0.001)	
Credential in ELL	-0.001*	-0.001*	-0.002**	
	(0.001)	(0.001)	(0.001)	
Credential in Special Education	-0.005**	-0.003	-0.005**	
	(0.002)	(0.002)	(0.002)	
Credential in English	0.002*	0.002*	0.003**	
	(0.001)	(0.001)	(0.001)	
Credential in Math	-0.000	-0.000	-0.000	
	(0.001)	(0.001)	(0.001)	
Credential in Science	0.001	0.001	0.001	
	(0.001)	(0.001)	(0.001)	
Lagged Student Test Score		-0.003***		
		(0.000)		
Class Size	-0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Class Average Lagged GPA	0.015***	0.025***	0.019***	
	(0.001)	(0.001)	(0.001)	
Class Average Lagged Referral rate	0.945***	0.951***	1.003***	
	(0.011)	(0.011)	(0.012)	
Average Referral Rate		0.028		
Controls for:				
Time-Varying Controls	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	
Lagged Student Achievement		✓		
Student by Year FEs			✓	✓
Class Period Indicators	✓	✓	✓	
Classroom FEs				✓
Joint significance tests (p-values)				
Student-Teacher Match Indicators	0.000***	0.000***	0.000***	0.001***
Adjusted R-squared	0.244	0.247	0.317	0.313
Observations	719,096	719,096	719,096	719,096

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.

Table A5: Regressions on the Likelihood of Referring Using Subject-Specific Lag Test Scores

	All Students		Black Students Only	
	Excludes Lag	Includes Lag	Excludes Lag	Includes Lag
	(1)	(2)	(3)	(4)
Same Race and Gender	-0.003**	-0.003*	-0.021	-0.021
	(0.002)	(0.002)	(0.014)	(0.014)
Same Race Only	-0.001	-0.001	-0.014	-0.014
	(0.002)	(0.002)	(0.016)	(0.016)
Same Gender Only	-0.003***	-0.003***	-0.006	-0.005
	(0.001)	(0.001)	(0.006)	(0.007)
Zero Experience at Current School	0.003*	0.003*	0.001	0.001
	(0.002)	(0.002)	(0.010)	(0.010)
Experience at Current School	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.001)
Master's Degree	0.001	0.001	-0.009	-0.009
	(0.001)	(0.001)	(0.011)	(0.011)
Credential in ELL	-0.001	-0.001	-0.008	-0.008
	(0.001)	(0.001)	(0.008)	(0.008)
Credential in Special Education	-0.017***	-0.017***	-0.055***	-0.055***
	(0.006)	(0.006)	(0.015)	(0.015)
Credential in English	0.008***	0.008***	0.016	0.016
	(0.003)	(0.003)	(0.012)	(0.012)
Credential in Math	-0.001	-0.001	-0.010	-0.010
	(0.003)	(0.003)	(0.013)	(0.013)
Credential in Science	0.004	0.004	0.029*	0.029*
	(0.003)	(0.003)	(0.016)	(0.016)
Lagged Student Test Score		-0.001		-0.013
		(0.002)		(0.011)
Class Size	-0.000	-0.000	0.001	0.001
	(0.000)	(0.000)	(0.001)	(0.001)
Class Average Lagged GPA	0.019***	0.019***	0.040***	0.040***
	(0.002)	(0.002)	(0.009)	(0.009)
Class Average Lagged Referral rate	0.957***	0.957***	1.267***	1.268***
	(0.018)	(0.018)	(0.043)	(0.043)
Average Referral Rate	0.032		0.127	
Controls for:				
Time-Varying Controls	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	✓
Lagged Student Achievement		✓		✓
Student by Year FEs	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓
Joint significance tests (p-values)				
Student-Teacher Match Indicators	0.025**	0.027**	0.430	0.450
Adjusted R-squared	0.383	0.383	0.432	0.432
Observations	212,450	212,450	14,497	14,497

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. The sample is restricted to classrooms teaching either a math or ELA subject All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. $p<0.10^*$ $p<0.05^{**}$ $p<0.01^{***}$.

Table A6: Regressions on the Likelihood of Referring by Student Race and Gender

	White	Black	Hispanic	Asian	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race and Gender	-0.004*	-0.046***	-0.008***	-0.001	-0.001	-0.009***
	(0.002)	(0.009)	(0.003)	(0.001)	(0.001)	(0.001)
Same Race Only	-0.004*	-0.029***	-0.005**	0.001	-0.003**	-0.004***
	(0.002)	(0.009)	(0.002)	(0.001)	(0.001)	(0.001)
Same Gender Only	-0.003***	-0.004	-0.004***	-0.001	0.001	-0.004***
	(0.001)	(0.003)	(0.001)	(0.000)	(0.001)	(0.001)
Zero Experience at Current School	0.003*	0.004	0.008***	0.001	0.004***	0.003**
	(0.002)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)
Experience at Current School	-0.000	-0.001**	-0.000	-0.000**	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Master's Degree	0.002	-0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.006)	(0.002)	(0.000)	(0.001)	(0.001)
Credential in ELL	-0.001	-0.007**	-0.003*	-0.001***	-0.001	-0.003***
	(0.001)	(0.004)	(0.001)	(0.000)	(0.001)	(0.001)
Credential in Special Education	0.000	-0.005	-0.008**	0.002	-0.006**	-0.005*
	(0.004)	(0.007)	(0.004)	(0.002)	(0.002)	(0.003)
Credential in English	0.000	0.011**	0.005**	0.001	0.002***	0.002
	(0.002)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)
Credential in Math	-0.001	0.008	-0.002	-0.000	-0.001	0.000
	(0.002)	(0.006)	(0.003)	(0.001)	(0.001)	(0.002)
Credential in Science	0.000	0.002	0.004	-0.000	-0.001	0.002
	(0.002)	(0.006)	(0.003)	(0.001)	(0.001)	(0.002)
Class Size	0.000	0.000*	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Class Average Lagged GPA	0.012***	0.045***	0.020***	0.007***	0.017***	0.021***
	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)
Class Average Lagged Referral rate	0.741***	1.276***	1.048***	0.428***	0.876***	1.084***
	(0.044)	(0.025)	(0.019)	(0.025)	(0.020)	(0.014)
Average Referral Rate	0.013	0.109	0.046	0.006	0.019	0.037
Controls for:						
Time-varying Controls	✓	✓	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓
Joint significance tests (p-values)						
Student-Teacher Match Indicators	0.049**	0.000***	0.001***	0.071*	0.029**	0.000***
Adjusted R-squared	0.209	0.395	0.292	0.152	0.298	0.324
Observations	77,679	52,162	196,924	307,668	346,342	372,754

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. $p<0.10$* $p<0.05$** $p<0.01$***.

Table A7: Regressions on the Likelihood of Referring by School Level Characteristics

	Grade Level		Average Test Scores		Poverty Level	
	Middle	High	Quartile 1	Quartile 4	Quartile 4	Quartile 1
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race and Gender	-0.009***	-0.003**	-0.011***	-0.001*	-0.011***	-0.001
	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)
Same Race Only	-0.008***	-0.000	-0.009***	-0.000	-0.006**	0.000
	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)
Same Gender Only	-0.004***	-0.002***	-0.004***	-0.001	-0.003**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Zero Experience at Current School	0.010***	-0.001	0.008***	-0.001	0.005**	-0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Experience at Current School	-0.000**	-0.000	-0.000*	-0.000**	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Master's Degree	-0.002	0.001	0.001	0.001**	0.003	0.001**
	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)
Credential in ELL	-0.004***	-0.000	-0.001	-0.003***	-0.002	-0.002***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Credential in Special Education	-0.006	-0.005**	-0.008	0.005	-0.005	0.009*
	(0.005)	(0.002)	(0.005)	(0.006)	(0.004)	(0.005)
Credential in English	0.006***	-0.000	0.008***	0.001	0.004**	-0.000
	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
Credential in Math	0.001	-0.000	0.002	-0.001	0.003	-0.001
	(0.002)	(0.001)	(0.004)	(0.001)	(0.003)	(0.001)
Credential in Science	0.001	-0.001	0.004	-0.001	-0.001	-0.001
	(0.002)	(0.001)	(0.004)	(0.001)	(0.003)	(0.001)
Class Size	-0.000	0.000	-0.000*	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Class Average Lagged GPA	0.051***	0.013***	0.039***	0.018***	0.021***	0.009***
	(0.003)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
Class Average Lagged Referral rate	1.002***	1.011***	1.020***	0.906***	1.005***	0.947***
	(0.019)	(0.013)	(0.017)	(0.035)	(0.017)	(0.033)
Average Referral Rate	0.041	0.020	0.064	0.008	0.044	0.005
Controls for:						
Time-varying Controls	✓	✓	✓	✓	✓	✓
School by Year FEs	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓
Joint significance tests (p-values)						
Student-Teacher Match Indicators	0.000***	0.002***	0.001***	0.251	0.001***	0.528
Adjusted R-squared	0.333	0.289	0.338	0.278	0.361	0.280
Observations	278,098	440,998	177,073	180,309	178,426	174,260

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The omitted student-teacher match group are students with a different gender and race than their teacher. Quartile 1 represents observations from schools with below the 25th percentile for a particular variable. Quartile 4 include observations from schools with above the 75th percentile for a particular variable. $p<0.10$* $p<0.05$** $p<0.01$***.

Table A8: Regressions on the Likelihood of At Least One Referral Converted into a Suspension

	Panel A – By Referral Reason					
	All	Violence	Drugs	Interper	Defiance	Walkout
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race and Gender	-0.0004*	-0.0001	-0.0000	0.0000	-0.0002*	-0.0000
	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Same Race Only	-0.0002	-0.0001	-0.0000	-0.0000	-0.0002	0.0001
	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Same Gender Only	-0.0001	0.0002**	-0.0000	-0.0001	-0.0002***	0.0000
	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Average Suspension Rate	0.0023	0.0005	0.0001	0.0007	0.0007	0.0003
Adjusted R-squared	0.0759	0.0362	0.00500	0.0370	0.00915	0.0112
Observations	719,096	719,096	719,096	719,096	719,096	719,096

	Panel B – By Race and Gender					
	All	Black		Hispanic		
		Male	Female	Male	Female	
	(1)	(2)	(3)	(4)	(5)	
Same Race and Gender	-0.0004*	-0.0117***	0.0062*	-0.0003	0.0000	
	(0.0002)	(0.0037)	(0.0036)	(0.0007)	(0.0006)	
Same Race Only	-0.0002	-0.0073*	0.0022	0.0006	0.0002	
	(0.0002)	(0.0039)	(0.0040)	(0.0008)	(0.0007)	
Same Gender Only	-0.0001	0.0007	-0.0009	0.0003	-0.0002	
	(0.0001)	(0.0018)	(0.0013)	(0.0005)	(0.0003)	
Average Suspension Rate	0.0023	0.0155	0.0087	0.0045	0.0017	
Adjusted R-squared	0.0759	0.0952	0.0621	0.0788	0.0343	
Observations	719,096	26,256	25,906	106,733	90,191	
Controls for:						
School by Year FEs	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.

Table A9: Regressions on the Total Referrals Converted into Suspension

	Panel A – By Referral Reason					
	All	Violence	Drugs	Interper	Defiance	Walkout
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race and Gender	-0.0006**	-0.0002	-0.0000	0.0001	-0.0003***	-0.0001
	(0.0003)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Same Race Only	-0.0005*	-0.0002**	-0.0000	0.0000	-0.0002*	-0.0000
	(0.0003)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Same Gender Only	-0.0002	0.0001	-0.0000	-0.0001	-0.0002**	-0.0000
	(0.0002)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Average # of Suspensions	0.0026	0.0006	0.0001	0.0008	0.0007	0.0004
Adjusted R-squared	0.0802	0.0522	0.0210	0.0399	0.0099	0.0110
Observations	719,096	719,096	719,096	719,096	719,096	719,096

	Panel B – By Race and Gender					
	All	Black		Hispanic		
		Male	Female	Male	Female	
	(1)	(2)	(3)	(4)	(5)	
Same Race and Gender	-0.0006**	-0.0158***	0.0077*	-0.0001	0.0000	
	(0.0003)	(0.0051)	(0.0043)	(0.0009)	(0.0007)	
Same Race Only	-0.0005*	-0.0092**	0.0022	0.0004	0.0000	
	(0.0003)	(0.0041)	(0.0040)	(0.0009)	(0.0007)	
Same Gender Only	-0.0002	-0.0012	-0.0005	0.0006	-0.0003	
	(0.0002)	(0.0023)	(0.0014)	(0.0006)	(0.0004)	
Average # of Suspensions	0.0026	0.0180	0.0094	0.0050	0.0019	
Adjusted R-squared	0.0802	0.0916	0.0652	0.0831	0.0467	
Observations	719,096	26,256	25,906	106,733	90,191	
Controls for:						
School by Year FEs	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓

Notes: Clustered-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for student, teacher, and classroom characteristics. Data come from a large urban school district in California between 2016-17 to 2019-20 school years. The unit of analysis is the student-by-teacher-by-year level. The omitted student-teacher match group are students with a different gender and race than their teacher. p<0.10* p<0.05** p<0.01***.