Economics of Education Review 93 (2023) 102376

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Who refers whom? The effects of teacher characteristics on disciplinary office referrals

Michael S. Hayes [a], Jing Liu [b], Seth Gershenson [c,*]

[a] Rutgers University – Camden, United States
[b] University of Maryland – College Park and IZA, United States
[c] American University and IZA, United States

ARTICLE INFO

Keywords:
Exclusionary discipline
Teacher effectiveness
Office referrals

ABSTRACT

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 same-race teachers and other observed teacher 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 that Black students' probability of receiving at least one referral is about 3 percentage points (26.6% of Black students' base rate) smaller than for white students when they have a Black teacher versus a white teacher. The reduced likelihoods of receiving referrals from same-race teachers also convert to reduced likelihoods of being suspended. These results are mostly driven by referrals for violence, interpersonal offences, and walkout infractions, middle school students, and students from high-poverty schools. 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. 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.

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 et al., 2022a). Specifically, the 2013–14 Civil Rights Data Collection finds that Black students accounted for 40% of suspensions but only 16% of enrollments. 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, Billings, & Deming, 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 et al., 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 disciplinary decisions, in shaping racial disparities in exclusionary discipline; however, 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,

JEL Classification: I2The authors thank APPAM participants and seminar participants at Syracuse University for helpful discussions. Wenjing Gao and Joyce Koo provided 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.

* Corresponding author.
E-mail address: gershens@american.edu (S. Gershenson).

https://doi.org/10.1016/j.econedurev.2023.102376
Received 20 October 2022; Received in revised form 15 January 2023; Accepted 13 February 2023
Available online 24 February 2023

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–and specifically that on the impact of same-race teachers–by examining teachers' impacts on office referrals. There are two reasons that teachers might vary in their 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).

Teachers vary in how they perceive student (mis)behavior (Dee, 2005). However, 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 made 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 et al. (2022b) 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 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 causal quantitative evidence on how having a teacher of the same race affects a student's likelihood of receiving an office referral. We also investigate how these effects vary by student subgroups and how some other observable teacher characteristics, such as experience and credentials, affect referral probabilities. 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., 2022, 2021). 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 track 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 who initiated the referral. These data allow us to estimate stacked two-way fixed effect (FE) regression models that compare (i) students across subjects (classrooms) within a given year and (ii) students within a single classroom (Fairlie et al., 2014). The student-by-year FE controls for unobserved time-invariant student traits as well as idiosyncratic shocks that might influence classroom assignments. Classroom FE control for other teacher or classroom characteristics that may be associated with teacher race.

Our findings suggest that Black students are significantly less likely to receive a disciplinary referral from Black teachers than from teachers of other racial backgrounds. We also provide suggestive evidence that having teachers who are more experienced or who hold a credential in teaching English language learners (ELL) or special education reduce referral rates. The race-match results are mainly driven by referrals for violence or interpersonal infractions in middle schools and in schools serving low-income neighborhoods. These 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 et al., 2022a). 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 and white teachers in diverse schools, 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, on Monday April 2nd, in the library). 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 pol- icy relevant context and provides adequate identifying variation. For example, during the 2017–18 school year, the average middle

[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).

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 present statistics 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 remainder being 11% white, 7% Black, and 12% who self identify as multi-racial or for whom we are 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 (in- dependent 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., 2022). In our sample, about 21% of student-teacher pairs are of the same race each year. 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, 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 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 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.

3. Methods

Our primary objective is to estimate the causal effect of having a same-race teacher on a student's likelihood of receiving an office referral. There are two main threats to identification. First, as in any analysis of how teachers or teachers' characteristics affect student outcomes, we worry about the non-random sorting of students and teachers into particular classrooms (Kalogrides & Loeb, 2013; Kalogrides et al., 2013). Second, teachers' race and ethnicity may be correlated with other teacher or classroom characteristics that influence referral rates. Following Fairlie et al. (2014), we account for these concerns by conditioning on student-by-year and classroom fixed effects (FE), respectively.

Specifically, we estimate two-way FE models of the form:

$$R_{ijct} = \beta Match_{ijt} + \gamma_c + \theta_{it} + \varepsilon_{ijct}, \tag{1}$$

where i, j, c, and t index students, teachers, classrooms, 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 4×4 set of student-teacher race interactions (including Black, Hispanic, Asian, and "other"), where white is the omitted reference group. Five mutually exclusive race categories for both students and teachers mean that 25 unique student-teacher race combinations exist in the data. However, the student-by-year (θ) and classroom (γ) FE subsume the individual student and teacher demographic indicators, respectively, along with other observed and unobserved year-specific characteristics that might otherwise confound our estimate of β.[3] This means that only 16, not 24, differential effects are identified via the 16 student-teacher race interaction terms in Eq. (1), which are all interpreted relative to white teachers and students (Fairlie et al., 2014). For example, the coefficient on the Black-student $\times$ Black-teacher interaction term represents how the effect of having a Black teacher relative to a white teacher differs for Black students relative to white students.

Eq. (1) is identified because, like in the community college context studied by Fairlie et al. (2014), in any given school year the middle and high school students in the district take multiple unique classes, each typically taught by a different teacher. We therefore exploit both within student-year variation in student i's exposure to different teachers *and* within- classroom variation in whether students share the same race as the teacher. Standard errors are two-way clustered at the teacher level and student level (Cameron et al., 2011).

The validity of OLS estimates of Eq. (1) requires that sorting into classrooms is conditionally random. The student-year FE account for both time-invariant and year-specific factors at the student level that influence sorting uniformly across classes. The classroom FE accounts for potential time-of-day sorting of students and teachers into specific class *periods* (Williams & Shapiro, 2018) as well as more general sorting into certain types of classrooms (e.g., subjects, peer groups, class size). An additional benefit of the classroom FE is that they control away any unobserved teacher qualities that may be correlated with teacher demographics (Dee, 2004b). The cost of doing so, of course, is that the

[2] Both measures yield similar results, so we report estimates for experience in the current school.

[3] Student-year FE make student, school, and year FE redundant, as well as other student and school controls that are constant within a given school-year or student-year. Similarly, classroom FE subsume controls for class size, composition, and so on.

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	7774	29,500	45,783	12,553
Panel B – Student-teacher-year level						
Same student-teacher race	0.21	0.51	0.09	0.17	0.24	0.00
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 and 17 and 2019–20. The “other” race cat- egory includes multiracial individuals and students missing race data. Statistics are reported as proportions except for the case of non-binary variables.

effect of year-invariant characteristics like teaching experience are not identified because they do not vary across the students within a classroom. In Section 4.3, we estimate variants of Eq. (1) that replace the classroom FE with a vector of observed teacher and classroom characteristics and provide suggestive evidence on how a handful of teacher qualifications affect referrals.

4. Results

4.1. Main results

Table 2 reports our baseline estimates of Eq. (1). Panel A reports estimates for any referral while subsequent panels report estimates for specific types of referrals. Many referrals are the result of multiple infractions, so we follow Lindsay and 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 was violent and defiant would be coded as a violence referral. The impact of teacher representation 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.

Each panel reports a matrix of nine interaction coefficients; the other-race student and other-race teacher interactions are omitted for brevity but are identified and estimated in the model. Estimates on the diagonal of each matrix, in bold, are the same-race interactions of primary interest. The Black-Black interaction term in the top left corner of Panel A is both large and statistically significant. It indicates that the effect of having a Black teacher (versus a white teacher) on the probability of receiving at least one referral is about 3 percentage points (26.6% of Black students’ base referral rate) smaller for Black students than for white students. Similarly, replacing a white teacher with a Black teacher significantly reduces Hispanic students’ referral likelihoods relative to those of white students. However, there is no discernible race-match effect on Hispanic students’ referrals.[4]

Panels B through F of Table 2 show that the race-match effect for Black students observed in Panel A was approximately evenly due to referrals for all types of infractions except for drug use/possession, perhaps because this is the most objectively observable infraction type. The Black teacher-Hispanic student effect was largely driven by infractions due to defiance and walkout.

Interestingly, the coefficient on the Asian student-Asian teacher interaction term in Panel A is positive and statistically significant. This seemingly contradicts general evidence of student-teacher race match improving educational outcomes. On the one hand, this might be a real effect, perhaps driven by growing evidence that white teachers tend to hold a positive “model minority” bias toward Asian students (Shi & Zhu, 2021) that leads Asian teachers to be harsher toward Asian students for whom they hold unrealistically high expectations. On the other hand,

[4] To examine the intensive margin, we use total referrals from the teacher as the dependent variable and find similar results; see Appendix Table A2.

Table 2
Estimated effect of teacher race/ethnicity on likelihood of referral by referral type.

	Panel A - All Referrals			Panel B - Violence Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.030*****	−0.004	0.001	**−0.006****	−0.001	0.002
Hispanic	−0.009**	**0.001**	−0.001	−0.001	**−0.000**	0.001
Asian	0.005	0.003	**0.003****	0.000	−0.000	**0.001**
White Referral Rate	0.018	0.016	0.011	0.002	0.002	0.003
F-test: (p-value)		0.004			0.090	
	Panel C - Drug Referrals			Panel D - Interpersonal Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**0.001**	−0.000	−0.000	**−0.011*****	0.002	−0.004
Hispanic	−0.000	**−0.000**	−0.001*	0.001	**0.002**	−0.001
Asian	0.000	−0.000	−0.000	0.003**	0.002**	**0.001****
White Referral Rate	0.000	0.000	0.000	0.003	0.004	0.003
F-test: (p-value)		0.346			0.002	
	Panel E – Defiance Referrals			Panel F - Walkout Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.006**	−0.003	0.001	**−0.007****	−0.001	0.001
Hispanic	−0.005*	**−0.002**	−0.001	−0.003**	**0.001**	0.000
Asian	0.001	0.001	**0.000**	−0.000	0.000	**0.000**
White Referral Rate	0.011	0.008	0.004	0.003	0.002	0.002
F-test: (p-value)		0.535			0.201	

Notes: The unit of analysis is the student-by-teacher-by-year level. The analytical sample contains 719,096 observations. Each panel reports estimates from regressions that interact all observed student and teacher races/ethnicity indicators and condition on both classroom and student-year fixed effects. We do not report identified interactions for "other race" teachers and students. Same race/ethnicity interactions are in bold. P-values for an F-test of the existence of same-race/ethnicity interactions are reported. The white referral rate is conditional on teacher race type. Standard errors are clustered at both the teacher and student levels. $p<0.10$* $p<0.05$** $p<0.01$***.

however, we do not want to over-interpret this coefficient given its small magnitude. Indeed, it may be a precisely estimated zero, given that Asians are the biggest student subgroup (43%) and second biggest teacher subgroup. The latter notion is supported by the fact that in subsequent panels of Table 2 all but one of the Asian-Asian interaction terms are small and statistically indistinguishable from zero.

In sum, the strong result for Black students in Table 2 is broadly consistent with prior research on suspensions (Holt & Gershenson, 2019; Lindsay & Hart, 2017). To cross validate our referral results, and to extend the evidence of same-race teacher effects on suspensions outside of North Carolina, we now re-estimate the baseline model given by Eq. (1) using an indicator for "ever suspended" as the outcome variable, which takes the value of 1 if at least one referral issued by a given teacher to the focal student converted to a suspension, and takes the value of 0 otherwise. This exercise is also interesting in the sense that it shows whether the "marginal" referrals created by student-teacher race match manifest into additional suspensions. These results are presented in Table 3.

Panel A estimates the baseline model using all suspensions regardless of the reason for the referral that led to the suspension. Mirroring the referral results, the Black student-Black teacher interaction term is large in magnitude and strongly statistically significant. It indicates that the effect of having a Black teacher (versus a white teacher) on the probability of receiving at least one suspension is about half of a percentage point (41.7% of Black students' base suspension rate) smaller for Black students than for white students. This means that some of the "extra" referrals received by Black students as a result of teacher race did in fact convert into suspensions. However, the Asian student-Asian teacher interaction term is approximately zero and no longer statistically significant. This confirms that the small uptick in referrals of Asian students assigned to Asian teachers is substantively inconsequential as they did not result in any additional suspensions.

As in Table 2, subsequent panels of Table 3 estimate the model separately for suspensions for different types of infractions and show that the general race-match suspension effect for Black students is driven by interpersonal and defiance infractions. These results are subtly different from the analysis of referrals by infraction type, which show effects on violence and walkout referrals as well. One possible explanation is that principals' decision-making on suspensions is a complex process that is not solely based upon teachers' referrals for a given infraction but depends on many other factors, such as a student's prior discipline history (Liu et al., 2022a). Thus, we do not observe a one-to-one mapping based on infraction types between Tables 2 and 3. The result for defiance is troubling, as it suggests that a subjective measure of misbehavior that is prone to teacher bias is creating additional referrals *and* suspensions for Black students.

4.2. Heterogeneity

We allow for heterogeneity along several dimensions by estimating the baseline model on different subsamples in Table 4. In panels A and B, we estimate Eq. (1) separately for male and female students, respectively. This is motivated by a baseline gender gap in referral rates and the general finding that boys and girls are affected differently by school quality and their household's socioeconomic status (Bertrand & Pan, 2013; Figlio et al., 2016, 2019). Consistent with this literature, the

Table 3
Estimated effect of teacher race/ethnicity on likelihood of suspension by referral type.

	Panel A - All Referrals			Panel B - Violence Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.005****	−0.000	−0.001	**−0.000**	0.001	0.001
Hispanic	0.000	**0.001**	−0.000	−0.000	**−0.000**	0.000
Asian	0.001*	−0.000	**−0.000**	0.000	−0.000	**0.000**
White Suspension Rate	0.001	0.001	0.001	0.000	0.000	0.000
F-test: (p-value)		0.084			0.278	
	Panel C - Drug Referrals			Panel D - Interpersonal Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**0.001**	−0.000	0.000	**−0.002***	−0.000	−0.000
Hispanic	0.000	**0.000**	−0.000	0.000	**0.000**	−0.000
Asian	0.000**	0.000	**−0.000**	0.001**	0.000	**0.000**
White Suspension Rate	0.000	0.000	0.000	0.000	0.000	0.000
F-test: (p-value)		0.557			0.1456	
	Panel E – Defiance Referrals			Panel F - Walkout Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.003****	−0.001	0.000	**−0.001**	−0.000	−0.002***
Hispanic	−0.000	**0.000**	−0.000	0.000	**0.000**	0.000
Asian	−0.000	−0.000	**−0.000**	0.000	−0.000	**−0.000**
White Suspension Rate	0.001	0.000	0.000	0.000	0.000	0.000
F-test: (p-value)		0.097			0.180	

Notes: The unit of analysis is the student-by-teacher-by-year level. The analytical sample contains 719,096 observations. Each panel reports estimates from regressions that interact all observed student and teacher races/ethnicity indicators and condition on both classroom and student-year fixed effects. We do not report identified interactions for "other race" teachers and students. Same race/ethnicity interactions are in bold. P-values for an F-test of the existence of same-race/ethnicity interactions are reported. The white suspension rate is conditional on teacher race type. Standard errors are clustered at both the teacher and student levels. $p<0.10$* $p<0.05$** $p<0.01$***.

race-match effect for Black males is slightly larger than that for Black females, though the difference is not statistically significant.

Similarly, panels C and D of Table 4 estimate the model separately for students in schools in the top and bottom quartiles of the poverty rate distribution based on students' neighborhood economic conditions. Again, this exercise is motivated by a baseline gap between school type in referral rates and the fact that student-teacher race-match effects tend to be largest for the most economically disadvantaged students (Gershenson et al., 2022). As expected, the race-match effect is nearly three times larger, and only statistically significant, in the schools serving students from the most disadvantaged neighborhoods. A similar result is found if we instead estimate the model separately for the top and bottom quartiles of the achievement distribution.[5]

Finally, panels E and F of Table 4 estimate the model separately for middle and high schools, respectively. This exercise is motivated by the fact that the baseline level of disciplinary infractions is significantly higher in middle schools and that the mix of infraction types differs by grade level (Liu et al., 2022a). The race-match effect is twice as large in middle schools as in high schools, though it is sizable and statistically significant in both. This is consistent with the fact that the types of disciplinary infractions most common in middle school are more subjective (e.g., defiance and noncompliance) (Liu et al., 2022a). It could also be that older students' behavior and other outcomes are less responsive to teacher effectiveness and schooling inputs generally (Jackson, 2014).

Once again, we now cross validate the heterogeneous referral results by re-estimating the heterogeneity models for suspensions rather than referrals. These results are presented in Table 5. They are broadly consistent with the referral results discussed in Table 4, though in some cases yield even larger differences across groups. This is most apparent when comparing genders, as the race-match effect on suspensions for Black boys is eight times larger than for Black girls. The effect on Black girls approaches zero and is no longer statistically significant. The results by school type parallel those for referrals: student-teacher race-match effects on Black students' suspension rates are almost entirely due to effects in schools serving low- income neighborhoods (and in lower-achieving schools). And the effects in middle schools are more than four times larger than those in high schools.

4.3. Other teacher qualifications

We conclude our empirical analysis by providing some suggestive evidence that several other observable teacher characteristics affect the likelihood that a student receives a disciplinary referral. This builds on an existing literature that documents effects of various teacher credentials and qualifications, most notably teacher experience, on student achievement (Clotfelter et al., 2007, 2010; Papay & Kraft, 2015; Wiswall, 2013). We do so by estimating versions of Eq. (1) that replace the classroom fixed effect with a vector of observed teacher and classroom covariates. The resulting estimates are suggestive in the sense that removing the classroom FE opens the door for potential confounding factors to bias the estimates.

[5] This is unsurprising because school poverty rates are strongly correlated with school achievement levels.

Table 4
Estimated Effect of Teacher Race/Ethnicity on Likelihood of Referral by Student Type.

	Panel A - Males Only			Panel B - Females Only		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.034*****	−0.004	0.001	**−0.029****	−0.004	0.002
Hispanic	−0.009*	**0.001**	−0.000	−0.012**	**−0.001**	−0.002
Asian	0.006	0.007**	**0.004***	0.003	−0.001	**0.001**
Observations		372,754			346,342	
White Referral Rate	0.024	0.023	0.016	0.012	0.010	0.007
F-test: (p-value)		0.009			0.078	
	Panel C – Highest Poverty			Panel D - Lowest Poverty		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.035****	−0.008	0.003	**−0.014**	0.004	0.003
Hispanic	−0.004	**0.003**	−0.002	−0.000	**−0.002**	−0.001
Asian	0.021**	0.010	**0.003**	−0.002	−0.000	**0.001**
Observations		174,260			178,274	
White Referral Rate	0.018	0.039	0.020	0.009	0.003	0.005
F-test: (p-value)		0.120			0.636	
	Panel E - Middle School			Panel F - High School		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.046****	0.000	0.011	**−0.019***	−0.007	−0.005
Hispanic	−0.009	**−0.001**	0.002	−0.009**	**0.001**	−0.003
Asian	0.011	0.002	**0.005***	0.003	0.004*	**0.001**
Observations		278,098			440,998	
White Referral Rate	0.018	0.019	0.017	0.018	0.014	0.007
F-test: (p-value)		0.021			0.213	

Notes: The unit of analysis is the student-by-teacher-by-year level. Each panel reports estimates from regressions that interact all observed student and teacher races/ethnicity indicators and con- dition on both classroom and student-year fixed effects. We do not report identified interactions for "other race" teachers and students. Same race/ethnicity interactions are in bold. P-values for an F-test of the existence of same-race/ethnicity interactions are reported. The white referral.

We estimate this specification for the full sample and on the sub-sample of Black students, as the student-teacher race-match effects were most pronounced for Black students and referral rates were highest for Black students. The estimated coefficients for the teacher characteristics of interest, along with two-way clustered 95% confidence intervals, are plotted in Appendix Fig. A1.[6] For the most part, effects are similar in the full sample and Black student samples, though estimates for the latter are less precisely estimated due to the dramatic drop in sample size.

Teaching experience significantly predicts student referrals. We include a linear "years at school" variable as well as an indicator for "new to school." This allows for an initial learning curve with linear returns thereafter.[7] There is a notable increase in the chances of being referred when assigned to a novice (new to school) teacher of about 0.4 percentage points; thereafter, the chances drop by about 0.01 percentage points per year. This is likely due to some combination of teachers' classroom management practices improving and teachers relaxing their threshold for issuing referrals over time. The effects for Black students are quite similar, though less precisely estimated.

The other notable teacher qualifications associated with lower referral rates are credentials in English as a second language and special education, which reduce the likelihood a student receives a referral by 0.2 and 0.5 percentage points, respectively. 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 that is even larger for Black students, though it is unclear why this is the case. There is no evidence that certifications in math or science systematically affect student referrals. Table A1, Table A2, Table A3 and Table A4

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 race match, teacher experience, and teacher credentials on the likelihood of receiving a referral, the total number of referrals, and the likelihood of having at least one suspension. Black students paired with a Black teacher are significantly less likely to receive disciplinary referrals *and* suspensions than from other teachers during the same school year. Heterogeneity analyses show that this effect is largest for male students, in middle schools, and in lower achieving and more economically disadvantaged schools. These findings are broadly 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).

We also find suggestive evidence that teaching experience affects the likelihood that students receive office referrals. Novice teachers and

[6] Exact point estimates and standard errors are reported in Appendix Table A3. Appendix Table A4 reports analogous regression results separately by referral reason.

[7] Using total teaching experience and a novice teacher indicator yield similar results. Non-parametric specifications in Appendix Figure A2 show that after the first year the effect is approximately linear.

Table 5
Estimated Effect of Teacher Race/Ethnicity on Likelihood of Suspension by Student Type.

	Panel A - Males Only			Panel B - Females Only		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.008*****	−0.000	0.001	**−0.001**	0.001	−0.000
Hispanic	0.001	**0.001**	0.001	0.000	**−0.000**	−0.001
Asian	0.001	0.000	**0.001**	0.001**	−0.001	**−0.000**
Observations		372,754			346,342	
White Suspension Rate	0.001	0.001	0.001	0.001	0.000	0.001
F-test: (p-value)		0.009			0.797	
	Panel C – Highest Poverty			Panel D - Lowest Poverty		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.008****	−0.002	0.001	**−0.002**	0.000	−0.001
Hispanic	−0.002	**0.002**	−0.002	−0.001	**0.000**	0.001
Asian	0.001	0.002	**−0.000**	−0.001	−0.000	**−0.000**
Observations		174,260			178,274	
White Suspension Rate	0.003	0.001	0.002	0.001	0.000	0.000
F-test: (p-value)		0.034			0.441	
	Panel E - Middle School			Panel F - High School		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.009****	−0.000	0.000	**−0.002**	−0.001	−0.000
Hispanic	−0.001	**−0.001**	0.001	0.001	**0.001***	−0.001
Asian	0.001	−0.000	**0.000**	0.001	0.000	**−0.000**
Observations		278,098			440,998	
White Suspension Rate	0.001	0.001	0.001	0.001	0.001	0.001
F-test: (p-value)		0.144			0.150	

Notes: The unit of analysis is the student-by-teacher-by-year level. Each panel reports estimates from regressions that interact all observed student and teacher races/ethnicity indicators and con- dition on both classroom and student-year fixed effects. We do not report identified interactions for "other race" teachers and students. Same race/ethnicity interactions are in bold. P-values for an F-test of the existence of same-race/ethnicity interactions are reported. The white suspension rate is conditional on teacher race type. Highest and lowest poverty indicates the top and bottom quartiles of the school-level "neighborhood poverty distribution." Standard errors are clustered at both the teacher and student levels. $p<0.10^{*}$ $p<0.05^{**}$ $p<0.01^{***}$.

teachers who are new to their school in particular are significantly more likely to make referrals. Interestingly, the effect of experience fades out fairly quickly and approaches zero after about five years of teaching experience. There is also 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.

A limitation of our analysis is that we cannot identify the precise mechanisms through which student-teacher race match and other observed teacher 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 et al., 2022a). This further bolsters the importance of recruiting and retaining a diverse teaching force that is representative of the student body in its charge and that is able to effectively teach and communicate with an increasingly diverse student body (Gershenson et al., 2021). The findings on teaching experience 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.

Author statement

All three authors contributed equally to all phases of the project.

Data availability

The authors do not have permission to share data.

Appendix

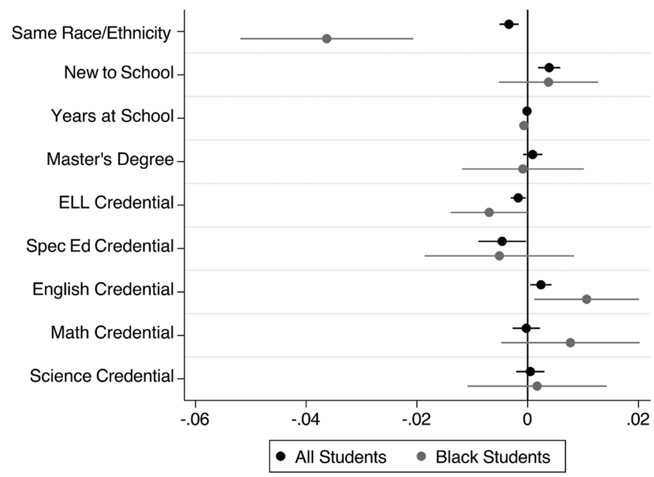

Fig. A1. The Effects of Teacher Characteristics on Disciplinary Office Referrals. Notes: Data come from a large urban school district in California between 2016 and 17 to 2019–20 school years. The unit of analysis is at the student-teacher-year level. Estimates come from a version of the baseline model that replaces classroom fixed effects with teacher and classroom characteristics. The latter include class size, classroom racial composition, classroom average GPA, lagged classroom average GPA, classroom subject, and class period indicators.

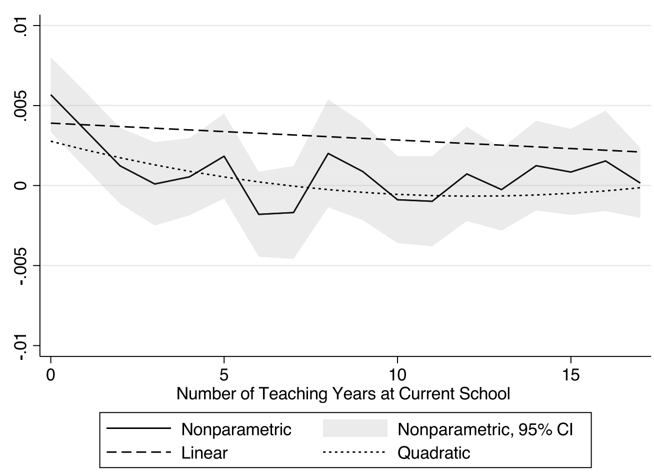

Fig. A2. The Effect of Teaching Experience on Disciplinary Office Referrals. Notes: Data come from a large urban school district in California between 2016 and 17 and 2019–20. The unit of analysis is at the student-teacher-year level. The omitted reference group is teachers with more than 17 years of 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 A1
Teacher Characteristics at the Teacher-Year Level and at the Student-Teacher-Year Level by Student Race.

	All Students	Studen t Race Comparison				
		White	Black	Hispanic	Asian	Other
Female	0.54	0.53	0.52	0.53	0.52	0.53
White	0.48	0.51	0.47	0.47	0.49	0.49
Black	0.06	0.04	0.09	0.07	0.04	0.05
Hispanic	0.14	0.13	0.16	0.17	0.10	0.12
Asian	0.19	0.18	0.15	0.17	0.24	0.20
Other Race	0.13	0.14	0.13	0.13	0.13	0.13
New to School	0.19	0.16	0.22	0.20	0.14	0.17
Years at School	7.15	7.81	6.39	6.65	8.51	7.74
	[6.33]	[6.46]	[6.03]	[6.07]	[6.67]	[6.50]
Master's degree	0.11	0.13	0.10	0.10	0.13	0.12
Missing data on teacher education	0.06	0.05	0.06	0.05	0.05	0.05
Credential in ELL	0.49	0.50	0.50	0.50	0.51	0.51
Credential in special education	0.13	0.02	0.07	0.04	0.02	0.02
Credential in English	0.22	0.23	0.22	0.24	0.25	0.24
Credential in math	0.17	0.21	0.17	0.19	0.22	0.21
Credential in science	0.14	0.17	0.16	0.16	0.19	0.18
Missing data on teacher credential	0.01	0.00	0.01	0.01	0.00	0.00
Teacher-Year Observations	6397					
Student-Teacher-Year Observations		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 and 17 and 2019–20. The unit of analysis is the teacher-year level. The "other" race category includes multiracial individuals and individuals missing race data; for this reason, "other" race students are never coded as "same race". New to school is a indicator that equals 1 if this is the first teaching year at their current school and 0 otherwise. Years at schools is the number of teaching years at current school. All statistics are reported as proportions, except for the case of non-binary variables.

Table A2
Estimated Effect of Teacher Race/Ethnicity on Total Referrals by Referral Type.

	Panel A - All Referrals			Panel B - Violence Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.126*****	−0.013	0.005	**−0.023*****	−0.001	0.002
Hispanic	−0.008	**0.011**	0.007	0.001	**−0.000**	0.006***
Asian	0.017*	0.008*	**0.010*****	0.003	−0.001	**0.002****
White Avg. Referrals	0.028	0.026	0.017	0.002	0.004	0.004
F-test: (p-value)	0.000		0.000			
	Panel C - Drug Referrals			Panel D - Interpersonal Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.001**	−0.001*	−0.001	**−0.043*****	−0.004	−0.011
Hispanic	−0.001	**−0.001***	−0.001**	0.001	**0.007****	−0.001
Asian	0.000	−0.000	−0.000	0.008***	0.005**	**0.003****
White Avg. Referrals	0.000	0.000	0.000	0.004	0.007	0.005
F-test: (p-value)	0.261		0.000			
	Panel E – Defiance Referrals			Panel F - Walkout Referrals		
	Teacher race			Teacher race		
	Black	Hispanic	Asian	Black	Hispanic	Asian
Student race						
Black	**−0.033****	−0.008	0.006	**−0.032****	−0.006	0.011
Hispanic	−0.004	**0.002**	0.001	−0.005	**0.005***	0.001
Asian	0.005	0.004*	**0.002**	0.000	0.001	**0.002****
White Avg. Referrals	0.016	0.011	0.006	0.006	0.003	0.003
F-test: (p-value)	0.035		0.004			

Notes: The unit of analysis is the student-by-teacher-by-year level. The analytical sample contains 719,096 observations. Each panel reports estimates from regressions that interact all observed student and teacher races/ethnicity indicators and condition on both classroom and student-year fixed effects. We do not report identified interactions for "other race" teachers and students. Same race/ethnicity interactions are in bold. P-values for an F-test of the existence of same- race/ethnicity interactions are reported. The white average number of referrals is conditional on teacher race. Standard errors are clustered at both the teacher and student levels. $p<0.10$* $p<0.05$** $p<0.01$***.

Table A3
Estimated Effects of Teacher Qualifications on Likelihood of Referral by Student Type.

	Student Race Category				
	All	White	Black	Hispanic	Asian
	(1)	(2)	(3)	(4)	(5)
Same Race/Ethnicity	−0.003***	−0.002	−0.036***	−0.005**	0.000
	(0.001)	(0.002)	(0.008)	(0.002)	(0.000)
New To School	0.004***	0.003*	0.004	0.008***	0.001
	(0.001)	(0.002)	(0.005)	(0.002)	(0.001)
Years at School	−0.000***	−0.000	−0.001**	−0.000	−0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Master's Degree	0.001	0.002	−0.001	0.002	0.001
	(0.001)	(0.001)	(0.006)	(0.002)	(0.000)
Credential in ELL	−0.002**	−0.001	−0.007*	−0.003*	−0.001***
	(0.001)	(0.001)	(0.004)	(0.001)	(0.000)
Credential in Special Education	−0.005**	0.000	−0.005	−0.008**	0.002
	(0.002)	(0.004)	(0.007)	(0.004)	(0.002)
Credential in English	0.002**	0.000	0.011**	0.005**	0.001
	(0.001)	(0.002)	(0.005)	(0.002)	(0.001)
Credential in Math	−0.000	−0.001	0.008	−0.003	−0.000
	(0.001)	(0.002)	(0.006)	(0.003)	(0.001)
Credential in Science	0.000	0.000	0.002	0.004	−0.000
	(0.000)	(0.002)	(0.006)	(0.003)	(0.001)
Average Referral Rate	0.028	0.013	0.109	0.046	0.006
Controls for:					
Time-varying controls	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓
Adjusted R-squared	0.317	0.211	0.398	0.292	0.153
Observations	719,096	77,679	52,162	196,924	307,668

Notes: Cluster-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for teacher and classroom characteristics. Classroom characteristics include class size, classroom racial composition, classroom average GPA, lagged classroom average GPA, classroom subject, and class period indicators. Data come from a large urban school district in California between 2016 and 17 to 2019–20 school years. The unit of analysis is the student-by-teacher-by-year level. The estimates in columns 1 and 3 are plotted in Appendix Fig. A1. $p<0.10$* $p<0.05$** $p<0.01$***.

Table A4
Estimated Effects of Teacher Qualifications on Likelihood of Referral by Referral Type.

	Referral Reason					
	All	Violence	Drugs	Interpersonal	Defiance	Walk out
	(1)	(2)	(3)	(4)	(5)	(6)
Same Race/Ethnicity	−0.003***	−0.001***	−0.000*	0.000	−0.001**	−0.001**
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
New To 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)
Years at 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
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Average Referral Rate	0.028	0.004	0.001	0.007	0.011	0.005
Controls for:						
Time-varying controls	✓	✓	✓	✓	✓	✓
Student by Year FEs	✓	✓	✓	✓	✓	✓
Class Period Indicators	✓	✓	✓	✓	✓	✓
Adjusted R-squared	0.317	0.066	0.023	0.105	0.117	0.082
Observations	719,096	719,096	719,096	719,096	719,096	719,096

Notes: Cluster-robust standard errors at the teacher level and student level are in parentheses. All regressions include time-varying controls for teacher and classroom characteristics. Classroom characteristics include class size, classroom racial composition, classroom average GPA, lagged classroom average GPA, classroom subject, and class period indicators. Data come from a large urban school district in California between 2016 and 17 to 2019–20 school years. The unit of analysis is the student-by-teacher-by-year level. $p<0.10^{*}$ $p<0.05^{**}$ $p<0.01^{***}$.

References

Bacher-Hicks, A., Billings, S.B., & Deming, D.J. (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime.National Bureau of Economic Research Working Paper No. w26257.

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.

Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics, 5*(1), 32–64.

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), 2633–2679.

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

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., . . ., et al. (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.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95*(2), 158–165.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review, 104*(8), 2567–2591.

Figlio, D., Karbownik, K., Roth, J., Wasserman, M., et al. (2016). School quality and the gender gap in educational achievement. *American Economic Review, 106*(5), 289–295.

Figlio, D., Karbownik, K., Roth, J., Wasserman, M., et al. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics, 11*(3), 338–381.

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. A., & Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy, 14*(4), 300–342.

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

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. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics, 32*(4), 645–684.

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., & Běteille, 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., & 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., Hayes, M. S., & Gershenson, S. (2022a). JUE Insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline. *Journal of Urban Economics*, Article 103453.

Liu, J., Penner, E.K., & Gao, W. (2022). Troublemakers? The Role of Frequent Teacher Referrers in Expanding Racial Disciplinary Disproportionalities. EdWorkingPaper No. 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. (2021). Model minorities in the classroom? Positive bias towards Asian students and its consequences. Center for Policy Research Working Paper No. 253, Center for Policy Research, Maxwell School, Syracuse University.

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

Shirrell, M., Bristol, T. J., & Britton, T. A. (2021). The effects of student-teacher ethno-racial matching on exclusionary discipline for Asian American, black, and Latinx students: Evidence from new york city. ed working paper 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).

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.

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.

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.