Strictly speaking: Examining teacher use of punishment and student outcomes

Stephen B. Holt* Katie Vinopal† Heasun Choi‡ Lucy C. Sorensen§

July 22, 2022

Abstract

While a growing body of literature has documented the negative impacts of exclusionary punishments, such as suspensions, on academic outcomes, less is known about how teachers vary in disciplinary behaviors and the attendant impacts on students. We use administrative data from North Carolina elementary schools to examine the extent to which teachers vary in their use of referrals and investigate the impact of more punitive teachers on student attendance and achievement. We also estimate the effect of teachers' racial bias in the use of referrals on student outcomes. We find more punitive teachers increase student absenteeism and reduce student achievement. Moreover, more punitive teachers negatively affect the achievement of students who do not receive disciplinary sanctions from the teacher. Similarly, while teachers with racial bias in the use of referrals do not negatively affect academic outcomes for White students, they significantly increase absenteeism and reduce achievement for Black students. We find the negative effects of both more punitive and more biased teachers persist into middle school and beyond. The results suggest punitive disciplinary measures do not aid teachers in productively managing classrooms; rather, teachers taking more punitive stances may undermine student engagement and learning in both the short- and long-run. Furthermore, bias in teachers' referral usage contributes to inequities in student outcomes.

Keywords: Teacher effectiveness, Absences, Student discipline

*Corresponding author. University at Albany, SUNY. Email: sbholt@albany.edu
†The Ohio State University
‡University at Albany, SUNY
§University at Albany, SUNY

1 Introduction

The negative effects of exclusionary punishments, such as suspensions and expulsions, on students' academic and long-term outcomes (Bacher-Hicks et al., 2019; Noltemeyer et al., 2015; Sorensen et al., 2021; Hinze-Pifer and Sartain, 2018), as well as large gaps in the use and effects of such practices by student race (Morris and Perry, 2016), are well documented.

Most of this existing work has focused on exclusionary punishment, generally implemented by principals, and not on teachers' discretion in referring students for punishment in the first place. In fact, while racial disproportionality in teachers' use of referrals has been documented (Lindsay and Hart, 2017; Liu et al., 2021a; Skiba et al., 2002), no research to date has estimated how teacher referrals, separate from whether or not they result in suspension, affect student outcomes. Better understanding the extent, variation, and effects of teacher punitiveness is important for at least two reasons.

First, it may have implications for how we train effective teachers. Recent research shows that school-wide reforms to relax discipline can improve safety, student-teacher relationships, and test scores (Craig and Martin, 2019). Similar mechanisms may be operating at the classroom level, even absent such school level reforms: perhaps teachers with less punitive approaches to discipline produce better student outcomes than teachers more likely to rely on referrals, in part due to improved teacher-student relationships. Compared to other more difficult-to-target teacher characteristics, punitiveness is a mutable classroom management practice; optimizing it represents a relatively straightforward opportunity to improve student outcomes.

Second, it has implications for racial equity in the classroom. While existing research shows that Black students are more likely to be suspended and referred than White students, less is known regarding the impact of teacher racial bias in referral behavior.

The current study uses value-added methods with administrative data from North Carolina elementary schools to answer two primary research questions. First, what is the overall effect of teacher punitiveness on student attendance and achievement? Second, what is the

effect of racial bias in teacher punitiveness on those student outcomes? For both questions we do not focus our analysis on attendance and achievement for referred students only, but rather for all students in the class. Thus, our analysis aims to understand the effects of punitiveness and racial bias in that punitiveness on overall classroom outcomes, not just on the outcomes for referred students. We also examine whether the short-term impacts of teacher referral practices on student attendance and achievement persist in the long term.

Our findings indicate that teachers with higher tendency to use disciplinary referrals for subjective infractions increase student absenteeism and reduce standardized math and reading scores. These effects persist for multiple years and hold even when we control for other measures of teacher effectiveness such as value-added scores. Higher teacher punitiveness in elementary school even leads to slight reductions in later high school graduation likelihood and indicators of college enrollment. We further uncover modest negative spillovers of teacher punitiveness onto other students in the classroom, suggesting that more frequent usage of disciplinary referrals has no perceivable benefits for broader classroom management goals. Finally, our estimated measure of teacher racial bias in subjective disciplinary referrals has no short- or long-term impacts on White students, but has negative impacts on the achievement, attendance, high school graduation, and four-year college intentions of Black students. These findings demonstrate the need for more policy attention to teacher disciplinary management issues as one method to improve equity in educational achievement and attainment.

2 Literature Review

Exclusionary school disciplinary practices — primarily suspensions and expulsions — have been associated with negative outcomes for the disciplined student, including lower educational achievement and attainment, weaker attendance, higher dropout rates, and higher rates of criminal activity and unemployment, among others (Chu and Ready, 2018; Lacoe

and Steinberg, 2018; Noltemeyer et al., 2015; Novak, 2021; Pyne, 2019; Sorensen et al., 2021; Wolf and Kupchik, 2017; Anderson et al., 2019). Further, racial disproportionality in school discipline is long- and well-documented, with Black students more likely to be reported for minor infractions as well as exposed to harsh discipline consequences compared to White students, even controlling for student behavior (Amemiya et al., 2020; Barrett et al., 2021; Gregory, 1995; Kinsler, 2011; Skiba et al., 2002, 2011; Shi and Zhu, 2021). Such racial discipline gaps likely contribute to racial achievement gaps (Pearman et al., 2019).

Harsh discipline policies might also positively (through a deterrence effect and the removal of disruptive students) or negatively (through worsening of school climate) affect the broader student body through a spillover effect (Bacher-Hicks et al., 2019; Sorensen et al., 2021). On balance, research has found negative spillover effects (Perry and Morris, 2014). For example, Bacher-Hicks et al. (2019) find that attending schools with high suspension rates is associated with higher rates of arrest and incarceration as adults, as well as lower educational attainment. Craig and Martin (2019) find that the rollback of a strict school discipline policy in New York City led to significant test score gains, and that these gains were driven by improvements in school culture, which benefited all students. And Sorensen et al. (2021) find that students in schools with principals who have a higher propensity to suspend have increased rates of juvenile justice complaints and higher dropout rates.

However, some work suggests the spillover effects of exclusionary discipline may benefit students, and therefore potentially be worth their cost to disciplined students. For example, Lacoe and Steinberg (2018) find that when Philadelphia relaxed its discipline policy, truancy increased and test scores declined. Similarly, Eden (2017) argues that as New York City reformed its discipline policy to reduce suspensions, schools experienced deteriorations in climate, as well as more violence. Using a structural approach, Kinsler (2013) argues that discipline has a positive effect on student achievement, due to the negative spillover effects from uncontrolled disruptive behavior.

While the patterns and effects of school-level disciplinary decisions are increasingly well-

studied, less attention has been paid to the classroom level. In particular, information on how teachers vary in their punitiveness — measured as their use of disciplinary referrals — is limited, and almost nothing is known about the resulting impacts on referred students and their peers. Although teachers may have no choice but to refer students to the principal for very serious offenses, such as assault or the use of a weapon, they exercise considerable discretion over their use of disciplinary referrals for more subjective behavioral infractions, such as tardiness or insubordination. They may also use classroom-level "soft" exclusionary discipline strategies, such as isolated seating or a break outside the classroom (Williford et al., 2021). Thus, even within a school, teachers may be more or less punitive and exhibit more or less racial bias in that punitiveness.

Several studies have documented racial disproportionality in teachers' use of referrals. A number document that Black students are more likely to be referred to the office compared to White students (Santiago-Rosario et al., 2021; Skiba et al., 2002, 2011). Most recently, Liu et al. (2021a) found that Black students are 12 percentage points more likely to receive a referral compared to White students at the same school, suggesting that documented racial disproportionality in suspensions is partially driven by bias in referrals. Further, a large part of disciplinary racial disparities can be attributed to differences in teachers' discretionary responses to subjectively defined behaviors, such as defiance, as opposed to objectively defined behaviors, such as truancy, suggesting bias on the part of teachers and principals (Girvan et al., 2017; Skiba et al., 2002; Smolkowski et al., 2016; Okonofua and Eberhardt, 2015). Work on the effects of racial representation in the teacher workforce implies a similar narrative. For example, Lindsay and Hart (2017) find that having a Black teacher is associated with a reduction in referrals, especially for subjective offenses, and rates of exclusionary discipline for Black students. And Holt and Gershenson (2017) find that students with a teacher of a difference race have more suspensions and a higher likelihood of being suspended at least once, compared to peer with a same-race teacher. More recently, Tran and Gershenson (2021) use data from Project STAR in which students are randomly

assigned to teachers to confirm that assignment to a same-race teacher causally reduces absenteeism among Black students.

Still, research has yet to causally investigate the effects of teachers' punitiveness and bias in that punitiveness on student outcomes. Just as most research has shown that punitive principals and schools — those with high rates of suspensions — lead to a negative school culture and, in turn, negative student outcomes, we might expect punitive teachers — those with high rates of referrals — to lead to a distrustful classroom climate and low-quality teacher-student relationships, and thus worse student outcomes. Indeed, a large body of work has found that positive student-teacher relationships are correlated with improved behavioral and cognitive outcomes for students (Gregory and Korth, 2016; Pianta, 2016; Roorda et al., 2011, 2017; Sabol and Pianta, 2012; Wubbels et al., 2016) and therefore fundamental to student success. Relatedly, a review of the classroom management literature reveals that more relational classroom management styles – those emphasizing building relationships with students and creating positive classroom environments through cooperation and respect — are correlated with higher student achievement (Djigic and Stojiljkovic, 2011), as well as higher observational ratings of teacher performance (Kwok, 2017).

The lack of evidence connecting teachers' use of referrals to student outcomes has partially been due to data limitations. In many states and localities, schools only report suspensions and not the disciplinary referrals that may precede those suspensions. Further, even when disciplinary referrals information is available, most student-level data does not link student referrals to the teacher that assigned them, thus making analyses of teacher-level punitiveness impossible. We overcome this challenge by focusing on self-contained classrooms in grades three through five. Since these classrooms have only one lead teacher, we assume that students' behavioral referrals come from that teacher.[1]

In particular, we begin to fill this gap in the literature by documenting the extent to which teachers vary in their use of referrals and producing estimates of 1) the effect of

[1]Our approach has the added contribution of focusing on younger students, who have typically received less attention in the school discipline literature.

teacher punitiveness on student attendance and test scores and 2) the effect of racial bias in teacher punitiveness on student attendance and test scores. For both, we estimate the short- and long-term direct effects on referred students, as well as the spillover effects on their classmates.

Given that disciplinary referrals are a common tool of classroom management and racial bias in those referrals is documented, our results have important implications for increasing teacher effectiveness as well as racial equity in education.

3 Data

We investigate the academic effects of teachers' use of referrals using teacher- and student-level administrative data in grades 3 through 5 from all North Carolina public schools from 2008 to 2013 obtained from the North Carolina Education Research Data Center (NCERDC). The NCERDC data offers three advantages key to identifying the effect of teachers' discretionary use of disciplinary referrals in the classroom. First, the NCERDC data provides disciplinary referral information for each referral a student receives in a given academic year, including the type of infraction that triggered the referral. In North Carolina, state law mandates that schools report incidents leading to removal from school (suspension, expulsion, or assignment to an alternative school), corporal punishment, or any offense from a list of 12 serious infractions (e.g., sexual assault, arson, use of weapons; see appendix Table A.1 for a list of offenses) (Sorensen et al., 2021). School districts also maintain discretion over reporting referrals assigned for a range of less serious "unacceptable behaviors," such as skipping school, tardiness, insubordination, and a variety of other behaviors (see appendix Table A.1). Second, we can link teachers and students to specific classrooms and observe patterns in teachers' use of referrals in a given classroom. Finally, we observe teachers and students for multiple years, a feature which allows us to account for potentially endogenous student assignment to teachers of varying characteristics.

Unfortunately, the NCERDC data does not link students' referrals to the specific teacher that assigned the referral. We therefore restrict our analysis to self-contained classrooms in elementary schools in grades 3 through 5. Since one lead teacher manages self-contained classrooms, we adopt a modest assumption that the behavioral referrals assigned to students in such classrooms come predominantly from the lead teacher of the class. Our final analytic sample contains 32,373 classrooms, 15,237 teachers, 330,417 students, and 1,271 schools.

Table 1: Summary of students in self-contained classrooms in grades 3 through 5 in North Carolina Public Schools, 2008-2013

	(1) Total sample	(2) Black student	(3) White student
Absences	5.79	5.11***	6.28
	(5.33)	(5.26)	(5.42)
Chronically absent	0.03	0.03***	0.04
Math score (std)	0.13	-0.38***	0.36
	(0.97)	(0.90)	(0.91)
ELA score (std)	0.09	-0.36***	0.33
	(0.96)	(0.91)	(0.90)
Disciplinary offenses $(N\|N > 0)$	2.09	2.38***	1.89
	(2.17)	(2.50)	(1.83)
Any disciplinary offenses	0.10	0.18***	0.07
Subjective referrals $(N\|N > 0)$	1.84	2.03***	1.68
	(1.75)	(1.95)	(1.49)
Any subjective referrals	0.06	0.11***	0.04
Male student	0.50	0.49***	0.51
White student	0.56	0.00	1.00
Black student	0.21	1.00	0.00
Hispanic student	0.12	0.00	0.00
Asian student	0.02	0.00	0.00
Native American student	0.01	0.00	0.00
Other race student	0.04	0.00	0.00
Economic disadvantage	0.47	0.73***	0.30
Observations	313,280	67,128	174,700

Note: Standard deviations in parentheses; *p < .10, **p < .05, ***p < .01 for t-test of difference in means between columns 2 and 3.

3.1 Measuring Teacher Punitiveness

We aim to examine the effects of teachers' punitiveness in use of referrals on student learning and engagement. If strict management of misbehavior through the use of referrals plays an important role in productive teaching, we anticipate students with more punitive teachers would learn more and attend school at higher rates. However, our inability to directly observe punitiveness in teachers' decision-making and student interactions presents an empirical challenge. Teachers assigned to classrooms with more disruptive students may appear more punitive in their disciplinary referral rate than teachers assigned to classrooms with less disruptive students. Similarly, teachers who give more referrals may be systematically assigned students more likely to engage in disruptive behaviors. In such a scenario, the link between teacher use of referrals and student outcomes might simply reflect a comparison across teachers managing classrooms with different pre-existing behavioral dynamics. We approach these empirical problems by estimating a teacher's "value-added" to a student's referral production function. Our reduced form model of teacher punitiveness models the referrals (P) student i receives with teacher j as the linear function:

$$P_{ijct} = \alpha_{jt} + \gamma_1 P_{i,t-1} + \gamma_2 A_{i,t-1} + \varepsilon_{ijct} \tag{1}$$

where i, j, c, and t index students, teachers, classrooms, and years, respectively, and A represents student ability (measured using lagged standardized math and reading scores). In equation 1, α represents the teacher-specific contribution to their average students' number of referrals.[2] This assumes that accounting for the number of referrals students received the prior year and students' pre-existing academic ability removes the endogenous student assignment component of referral production in the classroom and identifies the as-if random teacher contribution (Chetty et al., 2014a).[3]

[2]We estimate equation 1 in a two-step process. First, we regress referrals on lagged achievement and lagged referrals. Second, we regress the residuals from the first step on teacher-year fixed effects.

[3]As a sensitivity check, we add controls for within-class average lagged number of referrals and lagged average test scores of student i's peers to our model of teacher punitiveness. Table A.9 reports the main

Of course, as previously noted, some behaviors require teachers to refer students for disciplinary actions or are unambiguous in the necessity of a referral response from a teacher. For instance, if a student physically assaults the teacher or brings a weapon to class, teachers would be bound by policy to refer that student to administration for discipline. Our approach might be compromised if students prone to engaging in such severe infractions were not randomly distributed across teachers. We account for this possibility by focusing our analysis instead on the subset of referrals teachers assign for behaviors in which the severity is subject to interpretation.[4] In equation (1), the outcome is restricted to subjective referrals while the lag includes the count of all referrals a student received the prior year to ensure more precise accounting for student sorting on prior behavioral issues.[5] We take the estimated teacher contribution to subjective referrals as a measure of teachers' punitiveness, our primary independent variable and the focus of our study.[6] A one unit increase in teacher punitiveness represents one additional reported disciplinary referral for the average student, conditional on prior behavior.

This measure has two potential serious limitations that we address. First, the reporting of disciplinary referrals does not occur at random for incidents that do not fall under mandatory reporting requirements in North Carolina. Instead, reporting choices reflect in part the discretion of school administration and could correlate with underlying student or teacher characteristics. To address this, our empirical models estimating effects of teacher punitiveness will control for school fixed effects and thereby any school-level differences in

effects using this alternative specification for equation (1). Accounting for class-level peer effects does not appreciably change the results.

[4]Here, we follow Sorensen et al. (2021) in the behaviors we consider subjective in interpretation. Such behaviors are detailed in Table A.1 and include behaviors like tardiness, talking back to teachers, being disrespectful, disruptive behavior, and other infractions that involved subjective interpretation on whether a behavior rises to the level of referral.

[5]Appendix Table A.3 shows that our primary results are robust to using total referrals when measuring teacher punitiveness.

[6]See Figure A.1 and Figure A.2 in the appendix for a depiction of the distribution of punitiveness and teacher bias in referral use in our analytic sample. Notably, teachers vary considerably in their use of discretion in assigning referrals for more subjective behavioral infractions. Also evident in the figures is the presence of potential outliers. In Figure A.3 and Figure A.4, we exclude the highest and lowest 2% of observations and in Table A.10 and Table A.11 we replicate out main analysis on the sample without outliers. The estimates are consistent, which confirms our results are not being driven by outliers.

reporting practices.

Second, it could be that differences in teacher use of referrals for a given classroom reflects differences in student behavior rather than differences in teacher behavior. Controlling for prior year student referrals and test scores helps to account for time-invariant student behavioral tendencies, but it does not account for the fact that one class of students may behave systematically different than another class of students, even conditional on prior year behaviors. We address this issue in two ways. First, in our preferred models reported in the main text, we include a vector of controls for classroom-level characteristics to account for observable classroom differences, including average prior behavior.[7] Second, in Table A.4 and Table A.5, we adopt the leave-year-out estimation described by Chetty et al. (2014a).[8] In these robustness checks, because we estimate teacher punitiveness based on disciplinary referrals during all years except year t, teacher punitiveness in year t is by definition unrelated to the qualities of students or classrooms in year t. This also removes the possibility of simultaneity bias that could occur if we measured student learning and attendance in the same year that we measure teacher punitiveness. Moreover, we implement our preferred model to examine the effects of teacher punitiveness in time t in future years $t+1$, providing additional evidence that our results are robust to potential simultaneity bias.

3.2 Teacher Bias in Referral Use

In addition to the potential academic effects of teachers' use of referrals in general, we investigate the possibility that racial bias enters into the use of referrals with distinct effects on academic outcomes. Racial disproportionality in suspensions and other forms of exclusionary discipline is well documented (Amemiya et al., 2020; Barrett et al., 2021; Gregory, 1995;

[7]We take this approach because referrals at the elementary level, while consequential for students, are assigned less frequently than in later years. As a consequence, there are many teacher-years in which zero referrals are assigned.

[8]The details of this procedure are described in the appendix of Chetty et al. (2014a). We implement this procedure using Michael Stepner's `vam` package using Stata16. We also estimate teacher value-added scores for math and ELA scores. We derive these estimates following the same approach with the exception that lagged referrals are not included.

Kinsler, 2011; Skiba et al., 2002, 2011). However, it is not exactly clear at what point during the disciplinary process – from incident, to observation, to referral, to administrative decision – that racial disproportionality or bias first emerges. A study by Okonofua and Eberhardt (2015) suggests that racial stereotypes lead teachers to perceive behaviors of Black students differently than behaviors of White students and to escalate their negative responses to Black students over time. We seek to understand both whether teacher use of disciplinary referrals differs by student race and how such differences affect student outcomes.

As with estimating overall punitiveness, we must account for non-random sorting of students to teachers. However, in detecting racial bias in the absence of direct observation, we must make the additional assumption that, conditional on prior behavior and pre-existing student behavior, a teacher without racial bias would write punitive referrals for Black and White students at similar rates. Thus, we can adopt a similar approach and estimate teacher bias with the linear function:

$$P_{ijct} = \rho_{1jt} black_i + \gamma_{1jt} P_{i,t-1} + \gamma_{2jt} A_{i,t-1} + \varepsilon_{ijct}, \forall t \in \{j\} \tag{2}$$

For this estimation, we identify differences in disciplinary referrals across groups of students *within* the same teacher instead of identifying differences in disciplinary referrals across teachers. To do so, we run a separate regression for each teacher and year of student referrals on a student race indicator, controlling for student-level lagged measures of the number of referrals received in the prior year and student achievement in the prior year. We estimate equation (2) on a subsample of self-contained classrooms with at least 1 Black and 1 White student and restrict the sample to only Black and White students.[9] Thus, in equation (2), ρ_{1jt} represents teacher j's difference in subjective referrals given to Black students relative to White students in year t, which - after accounting for students' prior behavior issues and ability - we take as an estimate of teacher racial bias in the use of referrals.[10]

[9]Table A.2 summarizes the characteristics of students in this subset of classrooms.

[10]We account for years in which teachers assigned no referrals by assigning the within-teacher average bias estimated across years to those years. Figure A.2 depicts the distribution of teachers' racial bias in the use

of referrals for subjectively identified behavior problems and Figure A.4 depicts the distribution excluding the highest and lowest 2% of observations.

Table 2: Summary of teachers in self-contained classrooms in grades 3 through 5 in North Carolina Public Schools, 2008-2013

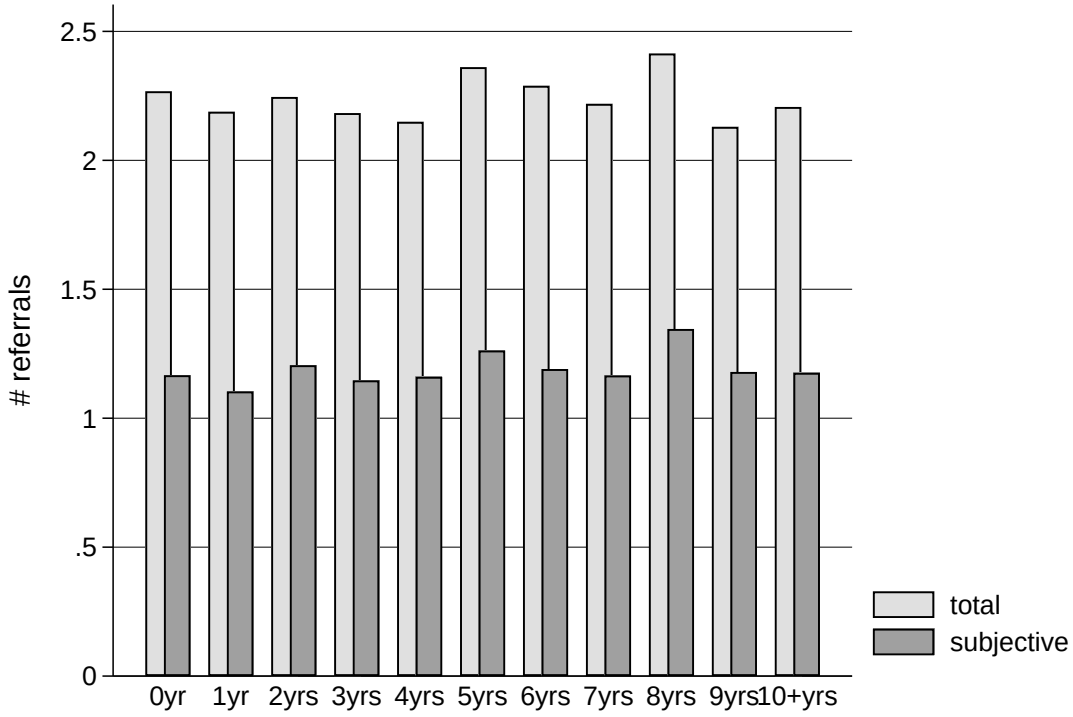
	All Classes			Mixed Race Classes		
	(1)	(2)	(3)	(4)	(5)	(6)
	Total sample	Black teacher	White teacher	Total sample	Black teacher	White teacher
Punitiveness in total referrals	-0.00	0.01	-0.00	0.01	0.01	0.01
	(0.45)	(0.55)	(0.44)	(0.42)	(0.50)	(0.41)
Punitiveness in subjective referrals	-0.00	0.01	-0.00	0.00	0.00	0.00
	(0.31)	(0.38)	(0.30)	(0.28)	(0.37)	(0.26)
Teacher bias in subjective referrals	-	-	-	0.14	0.09***	0.15
				(0.80)	(0.74)	(0.81)
VA ELA test scores	0.01	0.00**	0.01	0.00	0.00	0.00
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
VA math test scores	0.01	-0.00***	0.01	0.01	-0.00	0.01
	(0.13)	(0.14)	(0.13)	(0.13)	(0.14)	(0.13)
All referrals	2.66	3.17***	2.59	2.47	2.81**	2.44
	(4.61)	(5.27)	(4.52)	(4.41)	(5.14)	(4.32)
Subjective referrals	1.41	1.70***	1.37	1.31	1.51***	1.28
	(2.90)	(3.55)	(2.79)	(2.80)	(3.56)	(2.69)
Male	0.09	0.07***	0.09	0.09	0.06***	0.10
White	0.84	0.00	1.00	0.85	0.00	1.00
Black	0.13	1.00	0.00	0.12	1.00	0.00
Teacher Experience	11.24	13.08***	11.03	11.23	13.43***	11.02
	(8.61)	(9.38)	(8.48)	(8.63)	(9.42)	(8.51)
Advanced degree	0.31	0.33***	0.30	0.30	0.30	0.30
Lateral/Provisional license	0.02	0.05***	0.02	0.02	0.05***	0.02
Math license	0.03	0.02***	0.03	0.03	0.02***	0.03
ELA license	0.10	0.09***	0.10	0.11	0.10***	0.11
Class size	21.26	20.08***	21.44	21.58	20.69***	21.70
	(4.20)	(4.55)	(4.12)	(4.07)	(4.31)	(4.03)
Observations	28,370	3,690	23,901	18,186	2,173	15,517

Note: Unit of observation is teacher-years. Standard deviations in parentheses; *p < .10, **p < .05, ***p < .01 for t-test of difference in means between columns 2 and 3 and 5 and 6. VA = value-added.

Table 2 summarizes the teachers in our analytic sample. In our sample, the average elementary teacher writes about 2.7 disciplinary referrals per year and about half of those referrals - 1.4 per year - are for behavioral infractions with some measure of subjective interpretation. While the gap in punitiveness between Black and White teachers is small and statistically insignificant, Black teachers tend to give out more referrals on average (3.17 versus 2.59), an indication that Black teachers tend to be assigned students with more behavioral infractions in prior years. Moreover, among the subset of teachers who teach mixed-race classrooms as previously described (columns 4 through 6), Black teachers are less racially biased than White teachers in their assignment of referrals for subjective infractions. Overall, teachers show a fair amount of racial bias in their use of referrals. After accounting for prior behavioral problems and achievement, the average teacher teaching a class of 20 students with 10 Black students and 10 White students will give about 1.4 more referrals to Black students in their class than White students. A teacher one standard deviation above the mean in bias would give about 8 more referrals to Black students than White students in her classroom.

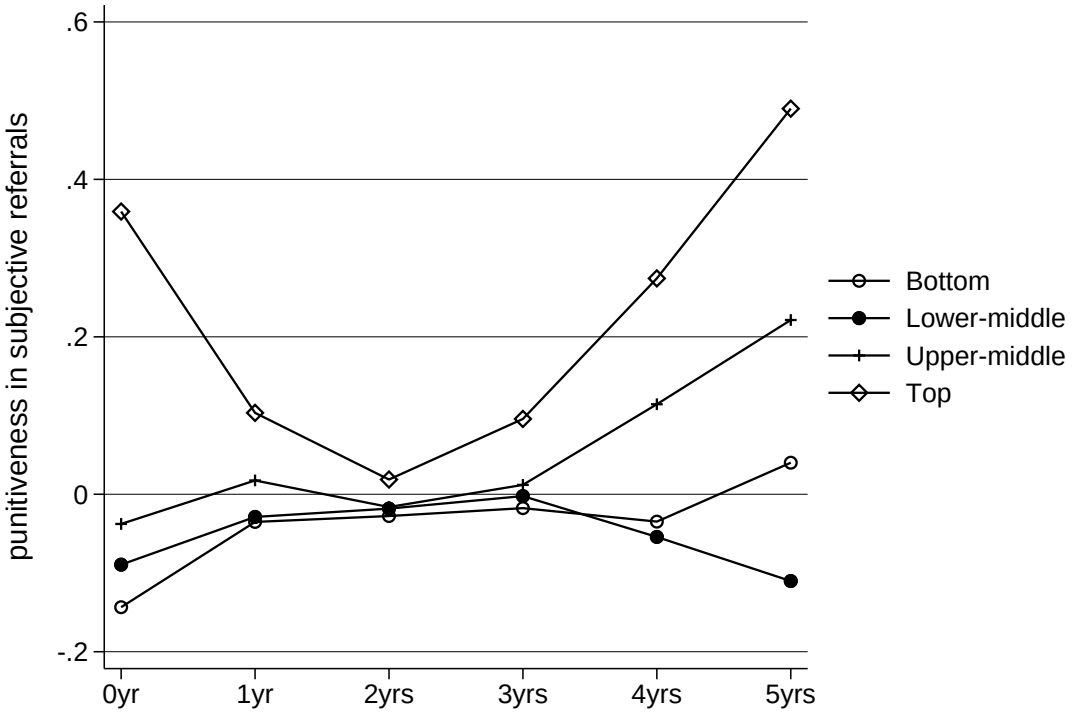
Teachers may change their use of referrals over time as they gain more experience with students, more comfort in the classroom, or receive more feedback from peers and principals. In Figure 1, we examine the average number of disciplinary referrals teachers assign by years of experience. Notably, teachers' use of referrals appears relatively stable over time. That is, more experienced teachers use referrals at the same rate, on average, as less experienced teachers. There is a second question of how stable teachers' approaches to disciplinary referrals is over time. In Figure 2, we sort teachers by their punitiveness score in their first year teaching in North Carolina public schools to examine the relative stability of teacher punitiveness over time. As Figure 2 shows, teachers in the top two quartiles of punitiveness in their first year of teaching tend to remain the most punitive teachers, in relative terms, throughout their early careers.

Figure 1: Number of referrals assigned per year by teacher years of experience



Note: Units of observation are teacher-years.

Figure 2: Punitiveness in the use of referrals by teacher years of experience, 2008 cohort of teachers



Note: Units of observation are teacher-years. Plots reflect quartile of punitiveness during teacher j's first year teaching in North Carolina public schools.

3.3 Effects of Teacher Referral Use on Student Outcomes

Our primary aim in measuring teachers' punitiveness and racial bias in the use of disciplinary referrals is to investigate the effect such potentially mutable characteristics have on student outcomes. We focus our study on student attendance and growth in achievement. Student attendance is an important indicator of engagement with school, and absence from school harms both short-term learning (Aucejo and Romano, 2016; Gershenson et al., 2017) and longer-term outcomes such as educational attainment (Liu et al., 2021b). Student test scores, though obviously imperfect proxies for learning, nonetheless again are strongly predictive of long-term educational attainment an success in the labor market (Chetty et al., 2014b). We

identify the impact of teachers' use of punishment (or bias) on these outcomes by adopting a value-added approach in models of the form:

$$Y_{ijcgst} = \beta_1 \widehat{\alpha}_{jt} + \beta_2 Z_j + \beta_3 X_i + \gamma_1 Y_{i,t-1} + \omega M_c + \varphi_s + \psi_g + \tau_t + \varepsilon_{ijcgst} \tag{3}$$

$$Y_{ijgst} = \beta_1 \widehat{\rho}_{jt} + \beta_2 Z_j + \beta_3 X_i + \gamma_1 Y_{i,t-1} + \omega M_c + \varphi_s + \psi_g + \tau_t + \varepsilon_{ijgst} \tag{4}$$

where i, j, c, g, s, and t index students, teachers, classrooms, grades, schools, and years, respectively, and Y represents the students' academic outcomes (absences, standardized math scores, and standardized ELA scores) at the end of their year with teacher j. Our main independent variables, punitiveness ($\widehat{\alpha}$ in equation (3)) and racial bias in use of referrals ($\widehat{\rho}$ in equation (4)), reflect estimated parameters as previously described. The variable M represents a vector of controls for classroom-level average student characteristics (% nonwhite, % free or reduced-price lunch eligible, class size, average lagged math and reading scores, and average lagged number of referrals) to account for class-specific variation in student composition. Finally, Z and X represent vectors of teacher characteristics (years of experience, master's degree indicator, license status, race, gender) and student characteristics (race, gender, economically disadvantaged), respectively.

Here, under the main identifying assumption that accounting for prior attendance and achievement ($Y_{i,t-1}$) controls for endogenous student sorting, β_1 provides a consistent estimate of the difference in attendance and learning between students facing more punitive (or biased) teachers and less punitive (or biased) teachers, ceteris paribus. We estimate equations (3) and (4) using OLS to allow for the inclusion of grade, school, and year fixed effects. We implement bootstrapped standard errors with 500 replications to account for the unknown sampling distribution of our estimated independent variables of interest.

4 Results

4.1 Effects of Teacher Punitiveness and Bias

We begin with our estimates of the main effects of teachers' punitiveness in their use of subjective referrals on student absences and achievement. A teacher one standard deviation above the mean in punitiveness assigns the average student in their class 0.31 more referrals than would be expected conditional on prior achievement and referrals. Scaled to the average class size of 21.26, each standard deviation increase in teacher punitiveness represents a teacher assigning about 6.59 more referrals, on average, than would be expected given their students' prior behavior and achievement. Table 3 presents the estimates of the effect of facing a teacher who is more punitive in this manner from our preferred model.

As the results in Table 3 show, students who face a more punitive teacher have more absences and lower achievement in both math and reading. The estimates reflect a modest effect on absences (about a 2.6% increase in absences from the sample average attributable to a 1 s.d. increase in teacher punitiveness). Given the relatively low rates of absenteeism and the role of family effort in getting children to school in elementary school, this effect may reflect marginal changes in attendance attributable to either student disengagement or parents' response to a stricter teacher.[11] As the results in columns 3 and 4 suggest, students have lower achievement in classrooms led by more punitive teachers, particularly in math. A student in a classroom with a teacher one standard deviation above average in punitiveness gains 2.4% of a standard deviation less in their math ability than a peer in a classroom with a less punitive teacher. For context, a teacher at the absolute mean in our measure of punitiveness assigns an average of 1.4 referrals for more subjective infractions in a given year. Our results suggest that, holding constant students' expected referrals, a teacher who gives out 6.59 more referrals over the course of the year leads their students to achieve 2.4% standard deviation less growth in math.

[11] In North Carolina schools, neither out-of-school suspension days nor in-school suspension days count as absences, and so increases in suspension likelihood should not directly explain the attendance findings.

An important and related question is whether teacher punitiveness operates independently of overall measures of teacher effectiveness. Put another way, it may be that teachers highly effective in other areas of instruction also happen to use referrals less, which would explain the above association. Alternatively, use of referrals might operate as a tool for managing a classroom adopted by all types of teachers and, as a result, the effects are unrelated to teachers' other dimensions of effectiveness in the classroom. In columns 5 through 8, we add controls for teacher effectiveness, measured as a teachers' value-added to math and reading achievement. Notably, the effects of punitiveness seem to operate independently of teacher effectiveness.[12] While the effect size of punitiveness on attendance and achievement shrinks slightly after accounting for teacher effectiveness, the effects remain substantive and significant. In Table A.6 in the appendix, we present estimates of our main results using a within-student approach, which identifies the effect of a student switching from less to more punitive teachers across years. The results are similar in magnitude and direction.

In short, the use of referrals as a disciplinary practice does not seem to lead to more productive classrooms. While the effects are modest, they indicate that teachers who rely more heavily on the use of referrals in their classrooms have less productive classrooms and may lead families to be less engaged with schooling, as the slight increase in absenteeism suggests. Even after accounting for other dimensions of teacher effectiveness, more punitive teachers have negative impacts on their students' learning and attendance.

Turning to teacher racial bias in the use of referrals for infractions with some subjective interpretation, Table 4 presents the estimated average impact of teacher bias on student absenteeism and achievement across all students. We again estimate the effect of bias in use of referrals with and without controls for teacher effectiveness. As the results demonstrate,

[12]However, we cannot completely eliminate the possibility that punitiveness is but one dimension of effectiveness. In Table A.4, we adopt a leave-year-out approach to estimating punitiveness. Under that specification of punitiveness, the impact of punitiveness on student outcomes shrinks and is no longer significant after accounting for teacher effectiveness. Indeed, the leave-year-out measure of punitiveness in referral use is negatively correlated with both value-added in reading ($-.105, t = -51.02$) and value-added in math ($-.108, t = -51.45$). However, this could be due to inefficiencies in measuring punitiveness using a leave-year-out approach related to the infrequent nature of referrals in elementary schools.

racial bias in the use of referrals operates independently of teacher effectiveness. Teachers who exhibit more racial bias in their referrals cause a modest decrease in student achievement, and the negative effect on math achievement remains even after for accounting for teacher effectiveness. Moreover, given that the average mixed-race classroom in our sample has only 6 Black students out of an average size of about 22 students, the average effect of teachers' racial bias in referral use might be driven downward if White students remain unaffected by teachers' racial bias.

Table 3: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
Punitiveness	0.492***	0.011***	-0.045***	-0.077***	0.485***	0.011***	-0.038***	-0.064***
	(0.06)	(0.00)	(0.01)	(0.01)	(0.06)	(0.00)	(0.01)	(0.01)
Value-Added Math	-	-	-	-	-0.475***	-0.008**	0.267***	0.778***
					(0.11)	(0.00)	(0.01)	(0.02)
Value-Added ELA	-	-	-	-	0.153	-0.007	0.329***	0.150***
					(0.19)	(0.01)	(0.03)	(0.03)
All teacher controls	✓	✓	✓	✓	✓	✓	✓	✓
All student controls	✓	✓	✓	✓	✓	✓	✓	✓
All classroom controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Absences	✓	✓			✓	✓		
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.363	0.139	0.643	0.662	0.363	0.139	0.645	0.673
Observations	313,280	313,280	313,280	313,280	313,280	313,280	313,280	313,280

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. Chronic = chronically absent. A student is considered chronically absent if they miss 18 school days or more in a given academic year.

Table 4: Effect of Teacher Bias in Referral Use for Subjective Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
Teacher Bias	0.049	0.000	-0.006*	-0.016***	0.047	0.000	-0.005	-0.013***
	(0.03)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)
Value-Added Math	-	-	-	-	-0.558***	-0.010**	0.254***	0.791***
					(0.13)	(0.00)	(0.02)	(0.02)
Value-Added ELA	-	-	-	-	0.380	-0.003	0.354***	0.169***
					(0.24)	(0.01)	(0.04)	(0.04)
All teacher controls	✓	✓	✓	✓	✓	✓	✓	✓
All student controls	✓	✓	✓	✓	✓	✓	✓	✓
All classroom controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Absences	✓	✓			✓	✓		
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.372	0.147	0.638	0.664	0.372	0.147	0.641	0.675
Observations	169,898	169,898	169,898	169,898	169,898	169,898	169,898	169,898

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. A student is considered chronically absent if they miss 18 school days or more in a given academic year.

Though the effects of teacher bias in use of referrals are modest, they reveal a policy-relevant problem - value-added measures of teacher effectiveness commonly used to evaluate teacher performance may not capture other important dimensions of performance, such as racial bias in the use of punishments. This is consistent with Jackson's (2018) evidence, from 9th graders in North Carolina, that the correlation between measures of teachers' effects on test scores and teachers' effects on other outcomes are only weakly correlated. Our results suggest that in addition to providing an incomplete picture of the full range of outcomes affected by teachers, value-added measures tied to student achievement miss dimensions of teacher performance in other important areas, such as equitable treatment of students, that also affect student outcomes. Teacher bias in the use of referrals is only modestly correlated with teacher effectiveness ($-0.051, t = -10.71$ for ELA, $-0.052, t = -10.76$ for math). Of course, the effect of both general harshness and racial bias in the use of referrals on student outcomes likely varies by both teacher and student characteristics. We examine heterogeneity in the effect by student and teacher race and ability.

4.2 Heterogeneity in Effects

In Table 5, we show the effect of teacher punitiveness on student outcomes separately in important subsamples of teachers. For instance, skilled teachers might offset the negative effects of adopting a harsher use of referrals in their classrooms, perhaps through cultivating a more trusting environment at the outset and lending credibility to their disciplinary decision-making. Similarly, racial dynamics in the classroom might lead to differences in student responses to harshness for Black teachers relative to White teachers. In Table 5, each panel presents the effect of punitive use of referrals on a given outcome across each subsample (the number of absences in an academic year, ELA scores, and math scores in panels A, B, and C, respectively).

Some notable patterns emerge from our analysis. First, across all three outcomes, harsher use of punishment has a similar impact on students among both Black and White teachers.

Second, the negative effect of teacher harshness in their use of referrals is generally slightly larger among above average value-added teachers, which suggests even effective teachers would be much more effective if they were less frequent in their use of referrals. Table A.7 in the appendix presents estimates of the effects of teacher punitiveness by student race and ability in reading and math. The results show the magnitude and direction of the effect is similar across student race, math ability, and reading ability. The results suggest the impact of facing a more punitive teacher is similar across all students.

Of course, we hypothesize that teachers' racial bias in the use of referrals likely impacts Black students more negatively than White students. In Table 6, we estimate the effect of teachers' racial bias in the use of referrals with subjective interpretation on student outcomes separately by student race and prior ability. We find that, perhaps unsurprisingly, negative outcomes for Black students drive our observed effects of teacher bias. Black students facing a more biased teacher increase absenteeism by about 9.5% from their base rate of absenteeism. Similarly, while White students show very little difference in learning when facing a more biased teacher, Black students' learning suffers when facing a more biased teacher. Teacher racial bias in disciplinary referrals also has stronger adverse impacts on absences and math test scores for students with below average prior achievement than for students with above average prior achievement. These results highlight a challenge facing administrators and policy-makers aiming to improve equity in educational quality - racial bias in teachers' classroom practice is often unobserved and even otherwise effective teachers may be biased.[13]

[13]We also estimate models in which punitiveness and bias are interacted with an indicator for whether the teacher is the same race as the students. We find no evidence of a race match effect for either punitiveness or racial bias for students overall or separately for Black students. Results from these models are available upon request. Together with our descriptive evidence that Black teachers show less racial bias in their use of referrals, we interpret this to suggest that race match effects decrease exposure, on average, to racial bias in referrals, but teachers' racial bias carries negative effects on Black student learning regardless of the teacher's race.

Table 5: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, separately by teacher characteristics, 2008-2013

	(1) White T	(2) Black T	(3) > Avg. VAR	(4) < Avg. VAR	(5) > Avg. VAM	(6) < Avg. VAM
PANEL A: Absences						
Punitiveness	0.487***	0.485***	0.397***	0.518***	0.529***	0.433***
	(0.07)	(0.19)	(0.09)	(0.08)	(0.10)	(0.08)
Adjusted R^2	0.363	0.364	0.355	0.368	0.357	0.366
Observations	271,041	33,623	169,743	143,537	164,535	148,745
PANEL B: ELA Scores						
Punitiveness	-0.048***	-0.022	-0.048***	-0.038***	-0.048***	-0.034***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.645	0.614	0.638	0.646	0.639	0.645
Observations	271,041	33,623	169,743	143,537	164,535	148,745
PANEL C: Math Scores						
Punitiveness	-0.074***	-0.075***	-0.085***	-0.057***	-0.084***	-0.052***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.665	0.645	0.664	0.662	0.666	0.668
Observations	271,041	33,623	169,743	143,537	164,535	148,745
All lagged controls	✓	✓	✓	✓	✓	✓
All teacher, class, & student controls	✓	✓	✓	✓	✓	✓
Teacher VAR & VAM						
School, grade, & year FE	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, an indicator for economically disadvantaged, and classroom characteristics. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

Table 6: Effect of Teacher Bias in Referrals for Subjective Infractions on Student Outcomes, separately by student characteristics, 2008-2013

	(1) White S	(2) Black S	(3) > Avg. ELA	(4) < Avg. ELA	(5) > Avg. Math	(6) < Avg. Math
PANEL A: Absences						
Teacher bias	-0.071***	0.384***	-0.021	0.125***	-0.017	0.110***
	(0.03)	(0.06)	(0.03)	(0.04)	(0.03)	(0.04)
Adjusted R^2	0.364	0.327	0.359	0.375	0.347	0.380
Observations	118,746	51,152	100,508	69,390	96,397	73,501
PANEL B: ELA Scores						
Teacher bias	-0.002	-0.014**	0.001	-0.004	-0.003	-0.006
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.609	0.590	0.393	0.388	0.492	0.488
Observations	118,746	51,152	100,508	69,390	96,397	73,501
PANEL C: Math Scores						
Teacher bias	-0.006	-0.033***	-0.007	-0.018***	-0.006	-0.012***
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
Adjusted R^2	0.647	0.613	0.567	0.504	0.442	0.380
Observations	118,746	51,152	100,508	69,390	96,397	73,501
All lagged controls	✓	✓	✓	✓	✓	✓
All teacher, class, & student controls	✓	✓	✓	✓	✓	✓
Teacher VAR & VAM	✓	✓	✓	✓	✓	✓
School, grade, & year FE	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

4.3 Spillover Effects

One potential explanation for our results is that the disciplinary actions taken following teacher referrals, such as out-of-school suspensions, directly reduce student achievement and increase absenteeism for the suspended student. In this way, more students receiving referrals may mechanically drive poorer student outcomes. However, it is also possible that highly punitive teachers could offset the lost learning among referred students through higher learning and more steady attendance from students who do not receive referrals. Similarly, if teacher bias effects are directly attributable to the effect of disciplinary actions on referred Black students, it is possible that this could have positive or negative spillover effects onto other students in the classroom. We seek to investigate this possibility by replicating our analysis after restricting the analytic sample to students who do not receive any referrals in year t in teacher j's classroom.

The results in Table 7 confirm that the effects of teachers' tendency in their use of referrals spill over into the classroom more broadly. Even students who do not receive referrals learn less when facing a more punitive teacher or a teacher who is more racially biased in their use of referrals. In Table A.8, in the appendix, we show that the spillover effects of teachers' punitiveness follow the same patterns as the main effects - among otherwise effective teachers, more punitive teachers reduce student achievement, which suggests teachers using a harsher approach to disciplinary matters is a dimension of overall efficacy in the classroom. Also, again consistent with our main findings, teacher racial bias in the use of referrals has a negative spillover effect on student learning even after accounting for teachers' overall effectiveness.

In Table 8, we estimate the spillover effects of teacher racial bias in referral use by student characteristics. The spillover effects again follow the same pattern as the main effects. Even among students who do not receive referrals, Black students facing a teacher whose use of referrals is racially biased experience lower achievement in both math and reading. Conversely, among White students who do not receive referrals, the effect of racial

bias on learning is negligible. The impacts are striking - even without the impact of receiving a referral, a Black student facing a teacher who averages 1 referral more for Black students than White students in a typical year scores 1.7% of a standard deviation lower on reading tests and 2.6% of a standard deviation lower on math tests.

In general, we find the negative impact on student learning of teachers' punitiveness and racial bias in their use of referrals spills over into students who do not receive referrals. However, the attendance patterns of students who do not receive referrals do not seem to be affected by teachers' referral usage tendencies. One possibility is that the absenteeism observed in the main model is driven entirely by students who receive referrals. Parents could receive cues or signals from teachers' use of referrals about their teachers' relationship with the classroom. Since attendance in elementary school is largely a function of parental effort, the attendance effects of teachers' bias or harshness in referral usage may operate through parents' opportunities to observe this signal. Our findings - a significant increase in absenteeism generally that is smaller and insignificant among students who do not receive referrals - lend some support to this possibility.

Table 7: Spillover Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
PANEL A: Punitiveness								
Punitiveness in subjective referrals	0.032	0.002	-0.033***	-0.066***	0.024	0.002	-0.026***	-0.051***
	(0.06)	(0.00)	(0.01)	(0.01)	(0.06)	(0.00)	(0.01)	(0.01)
Adjusted R^2	0.366	0.135	0.641	0.662	0.366	0.135	0.644	0.672
Observations	266,168	266,168	266,168	266,168	266,168	266,168	266,168	266,168
PANEL B: Bias in Referrals								
Teacher bias	-0.036	-0.001	-0.008**	-0.014***	-0.037	-0.001	-0.006*	-0.011***
	(0.03)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.378	0.142	0.634	0.662	0.378	0.142	0.636	0.673
Observations	142,626	142,626	142,626	142,626	142,626	142,626	142,626	142,626
All lagged controls	✓	✓	✓	✓	✓	✓	✓	✓
All teacher, class, & student controls	✓	✓	✓	✓	✓	✓	✓	✓
Teacher VAR & VAM					✓	✓	✓	✓
School, grade, & year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes students with no referrals in all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. A student is considered chronically absent if they miss 18 school days or more in a given academic year.

Table 8: Spillover Effect of Teacher Bias in Referrals for Subjective Infractions on Student Outcomes, separately by student characteristics, 2008-2013

	(1) White S	(2) Black S	(3) > Avg. ELA	(4) < Avg. ELA	(5) > Avg. Math	(6) < Avg. Math
PANEL A: Absences						
Teacher bias	-0.039	-0.022	-0.066**	0.013	-0.045	-0.031
	(0.03)	(0.06)	(0.03)	(0.04)	(0.03)	(0.04)
Adjusted R^2	0.363	0.335	0.362	0.384	0.350	0.391
Observations	104,854	37,772	89,804	52,822	86,322	56,304
PANEL B: ELA Scores						
Teacher bias	-0.004	-0.017*	-0.000	-0.003	-0.005	-0.007
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Adjusted R^2	0.606	0.592	0.394	0.379	0.487	0.487
Observations	104,854	37,772	89,804	52,822	86,322	56,304
PANEL C: Math Scores						
Teacher bias	-0.008*	-0.026***	-0.006	-0.017***	-0.005	-0.008
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
Adjusted R^2	0.646	0.620	0.565	0.503	0.445	0.369
Observations	104,854	37,772	89,804	52,822	86,322	56,304
All lagged controls	✓	✓	✓	✓	✓	✓
All teacher, class, and student controls	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes students with no referrals in all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

4.4 Long-Run Effects of Punitiveness and Bias

As Jacob et al. (2010) document, while effective teachers' impacts on future student learning persist, the share of lasting student learning attributed to measures of teacher effectiveness may be modest. In our case, while teachers' use of disciplinary referrals have detrimental immediate effects on student learning, the exposure to a more punitive teacher may lead to longer term changes in student behavior. Such teachers may trigger a long-run disengagement from education, leading to continued behavioral problems and long-run negative impacts on student learning. Alternatively, students may receive the strong signals of a punitive teacher and improve their behavior over time or experience only transient negative outcomes while with the teacher. We examine both the intermediate and long-run educational outcomes that can be attributed to facing more punitive and more biased teachers. To do so, we estimate our main models of punitiveness and bias on students' referral counts, absenteeism, and achievement in middle school to test for the persistence of the negative effects we document at the elementary level.[14] We follow a similar approach to examine the impact of facing more punitive/biased teachers in elementary school on high school graduation, behavioral indicators of college intention such as taking the PSAT and SAT, and self-reported college intentions from student surveys in fielded in 11th grade.

Table 9 presents our estimates of the intermediate impacts of facing a more punitive teacher in grades 4 or 5 on students' academic outcomes. We find evidence of a modest, persistent, negative impact of facing a more punitive teacher in elementary school. For instance, as columns 1 and 5 illustrate, rather than modifying student behavior going forward, students who faced a more punitive teacher in 4th or 5th grade than otherwise comparable peers receive more disciplinary referrals throughout middle school. Similarly, students with more punitive teachers in elementary school carry persistent attendance problems into

[14] In particular, we create separate fourth grade and fifth grade classroom samples and then within each sample regress student long-term outcomes on teacher punitiveness or bias in that grade, controlling for student prior year achievement/attendance, classroom characteristics, teacher characteristics, and school, grade, and year fixed effects. We acknowledge slight changes in the analytical sample for long-term outcomes due to sample attrition.

middle school. For instance, we find that a student facing a teacher in 4th grade who is 1 standard deviation above the mean in punitiveness has a 0.5% increase in absences in 7th grade from the baselines absence rate of 5.16 and 0.6% increase in 8th grade from a base rate of 5.72. While modest, the effects suggest more punitive teachers lead students to disengage from school and that disengagement persists into future years.

We also find evidence of modest, persistent, negative achievement effects as well. For instance, a student whose 4th grade teacher was 1 standard deviation above the mean in the harshness of their referral usage scores about 0.01 standard deviations lower in math in 6th grade, which suggests approximately half of the negative achievement effect persists two years out. We see similar patterns in the effect of 5th grade teachers' punitiveness on math achievement in 7th and 8th grade. Meanwhile, we find less consistent long-run impacts on students' ELA achievement, though the estimated impact is always negative.

Finally, Figure 3, Figure 4, and Figure 5 present our results for the effect of teacher punitiveness and teacher bias on the long-run educational outcomes. First, as Figure 3 shows, students who face a more punitive teacher in 4th or 5th grade are less likely to graduate high school and generally show a lower likelihood of attending college. For instance, students facing a more punitive teacher in 4th grade are significantly less likely to take the SAT and PSAT - first steps to college enrollment - and are less likely to report an intention to attend community college in 11th grade surveys. For 5th grade teacher punitiveness, we document similar patterns; however, the estimated effect on SAT/PSAT taking is less precise while the effect on 11th grade reported intentions of attending a 4-year college is negative and significant.

Given that the effect of teacher racial bias in their use of referrals carries heterogeneous effects by student race, we estimate the intermediate effects of teacher bias separately for Black and White students in Table 10 and Table 11 respectively. As with teacher punitiveness, we find that the negative effects of teachers' racial bias in the use of disciplinary referrals persist into future years. Black students who face a more racially biased teacher in

terms of disciplinary actions have more referrals and absences and significantly lower math and ELA achievement in middle school. We find that the negative effects on achievement fade and become imprecise by 8th grade. Meanwhile, for White students, we find generally no evidence of negative intermediate effects from facing a more racially biased teacher in elementary school. In fact, perhaps surprisingly, we find some evidence of a small positive impact on math achievement.

Figure 4 and Figure 5 reproduce this analysis to examine the impact of teachers' racial bias in referral use on long-run outcomes separately by Black and White students. As with both elementary and middle school outcomes, teachers' racial bias in referral use carries significant negative impacts on Black students and few meaningful effects on White students. Black students who face a more biased teacher in elementary school are about 1 percentage point less likely to graduate high school or report an intention to attend a 4-year university after high school. Meanwhile, White students generally see little benefit or harm from facing more racially biased teachers in elementary school.

Overall, our analysis of student outcomes in middle school attributable to the referral use patterns of their 4th and 5th grade teachers suggests that more punitive use of referrals carries lasting impacts on both students' future academic behaviors and future achievement. Similarly, the negative impact for Black students of facing a racially biased teacher - an early exposure to differential treatment in schools - lasts into middle and high school while White students are generally unaffected by such teachers in either the short or long term. The effects are generally consistent with a continued disengagement and disidentification with academic life after exposure to more punitive or biased teachers.

Table 9: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes in Middle School, 2008-2013

	4th Grade				5th Grade			
	(1) Referrals	(2) Absences	(3) Math	(4) ELA	(5) Referrals	(6) Absences	(7) Math	(8) ELA
Panel A: 6th Grade Outcomes								
Punitiveness	0.055***	0.117	-0.034**	-0.025*	0.066***	0.216**	-0.019	-0.008
	(0.01)	(0.10)	(0.01)	(0.01)	(0.01)	(0.09)	(0.01)	(0.01)
Adjusted R^2	0.141	0.175	0.588	0.595	0.157	0.213	0.636	0.629
Observations	131,042	149,540	148,879	148,689	129,724	148,197	147,889	147,693
Panel B: 7th Grade Outcomes								
Punitiveness	0.062***	0.314***	-0.019*	-0.034***	0.064***	0.289**	-0.033***	-0.007
	(0.01)	(0.10)	(0.01)	(0.01)	(0.01)	(0.12)	(0.01)	(0.01)
Adjusted R^2	0.134	0.153	0.563	0.549	0.146	0.183	0.615	0.590
Observations	132,806	145,173	143,960	143,837	131,512	144,394	143,197	143,065
Panel C: 8th Grade Outcomes								
Punitiveness	0.072***	0.356***	-0.015	-0.016	0.088***	0.446***	-0.024**	-0.018*
	(0.01)	(0.13)	(0.01)	(0.01)	(0.01)	(0.13)	(0.01)	(0.01)
Adjusted R^2	0.126	0.128	0.521	0.520	0.138	0.153	0.565	0.556
Observations	130,812	121,563	106,718	106,625	129,743	121,304	139,951	139,825
All teacher, teacher, and class controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
Lagged absences		✓				✓		
Lagged referrals	✓				✓			
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. All lags measured in t-1 relative to 4th grade (columns 1 through 4) or 5th grade (columns 5 through 8).

Table 10: Effect of Teacher Bias in Referrals for Subjective Infractions on Student Middle School Outcomes, Black Students, 2008-2013

	4th Grade				5th Grade			
	(1) Referrals	(2) Absences	(3) Math	(4) ELA	(5) Referrals	(6) Absences	(7) Math	(8) ELA
Panel A: 6th Grade Outcomes								
Teacher bias	0.042***	0.095	-0.029**	-0.018	0.052***	0.261***	-0.002	-0.004
	(0.01)	(0.08)	(0.01)	(0.01)	(0.01)	(0.09)	(0.01)	(0.01)
Adjusted R^2	0.080	0.131	0.566	0.576	0.099	0.155	0.633	0.627
Observations	20,519	24,866	24,743	24,703	19,473	23,689	23,643	23,600
PANEL B: 7th Grade Outcomes								
Teacher bias	0.050***	0.315***	-0.040***	-0.036***	0.054***	0.335***	-0.023**	-0.022**
	(0.01)	(0.09)	(0.01)	(0.01)	(0.01)	(0.12)	(0.01)	(0.01)
Adjusted R^2	0.074	0.093	0.477	0.492	0.089	0.121	0.542	0.540
Observations	20,864	24,140	23,881	23,856	19,809	23,140	22,925	22,905
PANEL C: 8th Grade Outcomes								
Teacher bias	0.052***	0.363***	-0.016	-0.022*	0.061***	0.151	-0.016*	-0.013
	(0.01)	(0.11)	(0.01)	(0.01)	(0.01)	(0.11)	(0.01)	(0.01)
Adjusted R^2	0.067	0.065	0.421	0.454	0.084	0.080	0.474	0.496
Observations	20,025	20,905	17,380	17,359	19,257	19,287	22,354	22,324
All teacher, teacher, and class controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
Lagged absences		✓				✓		
Lagged referrals	✓				✓			
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓

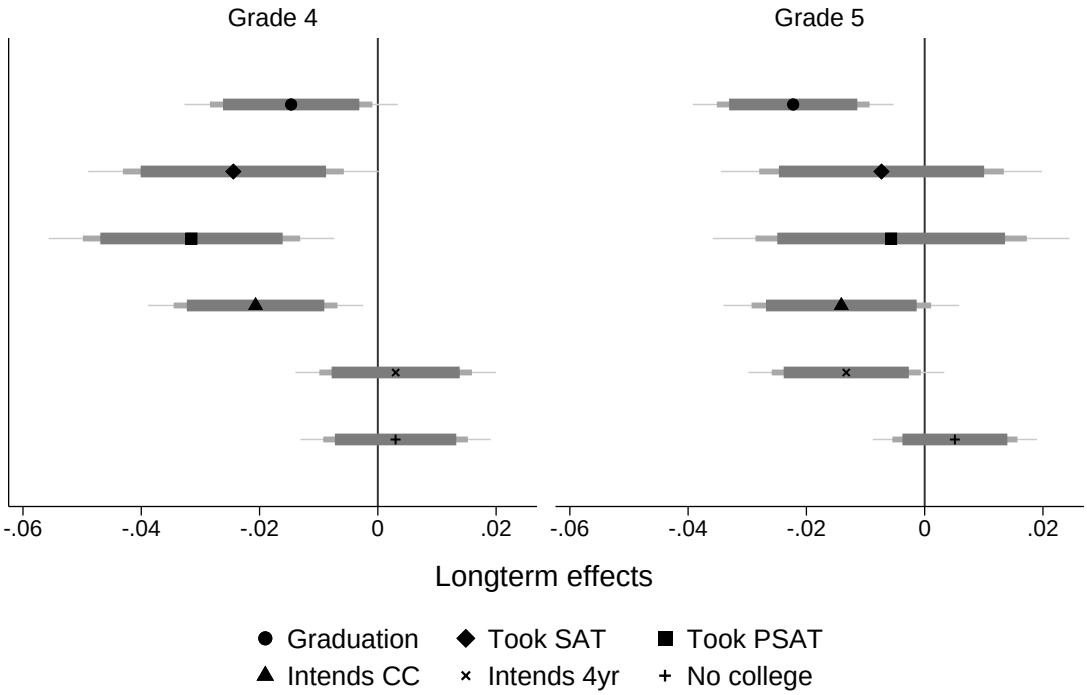
Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes students with no referrals in all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. All lags measured in t-1 relative to 4th grade (columns 1 through 4) or 5th grade (columns 5 through 8).

Table 11: Effect of Teacher Bias in Referrals for Subjective Infractions on Student Middle School Outcomes, White Students, 2008-2013

	4th Grade				5th Grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Referrals	Absences	Math	ELA	Referrals	Absences	Math	ELA
Panel A: 6th Grade Outcomes								
Teacher bias	-0.012***	-0.023	0.006	0.002	-0.001	-0.011	-0.006	0.000
	(0.00)	(0.04)	(0.01)	(0.01)	(0.00)	(0.04)	(0.01)	(0.01)
Adjusted R^2	0.108	0.170	0.554	0.551	0.127	0.219	0.604	0.587
Observations	51,592	57,039	56,797	56,743	50,419	55,951	55,896	55,839
PANEL B: 7th Grade Outcomes								
Teacher bias	-0.011**	-0.059	0.014**	-0.001	-0.006	-0.024	-0.004	0.003
	(0.00)	(0.05)	(0.01)	(0.01)	(0.00)	(0.04)	(0.01)	(0.01)
Adjusted R^2	0.099	0.149	0.545	0.511	0.113	0.187	0.592	0.553
Observations	52,063	55,294	54,910	54,883	50,864	54,393	54,076	54,042
PANEL C: 8th Grade Outcomes								
Teacher bias	-0.005	0.070	0.019**	0.011	-0.002	0.025	-0.002	0.009
	(0.00)	(0.08)	(0.01)	(0.01)	(0.00)	(0.07)	(0.01)	(0.01)
Adjusted R^2	0.091	0.127	0.494	0.483	0.103	0.152	0.541	0.524
Observations	51,515	46,166	40,322	40,290	50,593	44,706	52,865	52,827
All teacher, teacher, and class controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
Lagged absences		✓				✓		
Lagged referrals	✓				✓			
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓

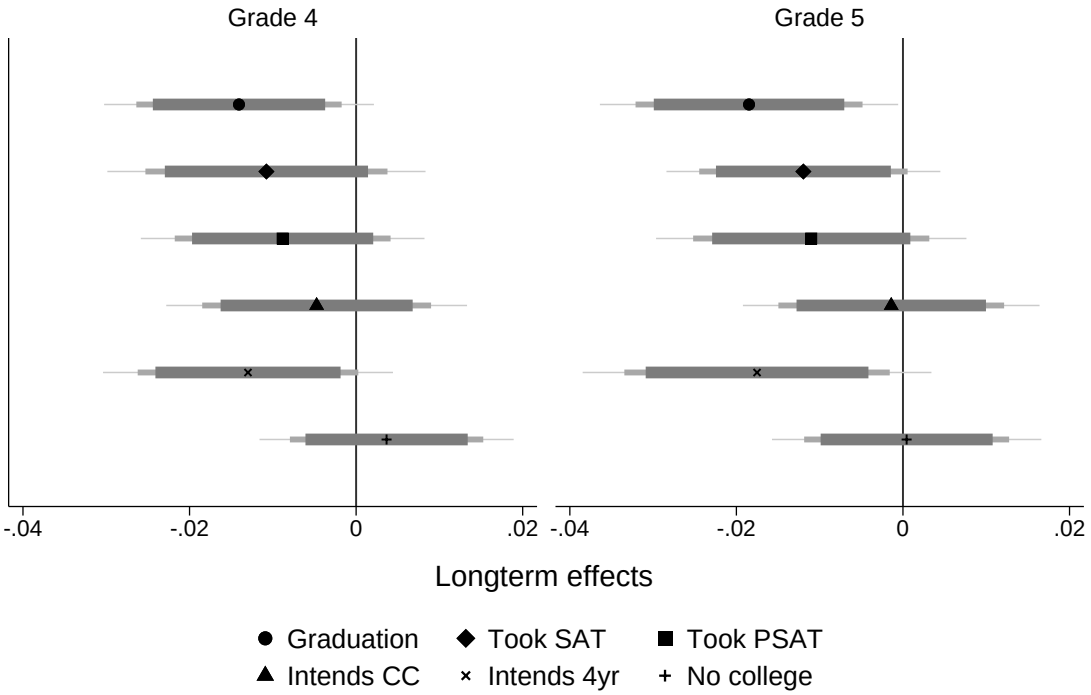
Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes students with no referrals in all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. All lags measured in t-1 relative to 4th grade (columns 1 through 4) or 5th grade (columns 5 through 8).

Figure 3: Effect of teacher punitiveness on long-run outcomes



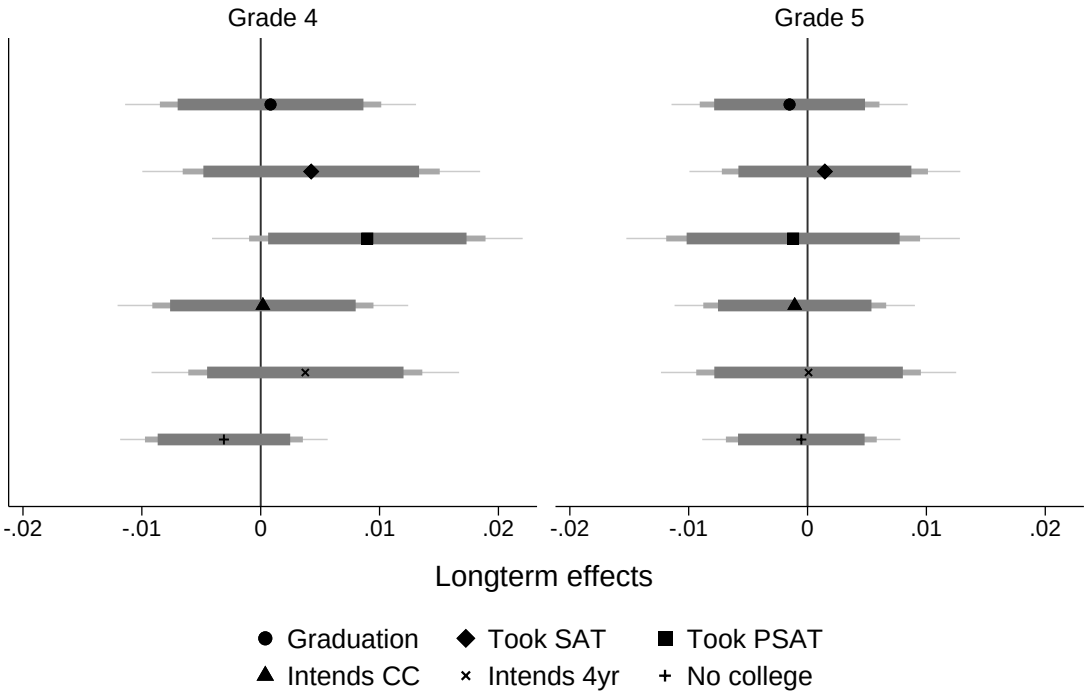
Note: Figure plots coefficient of teacher punitiveness in 4th and 5th grade on long-run educational outcomes. Outcomes are binary indicators for whether a student graduates high-school in North Carolina, took the SAT, took a PSAT, or reports an intent to attend a community college (CC), four-year university, or reports no college intentions. 4th grade sample N = 132,878. 5th grade sample N = 132,384. Model includes all student and teacher controls from primary effects models (columns 7 and 8 in Table 3). 90%, 95%, and 99% confidence intervals plotted in descending order of line thickness. Table A.12 in the appendix presents the coefficients.

Figure 4: Effect of teacher bias on long-run outcomes for Black students



Note: Figure plots coefficient of teacher bias in used of referrals in 4th and 5th grade on long-run educational outcomes. Outcomes are binary indicators for whether a student graduates high-school in North Carolina, took the SAT, took a PSAT, or reports an intent to attend a community college (CC), four-year university, or reports no college intentions. 4th grade sample N = 22,568. 5th grade sample N = 20,767. Model includes all student and teacher controls from primary effects models (columns 7 and 8 in Table 4). 90%, 95%, and 99% confidence intervals plotted in descending order of line thickness. Table A.13 in the appendix presents the coefficients.

Figure 5: Effect of teacher bias on long-run outcomes for White students



Note: Figure plots coefficient of teacher bias in used of referrals in 4th and 5th grade on long-run educational outcomes. Outcomes are binary indicators for whether a student graduates high-school in North Carolina, took the SAT, took a PSAT, or reports an intent to attend a community college (CC), four-year university, or reports no college intentions. 4th grade sample N = 51,030. 5th grade sample N = 49,188. Model includes all student and teacher controls from primary effects models (columns 7 and 8 in Table 4). 90%, 95%, and 99% confidence intervals plotted in descending order of line thickness. Table A.13 in the appendix presents the coefficients.

5 Discussion

Using a value-added approach to first quantify teachers' tendencies in use of disciplinary referrals and second identify the effects of these tendencies on student educational outcomes, we have shown that harsher uses of disciplinary actions lead to worse short- and long-term educational outcomes. Similarly, racially biased teachers on this dimension worsen educational

outcomes among Black students while White students seem relatively unaffected. General measures of teacher effectiveness, like value-added measures of a teachers' contribution to math and reading growth, do not absorb the effects of general punitiveness or teacher bias in the use of referrals, which suggests the effects of referral use operates independently of other dimensions of effectiveness.

We view teachers' tendencies in the use of referrals to be an approximate measure of their approaches to classroom management and relationships with their students. Indeed, given our sample of elementary school students, the increase in absenteeism among students assigned a more punitive teacher suggests teachers' harshness in their discretionary punishment decisions likely harms their relationships with students and/or their families. Broadly, we document a negative association between more punitive referral usage and classroom effectiveness, which calls into question the efficacy of referrals for more subjective infractions as a means to manage classrooms. A teacher's referral usage may also be a behavioral response to other problems - a symptom of poor classroom management skills rather than a cause.

Perhaps most troubling is the effect of racially biased use of disciplinary actions. Our findings suggest racially biased use of referrals has negative impacts on student achievement and attendance for Black students, but modest to no impact on White students. Worse, these negative effects on Black students persist into later years of schooling and eventually translate into lower likelihood of high school graduation and lower intentions to attend a four-year college. The effects of racial bias in teacher referrals remain even after accounting for teacher effectiveness, which suggests that our conventional means for evaluating teachers does not capture potential teacher-specific drivers of differences in student outcomes.

A growing consensus of research tells us that suspensions are bad for students (Lacoe and Steinberg, 2018; Bacher-Hicks et al., 2019; Anderson et al., 2019) and that schools suspend Black students at disproportionately high rates (Skiba et al., 2014; Kinsler, 2011; Barrett et al., 2021). However, the roots of this problem and of this disparity are still

not well understood. Cleary, school principals and administrators make any final decisions regarding whether and how to punish students for disciplinary infractions (Sorensen et al., 2021). However, long before this decision is made, teachers must make the choice of whether to refer a student. Assigning severity to a disciplinary offense in the classroom is highly subjective, with the majority of disciplinary offenses coded using phrases such as "disruptive behavior," "inappropriate language," or "disrespect of faculty or staff." Our findings show that teacher discretion in referring student behavior has immediate effects not only on the referred students, but on the engagement and learning of the entire classroom.

6 Conclusion

This study contains several limitations. First, in order to focus on settings where each student only sees a single teacher, we limit our sample to fourth and fifth grade self-contained classrooms. Disciplinary referrals and suspensions occur more rarely in elementary school grade levels than in middle school or high school grade levels, and therefore variation in teacher punitiveness is more limited. Nonetheless, we show that even in low-discipline settings, teacher referring behavior still matters for student outcomes.

Second, we cannot observe student behavior directly. We only observe student offenses that are reported within the administrative data system. This means that we are necessarily blind to any student behavioral incidents that teachers either choose to ignore or choose to handle internally. Nonetheless, by controlling for student prior incidents and prior test scores, we can estimate teacher punitiveness and bias conditional on prior student referrals and ability.

The take-away message from this paper may seem straightforward: teachers should use fewer referrals. However, this problem is not a simple one in practice. Many teachers use disciplinary referrals because they have run out of alternative options for classroom management. In a recent survey of Philadelphia teachers, many teachers reported feeling unsafe in

their classroom when they did not have the option to suspend disruptive or violent students (Griffith and Tyner, 2019). Although alternatives to traditional discipline abound – such as restorative justice and positive behavioral interventions – implementation of these programs is inconsistent and training often insufficient (Anderson and McKenzie, 2022; Wang, 2022; Augustine et al., 2018; Lacoe and Steinberg, 2019). Teachers should receive the training and resources they need to manage classroom conflict without the use of traditional discipline. Further, diversifying the teacher workforce could help to mitigate racial bias in teacher disciplinary actions (Lindsay and Hart, 2017). As a shorter-term option, enhancing teacher awareness around issues of implicit bias, or more generally encouraging empathetic responses to student misbehavior (Okonofua et al., 2022), could reduce the severity of racial disproportionality in discipline.

References

Amemiya, J., Mortenson, E., and Wang, M.-T. (2020). Minor infractions are not minor: School infractions for minor misconduct may increase adolescents' defiant behavior and contribute to racial disparities in school discipline. *American Psychologist*, 75(1):23.

Anderson, K. P. and McKenzie, S. (2022). Local implementation of state-level discipline policy: Administrator perspectives and contextual factors associated with compliance. *AERA Open*, 8:23328584221075341.

Anderson, K. P., Ritter, G. W., and Zamarro, G. (2019). Understanding a vicious cycle: The relationship between student discipline and student academic outcomes. *Educational Researcher*, 48(5):251–262.

Aucejo, E. M. and Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, 55:70–87.

Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., and Joseph, A. A. (2018). Can restorative practices improve school climate and curb suspensions. *An evaluation of the impact of restorative practices in a mid-sized urban school district*, pages 1–112.

Bacher-Hicks, A., Billings, S. B., and Deming, D. J. (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime. Technical report, National Bureau of Economic Research.

Barrett, N., McEachin, A., Mills, J. N., and Valant, J. (2021). Disparities and discrimination in student discipline by race and family income. *Journal of Human Resources*, 56(3):711–748.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*,

104(9):2593–2632.

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

Chu, E. M. and Ready, D. D. (2018). Exclusion and urban public high schools: Short-and long-term consequences of school suspensions. *American Journal of Education*, 124(4):479–509.

Craig, A. C. and Martin, D. C. (2019). Discipline reform, school culture, and student achievement. Technical report, unpublished working paper.

Djigic, G. and Stojiljkovic, S. (2011). Classroom management styles, classroom climate and school achievement. *Procedia-Social and Behavioral Sciences*, 29:819–828.

Eden, M. (2017). School discipline reform and disorder: Evidence from new york city public schools, 2012-16. *The Education Digest*, 83(1):22.

Gershenson, S., Jacknowitz, A., and Brannegan, A. (2017). Are student absences worth the worry in us primary schools? *Education Finance and Policy*, 12(2):137–165.

Girvan, E. J., Gion, C., McIntosh, K., and Smolkowski, K. (2017). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School psychology quarterly*, 32(3):392.

Gregory, A. and Korth, J. (2016). Teacher-student relationships and behavioral engagement in the classroom. *Handbook of social influences in school contexts: Social-emotional, motivation, and cognitive outcomes*, pages 178–191.

Gregory, J. F. (1995). The crime of punishment: Racial and gender disparities in the use of corporal punishment in us public schools. *Journal of Negro Education*, pages 454–462.

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

Hinze-Pifer, R. and Sartain, L. (2018). Rethinking universal suspension for severe student behavior. *Peabody Journal of Education*, 93(2):228–243.

Holt, S. B. and Gershenson, S. (2017). The impact of demographic representation on absences and suspensions. *Policy Studies Journal*.

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.

Jacob, B. A., Lefgren, L., and Sims, D. P. (2010). The persistence of teacher-induced learning. *Journal of Human Resources*, 45(4):915–943.

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

Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, 54(1):355–383.

Kwok, A. (2017). Relationships between instructional quality and classroom management for beginning urban teachers. *Educational Researcher*, 46(7):355–365.

Lacoe, J. and Steinberg, M. P. (2018). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes. *Peabody Journal of Education*, 93(2):207–227.

Lacoe, J. and Steinberg, M. P. (2019). Do suspensions affect student outcomes? *Educational Evaluation and Policy Analysis*, 41(1):34–62.

Lindsay, C. A. and Hart, C. M. (2017). Exposure to same-race teachers and student disci-

plinary outcomes for black students in north carolina. *Educational Evaluation and Policy Analysis*, 39(3):485–510.

Liu, J., Hayes, M., and Gershenson, S. (2021a). From referrals to suspensions: New evidence on racial disparities in exclusionary discipline.

Liu, J., Lee, M., and Gershenson, S. (2021b). The short-and long-run impacts of secondary school absences. *Journal of Public Economics*, 199:104441.

Morris, E. W. and Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, 63(1):68–86.

Noltemeyer, A. L., Ward, R. M., and Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, 44(2):224–240.

Novak, A. (2021). Trajectories of exclusionary discipline: Risk factors and associated outcomes. *Journal of School Violence*, 20(2):182–194.

Okonofua, J. A. and Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological science*, 26(5):617–624.

Okonofua, J. A., Goyer, J. P., Lindsay, C. A., Haugabrook, J., and Walton, G. M. (2022). A scalable empathetic-mindset intervention reduces group disparities in school suspensions. *Science Advances*, 8(12):eabj0691.

Pearman, F. A., Curran, F. C., Fisher, B., and Gardella, J. (2019). Are achievement gaps related to discipline gaps? evidence from national data. *Aera Open*, 5(4):2332858419875440.

Perry, B. L. and Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, 79(6):1067–1087.

Pianta, R. C. (2016). Teacher–student interactions: Measurement, impacts, improvement, and policy. *Policy Insights from the Behavioral and Brain Sciences*, 3(1):98–105.

Pyne, J. (2019). Suspended attitudes: Exclusion and emotional disengagement from school. *Sociology of Education*, 92(1):59–82.

Roorda, D. L., Jak, S., Zee, M., Oort, F. J., and Koomen, H. M. (2017). Affective teacher–student relationships and students' engagement and achievement: A meta-analytic update and test of the mediating role of engagement. *School Psychology Review*, 46(3):239–261.

Roorda, D. L., Koomen, H. M., Spilt, J. L., and Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of educational research*, 81(4):493–529.

Sabol, T. J. and Pianta, R. C. (2012). Recent trends in research on teacher–child relationships. *Attachment & human development*, 14(3):213–231.

Santiago-Rosario, M. R., Whitcomb, S. A., Pearlman, J., and McIntosh, K. (2021). Associations between teacher expectations and racial disproportionality in discipline referrals. *Journal of School Psychology*, 85:80–93.

Shi, Y. and Zhu, M. (2021). Equal time for equal crime? racial bias in school discipline.

Skiba, R. J., Arredondo, M. I., and Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education*, 47(4):546–564.

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

Skiba, R. J., Michael, R. S., Nardo, A. C., and Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The urban review*, 34(4):317–342.

Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N., and Horner, R. H. (2016). Vulnerable decision points for disproportionate office discipline referrals: Comparisons of discipline for african american and white elementary school students. *Behavioral Disorders*, 41(4):178–195.

Sorensen, L. C., Bushway, S. D., and Gifford, E. J. (2021). Getting tough? the effects of discretionary principal discipline on student outcomes. *Education Finance and Policy*, pages 1–74.

Tran, L. and 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.

Williford, A. P., Alamos, P., Whittaker, J. E., and Accavitti, M. R. (2021). Who's left out of learning? racial disparities in teachers' reports of exclusionary discipline strategies beyond suspensions and expulsions. *EdWorkingPaper*, (21-472).

Wolf, K. C. and Kupchik, A. (2017). School suspensions and adverse experiences in adulthood. *Justice Quarterly*, 34(3):407–430.

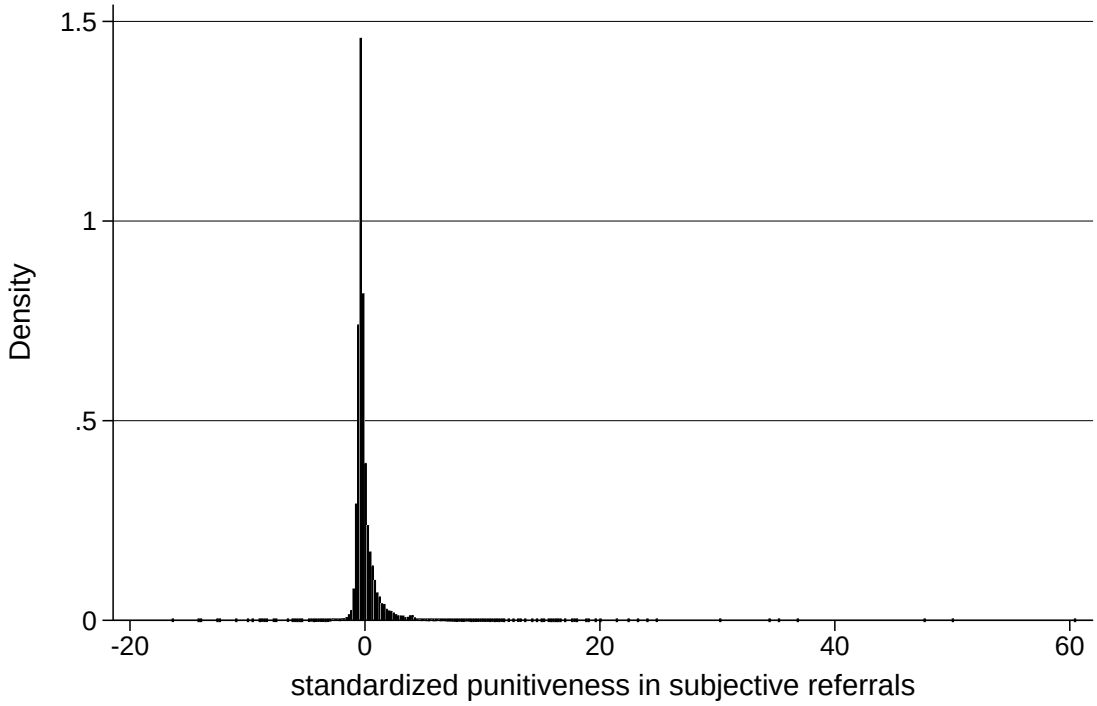
Wubbels, T., Brekelmans, M., and Mainhard, T. (2016). Teacher–student relationships and student achievement. In *Handbook of social influences in school contexts*, pages 137–152. Routledge.

Appendix A

Table A.1: Infractions coded as subjective referrals

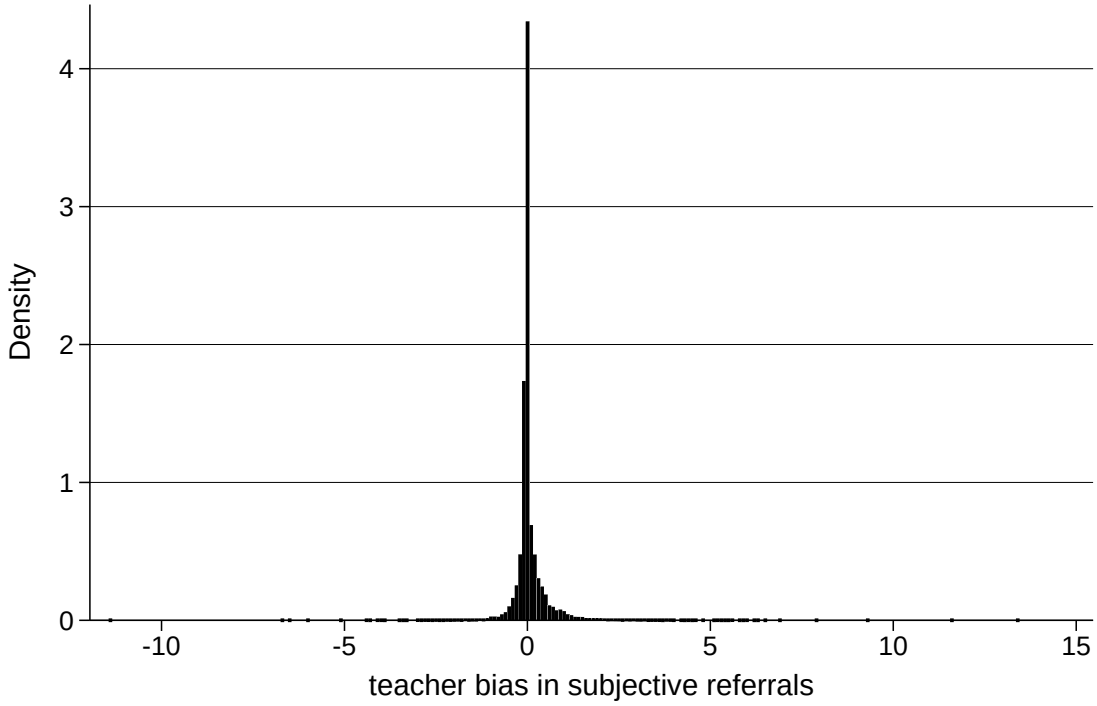
	(1) Subjective	(2) %	(3) Total
Possession of alcoholic beverage	No	0.000	0.001
Assault resulting in a serious injury	No	0.000	0.000
Assault on school personnel	No	0.000	0.000
Assault involving use of a weapon	No	0.000	0.000
Possession of a firearm or powerful explosive	No	0.000	0.000
Bomb threat	No	0.000	0.000
Burning of a school building	No	0.000	0.000
Possession of a weapon (not firearm)	No	0.002	0.017
Robbery without a dangerous weapon	No	0.000	0.000
Sexual assault (not involving rape or sexual offense)	No	0.000	0.000
Sexual offense	No	0.000	0.000
Possession of controlled substance in violation of law	No	0.000	0.001
Disruptive behavior	Yes	0.041	0.169
Unlawfully setting a fire	No	0.000	0.000
Insubordination	Yes	0.011	0.037
Fighting	No	0.021	0.148
Inappropriate language/disrespect	Yes	0.017	0.086
Aggressive behavior	Yes	0.029	0.159
Bus misbehavior	No	0.039	0.183
Disrespect of faculty/staff	Yes	0.006	0.019
Bullying	No	0.006	0.031
Late to class	No	0.008	0.020
Cutting class	No	0.000	0.000
Theft	No	0.005	0.035
Excessive tardiness	No	0.000	0.001
Disorderly conduct	Yes	0.003	0.014
Being in an unauthorized area	No	0.003	0.014
Dress code violation	No	0.000	0.001
Skipping school	No	0.000	0.001
Communicating threats	No	0.003	0.017
Cell phone use	No	0.000	0.002
Property damage	No	0.002	0.011
Assault on student not resulting in serious injury	No	0.002	0.010
Inappropriate items on school property	No	0.003	0.021
Observations	313,326	313,326	17,844

Figure A.1: Distribution of Teacher Punitiveness in Referrals for Subjective Infractions



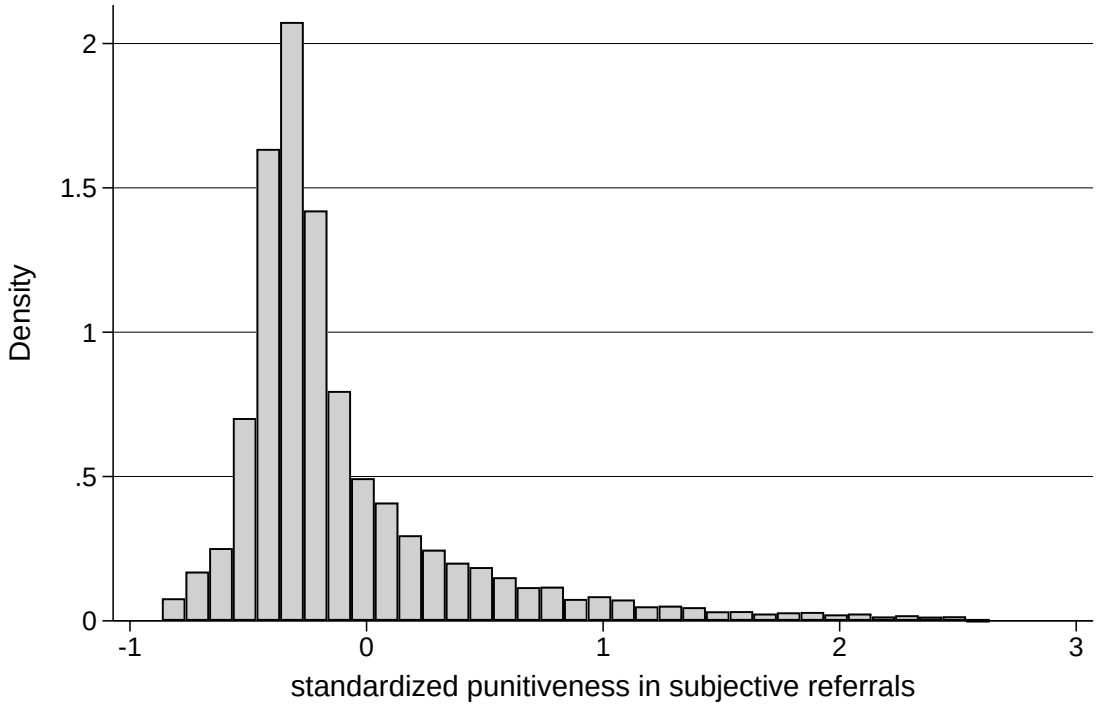
Note: Units of observation are teacher-years. Punitiveness estimate standardized.

Figure A.2: Distribution of Teacher Racial Bias in Referrals for Subjective Infractions



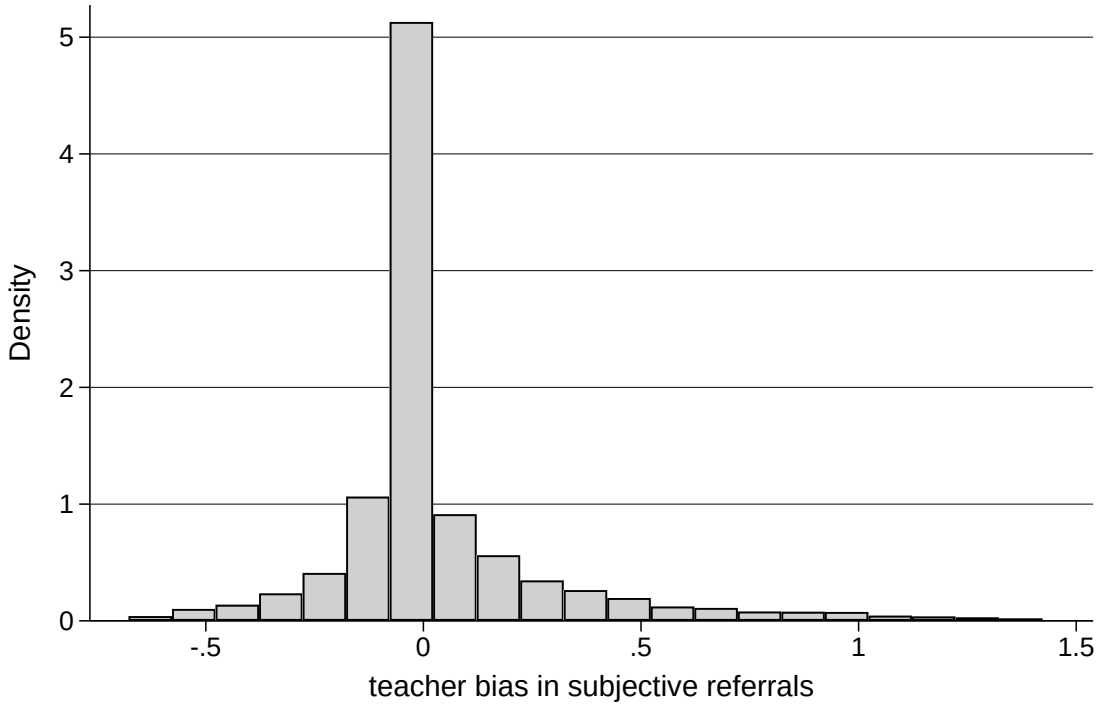
Note: Units of observation are teacher-years.

Figure A.3: Distribution of Teacher Punitiveness in Referrals for Subjective Infractions, no outliers



Note: Units of observation are teacher-years. Punitiveness estimate standardized. Highest and lowest 2% of observations dropped.

Figure A.4: Distribution of Teacher Racial Bias in Referrals for Subjective Infractions, no outliers



Note: Units of observation are teacher-years. Highest and lowest 2% of observations dropped.

Table A.2: Summary of students in mixed-race self-contained classrooms in grades 3 through 5 in North Carolina Public Schools, 2008-2013

	(1) Total sample	(2) Black student	(3) White student
Bias in subjective referrals	0.13	0.13	0.13
	(0.73)	(0.70)	(0.75)
Absences	5.93	5.00	6.33
	(5.41)	(5.11)	(5.48)
Chronically absent	0.04	0.03	0.04
Math score (std)	0.15	-0.36	0.37
	(0.97)	(0.90)	(0.92)
ELA score (std)	0.14	-0.33	0.34
	(0.95)	(0.90)	(0.90)
Disciplinary offenses ($N\|N > 0$)	2.18	2.44	1.91
	(2.24)	(2.52)	(1.86)
Any disciplinary offenses	0.10	0.18	0.07
	(0.31)	(0.38)	(0.26)
Subjective referrals ($N\|N > 0$)	1.89	2.07	1.70
	(1.79)	(2.00)	(1.51)
Any subjective referrals	0.06	0.11	0.04
	(0.24)	(0.31)	(0.20)
Male student	0.50	0.49	0.51
White student	0.70	0.00	1.00
Black student	0.30	1.00	0.00
Economic disadvantage	0.41	0.72	0.28
Observations	169,898	51,152	118,746

Note: Standard deviations in parentheses.

Table A.3: Effect of Teacher Punitiveness and Bias in Referrals for All Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
PANEL A: Overall effects								
Punitiveness (all referrals)	0.376***	0.007***	-0.037***	-0.058***	0.371***	0.007***	-0.032***	-0.047***
	(0.05)	(0.00)	(0.01)	(0.01)	(0.05)	(0.00)	(0.00)	(0.01)
Adjusted R^2	0.363	0.139	0.643	0.662	0.363	0.139	0.645	0.673
Observations	313,280	313,280	313,280	313,280	313,280	313,280	313,280	313,280
Teacher bias (all referrals)	0.063***	0.001	-0.006**	-0.011***	0.063***	0.001	-0.005**	-0.010***
	(0.02)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.372	0.147	0.639	0.664	0.372	0.147	0.641	0.675
Observations	169,898	169,898	169,898	169,898	169,898	169,898	169,898	169,898
PANEL B: Spillover effects								
Punitiveness (all referrals)	0.012	0.001	-0.028***	-0.049***	0.006	0.001	-0.022***	-0.038***
	(0.04)	(0.00)	(0.01)	(0.01)	(0.04)	(0.00)	(0.01)	(0.01)
Adjusted R^2	0.366	0.135	0.641	0.662	0.366	0.135	0.644	0.672
Observations	266,176	266,176	266,176	266,176	266,176	266,176	266,176	266,176
Teacher bias (all referrals)	-0.008	-0.000	-0.006**	-0.009***	-0.009	-0.000	-0.005**	-0.007***
	(0.02)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.378	0.142	0.634	0.662	0.378	0.142	0.636	0.673
Observations	142,626	142,626	142,626	142,626	142,626	142,626	142,626	142,626
All lagged controls	✓	✓	✓	✓	✓	✓	✓	✓
All teacher, class, & student controls	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, & year FE	✓	✓	✓	✓	✓	✓	✓	✓
Teacher VAR & VAM					✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. A student is considered chronically absent if they miss 18 school days or more in a given academic year. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

Table A.4: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
Punitiveness	0.028**	-0.000	-0.007***	-0.013***	0.021	-0.000	-0.000	0.000
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Value-Added Math	-	-	-	-	-0.067***	-0.001***	0.034***	0.103***
					(0.01)	(0.00)	(0.00)	(0.00)
Value-Added ELA	-	-	-	-	0.005	-0.001	0.023***	0.010***
					(0.01)	(0.00)	(0.00)	(0.00)
All teacher controls	✓	✓	✓	✓	✓	✓	✓	✓
All student controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Absences	✓	✓			✓	✓		
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.362	0.139	0.643	0.661	0.362	0.139	0.645	0.671
Observations	313,326	313,326	313,326	313,326	313,326	313,326	313,326	313,326

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Punitiveness measured using leave-year-out value-added model with shrinkage. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. FE = fixed-effect. Chronic = chronically absent. A student is considered chronically absent if they miss 18 school days or more in a given academic year.

Table A.5: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, separately by teacher characteristics, 2008-2013

	(1) White T	(2) Black T	(3) > Avg. VAR	(4) < Avg. VAR	(5) > Avg. VAM	(6) < Avg. VAM
PANEL A: Absences						
Punitiveness	0.034*	0.022	0.027	0.030	0.019	0.033*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Adjusted R^2	0.365	0.333	0.353	0.363	0.341	0.368
Observations	174,724	67,146	178,581	134,745	175,066	138,260
PANEL B: ELA Scores						
Punitiveness	-0.010***	-0.003	-0.005***	-0.004**	-0.006***	-0.004*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.610	0.589	0.393	0.400	0.505	0.494
Observations	174,724	67,146	178,581	134,745	175,066	138,260
PANEL C: Math Scores						
Punitiveness	-0.013***	-0.011***	-0.011***	-0.013***	-0.009***	-0.008***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.638	0.599	0.552	0.495	0.429	0.374
Observations	174,724	67,146	178,581	134,745	175,066	138,260
All lagged controls	✓	✓	✓	✓	✓	✓
All teacher & student controls	✓	✓	✓	✓	✓	✓
Teacher VAR & VAM						
School, grade, & year FE	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Punitiveness measured using leave-year-out value-added model with shrinkage. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. FE = fixed-effect. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

Table A.6: Effect of Teacher Punitiveness and Bias in Referrals for Subjective Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
PANEL A: Overall effects								
Punitiveness	0.427***	0.009***	-0.036***	-0.057***	0.423***	0.009***	-0.030***	-0.045***
	(0.09)	(0.00)	(0.01)	(0.01)	(0.09)	(0.00)	(0.01)	(0.01)
Adjusted R^2	0.621	0.386	0.816	0.838	0.621	0.386	0.818	0.844
Observations	313,280	313,280	313,280	313,280	313,280	313,280	313,280	313,280
Teacher bias	0.055	-0.000	-0.006	-0.012*	0.053	-0.000	-0.005	-0.009
	(0.05)	(0.00)	(0.01)	(0.01)	(0.05)	(0.00)	(0.01)	(0.01)
Adjusted R^2	0.633	0.401	0.810	0.836	0.633	0.401	0.811	0.843
Observations	137,731	137,731	137,731	137,731	137,731	137,731	137,731	137,731
PANEL B: Spillover effects								
Punitiveness	0.112	0.004	-0.036***	-0.057***	0.108	0.004	-0.028**	-0.043***
	(0.09)	(0.00)	(0.01)	(0.02)	(0.09)	(0.00)	(0.01)	(0.01)
Adjusted R^2	0.623	0.380	0.813	0.835	0.623	0.380	0.814	0.842
Observations	266,168	266,168	266,168	266,168	266,168	266,168	266,168	266,168
Teacher bias	-0.021	-0.001	-0.008	-0.014*	-0.023	-0.001	-0.007	-0.011
	(0.05)	(0.00)	(0.01)	(0.01)	(0.05)	(0.00)	(0.01)	(0.01)
Adjusted R^2	0.634	0.393	0.804	0.832	0.634	0.393	0.805	0.839
Observations	115,015	115,015	115,015	115,015	115,015	115,015	115,015	115,015
Student FE	✓	✓	✓	✓	✓	✓	✓	✓
All teacher & student controls	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, & year FE	✓	✓	✓	✓	✓	✓	✓	✓
Teacher VAR & VAM					✓	✓	✓	✓

Note: Cluster-robust standard errors (clustered at the class-level) in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available for estimating teacher punitiveness). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. A student is considered chronically absent if they miss 18 school days or more in a given academic year. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

Table A.7: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, separately by student characteristics, 2008-2013

	(1) White S	(2) Black S	(3) > Avg. ELA	(4) < Avg. ELA	(5) > Avg. Math	(6) < Avg. Math
PANEL A: Absences						
Punitiveness	0.351***	0.749***	0.354***	0.590***	0.333***	0.565***
	(0.08)	(0.10)	(0.07)	(0.08)	(0.07)	(0.08)
Adjusted R^2	0.366	0.335	0.354	0.364	0.341	0.369
Observations	174,700	67,128	178,564	134,716	175,054	138,226
PANEL B: ELA Scores						
Punitiveness	-0.058***	-0.044***	-0.014**	-0.038***	-0.034***	-0.040***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.610	0.589	0.393	0.400	0.505	0.495
Observations	174,700	67,128	178,564	134,716	175,054	138,226
PANEL C: Math Scores						
Punitiveness	-0.094***	-0.072***	-0.059***	-0.082***	-0.042***	-0.060***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.640	0.601	0.554	0.498	0.430	0.376
Observations	174,700	67,128	178,564	134,716	175,054	138,226
All lagged controls	✓	✓	✓	✓	✓	✓
All teacher, class, and student controls	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

Table A.8: Spillover Effect of Teacher Punitiveness on Student Outcomes, separately by teacher characteristics, 2008-2013

	(1) White T	(2) Black T	(3) > Avg. VAR	(4) < Avg. VAR	(5) > Avg. VAM	(6) < Avg. VAM
PANEL A: Absences						
Punitiveness	0.045	0.007	-0.052	0.046	0.078	-0.051
	(0.07)	(0.16)	(0.08)	(0.09)	(0.10)	(0.08)
Adjusted R^2	0.365	0.375	0.359	0.370	0.360	0.370
Observations	231,589	27,164	146,559	119,609	141,871	124,297
PANEL B: ELA Scores						
Punitiveness	-0.034***	-0.020	-0.045***	-0.019	-0.041***	-0.022*
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.642	0.620	0.635	0.646	0.636	0.645
Observations	231,589	27,164	146,559	119,609	141,871	124,297
PANEL C: Math Scores						
Punitiveness	-0.063***	-0.079***	-0.077***	-0.043***	-0.074***	-0.043***
	(0.01)	(0.03)	(0.02)	(0.01)	(0.02)	(0.01)
Adjusted R^2	0.664	0.650	0.663	0.662	0.664	0.668
Observations	231,589	27,164	146,559	119,609	141,871	124,297
All lagged controls	✓	✓	✓	✓	✓	✓
All teacher, class, and student controls	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. VAR = Teacher value-added score in ELA. VAM = Teacher value-added score in math.

Table A.9: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, 2008-2013

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
Punitiveness	0.475***	0.011***	-0.044***	-0.075***	0.469***	0.010***	-0.037***	-0.062***
	(0.06)	(0.00)	(0.01)	(0.01)	(0.06)	(0.00)	(0.01)	(0.01)
Value-Added Math	-	-	-	-	-0.475***	-0.008**	0.267***	0.778***
					(0.11)	(0.00)	(0.01)	(0.02)
Value-Added ELA	-	-	-	-	0.152	-0.007	0.329***	0.150***
					(0.19)	(0.01)	(0.03)	(0.03)
All teacher controls	✓	✓	✓	✓	✓	✓	✓	✓
All student controls	✓	✓	✓	✓	✓	✓	✓	✓
All classroom controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Absences	✓	✓			✓	✓		
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.363	0.139	0.643	0.662	0.363	0.139	0.645	0.673
Observations	313,280	313,280	313,280	313,280	313,280	313,280	313,280	313,280

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. Chronic = chronically absent. A student is considered chronically absent if they miss 18 school days or more in a given academic year. Class-level average lagged peer referrals and lagged average peer test scores included in punitiveness model.

Table A.10: Effect of Teacher Punitiveness in Referrals for Subjective Infractions on Student Outcomes, 2008-2013, Without Outliers

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
Punitiveness	0.710***	0.016***	-0.090***	-0.139***	0.702***	0.016***	-0.081***	-0.121***
	(0.11)	(0.00)	(0.01)	(0.02)	(0.11)	(0.00)	(0.01)	(0.02)
Value-Added Math	-	-	-	-	-0.501***	-0.008**	0.268***	0.780***
					(0.11)	(0.00)	(0.01)	(0.02)
Value-Added ELA	-	-	-	-	0.160	-0.007	0.328***	0.141***
					(0.19)	(0.01)	(0.03)	(0.03)
All teacher controls	✓	✓	✓	✓	✓	✓	✓	✓
All student controls	✓	✓	✓	✓	✓	✓	✓	✓
All classroom controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Absences	✓	✓			✓	✓		
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.363	0.138	0.643	0.662	0.363	0.138	0.645	0.673
Observations	300,776	300,776	300,776	300,776	300,776	300,776	300,776	300,776

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. Chronic = chronically absent. A student is considered chronically absent if they miss 18 school days or more in a given academic year. Highest and lowest 2% of punitiveness scores excluded.

Table A.11: Effect of Teacher Bias in Referral Use for Subjective Infractions on Student Outcomes, 2008-2013, Without Outliers

	(1) Absences	(2) Chronic	(3) ELA	(4) Math	(5) Absences	(6) Chronic	(7) ELA	(8) Math
Teacher Bias	0.188***	0.004**	-0.014**	-0.027***	0.187***	0.004**	-0.013**	-0.024***
	(0.05)	(0.00)	(0.01)	(0.01)	(0.05)	(0.00)	(0.01)	(0.01)
Value-Added Math	-	-	-	-	-0.582***	-0.011**	0.252***	0.797***
					(0.14)	(0.00)	(0.02)	(0.02)
Value-Added ELA	-	-	-	-	0.389	-0.002	0.362***	0.167***
					(0.26)	(0.01)	(0.04)	(0.04)
All teacher controls	✓	✓	✓	✓	✓	✓	✓	✓
All student controls	✓	✓	✓	✓	✓	✓	✓	✓
All classroom controls	✓	✓	✓	✓	✓	✓	✓	✓
Lagged Absences	✓	✓			✓	✓		
Lagged ELA and Math Scores	✓	✓	✓	✓	✓	✓	✓	✓
School, grade, and year FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.372	0.146	0.638	0.664	0.372	0.146	0.640	0.675
Observations	163,194	163,194	163,194	163,194	163,194	163,194	163,194	163,194

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Sample includes all standard, grade-level, self-contained classrooms from 4th and 5th grade (grades in which lagged test scores are available). Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect. A student is considered chronically absent if they miss 18 school days or more in a given academic year. Highest and lowest 2% of bias scores excluded.

Table A.12: Effect of Teacher Punitiveness on Long-Run Educational Outcomes

	(1) Graduation	(2) Took SAT	(3) Took PSAT	(4) Intends CC	(5) Intends 4 yr	(6) No college
Panel A. 4th Grade						
Punitiveness	-0.015**	-0.024**	-0.032***	-0.021***	0.003	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.019	0.214	0.123	0.019	0.138	0.033
Observations	132,878	132,878	132,878	132,873	132,873	132,873
Panel B. 5th Grade						
Punitiveness	-0.022***	-0.007	-0.006	-0.014*	-0.013**	0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.019	0.217	0.126	0.021	0.145	0.034
Observations	132,384	132,384	132,384	132,379	132,379	132,379
All teacher, student, and classroom controls	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p < .10, **p < .05, ***p < .01. Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged, and lagged ELA and math achievement. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect.

Table A.13: Effect of Teacher Bias on Long-Run Educational Outcomes, Separately by Race

	(1) Graduation	(2) Took SAT	(3) Took PSAT	(4) Intends CC	(5) Intends 4 yr	(6) No college
Panel A. 4th Grade, Black Students						
Teacher Bias	-0.014**	-0.011	-0.009	-0.005	-0.013*	0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	-0.022	0.123	0.047	-0.018	0.083	-0.008
Observations	22,568	22,568	22,568	22,568	22,568	22,568
Panel B. 5th Grade, Black Students						
Teacher Bias	-0.018***	-0.012*	-0.011	-0.001	-0.018**	0.000
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	-0.024	0.158	0.073	-0.019	0.093	-0.012
Observations	20,767	20,767	20,767	20,767	20,767	20,767
Panel C. 4th Grade, White Students						
Teacher Bias	0.001	0.004	0.009*	0.000	0.004	-0.003
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)
Adjusted R^2	0.013	0.171	0.078	0.008	0.127	0.024
Observations	51,030	51,030	51,030	51,028	51,028	51,028
Panel D. 5th Grade, White Students						
Teacher Bias	-0.002	0.001	-0.001	-0.001	0.000	-0.001
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.015	0.214	0.112	0.007	0.136	0.025
Observations	49,188	49,188	49,188	49,184	49,184	49,184
All teacher, student, and classroom controls	✓	✓	✓	✓	✓	✓

Note: Bootstrapped standard errors estimated using 500 replications in parentheses; *p $<$ .10, **p $<$ .05, ***p $<$ .01. Teacher controls include years of experience, an indicator for a graduate degree, indicators for licensure status, race, and gender. Student controls include race, gender, and an indicator for economically disadvantaged, and lagged ELA and math achievement. Class-level controls includes % economically disadvantaged, class size, % nonwhite students, % male, and lagged average absences, test scores, and referrals of student i's peers. FE = fixed-effect.