ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-216

Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes

Lucy C. Sorensen
University at Albany, SUNY

Shawn D. Bushway
RAND Corporation

Elizabeth J. Gifford
Duke University

Nationwide, school principals are given wide discretion to use disciplinary tools like suspension and expulsion to create a safe learning environment. There is legitimate concern that this power can have negative consequences, particularly for the students who are excluded. This study uses linked disciplinary, education, and criminal justice records from 2008 to 2016 in North Carolina to examine the impact of principal driven disciplinary decisions on middle school student outcomes. We find that principals who are more likely to remove students lead to reductions in reported rates of minor student misconduct. However, this deterrence comes at a high cost – these harsher principals generate more juvenile justice complaints and reduce high school graduation rates for all students in their schools. Students who committed minor disciplinary infractions in a school with a harsh principal suffer additional declines in attendance and test scores. Finally, principals exhibiting racial bias in their disciplinary decisions also widen educational gaps between White and Black students.

VERSION: February 2021

Suggested citation: Sorensen, Lucy C., Shawn D. Bushway, and Elizabeth J. Gifford. (2021). Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes. (EdWorkingPaper: 20-216). Retrieved from Annenberg Institute at Brown University: https://www.edworkingpapers.com/ai20-216

Getting Tough? The Effects of Discretionary Principal Discipline on Student Outcomes

Lucy C. Sorensen[1], Shawn D. Bushway[1,2], and Elizabeth J. Gifford[3]

[1] University at Albany, SUNY

[2] RAND Corporation

[3] Duke University

Acknowledgements: This research is supported by Award 2017-CK-BX-0006 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice. The authors are grateful for comments from Philip Cook, John Engberg, Jason Grissom, Denise Gottfredson, Jeffrey Smith, Yinzhi Shen, the editors and anonymous reviewers, and seminar participants at the Institute for Research on Poverty Summer Research Workshop, Carnegie Mellon University's Heinz College Monday Faculty Seminar, and the AEFP Session on Changing Discipline Practices, Changing Outcomes.

Abstract

Nationwide, school principals are given wide discretion to use disciplinary tools like suspension and expulsion to create a safe learning environment. There is legitimate concern that this power can have negative consequences, particularly for the students who are excluded. This study uses linked disciplinary, education, and criminal justice records from 2008 to 2016 in North Carolina to examine the impact of principal-driven disciplinary decisions on middle school student outcomes. We find that principals who are more likely to remove students lead to reductions in reported rates of minor student misconduct. However, this deterrence comes at a high cost – these harsher principals generate more juvenile justice complaints and reduce high school graduation rates for all students in their schools. Students who committed minor disciplinary infractions in a school with a harsh principal suffer additional declines in attendance and test scores. Finally, principals exhibiting racial bias in their disciplinary decisions also widen educational gaps between White and Black students.

Introduction

Recent high-profile killings of unarmed black citizens by police and subsequent protest movements across the U.S. have precipitated a national reckoning with issues of racial injustice (Barber 2020). These movements have exposed deep-rooted flaws in policing practices, including police practices inside schools (Sosa 2020, Turner and Beneke 2020). Because the same social forces that shape the justice system also shape school disciplinary practices (Hirschfield 2008), it is not surprising that concerns driving the recent debates on racial injustice mirror similar concerns surrounding school discipline, particularly exclusionary discipline.

School administrators have dramatically increased their use of formal punishments such as out-of-school suspensions and expulsions over the last 40 years (Cook, Gottfredson and Na 2010; Losen et al. 2015). Widespread concern that these get-tough policies disproportionately affect students from underrepresented minority groups (Curran 2016; Hoffman 2014; Kinsler 2011; Losen and Martinez 2013; Skiba et al. 2014) led to a comprehensive "Rethink School Discipline" guidance package from the Obama Administration in 2014 designed to curb the disparate use of out-of-school suspensions and expulsions (U.S. Department of Education 2014). Much of the discussion anticipates the direct focus on racial disparity motivating the 2020 Black Lives matter protests against police. And, again in a way that anticipates the Trump Administration response to the 2020 protests (e.g., calls for "law and order") (Phillips and Johnson 2020, Schwartzapfel 2020), the Trump Administration pushed back against the "Rethink" guidance, rescinding it on the grounds that "the threat of investigations by the Office of Civil Rights (OCR) … has likely had a strong negative impact on school discipline and safety" (Federal Commission on School Safety 2018, p. 67).

These dueling directives highlight fundamentally different beliefs about the relative size and importance of potential safety benefits for the school as a whole, as compared to the costs to those individuals who are removed from school, costs that are borne disproportionately by members of underrepresented minority groups. These costs are potentially formidable and might accumulate over time, leading to lower educational attainment or earlier involvement in the criminal justice system (Wolf and Kupchik 2017). Indeed, being suspended from school has been associated with decreased school engagement (Pyne 2019) and increased association with deviant peers (Novak 2019; Hemphill et al. 2012) and increased antisocial behavior (Hemphill et. al. 2006; Hemphill et. al., 2012). Scholars have in turn hypothesized these factors serve as mechanisms by which suspension can lead to criminal justice involvement (Novak 2019; Hemphill 2012; Moore and McArthur 2014; Skiba, Arredondo, and Williams 2014).

Neither presidential policy was guided directly by empirical research exploring potential policy tradeoffs at the school level.[1] Most existing research has focused on either racial disparities in the use of exclusionary discipline (e.g., Barrett et al. 2019, Kinsler 2011, Skiba et al. 2014) or the effects of being suspended or expelled on one's own educational and life outcomes (e.g., Chu and Ready 2018, Lacoe and Steinberg 2018, Perry and Morris 2014). This latter form of research, which shows strong correlations between suspensions and problematic life outcomes (Kupchik 2010), is to varying degrees plagued by the inability to separate the impact of the decision to suspend by the school administrator – the margin of interest – from the impact of the underlying behavior that prompted the disciplinary decision (Anderson, Ritter, and Zamarro 2019).[2] As a result, most reviews stop short from concluding definitively that

[1] The Federal Commission report cited only an opinion survey of school superintendents.

[2] Recent studies have introduced models with student fixed effects to better distinguish between correlates of disciplinary behavior and effects of the disciplinary punishment (Hwang 2018; Lacoe and Steinberg 2018). Although this certainly helps to control for many confounding factors, it cannot account for time-varying factors that

suspensions and expulsions cause later life problems (e.g. Noltemeyer, Ward and Mcloughlin 2015), although causal research on this topic is growing (Bacher-Hicks, Billings, and Deming 2019).

A second limitation of the existing research is a focus only on the students who were subject to the suspension. This focus ignores the possibility of beneficial or harmful spillover effects from suspension on the entire student-body, which may be significant (Lacoe and Steinberg 2018). The Trump administration's decision to rescind the Obama-era policies was based partly on the notion that harsh disciplinary policies benefit the larger student body. To examine the spillover consequences of suspension rates, a new working paper by Bacher-Hicks, Billings, and Deming (2019) exploited a court decision which ended district-wide busing in the Charlotte-Mecklenburg, North Carolina school district – effectively changing which school students attended and hence the suspension policy to which they were exposed. The authors found that assignment to a school with a high suspension rate in the year before the policy change increased the chance of being suspended, dropping out from high school, and being arrested or incarcerated as an adult, and decreased the chance of attending college.

Bacher-Hicks et al. (2019) recognize that their findings may not be due to how school administrators respond to serious disciplinary infractions.[3] First, schools with harsher administrators do not necessarily have higher suspension rates. Economic models of behavior

could predict both a suspension and a poor educational outcome. For example, a student experiencing family problems at home or engaging with a new peer group at school may be both more likely to get suspended and more likely to have attendance and achievement issues that year, even if the student was formerly high-performing. A recent study by Anderson, Ritter, and Zamarro (2019) does include infraction type in their regression estimating the effects of suspension on short-term student academic outcomes – an important control which we also integrate into our school-level model.

[3] In their conclusion, they state that: "A key concern in this study is whether variation in schools' conditional suspension rates arises from policy choices made by administrators, or from underlying variation in school context. While the large exogenous change in peers caused by the redrawing of school boundaries partly addresses this concern, we ultimately cannot directly connect our estimates of school "strictness" to concrete policy change.

predict that schools that punish the same behavior more harshly will have less misbehavior, and therefore may have fewer suspensions. Data on student behavior is needed to distinguish between underlying student behavior and administrators' response to that behavior. Second, the authors acknowledge their identification strategy, which conditions on student demographics and baseline test scores, captures all factors in the school associated with higher suspension rates before the policy change. This includes differences in disciplinary practice but could also include other unobservable school factors.

Given the interest in understanding the impact of administrator decisions, Hinze-Pifer and Sartain (2018) argue for a distinct line of inquiry focused specifically on school-level "suspension policy", rather than on suspensions or expulsions *per se*. Hinze-Pifer and Sartain (2018) operationalized their independent variable as the propensity to suspend, rather than the school-level suspension rate (Lacoe and Steinberg 2018, Perry and Morris 2014). They find evidence suggestive of a tradeoff between the benefits and costs of suspension, with the results varying by school racial composition. Hinze-Pifer and Sartain's (2018) decision to focus on the propensity to suspend at the school level was "grounded in the understanding that school administrators establish suspension practice by the series of choices they make in response to specific behaviors – it captures an extent to which the discipline regime is punitive, *conditional on the antecedent behavior (emphasis added)*" (p. 236).

School principals and members of their leadership team (i.e. assistant principals) are primarily responsible for making decisions about disciplinary responses to behavioral issues in a school (Bacher-Hicks, Billings and Deming 2019; Hartzell and Petrie 1992). Principals are responsible for the structural operation of the school, which includes maintaining order. In larger schools, assistant principals often take on the day to day responsibility of these operations,

particularly around student conduct and behavior order (Niewenhuizen and Brooks 2013, Oleszewski, Shoho and Barnet 2012). However, in qualitative work centered on urban middle schools, Williams et al. (2020) document how the disciplinary discretion of assistant principals are framed by the decisions made by their principals.[4] Within the law, it is often clear that the authority for decisions come from the principal, even if the assistant principal makes the day to day decisions. For example, in North Carolina, state statutes explicitly give principals the authority to suspend students for up to ten days who violate the code of student (N.C. Gen. Stat. '115C-391(c).)

Typically, the administrative team has a wide range of options when responding to violations, included speaking to the student/parents, suspending students from the classroom, expelling students from school (if approved by the school district), or even reporting students to the police (DeMatthews et al. 2017, Gottfredson and Gottfredson 2001, Skiba et al 2014). Principals have different attitudes towards this responsibility, which is reflected in their behavior. Skiba and colleagues (2014) found that students whose principal favored school exclusion rather than prevention (e.g., peer mediation or in-school suspension) were more likely to receive an out-of-school suspension or expulsion than an in-school suspension. Even when states or districts implement policies to restrict the use of suspensions, evidence shows that principals can, and do, ignore these policies (Steinberg and Lacoe 2018; Anderson 2018). Bacher-Hicks et al. (2019) found that a change in principals substantially reduced the year-to-year correlation between suspension rates at the school level, even though the school principal is only one member of the leadership team. To a large extent, "the principal is the policy".

[4] For example in one school, the principal had established a student conduct council that helped the assistant principals look for alternatives to suspension for students, an option that was not available at other middle schools in the same district. Perhaps as a result, the rate of suspensions within that middle school was particularly low (Williams et al. 2020).

In this paper, we build on this insight with a two-part strategy based on principal turnover using a statewide, longitudinal, administrative data from 2008 to 2016 on North Carolina public middle school students. One advantage relative to related research is that we observe both reported disciplinary events and disciplinary decisions. This information, together with our focus on within-school changes associated with principal turnover, allows us to distinguish administrator decisions from student behavior through the creation of a measure of removal propensity for each principal. As a result, we can study whether disciplinary policy affects school safety, and we can distinguish between the direct impact of disciplinary decisions on the students who are disciplined and the indirect effects on the larger student body (i.e. spillovers). Finally, our matching to juvenile justice complaints and adult convictions records allows analysis of potential pathways between the school discipline system and the criminal justice system.

We find that principal removal propensity reduces the number of reported disciplinary offenses that occur at the school – suggesting a "deterrence" effect – but increases the likelihood of exclusionary discipline conditional on having a disciplinary referral. For the full student population, higher principal removal propensity has null effects on short-term academic outcomes, but increases the rate of juvenile justice complaints, the overall student removal rate, and the rate of school dropout. These impacts are not the same across all students. In particular, students who had at least one minor disciplinary event saw the detrimental effects on achievement and attainment. These findings on the importance of disciplinary practices of principals complement existing research on the more general importance of principals for student academic outcomes (Branch, Hanushek, and Rivkin 2012; Coelli and Green 2012; Grissom, Kalogrides, and Loeb 2015). Because our study implicitly involves principal turnover within schools, we return to the broader literature on determinants of and implications of principal

turnover in the methods section (e.g. Grissom and Bartanen 2019; Henry and Harbatkin 2019; Miller 2013).

In addition, the data allow us to create a novel measure of principal racial bias in disciplinary decisions that we can link to student outcomes. Qualitative work suggests that teachers and administrators disproportionately target black students at multiple stages of the disciplinary process (Bell 2020, Gibson et al. 2014). We show that at the conclusion of this process, one can compare the principal's propensity to remove a black student to his or her propensity to remove a white student, for the exact same offense type and offense history. This is similar in spirit to the work of Barrett et al. (2019) who identify racial discrimination in student discipline by comparing the punishments of black and white students who fight one another, and of Kinsler (2011) who identifies across- and within-school punishment gaps between black and white students. We ultimately find that the level of racial bias a principal exhibits in assigning disciplinary consequences predicts enormous later academic losses for non-white students, and corresponding gains for white students, enrolled at that school. These results, taken together, support a direct focus on both reducing school suspensions and racial disparity in these suspensions.

Disciplinary System in North Carolina Middle Schools

North Carolina state policy provides principals with ample discretion over suspension decisions (North Carolina State Board of Education and Department of Public Instruction 2008; 2018). For minor reported incidents, school leadership independently determines whether or not to suspend a student, and if a suspension occurs, whether the suspension should occur in or out of school, and how long it will last for up to 10 days. For more serious offenses, principals can

recommend to the superintendent that a student receive a consequence that lasts longer than 10 days and up to 365 days.

Prior to determining the consequence for a given disciplinary incident, a teacher or some other actor must refer the incident to a school administrator. To appear in our data, the incident and consequence must then be reported through a statewide administrative data reporting system. State and federal statutes and State Board of Education policies mandate that certain types of offenses are reported: (1) any act resulting in in-school suspension, out-of-school suspension, expulsion, or assignment to an alternative school or alternative learning program; (2) any use of corporal punishment;[5] and (3) any act on a pre-determined list of 12 "Reportable Offenses" (North Carolina Department of Public Instruction 2019). Table A1 lists these 12 more serious reportable offenses and their associated probabilities of leading to student removal (OSS, transfer, or expulsion). In North Carolina, most disciplinary incidents are classified as "Unacceptable Behaviors" (e.g., skipping school, disrespect of faculty/staff, inappropriate language, dress code violation, and fighting) and less commonly result in school removal. Due to the varying reporting requirements and norms for addressing different offense types, our approach strives to distinguish decision-making regarding (a) the assignment of consequences from (b) the reporting of disciplinary events.

Policy Changes and Trends

North Carolina's legislature mandated in 2001 that the state collect and report annual information on student disciplinary consequences. In the 2000-01 school year, for every 100,000 students there were: 17,182 short-term suspensions, 2,642 transfers to alternative schools, 214

[5] Although legal in some districts, corporal punishment in North Carolina is rarely used, with only 2% of schools reporting corporal punishment, and only 1% of children attending schools that report any corporal punishment (Gershoff & Font, 2016).

long-term suspensions, and 12 expulsions (NC State Board of Education 2002). These counts rose annually, peaking in the 2006-2007 school year (NC State Board of Education 2011). Consistent with national trends, North Carolina suspension rates were marked by large racial and ethnic disparities which prompted civil lawsuits against school systems (Eckholm 2010). Partly in response to these lawsuits and heightened public awareness, North Carolina was among five states in 2011 to pass laws aimed at reducing out-of-school suspension or expulsion (The Council of State Governments Justice Center 2017). The change increased local administrators' discretion for handling situations of minor student misconduct (Morton, 2014). From 2011-12 to 2015-16, North Carolina reduced the proportion of students receiving suspensions from 17.8% to 14.5% (Council of State Governments Justice Center 2017). Despite this reduction, North Carolina's suspension rate is still high relative to other states and racial/ethnic disparities have persisted.

Data

We use longitudinal administrative records of public middle school students from the North Carolina Education Research Data Center (NCERDC) between 2008 and 2016, matched with later educational attainment, juvenile justice records, and adult conviction records from the North Carolina Department of Public Safety. The disciplinary records include information on each referral of a student disciplinary offense, regardless of whether that referral resulted in a suspension. This makes our dataset unique from many other studies which only observe those disciplinary incidents that result in suspension. North Carolina distinguishes "reportable offenses," which statutorily must be reported to the central administration, from "unacceptable behaviors," for which reporting requirements are less regulated. We restrict disciplinary referrals

to the 12 most common reportable offenses and the 23 most common unacceptable behaviors.[6] These 35 offense types (Table A1) represent 92.1% of all reported disciplinary incidents for 6th through 8th grade students during the study period. Although the data contain many types of disciplinary consequences, we focus on whether student reported offenses result in an instance of exclusionary discipline ("removal"), including: out-of-school suspension, expulsion, or transfer to an Alternative Learning Program or alternative school. These forms of exclusionary discipline are considered more serious and more likely to have lasting consequences for the student (Skiba et al. 2014).

Importantly, the data include information on the student's principal at the time of the disciplinary event and we can track principals as they transfer across schools or change positions within schools. We estimate within-school principal value-added scores in reading and math to provide supplemental measures of principals' ability to enhance student learning (Branch, Hanushek, and Rivkin 2012; Dhuey and Smith 2018; Grissom, Kalogrides, and Loeb 2015). Table A2 provides a correlation matrix of principal characteristics and estimated metrics.

After estimating the principals' propensity to remove (PTR) (described below), we link principal PTR measures to a variety of disciplinary and academic outcomes for all students enrolled in that principal's school. These outcomes include: an indicator of any disciplinary incident, an indicator of a serious disciplinary incident, an indicator of any removal from a disciplinary incident, a count of number of days absent, an indicator of grade retention, standardized test scores in reading and math, and high school graduation. Reading and math scores were each normalized to have a mean of zero and standard deviation of one by grade level

[6] We remove the most serious reportable offenses, which must be reported to law enforcement: death by other than natural causes, kidnapping, rape, robbery with a dangerous weapon, and taking indecent liberties with a minor. Each type rarely occurs, representing no more than 1% of all disciplinary events.

and year, and then averaged across subjects. We constructed a measure of on-time high school graduation which equals one if the student graduated high school in North Carolina public schools within six years of the spring of his or her sixth-grade year, and zero otherwise.[7]

This study builds upon prior evidence exploring the educational impacts of disciplinary practices to explore impacts on involvement in the juvenile or adult criminal justice system. Records on juvenile complaints from the fall of 2007 through the fall of 2010 come from the North Carolina Department of Public Safety (DPS).[8] We restricted juvenile offenses to only those occurring during the typical academic school year. We developed a web scraper to retrieve online records on criminal convictions for the entire time period from the DPS publicly available offender search feature of their website (https://www.ncdps.gov/dps-services/crime-data/offender-search). During our study period, North Carolina courts considered all individuals as adults at the age of 16, and thus we observe two cohorts within our sample who become eligible for adult criminal conviction. We construct an indicator for whether or not the student received a criminal conviction in North Carolina by age 20.[9] Juvenile complaints and adult conviction records were both matched to student records using an iterative algorithm based on first name, middle name, last name, date of birth, and county of birth. Of all student-year observations, 2% were matched to juvenile justice complaints; of all students who entered sixth grade in time to be observed through age 20, [10] 3% were matched to the adult offender database.

[7] Possible counterfactuals to on-time high school graduation include: delayed high school graduation or GED, graduation from North Carolina private schools, graduation outside the state, or dropout. We cannot differentiate well among these alternative outcomes.
[8] Our data access agreement does not permit using data from after the 2010-11 school year.
[9] By restricting the age range in which individuals can appear in criminal justice records and restricting to only the state of North Carolina, we underestimate the true conviction rates of middle school students in North Carolina public schools. However, we have no reason to believe that this underestimation would be systematically related to PTR. The same general argument holds for our analysis of high school graduation rates.
[10] Our nine-year panel data allows us to examine on-time high school graduation for five cohorts and criminal conviction by age 20 for three cohorts of entering middle school students.

We also develop an expansive set of control variables, including measures at the student level such as race/ethnicity, gender, limited English proficiency, economic disadvantage (<185% of the federal poverty limit), and incidence of reported disciplinary offenses. Control variables also include time-varying school level measures including student enrollment, number of full-time equivalent teachers, percent of students by race/ethnicity, and percent of students by level of economic disadvantage.[11] Alternate specifications control for principal characteristics including race/ethnicity, gender, years of experience, and math and reading value-added scores (Table A3).

Methods

Step 1. Estimating Principal Propensity to Remove (PTR)

This section describes how we measured a given middle school principal's propensity to remove (PTR). We used 9 years of disciplinary referral data (2007-2008 through 2015-2016 school years) and observed 2.38 million disciplinary events decided by 1,753 principals. For each disciplinary event, we constructed a binary variable that indicated whether or not the disciplinary referral resulted in removing the student from school, coded as 1 (i.e., out-of-school suspension, expulsion, or transfer to an alternative program) or whether the outcome did not remove the child from school, coded as 0 (i.e., in school suspension, detention or other sanction).

For step one, our unit of analysis was a disciplinary event that was reported through the school system to the state Department of Public Instruction. These reported disciplinary events led to an administrator's decision regarding punishment type. Our goal was to isolate the component of that disciplinary consequence that was due to principal decision-making – and not

[11] We use percent of students in the school with free lunch as a proxy for percent high economic disadvantage, and the percent of students with reduced price lunch as a proxy for percent moderate economic disadvantage. These measures come directly from the Common Core of Data (CCD).

due to the nature of the student offense, to the student's prior record, or to fixed school policies or school environment. We also implemented leave-year-out estimation so as to remove the simultaneity problem that would exist if principals appear more or less harsh in response to a certain cohort of students in a certain year. In step two of our analysis, we used PTR matched to students to assess how principal harshness affects student-level outcomes and to identify "optimal" principal behavior.

We have found no direct analog to this model in the education literature.[12] The best analogy outside of the education literature comes from the recent economics and crime literature on judge effects for studies of the impact of incarceration on employment or crime. In the first step of the model, researchers identify the judge harshness for criminal sanctions or sentences (usually prison vs. probation) given to convicted defendants sentenced by the same judges. In the second step, researchers take this measure of judge harshness and look for its impact on a different outcome, like employment (Harding et al. 2018, Bhuller et al. 2018, Mueller-Smith 2015).[13] These models have shed light on the potentially deleterious consequences of policies that allowed judge's considerable discretion in sentencing decisions. More optimistically, these models allow for the identification of judge behavior that appears to be "optimal". The main difference between the judge selection models and our work is that judges are often randomly

[12] There have been attempts to use value-added models to study the impact of principals on outcomes like test scores (Grissom, Kalogrides and Loeb 2015) and to map value-added estimates from a particular staff member (teacher effects on test scores) to student outcomes (earnings, college attendance, teen birth) (e.g. Chetty, Friedman, and Rockoff 2014). Bacher-Hicks et al. (2019) estimates a principal value-added model for suspensions in the Charlotte-Mecklenburg School District. Our approach differs because it focuses on disciplinary events as the unit of analysis, rather than the student.

[13] These models are usually instrumental variable models identified on the fact that judges are assigned randomly to cases. However, random assignments are typically made at an initial step of the process (called arraignment), and not all cases ultimately lead to conviction. As a result, random assignment to judges at arraignment does not necessarily lead to balance across judges at the conviction step. In cases where researchers have only convicted samples, researchers need to identify their model conditional on observables (Harding et al. 2018).

assigned to cases – in our work, we need to control for the fact that principals are not randomly assigned to schools.

Our sample for estimating principal PTR encompassed 35 high-frequency types of reported disciplinary referrals for all 6th to 8th graders. These disciplinary offense types ranged in frequency from disruptive behavior (n=631,231 events) to robbery with no weapon (n=64). The median category had 21,728 events. Because we acknowledge that principals may have responded more harshly to some offense types than to others, and that each principal faced a different disciplinary environment in terms of offense types and frequencies, we estimated 35 different principal propensities to remove, one for each offense type recorded. Within these offense types, our estimation objective was simply: How much more or less likely was the specified type of disciplinary offense to result in a removal under a given principal, when compared to the average principal acting under similar circumstances?

Specifically, PTR was estimated first as a vector of principal-year-specific fixed effects ($\widehat{\mu_{1p}}$) for each distinct disciplinary incident type (*k=1, . . . , 35)* in the sample of referrals from disciplinary event-level estimating equation of the general form:

$$r_{ijkpt} = \beta_0 + H_{it}\beta_1 + \theta_g + \Phi_s + \delta_t + \mu_{pkT} + \epsilon_{ijkpt} \text{ for offense } k, t \neq T \quad (1)$$

The dependent variable (r_{ijkpt}) was an indicator of school removal for the disciplinary event *j* of type *k* (for student *i*) supervised by principal *p* in year *t*. This model contains vectors of grade level fixed effects (θ_g), school fixed effects (Φ_s), and year fixed effects (δ_t). The only other control was a vector describing the student's prior history of disciplinary events (H_{it}) at time of event *t,* including the number of reported offenses that school year, and the number of reported offenses in the year prior. The principal-year fixed effect μ_{pkT} represented a measure of likelihood of removal under the principal for that particular offense type. We used leave-year-out

estimation, which implied that for year T, the vector of principal propensity to remove estimates (μ_{pkT}) was derived from all years the principal appeared in the dataset *except for* year T.[14]

From model 1, we captured the raw estimates of principal PTR for each offense type *k* and year T ($\hat{\mu}_{pkT}$). If we observed a principal for four years, we would have, for each of those four years, up to 35 unique estimates of the removal likelihood for each disciplinary type *k* that was observed at that school. In an approach similar to that used by Branch et al. (2012), we formed empirical Bayes estimates of these principal-offense PTR scores so as to place greater weight on scores that were estimated with better precision. Each estimate equaled a weighted average of the estimated removal likelihood for principal *p* and the average removal likelihood across all principals in the population: $\hat{\mu}_{pkT}(EB) = \left(1 - \lambda_{pkT}\right)\overline{\mu_{pk}} + \lambda_{pkT}\hat{\mu}_{pkT}$. In this way, the more precise the estimate of the principal removal likelihood for type *k*, the more weight placed on the estimated individual principal estimate, rather than the average principal estimate across the full sample. We provide more details on PTR estimation and adjustment in Appendix B.

This process resulted in a set of estimated and shrunk removal likelihood scores for each principal, in each year (due to leave-year-out), for each of 35 disciplinary types. From these, we created a single measure to reflect an individual principal's removal likelihood for a randomly drawn disciplinary event. Specifically, we calculated a weighted average across offense types, weighing each offense type by the *sample* proportion of reported disciplinary events in that offense type. The *sample* proportion of reported disciplinary events is not equivalent to the

[14] This leave-year-out method assumes a constant principal effect over time. An alternative to this approach would be to allow principals to exhibit different disciplinary practices in each year, and to therefore estimate a separate principal PTR for each year. When we do so, we see no significant trends in principal PTR between their first year at a school and their 6+ year at the school. Further, this measure is much less precise because it is estimated over a smaller number of disciplinary referrals that occur in a single year in the school. Therefore, we prefer to use the leave-year-out constant principal effect approach.

proportion of events faced by the specific principal *p* for each offense type. As a result, every principal faced the same weight for each offense type, regardless of the relative frequency or rarity of that offense type in their particular school. This weighting scheme created a final removal likelihood that created a unified measure of removal likelihood for each principal *p* under the assumption that each principal faced the same reported offense mix, thereby not punishing principals who faced more serious reported offenses on average.

$$\hat{\mu}_{pT} = \sum_{k=1}^{35} \hat{\mu}_{pkT}\, w_k \text{ where } w_k = \frac{n_k}{n} \quad (2)$$

with n_k = number of disciplinary events in the sample for offense type *k* in years $t \neq T$

and n = total number of disciplinary events in the sample in years $t \neq T$

After weighting by offense type, we have a single PTR score for each principal-year. This PTR score represents the likelihood that a principal will assign removal (OSS, transfer, or expulsion) for a randomly drawn disciplinary incident from the whole sample, conditional on that student's prior offense history and all time-invariant characteristics of the school. The extremes represent a very severe principal who suspends or expels every student for every offense (PTR = 1) versus a very lenient principal who never suspends or expels any student for any offense (PTR = 0). We have conducted several alternative PTR estimation strategies, such as using random effects, or using across-school variation instead of within-school variation, or including in-school suspension as a form of removal. Appendix B describes these alternative strategies, Appendix Table A4 provides a correlation matrix of all PTR measures, and Appendix Table A5 shows the sensitivity of our results to alternative PTR measures.[15]

[15] Results are generally quite consistent across the alternative PTR measures. For the across-school and random effect PTR estimates, there are actually significant associations between principal PTR and adult criminal conviction, suggesting that increased principal discipline severity might increase eventual likelihood of criminal conviction. Looking at the effects of PTR when it includes in-school suspension, the effects on student outcomes are similar but slightly more modest in magnitude.

Both the principal's average response to a disciplinary incident, and differences in responses across different types of students, may matter for student outcomes. It could be the case that a principal has separate propensities to remove for different types of students, even conditional on the type of offense and offense history. This trait, similar to the "punishment gaps conditional on student referral" observed by Kinsler (2011), and punishment gap for students involved in altercations observed by Barrett et al. (2019), can be uniquely identified in our data. We focused on one particular comparison – between black and white students – both for simplicity and due to the policy relevance of this comparison.[16]

To estimate racial differences in principal PTR, we first restricted the sample to schools in which at least 10% of reported offenses are committed by white students and at least 10% of reported offenses are committed by black students. We then replicated the method described above and interact principal fixed effects with student race, such that each principal has two PTR estimates per offense type: one for white students and one for black students. Finally, we calculate a new measure of principal removal bias as the difference between the two: $\hat{\mu}_{pT,black} - \hat{\mu}_{pT,white}$. A positive value would indicate that the principal removes black students for the same offense at a higher rate than white students, and a negative value would indicate the opposite. Because teachers and school staff may use discretion in the original reporting of the offense, prior to the assignment of a disciplinary consequence, this measure of black-white difference only reflects bias in the punishment phase, conditional on the process that generates the reporting.

[16] Our samples of Hispanic and other race/ethnicity students are not large enough and not distributed evenly enough across schools to create multiple metrics of principal removal bias.

Step 2. Estimating the Effect of Principal PTR on Student Outcomes

Next, we study the effects principal PTR on student outcomes. This study focused on nine short-term outcomes, generically referred to as Y. Three measures focused on student discipline, including an indicator variable for (1) whether or not the student committed any minor disciplinary offense that year; (2) whether or not the student committed any serious disciplinary offense that year; and (3) whether or not the student received out-of-school suspension, expulsion, or transfer that year (1=yes, 0=no). Three measures represented academic outcomes: (4) an average of end-of-year reading and math test scores (standardized z-scores); (5) number of days absent; and (6) an indicator of grade retention. And the final three indicators reflected transfers to the juvenile justice system: (7) any juvenile justice complaint during the school year; (8) any misdemeanor juvenile justice complaint; and (9) any felony juvenile justice complaint.

Our preferred model, specified in equation 4, controlled for time-fixed school factors (through school fixed effect α_s), time-varying school factors X_{1st}, student factors like sex, race, and socioeconomic status (X_{2i}), and grade (γ_g) and year fixed effects (τ_t).

$$Y_{it} = \beta_0 + \beta_1 \hat{\mu}_{pt} + \beta_2 X_{1st} + \beta_3 X_{2i} + \alpha_s + \gamma_g + \tau_t + \varepsilon_{it} \quad (4)$$

The effect of principal PTR on student outcomes (β_1) is thereby identified using variation in PTR from within-school principal turnover. Between 2008 and 2016, less than 10 percent of schools in our sample had no principal turnover, approximately 30 percent had one principal transition, and the remaining 60 percent had two or more principal transitions (Table A2, a). Although principal switches within schools are therefore quite common, principal switches across middle schools occur less frequently (Table A2, b). Because we are particularly interested in knowing if principals' choices only affect students involved in the disciplinary system or if there are

spillover effects to all students in the school, we estimate this equation across samples of students with differing involvement in the disciplinary system.

The question of how often and under what circumstances principals should use exclusionary discipline becomes even more pressing if these principal decisions affect long-term student trajectories. To determine the relationship between principal PTR and long-term student outcomes, we slightly modified the model in equation 4. We collapsed the data from disciplinary offense-level to the student level and then tracked whether those students graduated from high school on time (seven years after the start of sixth grade) and whether they were convicted of a criminal offense by age 20. The estimation equation for these two binary outcomes was:

$$Y_{ic} = \omega_0 + \omega_1 \overline{\mu_i} + \omega_2 \overline{X}_{1c} + \omega_3 X_{2i} + \alpha_s + \gamma_c + \varepsilon_{ic} \quad (5)$$

Here, the principal PTR measure was replaced with a cumulative average principal PTR over middle school years $\overline{\mu_i}$. To account for multiple cohorts of students, we included school fixed effects and cohort fixed effects, defined by the year in which that student entered sixth grade. In this way, long-term impacts were identified through across-cohort differences in exposure to more or less severe principals within the same school.

Results

Descriptive Analysis of Principal PTR

We begin by describing the principal PTR measure, including how it is correlated with other principal characteristics and indicators of principal effectiveness. The average principal in our sample removes students for 32% of reported disciplinary events they would confront in a representative sample of referral types. This statistic varies meaningfully across principals, even once the less precise estimates have been shrunk towards the mean (Figure A1). Principals at the

10th percentile of PTR would remove 21% of students for the average offense, whereas principals at the 90th percentile would remove 41%. We also calculated principal black-white removal "bias" as the principal's average likelihood of removing a black student minus their likelihood of removing a white student, for the same reported offense and offense history. On average principals are 1 percentage point more likely to remove black students than white students for the same offense type and history based on the within-school PTR measure, and 6 percentage points more likely based on the across-school PTR measure.[17] Noticeable variation exists in the level and direction of racial bias across principals (Figure A2).

The education literature has examined the importance of other principal characteristics, including value-added in student test scores (e.g., Branch, Hanushek, and Rivkin 2012; Grissom, Kalogrides, and Loeb 2015). Overall, our measure of principal PTR appears weakly correlated with principal value-added measures for reading (-0.122) and math (-0.026) (Table A2). Principals with higher PTR tend to be slightly less experienced, with a correlation of -0.068 between PTR and years of experience. These weak correlations support the notion that (a) principal disciplinary practices are a unique dimension of behavior that deserves attention and (b) principals with higher PTR are not systematically "better" or "worse" principals in other dimensions.

Effects of Principal PTR on Academic Outcomes

Table 2 documents the impacts of principal PTR on students' disciplinary and academic behaviors. Each coefficient on principal PTR in columns 1-6 represents the effect of substituting a principal who never removes students (PTR = 0) with a principal who always removes students

[17] For each principal, we perform a test to determine whether their removal propensity for white students is significantly different from their removal propensity for black students, for the same offense type and history. We find that 13.9% of principals have significantly higher PTR for black students across all offense types, and 7.5% of principals have significantly higher PTR for white students across all offense types.

(PTR = 1) for the average offense in our sample. Although there are outliers in our sample who have scores near these extremes (Figure A1), these coefficients should nonetheless be interpreted as the result of an extreme change in principal PTR. We therefore also provide rescaled versions of effect sizes based on the hypothetical experiment of replacing a principal at the 10th percentile (21% removal rate) with a principal at the 90th percentile (41% removal rate) of PTR (Table A7) . This reflects a comparable though slightly larger change than one standard deviation in unadjusted PTR (16 p.p.) and a slightly larger change than the average PTR change associated with within-school principal turnover (14 p.p.).

We observe that a 100-percentage point increase in principal PTR reduces students' likelihood of having any minor disciplinary referral that year by 10.4 percentage points (column 1). More realistically, switching from a principal at the 10th percentile to the 90th percentile of PTR would decrease the probability of having a minor offense by 2.1 percentage points. There are two potential interpretations of this result. The first is that stricter principals have a "deterrent" effect on student misbehavior. The second is that stricter principals lead teachers to report fewer students for the same level of offense, perhaps in an attempt to protect students from harsher consequences. These reductions in minor reported disciplinary infractions do not translate into similar reductions in serious reported disciplinary incidents, as can be seen by the statistically insignificant coefficient on PTR in column 2. As expected, when controlling for changes to reported student offending behavior, higher principal PTR also translates into higher conditional likelihood of removal (column 3).[18]

[18] An alternative approach for this particular regression would be to not control for reported student offenses, like with the other regressions in Table 2. When we exclude the reported student offenses control, the effect on removal is still positive and significant, but smaller in magnitude (coef = 0.033, $p<0.01$). This is because the higher principal PTR is simultaneously reducing the reported student offense rate while also removing students at a higher rate for those offenses that do occur.

Although separately considering each of the 35 offense categories reduces power and creates multiple hypothesis testing issues, we do so to explore patterns of reported student behavior changes. The coefficient on principal PTR was negative for nearly all 35 reported disciplinary behaviors, although these negative coefficients were frequently small and not statistically significant. Figure 1 shows the six significant effects, reflecting reductions in reports of disruptive behavior, bus misbehavior, disrespect of faculty/staff, being late for class, cutting class, and property damage. We see relatively uniform drops in minor disciplinary reports under principals with higher PTR. To the best of our knowledge, this is the first paper to document that harsher discipline reduces minor reported disciplinary incidents.

These benefits come with a cost. Controlling for the student's disciplinary offense record, the 100-percentage point rise in principal PTR increases the likelihood of student removal by 7.3 percentage points (column 3), or 1.4 percentage points when moving between our hypothetical 10^{th} and 90^{th} percentile principal. A more severe principal has to deal with fewer reported minor student offenses but is more likely to assign out-of-school suspension or expulsion or transfer for the disciplinary offenses he or she is referred, all else held constant.

For three academic outcomes – absences, test scores, and grade retention – principal PTR has no discernable effect on the aggregate student population (Table 2, columns 4-6). The point estimates suggest potentially small increases in absences, decreases in test scores, and increases in grade retention, but none of these estimates are statistically significant. For one outcome – high school graduation – principal PTR has a marginally significant negative effect ($p<0.1$). Replacing a 10^{th} percentile PTR with a 90^{th} percentile PTR principal decreases student likelihood

of on-time graduation by 2 percentage points.[19] Given that students under harsher principals are more likely to be removed from school, but less likely to commit offenses that would lead them to be considered for removal, these modest effects on academic outcomes are perhaps unsurprising. We anticipate, however, that principal PTR could affect students involved in the disciplinary system quite differently than those not involved.

To investigate this question, Table 3 presents the original model estimated for three samples: (1) students with no reported disciplinary offenses during that school year; (2) students with at least one minor disciplinary offense but no serious offense reported during that school year; and (3) students with at least one serious offense reported during that school year.[20] This model is particularly useful for considering whether principal PTR has spillover effects on students who are not suspended. Recall that positive spillover is one of the arguments against restricting the use of out-of-school suspensions. Principal PTR has no significant effect on absences, test scores, or grade retention for students without any reports of misbehavior during the school year (row 1, columns 2-4), and a marginally significant negative effect on high school graduation. Thus, there are no positive spillover effects on academic outcomes, and suggestive evidence of negative spillovers.

For students with minor disciplinary records, however, the estimated coefficients on principal PTR indicate noticeably negative outcomes – the main concern driving advocates for reform. A student who commits a minor disciplinary infraction under a principal with the highest possible PTR will have 2.5 more absences than a student committing a disciplinary infraction under a principal with the lowest PTR, even within the same school environment. The student under the

[19] For graduation regressions, we use individual-level data and replace principal PTR in year *t* with average principal PTR faced by the student in 6^{th}-8^{th} grade. We also replace year and grade fixed effects with cohort fixed effects.
[20] Serious offenses are defined as "reportable offenses" from Table A1, and minor offenses are all other types.

harsher principal will also have one-tenth of a standard deviation lower test score achievement, a 2.3-percentage point greater likelihood of grade retention, and a 13.6-percentage point lower likelihood of high school graduation. Switching from a principal at the 10^{th} percentile to the 90^{th} percentile of PTR would thereby increase absences by 0.5 days, reduce test scores by 0.02 standard deviations, increase grade retention rates by 0.4 percentage points, and decrease graduation rates by 2.7 percentage points.

Principal PTR does not appear to have the same influence on academic outcomes for students committing serious disciplinary offenses (row 3). This may be because our measure of principal discretionary behavior is more likely to affect students "on the margin," than to affect students who have committed a serious offense, who already face a near certain out-of-school suspension under most principals. It also partly reflects the significant loss in sample size, captured by the higher standard errors.

Besides demonstrating heterogeneous effects across students, these findings support our argument that principal PTR directly measures discretionary disciplinary severity, and not some other unobservable principal quality or school-level change which would have affected students outside the disciplinary system. The magnitudes of the estimated impacts on students who commit minor disciplinary infractions are substantive (Table A7), particularly given the long-term implications for educational attainment. For example, substituting a harsher principal at the 90^{th} percentile of PTR with a more lenient principal at the 10^{th} percentile would improve graduation outcomes at a magnitude equivalent to over one-fourth of the national White-Black graduation rate gap, or at a magnitude equivalent to over one-third of the national White-Hispanic graduation rate gap (National Center for Education Statistics 2020). Such a policy would impressively have similar size effects on high school graduation likelihood as, for

example, the Head Start preschool program (Ludwig and Miller 2007), or as a targeted middle or high school dropout prevention program (Agodini and Dynarski 2004).

Effects of Principal PTR on Criminal Justice Outcomes

The next set of analyses tests the impacts of middle school principal PTR on juvenile justice complaints and criminal conviction in young adulthood (Table 4). We find in the first row, first column that students under a principal with the highest possible PTR will be charged with a juvenile crime at a rate 1.2 percentage points higher than students under a principal with the lowest PTR. Substituting a 10^{th} percentile PTR with a 90^{th} percentile PTR principal would therefore increase the rate of juvenile justice complaints by 0.3 percentage points, from a mean rate of 2 percent. In our data, 52 percent of all middle school juvenile complaints come directly from school-based offenses. The next two columns in Table 4 show the effect of PTR on juvenile complaints to be driven entirely by increases in school-related offenses, rather than by increases in juvenile crimes outside of schools. Principals with higher propensity to suspend students appear to therefore also refer a greater proportion of students to juvenile justice for incidents occurring in school, suggesting this act of referral as an integral mechanism through which school disciplinary processes seep into other justice system involvement. Columns four and five indicate that these increased juvenile justice referrals are for misdemeanor crimes, not felonies.

We next partition the sample by student disciplinary record to further explore mechanisms underlying increases in juvenile justice complaints. There are no discernable effects of principal PTR on student juvenile justice complaints either for students without a disciplinary record, or for students with minor disciplinary records. There are, however, large effects of principal PTR on juvenile justice complaints for students reported as committing serious disciplinary offenses. For this group of students, a 100-percentage point increase in PTR increases juvenile justice

complaints by 35.6 percentage points. That is, substituting a 10th percentile with a 90th percentile PTR principal would increase juvenile justice complaints by 7.1 percentage points. This finding suggests that, when given the option between two potential student punishments, high PTR principals universally tend to choose the harsher punishment, whether that means assigning an out-of-school suspension for a more minor offense (Table 3), or referring a student to the juvenile system for a more serious offense (Table 4).

We find no significant impact on adult criminal conviction (column 4). In the partitioned samples, we do find that principal PTR slightly reduces the likelihood of criminal conviction for students with no disciplinary record during middle school, a result that clashes somewhat with our earlier finding that principal PTR marginally decreases the graduation rate for this group. However, because this outcome requires a much longer follow-up period to trace any convictions by the age of 20, these results are based on only 2 cohorts of students – and may change with the addition of more cohorts, which would increase the amount of variation available for identification. Overall, the juvenile justice results support the school-to-prison pipeline metaphor, that punishments doled out within school could have farther-reaching consequences within the criminal justice system (Bacher-Hicks, Billing and Deming 2019; Skiba, Arrendodo, and Williams 2014), although it is unclear how long these consequences persist.

Effects of Principal Racial Removal Bias on Student Outcomes

We sought to determine if differential treatment in the disciplinary process by student race influences student outcomes. To do so, we estimated similar models as presented in Table 2 and interacted principal black-white removal bias with student race to examine specific effects on white, black, Hispanic, and other race students. Table 5 column 1 shows that increased principal removal bias leads to decreased absenteeism for white students, and substantively

increased absenteeism for black and Hispanic students. Specifically, a change from a principal at the 10th percentile level of removal bias (-0.033) to a principal at the 90th percentile level of removal bias (0.061) would decrease white absences by 0.25 days, increase black absences by 0.41 days, and increase Hispanic absences by 0.34 days (Figure 2a). The same replacement to a principal more biased against black students would increase white test scores by 0.025 standard deviations, and decrease black and Hispanic student test scores by 0.07 and 0.08 standard deviations (Table 5 column 2, Figure 2b). It would also have no impact on grade retention likelihood for white students, and increase grade retention likelihood for black and Hispanic students (Table 5 column 3, Figure 2c).

Although the causal logic for how principal black-white disciplinary bias affects black and white students is straightforward, it is less so for why principal black-white disciplinary bias affects Hispanic students. We see two possible ways to interpret these results. The first is that a principal's black-white removal bias (observed) is positively correlated with that same principal's Hispanic-white removal bias (not observed). This would generate the results we find, since effects of black-white removal bias are generally larger for black students than for Hispanic students, but in the same direction. The second interpretation is that principal black-white removal bias has spillovers onto students of other races/ethnicities. For example, a principal who makes biased disciplinary decisions against black students may create a negative school climate for other minority students, such as Hispanic students.

Finally, the degree of bias observed in the middle school principal even holds long-term inequality-widening ramifications as it causes increases in the likelihood of high school graduation for white students ($p<.1$) and decreases for black and Hispanic students. In particular, replacing a principal at the 10th percentile of racial removal bias with a principal at the 90th

percentile increases graduation of white students by 2.0 percentage points, but decreases that of black students by 2.6 percentage points and for Hispanic students by 1.1 percentage points (Table 5 column 4, Figure 2d). These findings as a whole reveal that principal discretion in assigning punishment can have serious consequences for students, particularly if principals use that discretion to treat students differently based on the student's race or ethnicity.

Robustness Tests

A key difference between our method and the judge effects method is that disciplinary cases are not randomly assigned to principals in the way that cases are sometimes randomly assigned to judges within a jurisdiction. All disciplinary cases in the school are handled by the same school principal and her team, and that principal was not assigned to the school randomly. Although school fixed effects account for any stable characteristics of the school that may be associated with principal PTR, transitions within schools from one type of principal to another could still occur non-randomly (Miller 2013, Grissom and Bartanen 2019, Henry and Harbatkin 2019). For example, if the hiring of a more severe principal is caused by a surge in student misconduct, we could incorrectly attribute changes in student outcomes to the principal instead of to unobservable changes within the school environment that led to principal turnover.

To understand more about the general sorting process of principals to schools based on their disciplinary behavior, we first conduct an across-school analysis. Specifically, we regress principal PTR and principal black-white PTR bias measures on a large set of school and student characteristics (Table A8, columns 1 and 3). Several significant predictors emerge. For example, schools with more black students have higher average PTR, and schools with more economically disadvantaged students have lower average principal racial bias. However, once we run the same regressions with school fixed effects included, these significant associations disappear. There is

not a single student or school characteristic significantly associated with principal PTR, and there is only one of thirteen characteristics significantly associated with principal bias (coef = 0.0004). In short, it appears that although principal sorting within schools over time based on disciplinary practices is plausible in theory, in practice this form of selection is unlikely to bias our results meaningfully.

To further examine school dynamics over time before and after principal transitions, we conducted several event study tests with proxy indicators of school environment and student composition: removal rates, minor disciplinary offense rates, serious disciplinary offense rates, and student test scores. We first examine trends before and after any principal turnover event. Figures A3a-d shows no apparent trends in either reported disciplinary incidents or student achievement correlated with timing of principal turnover.[21] Figures A4a-d focus on estimated trends around specific transitions from more lenient principals to more severe principals, and Figures A5a-d focus on estimated trends around transitions from more severe to more lenient principals. We define lenient-to-severe transitions as those that result in PTR increases of at least 1 standard deviation, and severe-to-lenient transitions as those that result in PTR decreases of at least 1 standard deviation. For these more specific types of transitions, we see some minor fluctuations in reported student misbehavior, but nothing statistically significant and nothing that would support a claim that school-specific trends are driving both changes in principal PTR and changes in student outcomes.

[21] This may seem contradictory to recent evidence from Bartanen, Grissom, and Rogers (2019) and Henry and Harbatkin (2019), which both find that school achievement drops in the year directly following principal turnover. This discrepancy could arise due to modeling differences: for example, Bartanen et al. compare schools with principal turnover to a matched sample of no-turnover schools, whereas we observe within-school trends over time. We also use student-level instead of school-level data and include a series of student-level control measures which could partially soak up any academic effects of turnover. Finally, the difference could be attributed to context differences between states, or, for the Henry and Harbatkin results in North Carolina, the fact that middle schools make up only 22% of their sample.

Although these examinations of the factors contributing to principal turnover are reassuring, the introduction of a principal with a higher PTR could still introduce a host of other changes for a school that could influence student outcomes.[22] For this reason, we perform an alternative estimation of the main model, controlling for a series of seven other principal characteristics: years of experience, value-added in reading, value-added in math, gender, and race/ethnicity. All but one of the statistically significant results presented in Table 2 and Table 4 are replicated with nearly identical effect sizes (Table A3). The one exception is the long-term effect of principal PTR on high school graduation, which decreases from 10.6 to 7.9 percentage points and becomes only marginally significant (Table A3, column 8). It could be that some effects of principal disciplinary decision-making operate through changes in student academic success, in which case it is to be expected that the results are slightly attenuated with inclusion of principal value-added measures. On the other hand, principal PTR leads to marginally significant increases in grade retention for the full population of students, once controlling for other principal characteristics (column 6).

Another potential concern relates to teacher and staff reporting behavior. In order for a student behavioral offense to appear in our data, a teacher or other actor must choose to report the student incident, and an administrator must choose to enter the incident into the administrative data system. Although the State Board of Education dictates certain standard reporting practices (North Carolina Department of Public Instruction, 2019), there is still considerable discretion in which student offenses get reported prior to assignment of a disciplinary consequence (Lindsay and Hart 2017). Our empirical approach already corrects for

[22] For example, Burkhauser (2017) uses North Carolina data to show that principals play a large role in determining teacher perceptions of school working conditions: teacher time use, teacher empowerment / school leadership, professional development, and physical environment.

this issue partially, by controlling for time-invariant school reporting practices using the school fixed effect, and by reweighting the PTR measure to equal the response to the average reported offense at *any* school, rather than the average reported offense at that particular school. Nonetheless, it remains possible that principal changes could lead to changes in within-offense-type reporting practices at the school. To test the sensitivity of our findings to such changes, we calculate a new measure, the "subjective reporting ratio," equal to the proportion of total disciplinary referrals in each school-year expected to have higher variance in reporting norms. Although no offense types are purely subjective, or purely objective, some will involve more discretionary reporting than others. We define this list to include: disruptive behavior, insubordination, inappropriate language/disrespect, aggressive behavior, disrespect of faculty, and disorderly conduct, which together represent the majority of incidents in the referral data. Adding this subjective reporting ratio as a control variable in our main model results in the nearly identical effects of PTR on all short- and long-term student outcomes (Table A9).[23] We interpret this as further support that PTR reflects decision-making about student consequences on the margin, not differences in reporting behavior.

Finally, we explored the extent to which assistant principals matter for driving disciplinary decisions. The literature on assistant principals indicates that assistant principals are often responsible for the day to day implementation of disciplinary policy, ostensibly under the auspices of the principal (Williams et al. 2020). Therefore, it is at least possible that the discretion of assistant principals is the key factor driving disciplinary decisions, rather than the discretion of the principal. To test for this possibility, we regressed student removal first on only

[23] Table A8 also provides two other alternative models: (i) controlling for the lagged dependent variable; and (ii) controlling for PTR black-white bias in the main PTR models. Results are generally consistent, although we lose sample size in both cases.

principal fixed effects, and then on principal and assistant principal fixed effects, in the disciplinary referral data. The proportion of additional variance in removal explained by assistant principals was only 0.018 (total R-squared increased from 0.346 to 0.364), which supports our belief that principals are primarily driving the disciplinary environment (see also Bacher-Hicks, Billings and Deming 2019).

Discussion

In the present public debate, the get-tough approach to school discipline is presented as a quandary: a principal must choose between creating a better, safer learning environment for all students at the cost of potential harm to the students who are suspended. Our results suggest this quandary is false, and potentially harmful. There are very few observable benefits from getting tough and removing students for minor infractions.

We found that schools with tougher principals faced fewer reported minor disciplinary offenses, but found no difference in reported serious crimes or violent acts. The reduction in minor disciplinary infractions could reflect either changes in student behavior as they react to a harsher school disciplinary regime, or changes in teacher behavior as they selectively choose which student offenses to report. We also found no short- or long-term academic benefits for the larger student body. Moreover, the costs of the get-tough approach were substantial. We found that students in schools with tougher principals were more likely to face a juvenile justice complaint and less likely to graduate from high school. These results support the results found by Bacher-Hicks et al. (2019) in Charlotte-Mecklenburg – one of the NC school districts in our sample - around the 2003 redistricting. They found that moving to a middle school that had more suspensions prior to the redistricting led to higher dropout rates, more involvement in the adult criminal justice system, and less enrollment in a four-year college for all students. This absence

of clear positive spillover effects from increased use of student removal also corroborates patterns described in other research (Hinze-Pifer and Sartain 2018; Lacoe and Steinberg 2018; Perry and Morris 2014).

Our paper also presents new evidence that those who experience the exclusion are negatively affected. The subset of students who commit any minor disciplinary infraction experienced increased absences, decreased test score performance, more grade retention, and higher chances of school dropout. Students who commit minor offenses in middle school represent a vulnerable population, at a transitional moment in their education. Middle schools have the highest rates of student disciplinary problems of any school level (National Center for Education Statistics 2019), making the nature of administrative response to those incidents particularly impactful. For many students, middle school years are also a critical time for determining future trajectories towards either school engagement or school dropout (Balfanz, Herzog, and Mac Iver 2007). Receiving an out-of-school suspension, expulsion, or transfer for a minor offense could foment a negative turning point in that student's trajectory.

The evidence of bias against minority students in our data heightens these more general concerns regarding harsh disciplinary practices. Principals with high PTR were not necessarily more biased, but the average principal did remove black students at a rate higher than the rate for white students, conditional upon offense type, offense history and school grade fixed effects. This estimated removal bias likely represents a lower bound on racial disproportionality, as teachers are more likely to report incidents involving black students to school leadership in the first place (Bell 2020, Gibson et al. 2014, Lindsay and Hart 2017, Okonofua and Eberhardt 2015). Kinsler (2011) similarly found that black students received more serious punishments than white students in a cross-sectional sample, and concluded that across-school variation

explained the majority of this discipline gap. Our study using a panel version of the same data as Kinsler (2011) illustrates how the across-school variation in disciplinary gaps reflects in large part differences in the decision-making of specific school leaders. Our findings demonstrate that within-school increases in principal removal bias lead to worse absenteeism, test score achievement, grade retention, and even high school graduation outcomes for black and Hispanic students, while simultaneously promoting the educational outcomes of white students.

Some have criticized the Obama administration's "Rethink School Discipline" framework for its focus on disparities in outcomes (e.g. suspension rates) rather than on disparities in school actions *conditional on* student behavior. Our results both confirm that the latter disparity does indeed exist, and documents that this disparity matters for students. Policies aimed at equalizing disciplinary treatment across students and/or enhancing transparency in disciplinary decisions could therefore go a long way towards closing not only racial gaps in exclusionary discipline, but also gaps in achievement, educational attainment, and potentially incarceration.

Broadly, our results support efforts to find alternative strategies to reduce minor offenses. Indeed, researchers and practitioners are hard at work implementing, studying, and scaling-up alternative approaches to discipline that emphasize prevention through positive behavioral support and social and emotional learning (e.g. Osher, Bear, Sprague, and Doyle 2010). A recent meta-analysis shows that non-punitive school interventions have the potential to effectively eliminate the need for exclusionary discipline (Valdebenito et al. 2019). More generally, given the high individual and societal costs associated with failure to graduate high school, policy makers should work to ensure that the immense power of the principal to remove students be applied in a fair and evidence-based manner.

In our introduction, we note that our focus on suspension policies by principals is motivated in large part by the large racial disparities in suspension rates, a concern that mirrors concerns about racial disparities in the justice system writ large. Although it is always risky to make inference to other contexts, our findings from the school-setting may suggest that directly limiting the discretion of powerful on-the-ground actors, like the police, to activate harsh punishment, particularly when that harsh punishment falls disproportionately on disadvantaged groups, has merit on empirical as well as normative grounds. Lessons learned from the experiments occurring in the school context may inform ongoing discussions regarding how to minimize the harms from harsh formal social control in the community context without jeopardizing public safety.

References

Agodini, Roberto and Mark Dynarski. 2004. "Are experiments the only option? A look at dropout prevention programs." *The Review of Economics and Statistics* 86(1): 180-194.

Anderson, Kaitlin P. 2018. "Inequitable compliance: Implementation failure of a statewide student discipline reform." *Peabody Journal of Education* 93, no. 2: 244-263.

Anderson, Kaitlin P., Gary W. Ritter, and Gema Zamarro. 2019. "Understanding a vicious cycle: The relationship between student discipline and student academic outcomes." *Educational Researcher* 48 (5): 251-262.

Bacher-Hicks, Andrew, Stephen Billings, and David Deming. 2019. The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime. NBER Working Paper.

Balfanz, Robert, Lisa Herzog, and Douglas J. Mac Iver. 2007. "Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions." *Educational Psychologist* 42, no. 4: 223-235.

Barber, Sharrelle. 2020. "Death by racism." *The Lancet Infectious Diseases* 20.8: 903

Barrett, Nathan, Andrew McEachin, Jonathan N. Mills, and Jon Valant. 2019. "Disparities and discrimination in student discipline by race and family income." *Journal of Human Resources*: 0118-9267R2.

Bartanen, Brendan, Jason A. Grissom, and Laura K. Rogers. 2019. "The Impacts of Principal Turnover." *Educational Evaluation and Policy Analysis*.

Bell, Charles. 2020. "`Maybe if they let us tell the story I wouldn't have gotten suspended': Understanding Black students' and parents' perceptions of school discipline." *Children and Youth Services Review* 110: 104757.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad. 2018. "Intergenerational effects of incarceration." In *AEA Papers and Proceedings*, 108: 234-40.

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. 2012. "Estimating the effect of leaders on public sector productivity: The case of school principals." No. w17803. National Bureau of Economic Research.

Burkhauser, Susan. 2017. "How much do school principals matter when it comes to teacher working conditions?" *Educational Evaluation and Policy Analysis* 39, no. 1: 126-145.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review* 104, no. 9: 2633-79.

Chu, Elizabeth M. and Douglas D. Ready. 2018. "Exclusion and urban public high schools: Short-and long-term consequences of school suspensions." *American Journal of Education* 124, no. 4: 479-509.

Coelli, Michael, and David A. Green. 2012. Leadership effects: School principals and student outcomes. *Economics of Education Review* 31(1): 92-109.

Cook, Philip J., Denise C. Gottfredson, and Chongmin Na. 2010. "School crime control and prevention." *Crime and Justice* 39, no. 1: 313-440.

Council of State Governments Justice Center. 2017. *Realizing the full vision of school discipline reform: A framework for statewide change.* Available at: https://csgjusticecenter.org/wp-

content/uploads/2017/09/JC_Realizing-the-Full-Vision-of-School-Discipline-Reform_A-Framework-for-Statewide-Change.pdf.

Curran, Chris. 2016. "Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student Behavior." *Educational Evaluation and Policy Analysis* 38, no. 4: 647-668.

DeMatthews, David E., Roderick L. Carey, Arturo Olivarez, and Kevin Moussavi Saeedi. 2017. "Guilty as charged? Principals' perspectives on disciplinary practices and the racial discipline gap." *Educational Administration Quarterly* 53, no. 4: 519-555.

Dhuey, Elizabeth and Justin Smith. 2018. "How school principals influence student learning." *Empirical Economics* 54, no. 2: 851-882.

Eckholm, Erik. 2010. *School suspensions lead to legal challenge*. New York Times. Available at: https://www.nytimes.com/2010/03/19/education/19suspend.html

Federal Commission on School Safety. 2018. *Final Report of the Federal Commission on School Safety*. U.S. Department of Education, Department of Justice, Department of Health and Human Services, Department of Homeland Security. Washington, D.C.

Gershoff, Elizabeth T. and Sarah A. Font. 2016. "Corporal punishment in US public schools: Prevalence, disparities in use, and status in state and federal policy." *Social policy report*, *30*.

Gibson, Priscilla A., Robert Wilson, Wendy Haight, Misa Kayama, and Jane M. Marshall. 2014. "The role of race in the out-of-school suspensions of black students: The perspectives of students with suspensions, their parents and educators." *Children and Youth Services Review* 47: 274-282.

Gottfredson, Gary D. and Denise C. Gottfredson. 2001. “What schools do to prevent problem behavior and promote safe environments.” *Journal of Educational and Psychological Consultation* 12, no. 4: 313-344. doi:10.1207/s1532768xjepc1204_02

Grissom, Jason A. and Brendan Bartanen. 2019. “Principal effectiveness and principal turnover.” *Education Finance and Policy* 14, no. 3: 355-382.

Grissom, Jason A., Demetra Kalogrides, and Susanna Loeb. 2015. “Using student test scores to measure principal performance.” *Educational Evaluation and Policy Analysis 37,* no. 1: 3-28.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway. 2018. “Imprisonment and labor market outcomes: Evidence from a natural experiment.” *American Journal of Sociology* 124, no. 1: 49-110.

Hartzell, Gary N. and Thomas A. Petrie. 1992. “The principal and discipline: Working with school structures, teachers and students.” *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* 65, no. 6:376-380.

Hemphill, S. A., Toumbourou, J. W., Herrenkohl, T. I., McMorris, B. J., and Catalano, R. F. 2006. The effect of school suspensions and arrests on subsequent adolescent antisocial behavior in Australia and the United States. *Journal of Adolescent Health*, *39*(5), 736-744.

Hemphill, Sheryl A., Todd I. Herrenkohl, Stephanie M. Plenty, John W. Toumbourou, Richard F. Catalano, and Barbara J. McMorris. 2012. "Pathways from school suspension to adolescent nonviolent antisocial behavior in students in Victoria, Australia and Washington State, United States." *Journal of community psychology 40, 3: 301-318.*

Henry, Gary T, and Erica Harbatkin. 2019. Turnover at the Top: Estimating the Effects of Principal Turnover on Student, Teacher, and School Outcomes. (Ed Working Paper: 19-95). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-95

Hinze-Pifer, Rebecca and Lauren Sartain. 2018. "Rethinking universal suspension for severe student behavior." *Peabody Journal of Education* 93, no. 2: 228-243.

Hirschfield, Paul J. 2008. "Preparing for prison? The criminalization of school discipline in the USA." *Theoretical Criminology* 12, 1: 79-101.

Hoffman, Stephen. 2014. "Zero benefit: Estimating the effect of zero tolerance discipline polices on racial disparities in school discipline." *Educational Policy* 28, no. 1: 69-95.

Hwang, NaYoung. 2018. "Suspensions and achievement: Varying links by type, frequency, and subgroup." *Educational Researcher* 47, no. 6: 363-374.

Kinsler, Josh. 2011. "Understanding the black–white school discipline gap." *Economics of Education Review* 30, no. 6: 1370-1383.

Kupchik, Aaron. 2010. *Homeroom security: School discipline in an age of fear.* New York: New York University Press.

Lacoe, Joanna and Matthew P. Steinberg. 2018. "Do suspensions affect student outcomes?" *Educational Evaluation and Policy Analysis* 41, no. 1: 34-62.

Lindsay, Constance A., and Cassandra MD Hart. 2017. "Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina." *Educational Evaluation and Policy Analysis* 39, 3: 485-510.

Losen, Daniel J., Cheri L. Hodson, I. I. Keith, A. Michael, Katrina Morrison, and Shakti Belway. 2015. “Are We Closing the School Discipline Gap?” *UCLA: The Civil Rights Project / Proyecto Derechos Civiles*. Retrieved from https://escholarship.org/uc/item/2t36g571

Losen, Daniel J. and Tia E. Martinez. 2013. “Out of school and off track: The overuse of suspensions in American middle and high Schools.” The UCLA Center for Civil Rights Remedies at The Civil Rights Project.

Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start improve children's life chances? Evidence from a regression discontinuity design." *The Quarterly journal of economics* 122 (1): 159-208.

Miller, A. 2013. Principal turnover and student achievement. *Economics of Education Review*, 36, 60–72. https://doi.org/10.1016/j.econedurev.2013.05.004

Moore, Tim, and Morag McArthur. 2014. "If only I, they, we had done things differently: Young people talk about school difficulties and crime." *Children and youth services review* 44: 249-255.

Morton, Rebecca. 2014. “Returning “decision” to school discipline decisions: An analysis of recent, anti-zero tolerance legislation.” *Washington University Law Review 91*: 757-784.

Mueller-Smith, Michael. 2015. “The criminal and labor market impacts of incarceration.” *Unpublished Working Paper*, *18*.

Niewenhuizen, Lisa and Jeffrey S. Brooks. 2013. “The assistant principals’ duties, training and challenges: From colorblind to critical race perspective. In Jeffrey Brooks and Noelle Arnold (eds.) *Anti-racist School Leadership: Toward Equity for America’s Students*. Pp. 185-2010. Information Age Publishing: Charlotte, N.C.

Noltemeyer, Amity L., Rose Marie Ward, and Caven McLoughlin. 2015. “Relationship between school suspension and student outcomes: A meta-analysis.” *School Psychology Review 44*, no. 2: 224-240.

Novak, Abigail. 2019. "The school-to-prison pipeline: An examination of the association between suspension and justice system involvement." *Criminal justice and behavior* 46, 8: 1165-1180.

National Center for Education Statistics (NCES). 2020. *Public High School Graduation Rates*. U.S. Department of Education. Available at: https://nces.ed.gov/programs/coe/indicator_coi.asp.

National Center for Education Statistics (NCES). 2019. *Indicators of School Crime and Safety: 2018*. U.S. Department of Education and U.S. Department of Justice.

North Carolina Department of Public Instruction. 2019. *North Carolina discipline data reporting procedures*. DPI Center for Safer Schools.

North Carolina State Board of Education. 2018. *Report to the North Carolina General Assembly Consolidated Data Report, 2016-17*. Department of Public Instruction.

North Carolina State Board of Education. 2011. *Report to the North Carolina General Assembly Consolidated Data Report, 2010-2011*. Department of Public Instruction.

North Carolina State Board of Education. 2008. *Report to the Joint Legislative Education Oversight Committee Consolidated Data Report, 2007-08*. Department of Public Instruction.

North Carolina State Board of Education. 2002. *Annual Study of Suspensions and Expulsions, 2000-2001*. Department of Public Instruction.

Okonofua, Jason A., and Jennifer L. Eberhardt. 2015. "Two strikes: Race and the disciplining of young students." *Psychological science* 26, 5: 617-624.

Oleszewski, Ashley, Alan Shoho, and Bruce Barnett. 2012. “The development of assistant principals: a literature review.” *Journal of Educational Administration 50, no. 3:264-286.*

Osher, David, George G. Bear, Jeffrey R. Sprague, and Walter Doyle. “How can we improve school discipline?” *Educational Researcher* 39(1): 48-58.

Perry, Brea L. and Edward W. Morris. 2014. “Suspending progress: Collateral consequences of exclusionary punishment in public schools.” *American Sociological Review 79*, no. 6: 1067-1087

Phillips, Kristine and Kevin Johnson. 2020. “Trump Emphasizes Violence in Cities: Are His ‘Law and Order’ Tactics Working.” USA Today.

Pyne, Jaymes. 2019. "Suspended attitudes: Exclusion and emotional disengagement from school." *Sociology of Education* 92, 1: 59-82.

Schwartzapfel, Beth. 2020. “What Trump Really Means When He Tweets ‘Law and Order’: A Brief History of a Political Dog Whistle.” The Marshall Project.

Skiba, Russell J., Mariella I. Arredondo, and Natasha T. Williams. 2014. “More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline.” *Equity & Excellence in Education 47*, no. 4: 546-564.

Skiba, Russell J., Choong-Geun Chung, Megan Trachok, Timberly L. Baker, Adam Sheya, and Robin L. Hughes. 2014. “Parsing disciplinary disproportionality.” *American Educational Research Journal 51*, no. 4: 640-670.

Sosa, Leticia Villarreal. 2020. School Resource Officers and Black Lives Matter Protests: It’s Time for School Social Work to Take a Stand. *Children & Schools* Online First

Steinberg, Matthew P. and Johanna Lacoe. 2018. "Reforming school discipline: school-level policy implementation and the consequences for suspended students and their peers." *American Journal of Education* 125, no. 1: 29-77.

Turner, Erica and Beneke, Abigail J. 2020. "'Softening' school resource officers: the extension of police presence in schools in an era of Black Lives Matter, school shootings, and rising inequality." *Race Ethnicity and Education*, *23*(2), 221-240.

U.S. Department of Education. 2014. *U.S. Department of Education and Justice Release School Discipline Guidance Package to Enhance School Climate and improve School Discipline Policies/Practices*. Available at: https://www.ed.gov/news/press-releases/us-departments-education-and-justice-release-school-discipline-guidance-package-

Valdebenito, Sara, Manuel Eisner, David P. Farrington, Maria M. Ttofi, and Alex Sutherland. 2019. "What can we do to reduce disciplinary school exclusion? A systematic review and meta-analysis." *Journal of Experimental Criminology* 9: 1-35.

Williams, John A., Chance Lewis, Tehia Glass, Bettie R. Butler, and Jae Hoon Lim. 2020. "The discipline gatekeeper: Assistant principals' experiences with managing school discipline in urban middle schools." *Urban Education Online first. 1-209.*

Wolf, Kerrin C. and Kupchik, Aaron. 2017. "School suspensions and adverse experiences in adulthood." *Justice Quarterly* 34, no. 3: 407-430.

Tables

Table 1. Descriptive Statistics of Student-Level Analytical Dataset

	N	Mean	Std Dev	Minimum	Maximum
Outcome variables					
Any minor offense	2,548,545	0.24	0.43	0.00	1.00
Any serious offense	2,548,545	0.01	0.09	0.00	1.00
Removal indicator	2,548,545	0.11	0.32	0.00	1.00
Days absent	2,318,333	6.20	6.87	0.00	170.00
Test scores (SD)	2,542,587	0.00	1.00	-3.00	2.99
Grade retention	2,219,268	0.01	0.09	0.00	1.00
Juvenile complaint[1]	1,109,364	0.02	0.13	0.00	1.00
Misdemeanor juvenile[1]	1,109,364	0.01	0.12	0.00	1.00
Felony juvenile[1]	1,109,364	0.00	0.05	0.00	1.00
HS graduation[2]	423,237	0.71	0.45	0.00	1.00
Adult conviction[2]	228,907	0.04	0.19	0.00	1.00
Principal variables					
Principal PTR	2,548,545	0.32	0.08	-0.01	0.83
Principal removal bias	1,761,269	0.01	0.04	-0.17	0.18
Principal VA (reading)	2,547,842	0.00	0.08	-1.01	0.51
Principal VA (math)	2,548,545	0.00	0.09	-0.89	0.45
Principal experience	2,443,450	21.99	7.50	0.00	48.00
Principal female	2,548,283	0.51	0.50	0.00	1.00
Principal Black	2,548,545	0.25	0.43	0.00	1.00
Principal Hispanic	2,548,545	0.00	0.04	0.00	1.00
Principal other race	2,548,545	0.02	0.13	0.00	1.00
Student variables					
Student black	2,548,545	0.21	0.41	0.00	1.00
Student Hispanic	2,548,545	0.10	0.30	0.00	1.00
Student other race	2,548,545	0.06	0.24	0.00	1.00
Student female	2,548,347	0.49	0.50	0.00	1.00
Student LEP	2,548,545	0.06	0.23	0.00	1.00
Student ED	2,548,435	0.57	0.50	0.00	1.00
School variables					
Assistant principals	2,548,545	2.07	1.07	0.00	7.00
FTE teachers	2,548,545	48.69	15.54	0.00	106.00
Percent high ED	2,548,545	46.01	25.80	0.00	100.00
Percent moderate ED	2,548,545	7.82	6.57	0.00	100.00
Enrollment	2,548,545	735.45	296.62	10.00	1781.00
Percent black	2,548,545	20.95	19.53	0.00	97.58
Percent Hispanic	2,548,545	9.40	8.33	0.00	68.68
Percent other race	2,548,545	6.07	6.60	0.00	93.95

[1]Juvenile justice data is only available for four of the nine years.
[2]These long-term outcome measures are available in a collapsed student-level dataset (rather than student-year-level) for certain cohorts of students.

Note. PTR = propensity to remove; VA = value-added score; LEP = Limited English Proficiency; ED = economically disadvantaged; FTE = full time equivalent.

Table 2. Effects of Principal PTR on Disciplinary and Academic Outcomes

Variables	Minor Offense	Serious Offense	Any Removal	Days Absent	Test Scores (SDs)	Grade Retention	H.S. Graduation
Principal PTR	-0.1032**	-0.0012	0.0737**	0.3261	-0.0238	0.0058	-0.1060*
	(0.027)	(0.002)	(0.014)	(0.329)	(0.034)	(0.004)	(0.041)
Controls	✔	✔	✔	✔	✔	✔	✔
Year and Grade FE	✔	✔	✔	✔	✔	✔	--
Cohort FE	--	--	--	--	--	--	✔
School FE	✔	✔	✔	✔	✔	✔	✔
Observations	2,548,301	2,548,301	2,548,301	2,318,308	2,542,365	2,219,258	423,208
R-Squared	0.144	0.031	0.388	0.061	0.283	0.017	0.065

** p<0.01, * p<0.05, + p<0.1.

Note. Robust standard errors in parentheses, clustered by school. Number of student minor and serious offenses controlled for in "Any Removal" regression. Coefficients on control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Table 3. Effects of Principal PTR on Academic Outcomes, by Student Disciplinary Record

Variables		Any Removal	Days Absent	Test Scores (SDs)	Grade Retention	High School Graduation
No Disciplinary Record	β	--	0.0969	-0.0613+	0.0025	-0.1048*
	S.E.		(0.281)	(0.036)	(0.002)	(0.051)
	N		1,755,013	1,931,208	1,685,273	268,939
Minor Disciplinary Record	β	0.3766**	2.4597**	-0.1040*	0.0225*	-0.1377*
	S.E.	(0.071)	(0.729)	(0.046)	(0.010)	(0.061)
	N	594,980	546,005	592,730	518,095	146,962
Serious Disciplinary Record	β	-0.0246	0.2214	-0.1779	-0.0023	-0.0716
	S.E.	(0.044)	(2.022)	(0.111)	(0.025)	(0.127)
	N	18,656	17,255	18,394	15,854	7,236
Controls		✔	✔	✔	✔	✔
Year and Grade FE		✔	✔	✔	✔	--
Cohort FE		--	--	--	--	✔
School FE		✔	✔	✔	✔	✔

** p<0.01, * p<0.05, + p<0.1.

Note. Robust standard errors in parentheses, clustered by school. Each cell represents the coefficient on Principal PTR from a separate regression. Coefficients on control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Table 4. Effects of Principal PTR on Criminal Justice Outcomes, by Student Disciplinary Record

Effect of Principal PTR	Any Juvenile Complaint	School-Based Juvenile Complaint	Non-School Juvenile Complaint	Misdemeanor Juvenile Complaint	Felony Juvenile Complaint	Adult Criminal Conviction
All Students	0.0126**	0.0097*	0.0033	0.0102*	-0.0003	-0.0255
	(0.005)	(0.004)	(0.003)	(0.004)	(0.001)	(0.023)
	N=1,109,200	N=1,109,200	N=1,109,200	N=1,109,200	N=1,109,200	N=228,904
No Disciplinary Record	0.0005	0.0006	-0.0001	0.0007	0.0000	-0.0385**
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.014)
	N=832,389	N=832,389	N=832,389	N=832,389	N=832,389	N=155,183
Minor Disciplinary Record	0.0286	0.0209	0.0113	0.0221	-0.0028	-0.0698
	(0.020)	(0.018)	(0.008)	(0.019)	(0.004)	(0.047)
	N=268,251	N=268,251	N=268,251	N=268,251	N=268,251	N=70,247
Serious Disciplinary Record	0.3560**	0.3559**	-0.1001+	0.3364**	-0.0180	-0.1056
	(0.110)	(0.107)	(0.056)	(0.116)	(0.051)	(0.206)
	N=8,510	N=8,510	N=8,510	N=8,510	N=8,510	N=3,417
Controls	✔	✔	✔	✔	✔	✔
Year and Grade FE	✔	✔	✔	✔	✔	--
Cohort FE	--	--	--	--	--	✔
School FE	✔	✔	✔	✔	✔	✔

** p<0.01, * p<0.05, + p<0.1.

Note. Robust standard errors in parentheses, clustered by school. Coefficients on control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Table 5. Effects of Principal Bias on Academic Outcomes, by Student Race

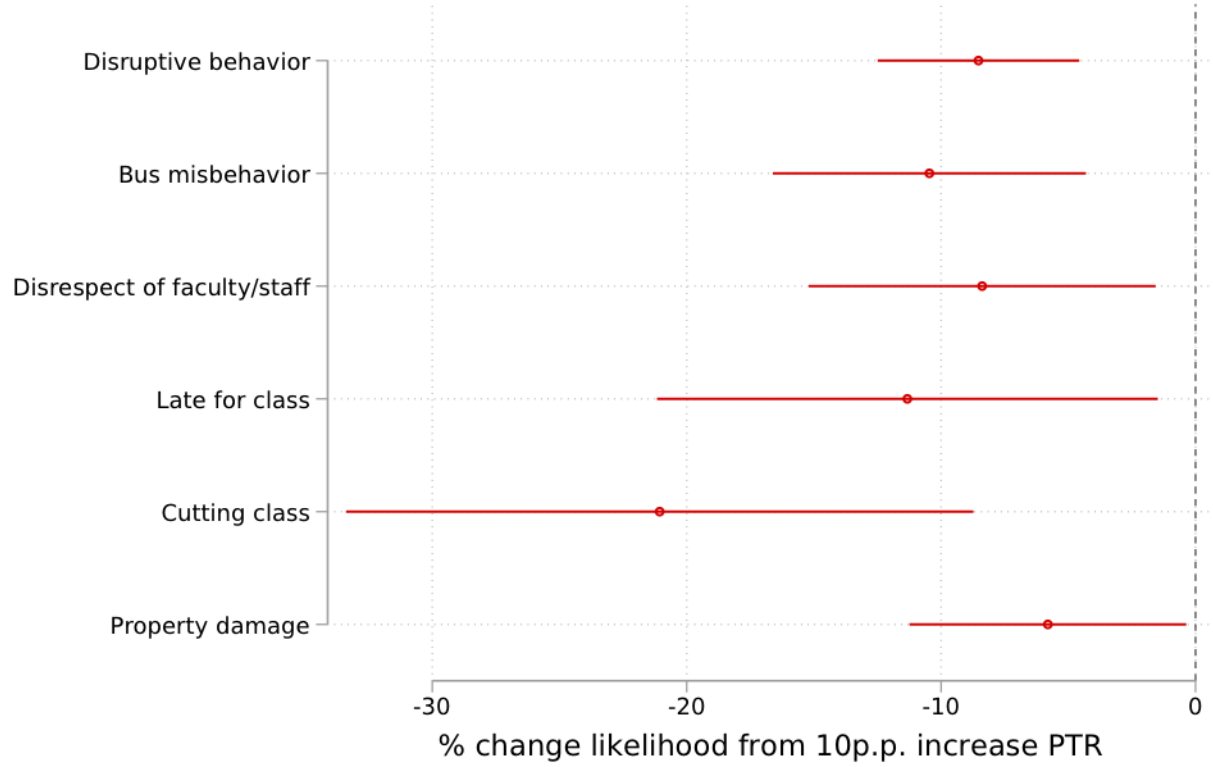
Variables	Days Absent	Test Scores (SD)	Grade Retention	High School Graduation
Principal Bias	-2.6312**	0.2813*	-0.0086	0.1930
	(0.951)	(0.112)	(0.006)	(0.124)
Bias x Black	7.0277**	-1.0024**	0.0386**	-0.4882**
	(1.111)	(0.185)	(0.010)	(0.114)
Bias x Hispanic	6.2824**	-1.1264**	0.0254**	-0.3338*
	(1.199)	(0.236)	(0.009)	(0.141)
Bias x Other	1.8847+	-0.1374	0.0010	-0.4198**
	(1.072)	(0.216)	(0.008)	(0.126)
Student Black	-2.0343**	-0.4323**	-0.0002	0.0658**
	(0.050)	(0.008)	(0.000)	(0.005)
Student Hispanic	-1.6889**	0.0613**	-0.0032**	0.0285**
	(0.056)	(0.010)	(0.000)	(0.006)
Student Other	-1.2706**	0.0516**	-0.0008*	0.0094+
	(0.055)	(0.013)	(0.000)	(0.005)
Year and Grade FE	✔	✔	✔	--
School FE	✔	✔	✔	✔
Cohort FE	--	--	--	✔
Observations	1,607,988	1,757,339	1,538,647	306,613
R-Squared	0.061	0.281	0.016	0.059

** p<0.01, * p<0.05, + p<0.1.

Note. Robust standard errors in parentheses, clustered by school. Coefficients on control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Figures

Figure 1. Effects of Principal PTR on Selected Student Disciplinary Offenses



Note. This figure shows effects of a 10-percentage point increase in principal PTR on an indicator of whether the student has each particular disciplinary incident type, in a linear probability model with school, grade, and year fixed effects, and controls. Of the 35 incident types, only those that had statistically significant effects are included. Control variables include: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Figure 2. Academic Effects of Principal Black-White PTR Bias by Student Race/Ethnicity

Figure 2a. Student Absences

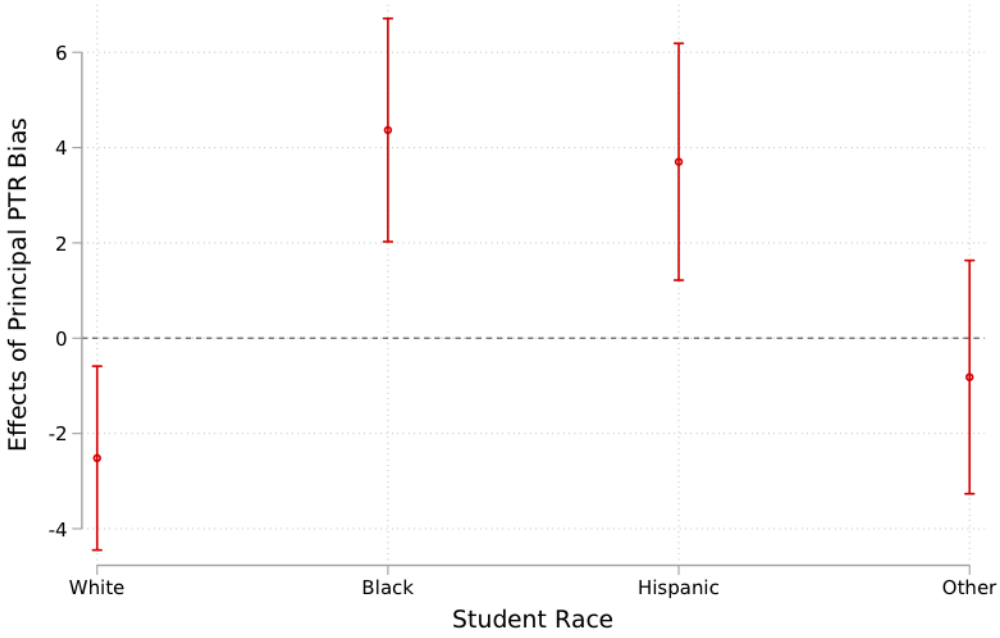


Figure 2b. Student Test Scores (SDs)

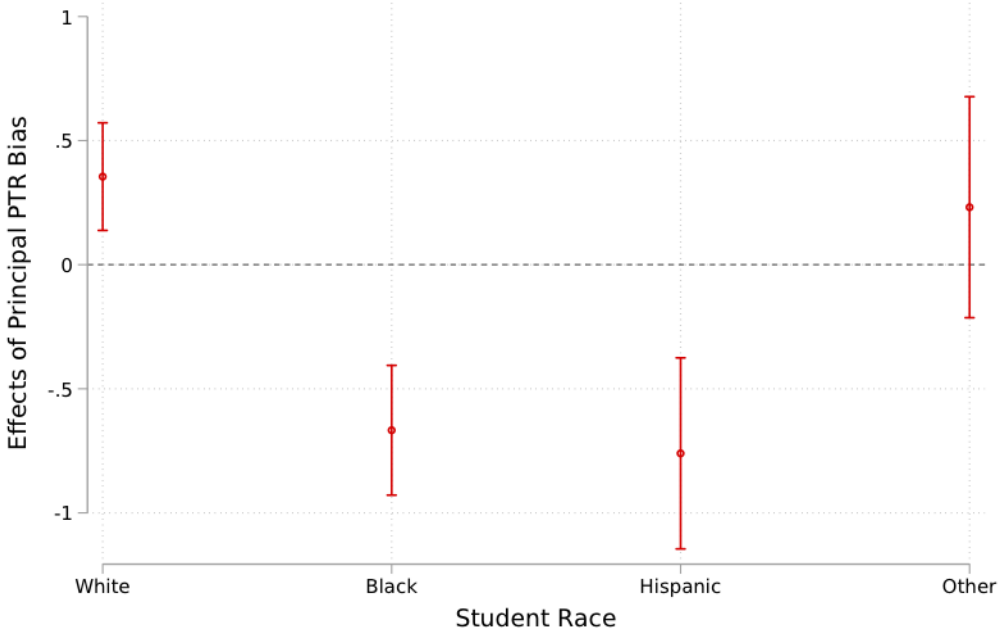


Figure 2c. Student Grade Retention

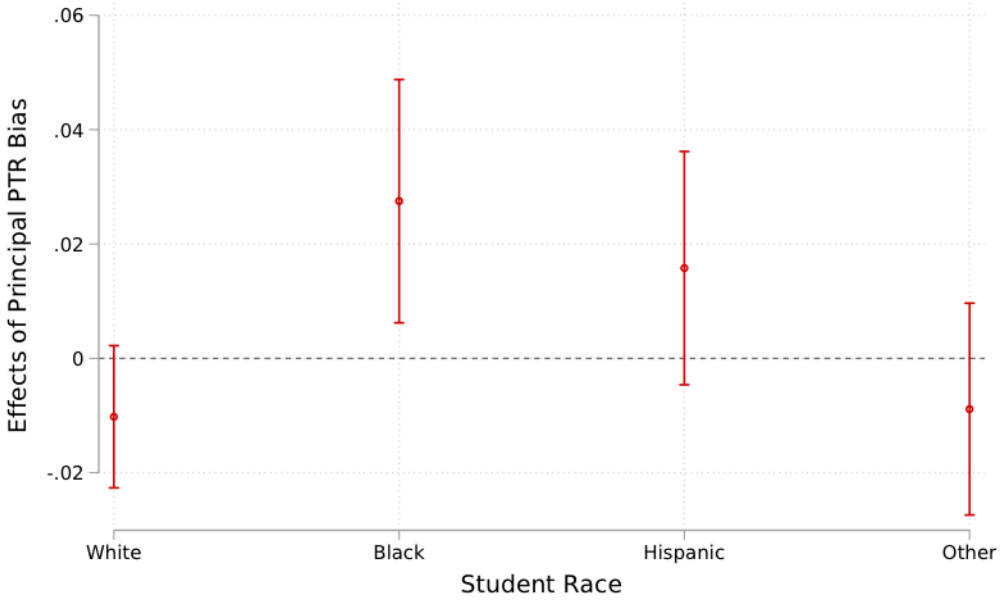
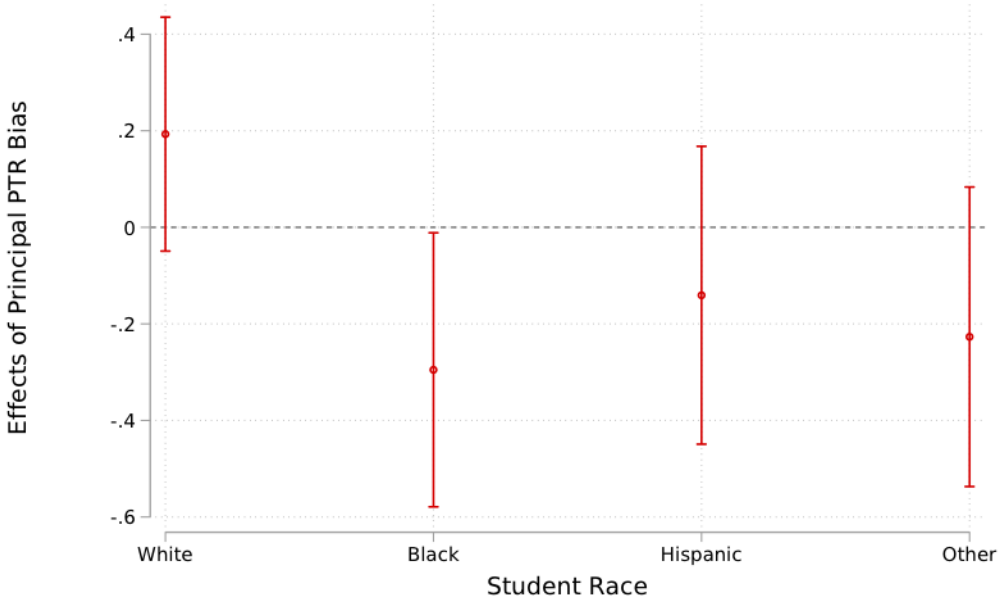


Figure 2d. High School Graduation



Note. Marginal effects are presented with 95% confidence intervals. All models are estimated with school fixed effects, year and grade fixed effects for short-term outcomes, cohort fixed effects for high school graduation, and controls. Control variables include: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

APPENDIX A. SUPPLEMENTAL TABLES AND FIGURES

Figure A.1. Distribution of Estimated Principal Propensity-to-Remove (PTR)

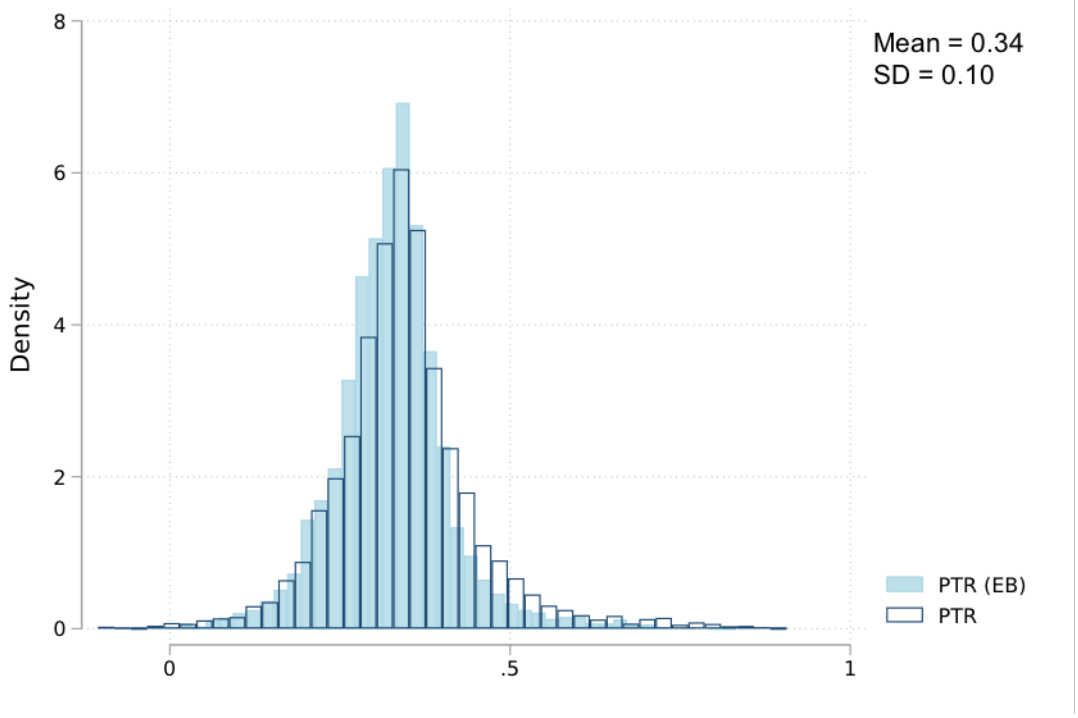


Figure A.2. Distribution of Estimated Principal Black-White PTR Bias

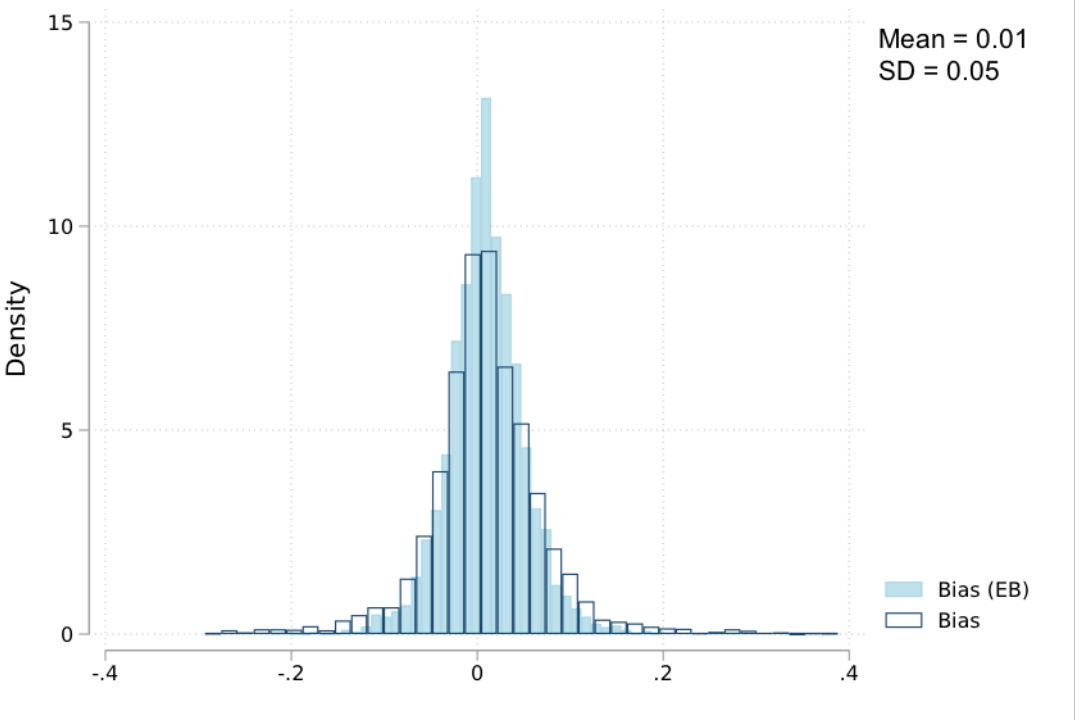


Figure A.3. Trends Before and After Any Principal Transition

Figure .A3a. Student Removal

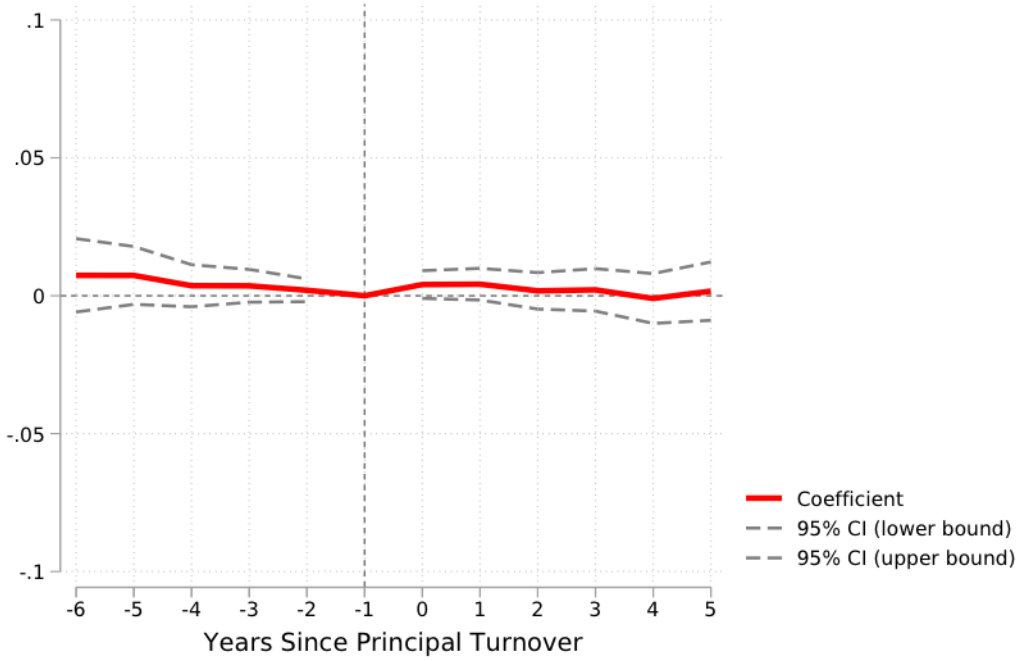


Figure A.3b. Minor Student Disciplinary Offense

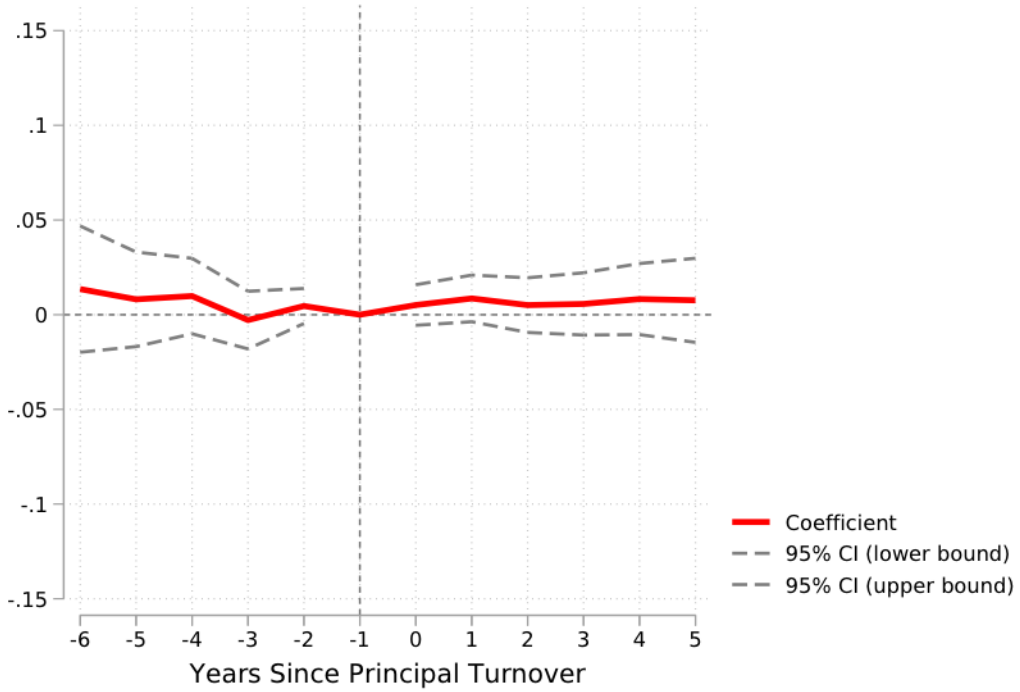


Figure A.3c. Serious Student Disciplinary Offense

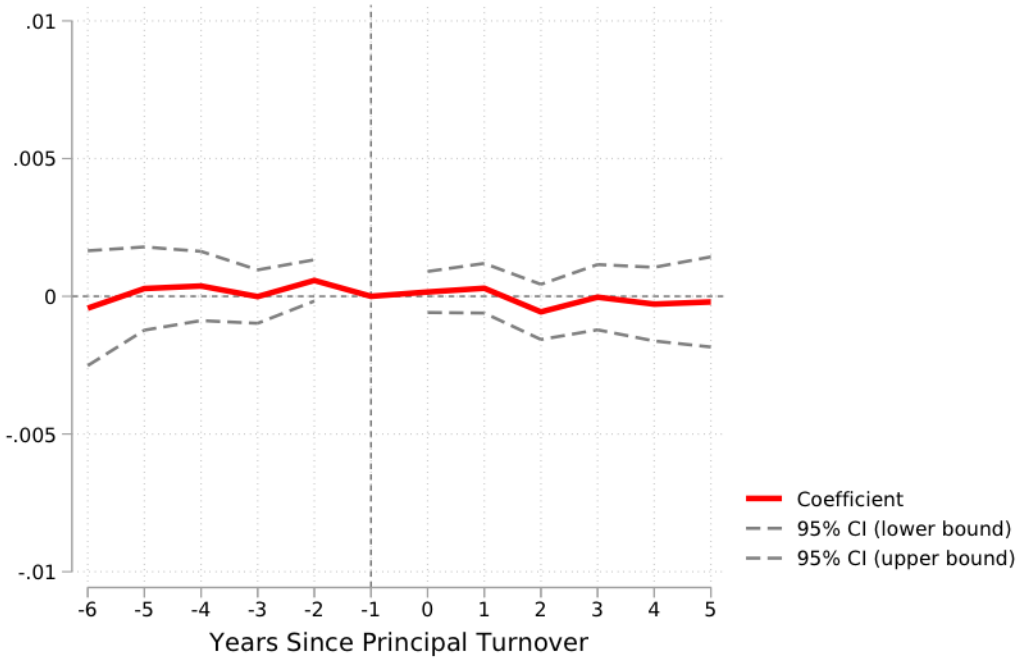
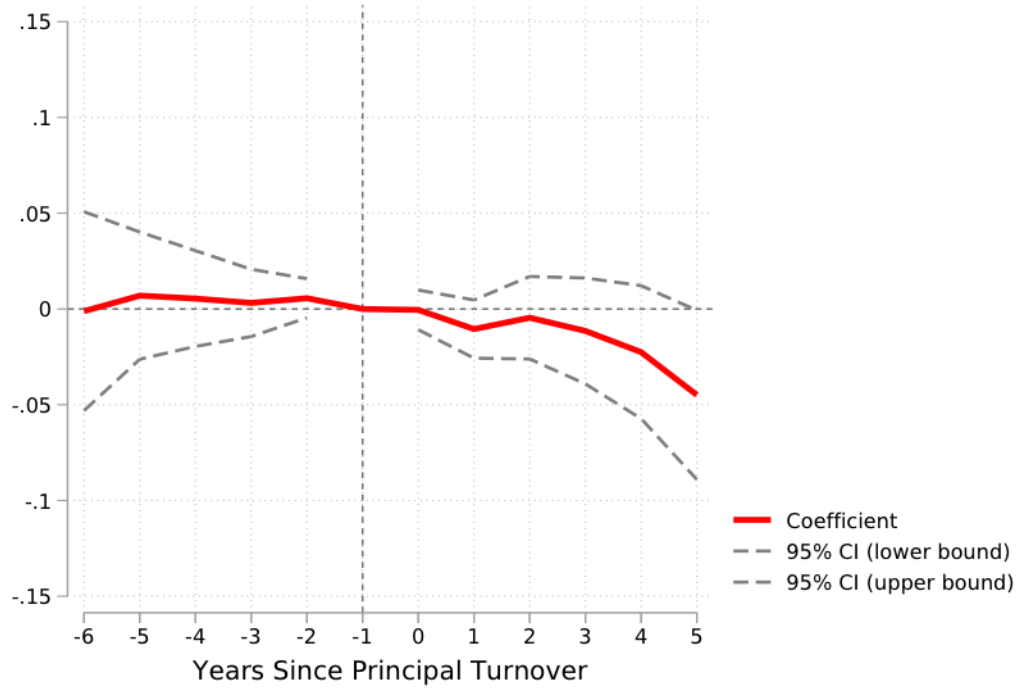


Figure A.3d. Student Test Scores



Note. These models are estimated with school, grade, and year fixed effects, and a series of student- and school-level controls. Principal turnover is defined as the first principal transition observed in the school.

Figure A.4. Trends Before and After Transition from Lenient to Severe Principal

Figure A.4a. Student Removal

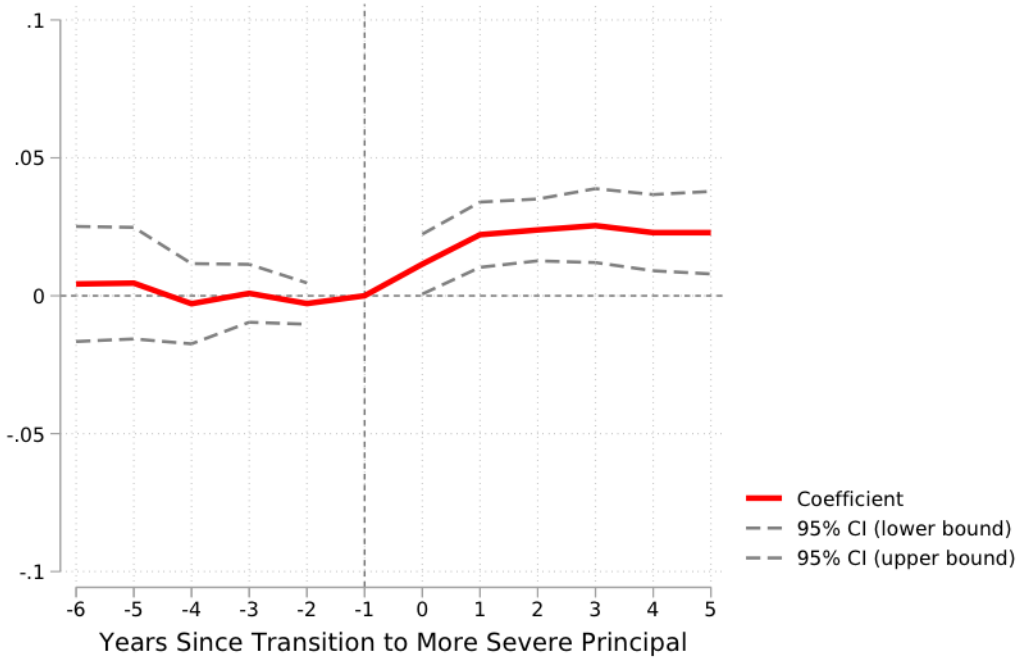


Figure A.4b. Minor Student Disciplinary Offense

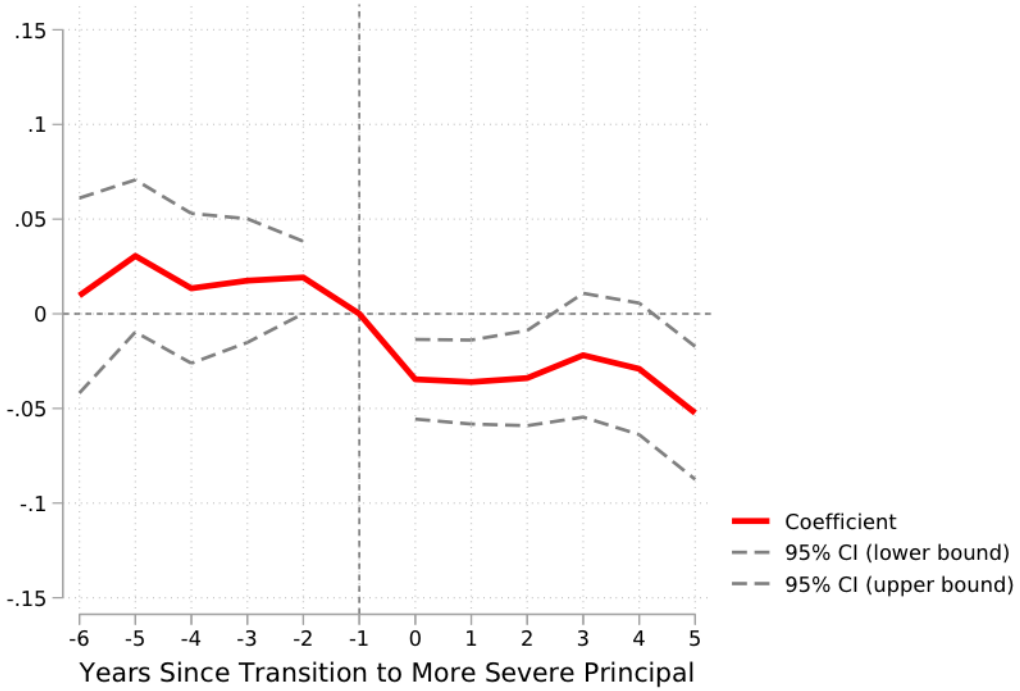


Figure A.4c. Serious Student Disciplinary Offense

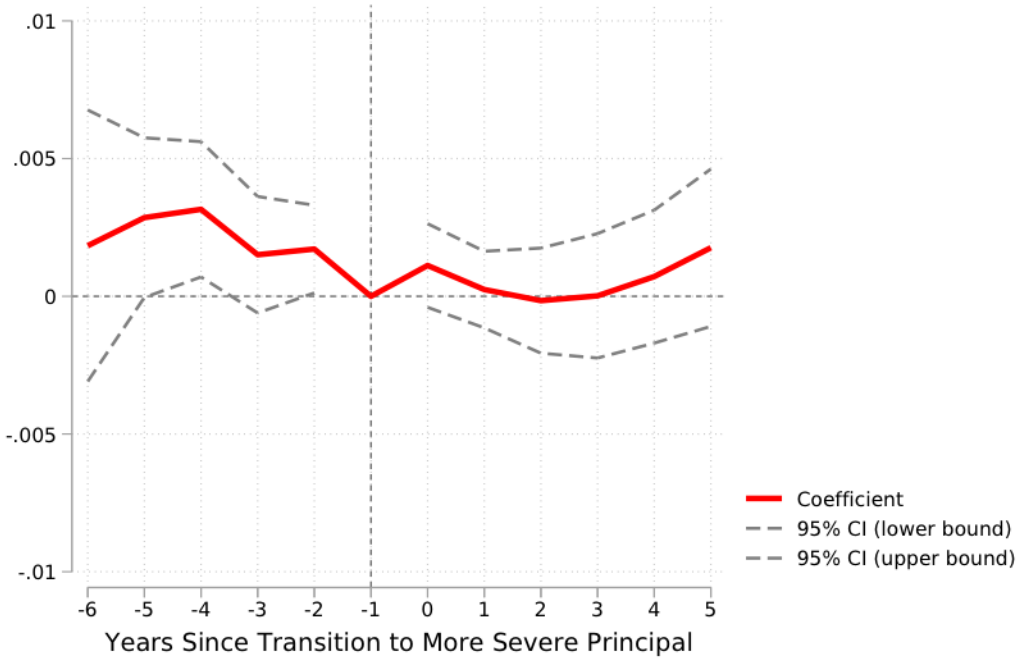
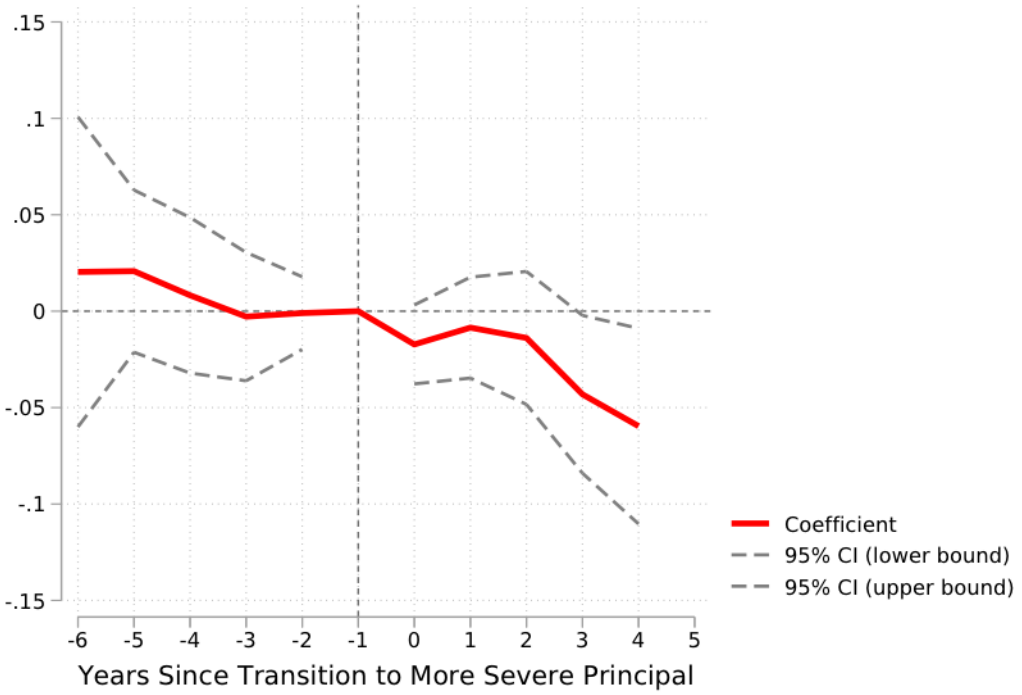


Figure A.4d. Student Test Scores



Note. These models are estimated with school, grade, and year fixed effects, and a series of student- and school-level controls. Transition to a severe principal is defined as the first principal transition observed in the school where PTR of the departing principal is 1SD (~8 p.p.) below the PTR of the incoming principal.

Figure A.5. Trends Before and After Transition from Severe to Lenient Principal

Figure A.5a. Student Removal

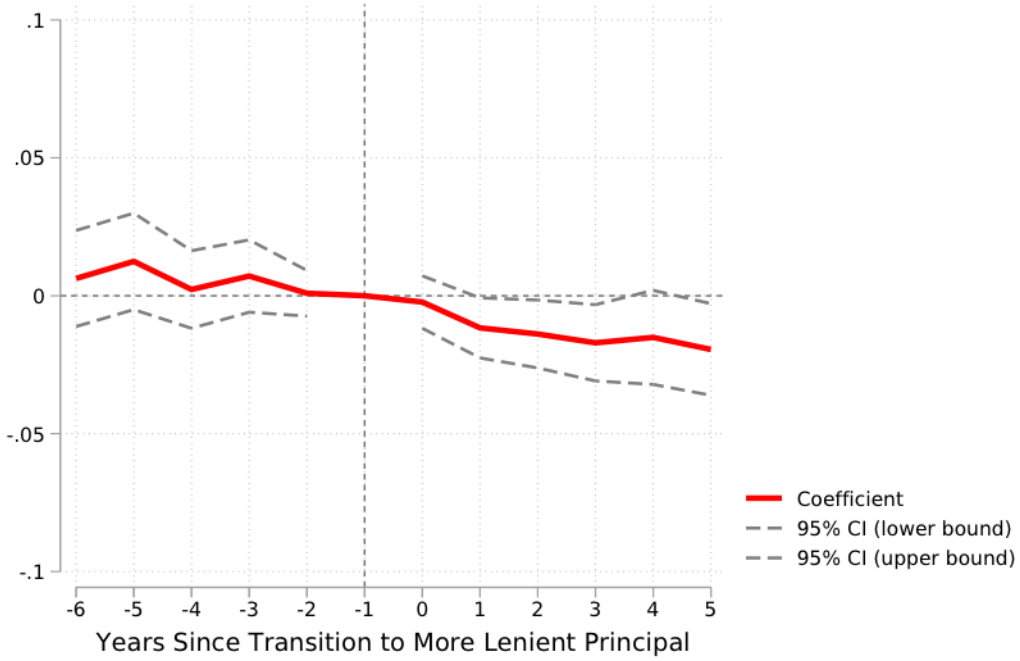


Figure A.5b. Minor Student Disciplinary Offense

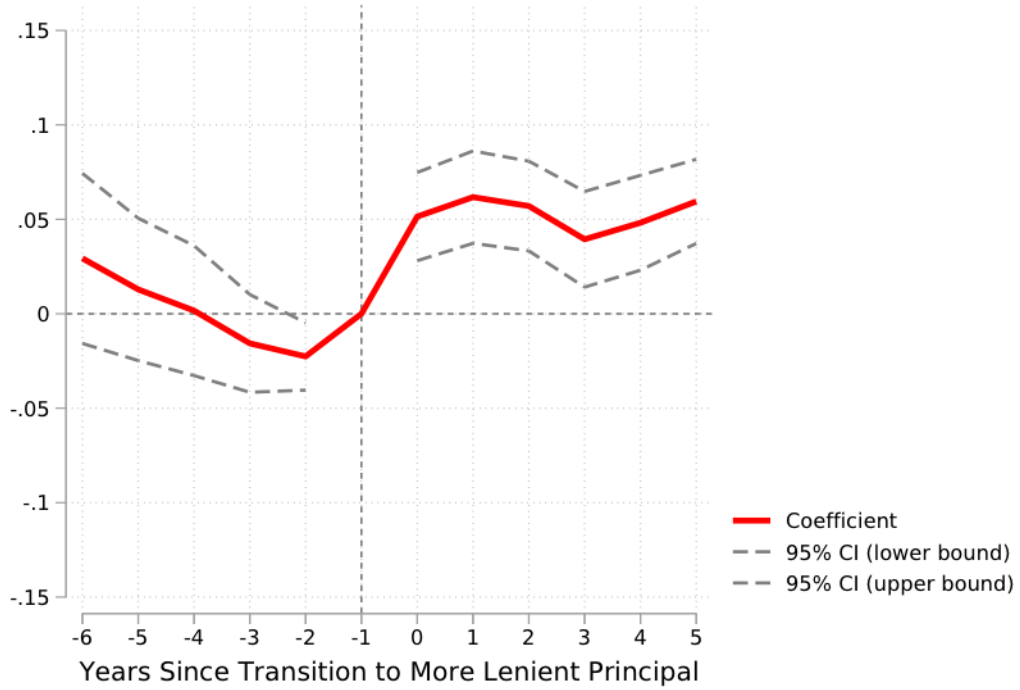


Figure A.5c. Serious Student Disciplinary Offense

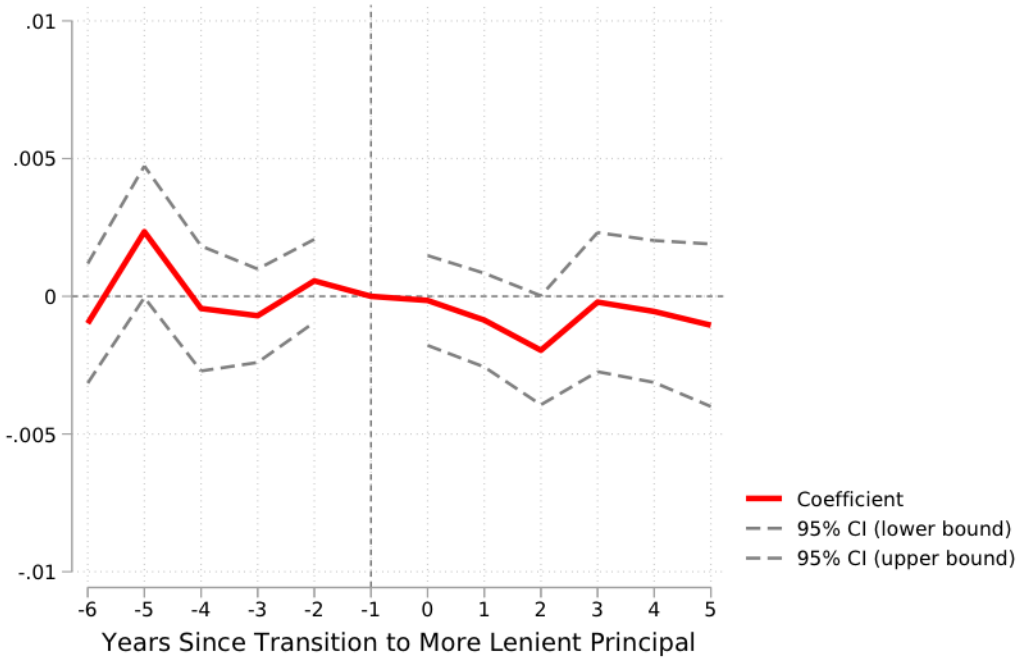
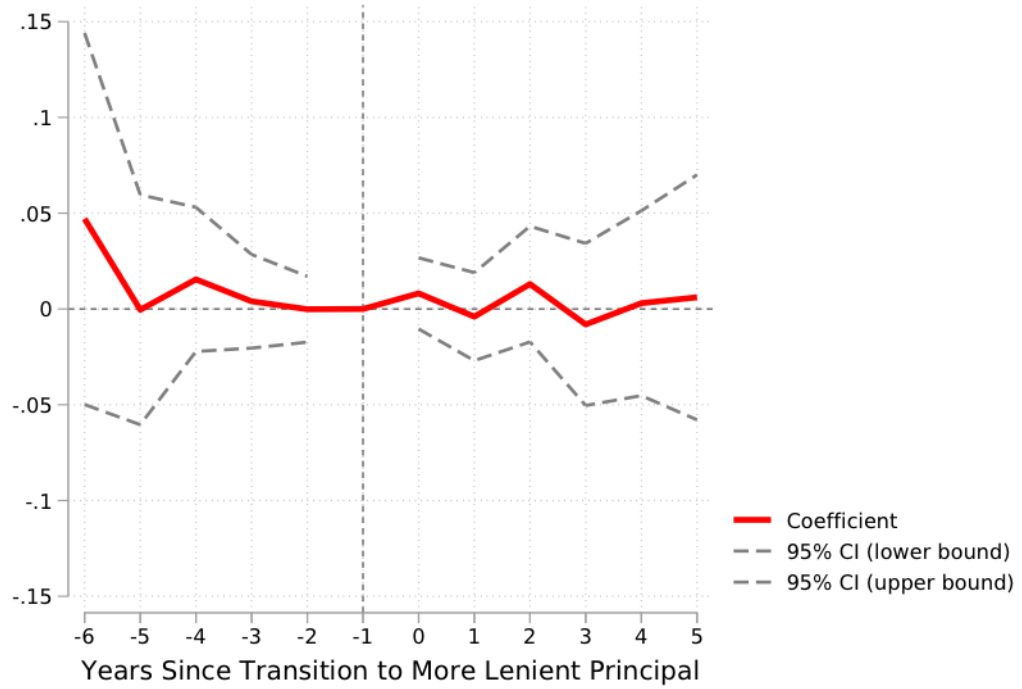


Figure A.5d. Student Test Scores



Note. These models are estimated with school, grade, and year fixed effects, and a series of student- and school-level controls. Transition to a lenient principal is defined as the first principal transition observed in the school where PTR of the departing principal is 1SD (~8 p.p.) above the PTR of the incoming principal.

Table A.1. Event Frequency and Probability of Removal by Offense Type

Offense Type	Reportable Offense	N	Pr(Removal)
Excessive tardiness	No	34,811	0.028
Bus misbehavior	No	184,140	0.035
Cutting class	No	44,559	0.129
Cell phone use	No	19,964	0.129
Skipping school	No	42,678	0.179
Dress code violation	No	30,690	0.201
Being in an unauthorized area	No	29,566	0.203
Other school defined offense	No	86,592	0.218
Disruptive behavior	No	631,231	0.252
Insubordination	No	278,488	0.268
Disrespect of faculty/staff	No	109,461	0.299
Late to class	No	67,918	0.307
Bullying	No	68,203	0.342
Inappropriate language/disrespect	No	201,577	0.347
Property damage	No	17,687	0.399
Aggressive behavior	No	194,624	0.467
Disorderly conduct	No	32,439	0.489
Inappropriate items on school property	No	15,966	0.546
Theft	No	33,946	0.574
Communicating threats	No	21,728	0.753
Sexual offense	Yes	145	0.766
Robbery without a dangerous weapon	No	64	0.766
Assault on student not resulting in serious injury	No	17,085	0.780
Sexual assault (not involving rape or sexual offense)	Yes	613	0.819
Burning of a school building	Yes	419	0.823
Assault resulting in a serious injury	Yes	235	0.838
Bomb threat	Yes	257	0.848
Possession of a weapon (not firearm)	Yes	10,529	0.858
Possession of alcoholic beverage	Yes	2,637	0.864
Assault on school personnel	Yes	3,077	0.865
Unlawfully setting a fire	No	561	0.866
Possession of controlled substance in violation of law	Yes	10,196	0.868
Fighting	No	184,240	0.871
Possession of a firearm or powerful explosive	Yes	185	0.881
Assault involving use of a weapon	Yes	257	0.891

Note. Table is sorted from offenses least likely to result in student removal to offenses most likely to result in removal. Removal is defined as out-of-school suspension, expulsion, or transfer to an alternative school. Reportable offenses are classified by NC DPI as those requiring mandatory reporting to the state and/or law enforcement.

Table A.2. Summary of Principal Turnover Within and Across Schools

(a) Count of Principals Observed in Each School 2008-2016

# Principals per School	Frequency	% Observations
Full Sample		
1	229,838	8.18
2	912,761	32.47
3	944,723	33.61
4	526,570	18.73
5	171,584	6.10
6	20,758	0.74
7	4,907	0.17
Principals with 2+ Years in School		
1	550,779	21.61
2	1,330,445	52.2
3	614,693	24.12
4	52,628	2.07

(b) Count of Schools in Which Each Principal Observed 2008-2016

# Schools per Principal	Frequency	% Observations
Full Sample		
1	2,214,675	78.78
2	539,868	19.20
3	45,039	1.60
4	11,559	0.41
Principals with 2+ Years in School		
1	2,201,590	86.39
2	337,186	13.23
3	9,769	0.38

Note. In both tables above, the first panel describes the number of principals observed in each school (or number of schools observed for each principal) prior to the sample restriction that each principal must serve at least two years in the school to be included. The second panel represents these counts post-restriction.

Table A.3. Correlation Matrix of Alternative PTR and Removal Bias Estimates

	Propensity to Remove (PTR)					Black-White Removal Bias			
	Preferred	Across-School	No EB Shrinkage	Random Effects	ISS Included	Preferred	Across-School	No EB Shrinkage	Random Effects
PTR									
Preferred	1.000								
Across-School	0.268***	1.000							
No EB Shrinkage	0.985***	0.281***	1.000						
Random Effects	0.241***	0.969***	0.254***	1.000					
ISS Included	0.600***	0.104***	0.574***	0.091**	1.000				
Removal Bias									
Preferred	0.0672*	0.0983***	0.0667*	0.100***	0.0206	1.000			
Across-School	0.0304	0.198***	0.0299	0.192***	0.00597	0.861***	1.000		
No EB Shrinkage	-0.0175	0.0570*	-0.0235	0.0606*	-0.0199	0.916***	0.825***	1.000	
Random Effects	0.0399	0.241***	0.0462	0.238***	-0.00710	0.752***	0.867***	0.776***	1.000

** p<0.01, * p<0.05, + p<0.1.

Note. Each cell is the pairwise correlation between the two measures within the disciplinary referral dataset. All models residualize the removal indicator by student grade level and student offense history during that school year and in the prior three school years. All estimates are weighted averages of estimates by offense type. PTR Preferred and Bias Preferred are the measures used in the main analysis of student outcomes. “Across School” estimates PTR after residualizing removal indicators by school and year; “No EB Shrinkage” does not apply empirical Bayes shrinkage; “Rand Effects” uses a random effects leave-year-out estimate with EB shrinkage; and “Non-Discretion” restricts the sample of disciplinary events to those with less discretionary reporting practices.

Table A.4. Effects of PTR, Comparison of Alternative PTR Estimates

PTR Measure	Minor Offense	Serious Offense	Any Removal	Days Absent	Test Scores (SD)	Grade Retention	Juvenile Complaint	High School Graduation	Adult Conviction
Preferred	-0.1032**	-0.0012	0.0737**	0.3261	-0.0238	0.0058	0.0126**	-0.1060*	-0.0255
	(0.027)	(0.002)	(0.014)	(0.329)	(0.034)	(0.004)	(0.005)	(0.041)	(0.023)
Across School	-0.1010**	-0.0007	0.0646**	0.3976	-0.0056	0.0042	0.0091*	-0.1455**	0.0468*
	(0.024)	(0.001)	(0.012)	(0.273)	(0.028)	(0.003)	(0.004)	(0.034)	(0.020)
No EB Shrinkage	-0.0949**	-0.0009	0.0636**	0.3198	-0.0202	0.0049	0.0099**	-0.1011**	-0.0161
	(0.023)	(0.001)	(0.012)	(0.270)	(0.028)	(0.003)	(0.004)	(0.032)	(0.019)
Random Effects	-0.1205**	-0.0007	0.0859**	0.3556	-0.0058	0.0043	0.0099*	-0.0631+	0.0495*
	(0.026)	(0.002)	(0.013)	(0.287)	(0.028)	(0.003)	(0.004)	(0.037)	(0.022)
ISS Included	-0.0463	-0.0008	0.0458**	-0.1540	0.0072	0.0064+	0.0076	-0.0908*	0.0037
	(0.029)	(0.002)	(0.013)	(0.380)	(0.033)	(0.003)	(0.005)	(0.045)	(0.017)
Controls	✔	✔	✔	✔	✔	✔	✔	--	--
Year and Grade FE	✔	✔	✔	✔	✔	✔	✔	--	--
School FE	✔	✔	✔	✔	✔	✔	✔	✔	✔
Cohort FE	--	--	--	--	--	--	--	✔	✔

** p<0.01, * p<0.05, + p<0.1.

Note. Each cell in this table represents the estimated coefficient on PTR in a model with school fixed effects and the full set of control variables. The alternative PTR estimates are described in Appendix B, and a correlation matrix of these measures is shown in Appendix Table A3. Coefficients on control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Table A.5. Correlation Matrix of Principal Characteristics

Principal Characteristic	PTR	PTR Bias	VA Reading	VA Math	Experience	Female	Black	Hispanic	Other Race
PTR	1.000								
PTR Bias	0.009	1.000							
VA Reading	-0.1216	0.1044	1.000						
VA Math	-0.0258	-0.0137	0.402	1.000					
Experience	-0.0666	0.0522	0.0411	-0.0219	1.000				
Female	-0.043	0.0154	0.1005	0.0271	0.1082	1.000			
Black	-0.0458	-0.0607	0.0215	-0.0979	-0.0331	0.0992	1.000		
Hispanic	0.0244	0.0277	0.0472	0.0238	-0.0312	-0.027	-0.0223	1.000	
Other Race	0.0335	-0.0239	-0.0204	-0.0502	-0.0507	0.0045	-0.0735	-0.005	1.000

Note. VA Reading and VA Math are principal value-added scores estimated within-school.

Table A.6. Effects of PTR on Student Outcomes, Controlling for Principal Characteristics

Variables	Minor Offense	Serious Offense	Any Removal	Days Absent	Test Scores (SD)	Grade Retention	Juvenile Complaint	High School Graduation	Adult Conviction
Principal PTR	-0.0941**	-0.0013	0.0688**	0.3218	0.0192	0.0062+	0.0117*	-0.0820+	-0.0381
	(0.027)	(0.002)	(0.015)	(0.339)	(0.030)	(0.004)	(0.005)	(0.042)	(0.025)
Principal Experience	0.0815*	0.0063*	-0.0139	-0.4152	0.0718	0.0028	-0.0110	-0.0008	-0.0002
	(0.035)	(0.003)	(0.017)	(0.425)	(0.045)	(0.004)	(0.015)	(0.001)	(0.000)
Principal V-A Read	0.0064	-0.0091**	-0.0322+	-0.1555	0.7253**	0.0116*	0.0138	0.4714**	-0.1831**
	(0.039)	(0.003)	(0.018)	(0.517)	(0.052)	(0.006)	(0.014)	(0.079)	(0.050)
Principal V-A Math	0.0002	0.0000	-0.0004**	-0.0042	0.0000	0.0000	-0.0001	0.0969	-0.2055**
	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.076)	(0.050)
Principal Female	-0.0014	-0.0002	0.0043+	-0.0330	-0.0035	-0.0001	0.0014	-0.0022	0.0028
	(0.005)	(0.000)	(0.002)	(0.075)	(0.007)	(0.001)	(0.001)	(0.012)	(0.007)
Principal Black	-0.0042	0.0013*	-0.0003	0.0369	-0.0135	-0.0014	0.0011	0.0032	0.0038
	(0.007)	(0.001)	(0.003)	(0.120)	(0.009)	(0.001)	(0.002)	(0.014)	(0.009)
Principal Hispanic	0.0095	0.0035**	0.0319**	0.2690*	0.0267	-0.0066**	Omitted	0.0764	0.0442
	(0.019)	(0.001)	(0.006)	(0.109)	(0.029)	(0.001)		(0.104)	(0.040)
Principal Other	0.0411	-0.0004	0.0158	-0.4200*	0.0283	-0.0041	-0.0002	-0.0425	-0.0034
	(0.033)	(0.001)	(0.022)	(0.202)	(0.025)	(0.005)	(0.005)	(0.047)	(0.035)
Controls	✔	✔	✔	✔	✔	✔	✔	✔	✔
Year and Grade FE	✔	✔	✔	✔	✔	✔	✔	--	--
School FE	✔	✔	✔	✔	✔	✔	✔	✔	✔
Cohort FE	--	--	--	--	--	--	--	✔	✔
Observations	2,442,339	2,442,339	2,442,339	2,213,096	2,436,662	2,119,693	1,012,669	403,257	214,514
R-Squared	0.144	0.029	0.388	0.061	0.283	0.017	0.050	0.065	0.047

Note. Models identical to those presented in Tables 2 and 4 apart from inclusion of principal covariates. Coefficients on student-level and school-level control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

Table A.7. Benchmarks for PTR Effect Sizes

Outcome		Magnitude change in PTR: 100 percentage point	10th to 90th percentile	1 standard deviation	Average turnover change
All students					
Minor offense	Effect size	-0.103	-0.021	-0.017	-0.014
	Percent change	-43.0%	-8.6%	-6.9%	-6.0%
Any removal	Effect size	0.074	0.015	0.012	0.010
	Percent change	67.0%	13.4%	10.7%	9.4%
H.S. graduation	Effect size	-0.106	-0.021	-0.017	-0.015
	Percent change	-14.9%	-3.0%	-2.4%	-2.1%
Juvenile comp	Effect size	0.013	0.003	0.002	0.002
	Percent change	63.0%	12.6%	10.1%	8.8%
No record					
H.S. graduation	Effect size	-0.105	-0.021	-0.017	-0.015
	Percent change	-13.6%	-2.7%	-2.2%	-1.9%
Conviction	Effect size	-0.039	-0.008	-0.006	-0.005
	Percent change	-385.0%	-77.0%	-61.6%	-53.9%
Minor record					
Any removal	Effect size	0.377	0.075	0.060	0.053
	Percent change	121.5%	24.3%	19.4%	17.0%
Days absent	Effect size	2.460	0.492	0.394	0.344
	Percent change	28.8%	5.8%	4.6%	4.0%
Test scores	Effect size	-0.104	-0.021	-0.017	-0.015
	Percent change	-19.6%	-3.9%	-3.1%	-2.7%
Grade retention	Effect size	0.023	0.005	0.004	0.003
	Percent change	112.5%	22.5%	18.0%	15.8%
H.S. graduation	Effect size	-0.138	-0.028	-0.022	-0.019
	Percent change	-22.6%	-4.5%	-3.6%	-3.2%
Serious record					
Juvenile comp	Effect size	0.356	0.071	0.057	0.050
	Percent change	79.1%	15.8%	12.7%	11.1%

Note. This table compares effect sizes based on different magnitude changes in principal PTR. It only does so for outcomes in the main analysis (Tables 2-4) with statistically significant effects at the 95% confidence level. Column 1 corresponds to original results. Column 2 shows effects of moving from 10th percentile PTR (21% removal) to 90th percentile PTR (41% removal). Column 3 shows effects of a 1 standard deviation change in across-school principal PTR. Column four shows effects of the average magnitude change in PTR during principal turnover. Percent change is the percent change in the outcome from the outcome mean.

Table A.8. Student and School Predictors of Principal PTR and Bias

	Propensity to Remove		Removal Bias	
Variables	Across school	Within school	Across school	Within school
Student measures				
White (omitted)	--	--	--	--
Black	0.0007	-0.0002	0.0008*	0.0001
	(0.001)	(0.000)	(0.000)	(0.000)
Hispanic	-0.0011	-0.0004	-0.0000	0.0000
	(0.001)	(0.000)	(0.000)	(0.000)
Other race/ethnicity	0.0023+	-0.0003	0.0005*	0.0000
	(0.001)	(0.000)	(0.000)	(0.000)
Female	-0.0004	0.0000	-0.0001	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Limited English	0.0018	0.0006	0.0020*	-0.0001
	(0.002)	(0.001)	(0.001)	(0.000)
Econ Disadvantage	0.0020	-0.0001	-0.0023**	-0.0002
	(0.001)	(0.000)	(0.001)	(0.000)
School measures				
FTE teachers	-0.0015*	0.0002	0.0003	-0.0005
	(0.001)	(0.001)	(0.000)	(0.000)
Percent free lunch	0.0005**	0.0001	-0.0002**	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Percent reduced price	-0.0010*	0.0004	-0.0007**	-0.0002+
	(0.000)	(0.000)	(0.000)	(0.000)
Student enrollment	0.0001*	0.0000	-0.0000	0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Percent students Black	0.0035**	-0.0001	0.0002+	0.0004*
	(0.000)	(0.000)	(0.000)	(0.000)
Pct students Hispanic	0.0000	0.0006	0.0000	-0.0004+
	(0.001)	(0.001)	(0.000)	(0.000)
Pct students other race	0.0002	-0.0000	-0.0007+	-0.0005
	(0.001)	(0.001)	(0.000)	(0.000)
Rural (omitted)	--		--	
Urban	-0.0106		-0.0050	
	(0.013)		(0.004)	
Suburban	-0.0127		-0.0043	
	(0.011)		(0.004)	
Town	0.0001		-0.0076	
	(0.014)		(0.005)	

	Propensity to Remove		Removal Bias	
Variables	Across school	Within school	Across school	Within school
Lowest grade offered	-0.0113**		-0.0015	
	(0.003)		(0.001)	
Highest grade offered	0.0157*		-0.0012	
	(0.007)		(0.002)	
Title 1 eligible	0.0087		-0.0050	
	(0.010)		(0.004)	
Magnet school	0.0203		0.0249*	
	(0.021)		(0.011)	
Year FE	✔	✔	✔	✔
Grade FE	✔	✔	✔	✔
School FE	--	✔	--	✔
Observations	2,531,399	2,548,301	1,752,649	1,761,109
R-squared	0.248	0.425	0.064	0.504

** p<0.01, * p<0.05, + p<0.1

Note. Robust standard errors in parentheses, clustered by school. FTE = full-time equivalent. The within-school PTR and removal bias dependent variables are our preferred estimates. The across-school variables are those estimated without school fixed effects in the first stage.

Table A.9. Effects of PTR on Student Outcomes, Alternative Model Specifications

Variables	Minor Offense	Serious Offense	Any Removal	Days Absent	Test Scores (SD)	Grade Retention	Juvenile Complaint	High School Graduation	Adult Conviction
Reporting Control									
Principal PTR	-0.1010**	-0.0013	0.0730**	0.3259	-0.0241	0.0057	0.0125**	-0.1010*	-0.0261
	(0.026)	(0.002)	(0.014)	(0.329)	(0.034)	(0.004)	(0.005)	(0.042)	(0.023)
Subjective Reporting	0.0533**	-0.0034**	-0.0177*	-0.0041	-0.0078	-0.0019	-0.0045*	0.1083**	0.0193+
	(0.018)	(0.001)	(0.007)	(0.145)	(0.018)	(0.001)	(0.002)	(0.025)	(0.011)
Observations	2,548,301	2,548,301	2,548,301	2,318,308	2,542,365	2,219,258	1,109,200	423,208	228,904
Lagged DV Control									
Principal PTR	-0.0115	-0.0026	0.0577**	0.0426	-0.0236	0.0076+	0.0132+	--	--
	(0.027)	(0.002)	(0.013)	(0.360)	(0.032)	(0.004)	(0.008)		
Lagged Outcome	0.4238**	0.0453**	0.1658**	0.5711**	0.9873**	-0.0067**	0.1849**	--	--
	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.006)		
Observations	1,354,669	1,489,905	1,489,905	1,323,837	1,485,886	1,278,968	491,767		
Racial Bias Control									
Principal PTR	-0.0891*	-0.0046*	0.0613**	0.6948+	-0.0796+	0.0064	0.0095+	-0.0587	-0.0297
	(0.041)	(0.002)	(0.021)	(0.414)	(0.045)	(0.005)	(0.006)	(0.054)	(0.031)
PTR Removal Bias	-0.0318	-0.0020	0.0222	-0.3294	-0.0388	0.0021	-0.0055	0.0641	-0.0080
	(0.067)	(0.004)	(0.028)	(0.884)	(0.089)	(0.007)	(0.011)	(0.125)	(0.047)
Observations	1,761,109	1,761,109	1,761,109	1,607,988	1,757,339	1,538,647	777,881	306,613	161,475
Controls	✔	✔	✔	✔	✔	✔	✔	✔	✔
Year and Grade FE	✔	✔	✔	✔	✔	✔	✔	--	--
School FE	✔	✔	✔	✔	✔	✔	✔	✔	✔
Cohort FE	--	--	--	--	--	--	--	✔	✔

Note. Models identical to those presented in Tables 2 and 4 apart from inclusion of specific controls. Coefficients on most control variables not shown: student race/ethnicity indicators, student gender, student limited English proficiency, student economic disadvantage, number of full-time-equivalent teachers, school enrollment, percent of students high economic disadvantage, percent of students moderate economic disadvantage, percent of students by race/ethnicity.

APPENDIX B. ESTIMATION OF PROPENSITY TO REMOVE (PTR)

This appendix provides detailed information on the specific steps for estimating PTR.

Step 1: Adjust for student and school factors. In database of all disciplinary referrals for all students grades 6-8 from 2008-2016, we perform the following regression:

$$r_{ijkpst} = \beta_0 + H_{it}\beta_1 + \gamma_s + \theta_g + \delta_t + \epsilon_{ijkpst}$$

where r_{ijkpst} is an indicator variable of whether or not removal (out-of-school suspension, expulsion, or transfer to alternative school) was assigned for disciplinary event *j* of offense type *k* involving student *i* under principal *p* in school *s*, year *t*; H_{it} is a vector containing one count variable of number of offenses committed by student *i* in year *t* and one count variable of number of offenses committed by student *i* in years *t-3* through *t-1*; γ_s are school fixed effects; θ_g are grade fixed effects; and δ_t are year fixed effects.

From this equation, we capture residuals:

$$e_{ijkpst} = r_{ijkpst} - \widehat{r_{ijkpst}}$$

Step 2: Capture principal fixed effects for each offense type. Then we loop through each offense type *k* and year *T* to estimate offense-type-specific principal removal propensities, averaging across all years the principal is observed except for year *T*. (This constructs the "leave-year-out" measure). For each offense type *k* and year *T* we therefore regress residuals from the equation above on the full set of principal fixed effects without a constant:

$$e_{ijkpst} = \delta_{pkT} + \epsilon_{ijkpst} \text{ for all } t \neq T$$

where e_{ijpst} is the residualized removal indicator, and δ_{pkT} are principal fixed effects specific to offense type *k* and year *T*. We capture the coefficient $\widehat{\delta_{pkT}}$ and standard error $\widehat{\sigma_{pkT}}$ on each principal fixed effect in order to construct the final PTR measure.

Step 3. Shrink estimates using empirical Bayes adjustment. We follow the shrinkage approach of Branch, Hanushek, and Rivkin (2012, p. 12) which itself is based on prior research (e.g. Kane and Staiger 2002, Jacob and Lefgren 2005, Morris 1983). The aim of the shrinkage adjustment is to reduce variation in PTR that is due to noise, and weight more heavily estimates that are generated with greater precision:

$$\widehat{\delta_{pkT}}(EB) = \left(1 - \frac{\widehat{\sigma_{pkT}}^2}{\widehat{\sigma_{pkT}}^2 + Var(\widehat{\delta_{kT}})}\right)\widehat{\delta_{pkT}} + \frac{\widehat{\sigma_{pkT}}^2}{\widehat{\sigma_{pkT}}^2 + Var(\widehat{\delta_{kT}})}\overline{\delta_{kT}}$$

In the above equation, $\widehat{\delta_{pkT}}$ is the estimated fixed effect coefficient, $\widehat{\sigma_{pkT}}^2$ is the estimated standard error on the principal fixed effect squared, $Var(\widehat{\delta_{kT}})$ is the estimated variance of the set of principal fixed effects, and δ_{kT} is the mean principal fixed effect for offense type k.

Step 4. Take weighted average across all offense types of principal estimates. Our aim for taking the weighted average across offense types is to create a uniform metric by which principals can be evaluated in the same way regardless of the frequency or severity of reported disciplinary

events they face at their school. The first thing we need for this calculation is a count of the number of disciplinary events by type *k* and year *T*.

$$n_{kT} = \sum_{t \neq T} n_{kt}; \quad n_T = \sum_k n_{kT}$$

The first value above n_{kT} sums all offenses of type *k* across all years except the current year *T*. The second value n_T is the sum of all offenses in all years except the current year *T*, regardless of type. Therefore, the weight given to each offense type equals n_{kT}/n_T.

$$\widehat{PTR}_{pT}(EB) = \sum_{k=1}^{35} \frac{n_{kT}}{n_T} \widehat{\delta_{pkT}}(EB)$$

Estimating Principal PTR Bias

In order to estimate differences in principal PTR between white and black students in the same school, we follow the same steps 1 through 4 above, with modifications. First, we restrict the sample of disciplinary referrals to only school-years in which the percent of disciplinary referrals involving white students is at least 10%, and the percent of disciplinary referrals involving black students is at least 10%. Second, in step 2 we adapt the principal fixed effects estimation to:

$$e_{ijkpst} = (\delta_{pkT} \times white_i) + (\delta_{pkT} \times black_i) + \epsilon_{ijkpst} \text{ for all } t \neq T$$

where e_{ijkpst} are again residuals from step 1 (this time from the restricted sample), $\delta_{pkT} \times white_i$ is an interaction term of the principal fixed effect with an indicator of whether the student is white, and $\delta_{pkT} \times black_i$ is an interaction term of the principal fixed effect with an indicator of whether the student is black. This regression therefore estimates two separate PTRs for each principal, offense type, and year: one for their disciplinary interactions with white students and one for their disciplinary interactions with black students. We capture those as $\widehat{\delta^w_{pkT}}$ and $\widehat{\delta^b_{pkT}}$.

As above in step 3, we apply EB shrinkage to each term to generate $\widehat{\delta^w_{pkT}}(EB)$ and $\widehat{\delta^b_{pkT}}(EB)$. We then subtract the two to determine the degree of removal bias for each principal, year, and offense type *k*: $\widehat{bias_{pKT}} = \widehat{\delta^b_{pkT}}(EB) - \widehat{\delta^w_{pkT}}(EB)$. Finally, we repeat step 4 above to take a weighted average of all bias terms across offense types.

Alternative Estimation Strategies

We have replicated the above process with several modifications to determine the sensitivity of our main results to alternative PTR measures. These modifications include:

- Using random effects instead of fixed effects (introduces normality assumption)
- Estimating PTR across-schools instead of within-schools (i.e. excluding school FE)
- Not adjusting PTR using empirical Bayes shrinkage
- Allowing current year offenses to inform PTR (not leave-year-out)
- Including in-school suspension as a form of removal