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

Lucy C. Sorensen and Shawn Bushway

University at Albany, SUNY

Elizabeth J. Gifford

Duke University

DRAFT FOR AEFP CONFERENCE, MARCH 2019

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.

ABSTRACT

Nationwide, school principals struggle to maintain safe school environments while also taking care not to punish students too severely for minor disciplinary infractions. The current study examines the effects of having a principal that is more severe (or more lenient) in assigning out-of-school removal for the same type of disciplinary offense, on student outcomes. The analysis uses detailed data on disciplinary referrals matched to student and offender records from the North Carolina middle schools and the Department of Public Safety from 2008 to 2016. Our identification relies on within-school changes in disciplinary severity from principal turnover events. We uncover in aggregate mostly neutral effects of principal disciplinary severity on a series of student outcomes. There are, however, large detrimental impacts on particular student subgroups, such as students who have committed minor disciplinary infractions, and racial minority students in schools with disproportionately severe principals. These adverse impacts emerge across both short-term measures of academic behaviors and also long-term educational attainment and criminal conviction.

INTRODUCTION

On the first weekend in March 1989, *Lean on Me*, a movie starring Morgan Freeman as get-tough principal Joe Clark in Paterson, NJ, won the weekend box office race, a rare achievement for a bio-pic about a high school principal.[1] In both the movie and in real life, "Crazy" Joe Clark's defining act during his first week as principal was to expel over 300 students for fighting, vandalism, drug possession, and abusing teachers. In retrospect, the notoriety of Clark was a harbinger of a new "get tough" model of school discipline that expanded rapidly in the 1990's. Perhaps the most prominent example of these get-tough policies was the "zero-tolerance" policy that, in effect, sought to expand strict administrator responses by mandate. School-based zero tolerance policies, which first appeared in the same year as the movie (1989) required mandatory disciplinary actions – usually suspension, expulsion or arrest – in response to certain types of misconduct (Curran, 2019, Skiba & Peterson, 1999, Skiba & Knesting, 2001).[2] By 1997, 94% of all U.S. school principals reported that their schools had zero-tolerance policies for firearms, 88% for drugs, and 79% for violence or tobacco (U.S. Department of Education 1998).

There is now an ongoing backlash against get-tough policies, including zero tolerance policies, on the grounds that these harsh practices disproportionately affect or even target students from underrepresented minority groups (Curran, 2016; Hoffman, 2014; Kinsler, 2011; Losen & Martinez, 2013; Skiba et al, 2014). Moreover, opponents of school suspensions and

[1] According to Box Office Mojo, available at: https://www.boxofficemojo.com/weekend/chart/?yr=1989&wknd=09&p=.htm

[2] This effort, motivated by the belief that clear punishments would deter student misbehavior was accelerated by the passage of the federal Gun-Free Schools Act in 1994, which required zero tolerance policies resulting in a one year expulsion for having a weapon in school for schools requesting federal funding. For articles supporting the use of severe discipline to improve school safety, see Petrilli, 2014; Eden, 2017. The term zero-tolerance first appeared in 1986 in a DOJ program designed to impound boats carrying any amount of drugs (Skiba & Knesting, 2001).

expulsions believe that harsh discipline, particularly for minor offenses, harms educational outcomes and/or increases the likelihood of later interactions with the criminal justice system for students targeted by these actions.[3] As a result, many states, including North Carolina (Duke Law News 2011), have banned the use of zero tolerance policies, moving disciplinary discretion back into the hands of principals (U.S. Department of Education 2014).[4] Unfortunately, there is still little empirical evidence about causal consequences of these decisions to guide school principals and other administrators in their choices (Steinberg & Lacoe, 2017).

In part, this lack of empirical evidence arises because it is difficult to statistically decouple the consequences of the administrator's choice from the underlying student behavior that resulted in a disciplinary referral. The authors of the most recent meta-analysis on the consequences of suspensions concluded that "we cannot definitely conclude that suspensions caused the (negative) outcomes. It could be that students who are low achievers, for example, are more likely to misbehave in the classroom because of frustration and thus receive suspensions. … (M)ost studies did not control for pre-suspension academic or behavioral difficulties that might influence the outcome variables" (Noltemeyer, Ward, and McLoughlin, 2015 P.235). Not surprisingly, a recent paper which used more stringent controls of prior student behavior, including student fixed effects and a policy-based instrumental variable (Lacoe and Steinberg 2018), found more modest estimates of the connection between suspensions and student achievement than those reported by Noltemeyer, Ward and McLoughlin (2015). This

[3] For a review of this perspective and the relevant literature, see Noltemeyer, Ward and Mcloughlin 2015.
[4] This movement received considerable support by a wide range of guidelines issued in 2014 by the Obama administration which were widely interpreted as discouraging the continued use of zero tolerance policies outside the narrow confines of the requirements imposed by the Gun-Free Schools Act. A recent study by Curran (2019) using a different methodology than the earlier U.S. Department of Education Survey (1998) found a much lower rate of zero-tolerance and mandatory expulsion rules at both the district and state level.

problem of selection has also plagued research on racial disproportionality in discipline, where it is often difficult to disentangle racial differences in student behavior or school contexts from racial differences in administrative response (Skiba et al., 2011).

In our view, one solution to this selection problem is to focus on principals and their behavior, rather than the specific sanction. This solution arises from our observation that principal behavior, rather than suspension or expulsions, is the relevant policy lever. School principals in the U.S. have primary responsibility for creating a school's disciplinary climate, a responsibility that includes reacting to undesirable student conduct using approaches like speaking to the student/parents, suspending the student from the classroom, expelling the student from school, or even reporting the student to the police (DeMatthews, Carey, Olivarez, & Saeedi, 2017; Gottfredson & Gottfredson, 2001; Skiba et al, 2014). Empirically, a recent survey of principals showed that individual principal perspectives on discipline (in effect, the principal's *severity*) were strongly correlated with the use of school suspensions in those principals' schools, even net of individual characteristics like type of offense (Skiba et al., 2014).

In this paper, we study the effect of having a middle-school principal with a higher relative likelihood of suspending a student on students' academic and long-term outcomes. We first measure middle school principals' decisions when assigning consequences for specific types of offenses and types of students. We then use our measure of principal behavior to test whether a given principal's propensity – on the margin – to remove students from school (through out-of-school suspension, expulsion, or transfer) has any causal lasting effects on those students who were removed. We also test whether principal stringency has spillover effects on students not directly involved in the school's disciplinary system.

Specifically, we use detailed administrative data on individual disciplinary referrals and consequences from middle schools in the North Carolina public school system from 2008 to 2016. We focus on middle school grades because for many students their first disciplinary incident will occur sometime between grades 6 through 8, and behaviors that typically emerge first during middle school – including chronic absenteeism, failed courses, and suspensions – provide a reliable warning sign for eventually dropping out of school (Balfanz, Herzog, & Mac Iver, 2007). We create a novel measure of each principal's relative tendency to use exclusionary discipline in a given circumstance. For each disciplinary referral, we have information on specific type of offense(s), assigned consequence(s), student's disciplinary history, and student's grade level. By regressing (within the full sample of disciplinary reports) an indicator of student removal on student offense history information and a principal-specific random coefficient for each of 35 offense types, we create a conditional measure of principal severity that may vary by offense type (adjusted via Empirical Bayes shrinkage). Using a weighted average of these principal coefficients, each principal is assigned a single disciplinary severity score. As expected, principals varied substantially in the likelihood to suspend students for the same offense. For example, the principal at the 90th percentile has a 49.3 percent chance of choosing exclusionary discipline as the response to the median incident. The principal at the 10th percentile, in contrast, has a 12.8 percent chance of using exclusionary discipline for the same incident.

To understand how principals' decisions affect short and long-term student outcomes, we incorporate our measure of principal severity into a longitudinal panel of student-level data. These education records include information on student's reading and math achievement, absences, grade retention, and high school graduation/dropout. By linking the data to information collected by the North Carolina Department of Public Safety, we also examine the effect of

principal severity on conviction/incarceration in the adult criminal justice system. Our preferred model for short-term outcomes includes school fixed effects, year and grade fixed effects, school time-varying controls, and student controls. Our long-term outcomes model uses individual level data with middle school and cohort fixed effects. These models are identified from the variation of principal severity that occurs within schools over time due to principal turnover. We run a series of robustness tests to ascertain whether the effects of principal severity on students are due to this specific form of principal disciplinary decision-making or other qualities of the principal's leadership.

We find that principal severity reduces the number of 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, we uncover mostly neutral effects of principal severity on short-term academic outcomes, except perhaps a slight increase in absenteeism. For certain subgroups of students, however, our findings imply that principal severity has large and lasting detrimental effects. For example, replacing a principal who never removes students from school with a principal who always removes students would increase absences by 4.6 days (from a mean of 6.2 days) and increase likelihood of grade retention by 2.3 percentage points (from a mean of 1 percent). These adverse consequences that occur during middle school endure to young adulthood, as students assigned to the most severe middle school principals were 32.2% less likely to graduate high school and 12.9% more likely to be convicted of a crime than students assigned to the least severe principals. Finally, the level of racial bias a principal exhibits – defined as the difference in propensity to remove a black student for the same offense and same offense history as a white

student – predicts enormous academic losses for black and Hispanic students enrolled at that school.

Variation in Suspension Rates and the Role of School Principals in Student Discipline

As documented by the Office of Civil Rights, the percent of students who experienced a school suspension during a given school year nearly doubled from 1973 to 2006 from a rate of 3.7% to 6.9% (Losen & Skiba, 2010). During this same time, racial and ethnic disparities in suspensions grew tremendously resulting in Black students being suspended at three times the rate of White students by 2006 (Losen & Skiba, 2010). Moreover, rates of suspension vary enormously across schools, districts, and states (U.S. Government Accountability Office, 2018). For example, during the 2015-16 school year, North Carolina had a rate of 14.5% relative to Illinois which had a rate of 3.8% (Council of State Governments Justice Center, 2017).

This variation raises important issues for educators and policy makers. First, racial and ethnic disparities in suspension raise concerns about equity of access to education and the possibility of unjust loss of instructional time. Second, variation across schools and states draws into question how system-level actors and policies shape students' risk for being suspended (Kinsler, 2011). Finally, the existence of such variation across the U.S. could be viewed as a signal that there is uncertainty regarding what the most effective strategy for addressing student misconduct is, and what role suspension/expulsion plays in addressing student misconduct. Empirical evidence regarding the causal effect of disciplinary severity on both the students who are suspended and the classmates/schoolmates of students who are suspended would help to inform principals and school staff regarding what the "best" decision should be.

For most schools, principals are the local administrator with most direct responsibility for making disciplinary decisions. According to a 2014 national survey of superintendents, the ability to suspend students, in nearly all school districts, is limited to principals (95% of surveyed school districts), building administrative staff (25%), or central office staff such as superintendents (57%) (Pudelski, 2014). To decide whether or not to suspend a student, the principal and/or her staff gather information from teachers and other observers to understand the incident, and then guided by policy, determines the severity of the misconduct as well as what consequences to assign.

Qualitative studies highlight stark differences in how principals approach disciplinary decisions. For example, in a study of four middle schools, principals from schools with low suspension rates viewed the disciplinary policy as a guideline and described modifying their response based on the circumstances surrounding a particular event. In contrast, principals in schools with high suspension rates reported more rigidly adhering to policies (Mukaria, 2002). Some principals describe rigid rule adherence as a strategy for remaining neutral and consistent when addressing student misconduct (DeMatthews, Carey, Olivarez, & Saeedi, 2017). Other principals were more flexible in their approach and factored into their decision information about the child, the events surrounding the incident, and viewed suspension as one of several tools for addressing a particular disciplinary event (Mukaria, 2002).

These principal attitudes appear to translate directly in principal behavior. Russell Skiba and his colleagues (Skiba et al., 2014) found that when a student's principal favored school exclusion rather than prevention, the student had a greater odds of receiving an out-of-school suspension or expulsion (rather than an in-school-suspension). Heilbrun, Cornell and Lovegrove (2015) found that principal attitudes towards zero-tolerance policies were correlated with school-

level suspension rates. When regressions were estimated separately for black students and white students, principal endorsement of zero tolerance policies explained nearly 20 percent of the variance in suspension rates for black students relative to only four percent of the variance in white suspension rates. These studies appear to show empirically what most experts believe to be true -- principals drive disciplinary practices.

School Climate, Student Behavior, and Suspensions

The principal can translate her beliefs into actions through both formal and informal means. Formally, she chooses how to respond to incidents of student misbehavior; informally, she may set guidelines, cultural norms, and the overall disciplinary climate in the school. A related literature has examined how school disciplinary climate more generally relates to student behavior and outcomes (e.g., Gottfredson, Gottfredson, Payne, & Gottfredson, 2005). One classification for school disciplinary climate is based on student perceptions of "support" from their teachers and school staff for addressing their individual needs and the amount of "structure", or consistent enforcement of fair expectations (Lau et al., 2018). Schools where the climate is characterized as high in both support and structure, tend to outperform other schools in terms of lower rates of suspension (Heilbrun, Cornell, & Konold, 2018; Huang & Cornell, 2018), reduced teasing and bullying (Gregory et al., 2010), and reduced physical and verbal aggression with peers and teachers (Berg & Cornell, 2015). However, for classmates who were not suspended, more punitive contexts, measured by higher school-level suspension rates, have been associated with worse reading and math achievement (Perry & Morris, 2014). Marchbanks, Peguero, Varela, Blake, and Eason (2018) derived a measure of school leniency by comparing the school's actual suspension rate with the school's predicted suspension rate. Students in

schools with either particularly stringent or particularly lenient practices, relative to what might be expected, had higher chances of being referred to the juvenile justice system.

The Effects of Disciplinary Removals on Student Outcomes

Short-Term Academic Outcomes

Correlational research that relies on either cross-school variation in suspension rates (e.g., Rausch & Skiba, 2005) or cross-sectional variation in student behavior and outcomes (e.g., Kinsler, 2013) suggests that students who are suspended are at higher risk for a host of negative outcomes including worse attendance, lower reading achievement scores, lower grade-point-average, higher dropout rates, and increased contact with the juvenile justice system (for a review see Welsh & Little, 2018). Yet, causally determining whether or not suspensions and other exclusionary forms of discipline cause students who are suspended to have lower academic achievement is challenging. Exclusionary forms of discipline are typically given in response to student misbehavior. Underlying causes for student misbehavior may be driven by a host of family, peer, and life circumstances (e.g., family conflict, housing or food insecurity, peer conflict, etc.) that are not systematically documented and therefore unobservable to researchers.

To more carefully identify the effect of being suspended on school outcomes, recent studies have used student-level panel data to compare how a given student's performance varies in years in which s/he was suspended relative to years in which s/he was not suspended (Chu & Ready, 2018; Hwang, 2018, Lacoe & Steinberg, 2018). Such an approach was used by Hwang (2018) who employed a student-fixed effects approach with quarterly math and English Language Art test data from California. Results from this study revealed that the within-student effects of out-of-school suspension were associated with a decrease in student's English

Language Arts tests but not associated with a change in math scores. Similarly, in an analysis for public school students from New York City, Chu and Ready (2018) followed a cohort of first time 9th graders over a seven-year period. Within-student models revealed that suspensions were associated with modest decreases in the percent of math and English credits passed, and increases in days absent and dropping out of school the semester following the suspension. Lacoe and Steinberg (2018) found similar results using the same fixed effect approach in a sample of 3rd through 12th graders in Philadelphia. However, when they limited their models to suspensions resulting from minor classroom misconduct – failure to follow classroom rules and profane language – the effect of suspension was no longer statistically significant.

While the results of these studies suggest that students who are suspended may fare worse than they otherwise would have on academic outcomes, these studies still do not account for time varying co-occurring factors that could both affect a student's behavior in school and their ability to perform well on tests (e.g., parental divorce, family member illness or death, etc.). An instrumental variables approach that uses variation in suspension rates stemming from policy changes that focused on reducing suspensions offers a chance to control for these factors (Lacoe & Steinberg, 2018). The one instrumental variables study we found had some evidence that a policy change that reduced suspensions for minor offenses slightly improved math and reading scores for the affected students (Lacoe & Steinberg, 2018).

Kinsler (2013) conceptualized principals' disciplinary decisions in the context of optimizing the schooling environment for all students. Through this lens, the principal established a disciplinary strategy which weighed competing demands of maintaining order, minimizing loss of instructional time for students who misbehaved, and minimizing disruptions for classmates of students who misbehaved. Counter to the narrative that school suspensions

have enduring harmful effects on students who experience suspensions (Wald & Losen, 2003), results from this study suggested that schools with stricter suspension policies had fewer disciplinary infractions and higher levels of student achievement. Because these results were based on a single year of cross-sectional data and strong assumptions regarding the incoming students' knowledge of the principal practices, these results do not address the causal association between suspension and achievement. However, the contrasting findings highlight the need for rigorous investigations to inform principal decision making.

Young Adult Outcomes in the Criminal Justice System

Advocates, policymakers, and researchers commonly use the term "school-to-prison-pipeline" to capture associations between education policies and practices and long-term student outcomes in both the juvenile justice and adult criminal justice system (Skiba et. al,. 2014). Among the themes captured by the term is how school disciplinary practices, which disproportionately affect African American males, affect a student's long-term outcomes including imprisonment in adult correctional facilities. Despite the ubiquity of the term (Sykes et al. 2015), the actual evidence base for understanding whether and how specific school policies are indeed causally affecting student's probability of being involved in the criminal justice system is underdeveloped (Nicholson-Crotty et al., 2009, p.1006-1007).

To begin to build this evidence base, researchers have used large national longitudinal studies to document associations between being suspended in secondary school and one's risk of criminal justice involvement as a young adult. Findings from the National Longitudinal Survey of Adolescent to Adult Health (Add Health) and the National Longitudinal Survey of Youth (NLSY 97) have consistently found that relative to non-suspended peers, students who were suspended had higher levels of criminal justice involvement in young adulthood (Wolf &

Kupchik, 2017; Barnes & Motz, 2018; Mowen & Brent, 2016; Rosenbaum, 2018). For example, Wolf and Kupchik (2017) found that relative to their peers who had never been suspended, students who had been suspended by the first wave interview (grades 7-12), had a 31% greater odds of criminal activity and a 72% greater odds of incarceration (Wolf & Kupchik, 2017). Further analysis suggested that removing racial inequalities in school suspensions would reduce racial inequality in later arrest risk by as much as 16% (Barnes & Motz, 2018). Another investigation found that five years after an initial suspension, students who had been suspended were 40% more likely to have been arrested and 3.8 times more likely to have been convicted (Rosenbaum, 2018).

While findings from national longitudinal studies suggest that suspension may be a risk factor for long-term criminal involvement, they are designed to track individuals, rather than system-level actors. Determining the cumulative impacts of administrator decisions on the margin in assigning disciplinary consequences is necessary to move beyond associations of student suspension and student outcomes to identifying policy-relevant mechanisms. In this paper, we will study the impact of middle school principals in North Carolina on a variety of both short- and long-term outcomes, including involvement in the adult criminal justice system.

Disciplinary System in North Carolina Middle Schools

In North Carolina, state policy dictates that principals have a great deal of control over suspension decisions (North Carolina State Board of Education and Department of Public Instruction, 2008; 2018). For minor incidents, principals independently determine 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 for longer than 10 days and up to 365 days. For students who have received a long-term suspension, school districts may allow these students to attend an alternative learning program or alternative school. In recent years, state and district efforts to reduce long-term suspensions have incentivized superintendents to place students in alternative learning programs so as to avoid a long-term out-of-school suspension (NC DPI, 2000). Less commonly, a superintendent may recommend to the school board that a student, aged 14 years or older, be expelled from school. The local school board would then conduct a hearing to determine if the student's presence in school presented "a clear threat to the safety of other students or school staff" (North Carolina General Statutes, 2013). [5]

Prior to this determination of consequence for a given disciplinary incident, a teacher or some other actor must refer the student incident to a school administrator. To show up in our data, the incident and consequence must then be officially reported through a specific administrative data reporting system. State and federal statutes and State Board of Education policies mandate that certain types of offenses require reporting: (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[6]; and (3) any act on a pre-

[5] When students are recommended for a long-term suspension or expulsion, they must be offered a hearing (Wettach, 2011). Hearings aim to determine if the student's presence in school indicates a "clear threat to the safety of other students or school staff". Hearing procedures vary across local school boards and mimic several features of criminal court proceedings. Students have due process protections throughout these proceedings, including the right to be represented by an attorney, or, at the option of the district, a non-attorney advocate; the right to attend the hearing with his/her parents; the right to review evidence that will be presented during the hearing including audio or video recording; the right to question witnesses presenting evidence against the student, the right to present evidence on his/her own behalf; the right to have a recording made of the hearing and to make his/her own recording of the hearing and the right to have a written decision based on substantial evidence produced at the hearing. Initial decisions may be appealed to the local school board who must decide with 30 days the status of the appeal. Decisions made by the local school board are subject to judicial review.

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

determined list of "Reportable Offenses" (NC DPI, 2019) These 12 more serious reportable offenses are all listed with the RO code in Table 1, along with their frequencies and associated probabilities of leading to student removal (OSS, transfer, or expulsion). North Carolina classifies the majority of disciplinary incidents, however, as "Unacceptable Behaviors" (UB), which typically occur with much greater frequency and lower probability of a removal consequence (see Table 1). These include offenses such as skipping school, disrespect of faculty/staff, inappropriate language, dress code violation, and fighting. Because of these varying reporting requirements and norms for different offense types and offense consequences, our approach attempts to isolate principal decision-making on consequence assignment from decision-making on referring and reporting.

Policy Changes and Trends

The effects of zero tolerance policies were felt in North Carolina. Between 1999-2000 and the following school year, there was a 22% increase in the number of long-term suspensions in North Carolina and a 71% increase in the percent of students who were expelled (*c.f.,* Goodman, 2005). Consistent with national trends, North Carolina suspensions were marked by large racial and ethnic disparities in the use of suspension which prompted civil law suits against school systems. Notably, in *King v. Beaufort County Education Board,* the plaintiff argued that

her right to attend public school was being violated.[78] This lawsuit emerged from a minor fistfight that occurred in January 2008 involving two female students, neither of whom had a weapon or were seriously injured (Sullivan, 2011). Both students were suspended for the remainder of the year and denied access to alternative schooling. In the culmination of the ensuing law suit, the North Carolina Supreme Court decided that students "have a constitutional right to continued educational services during a long-term suspension unless the school board can establish that the board has an important or significant reason for refusing to provide some type of alternative education (Wettach, 2011).

In 2011, North Carolina was among five states to pass laws aimed at reducing out-of-school suspension or expulsion (The Council of State Governments Justice Center, 2017). The change was specifically designed to allow local administrators the ability to implement discretion for minor forms of misconduct in the school, a protection against severe implementation of zero tolerance policies (Morton, 2014). Specifically, the N.C. General Assembly repealed Article 27 of Chapter 115 of the N.C. General Statutes, which allowed zero tolerance practices, replacing it with the following language:

[7] The North Carolina state constitution grants children the right to attend public school (N.C. Const., Article I, Section 15 and Article IX, Section 2), the interpretation of this has been found in the courts and has implications for school exclusionary practices. In 1994, parties (including the Leandro family) from five low-wealth rural counties sued the state on the basis that the quality of a child's education should be independent of the wealth of the child's family and community. The Plaintiffs argued that the state had not taken adequate steps to equalize school funding across the state. In the 1997 *Leandro v. State* decision, the Supreme Court of North Carolina held that a free public education is a fundamental right guaranteed under the North Carolina Constitution, and that attempts to limit it will be subject to strict scrutiny (Goodman, 2005).

[8] In September 2010, a complaint was filed with the Office of Civil Rights arguing that the disciplinary policies and practices of Wake County Public Schools were discriminatory--highlighting that Black first-time offenders were more likely to receive an out-of-school suspension than White first-time offenders. Losen (2011) obtained North Carolina statewide data through a Freedom of Information Act request, and demonstrated that Black first time offenders had substantially higher rates of receiving out-of-school suspensions for minor offenses such as cell phone use, dress code, disruptive behavior, and public displays of affection than White students.

> "In order to create and maintain a safe and orderly school environment conducive to learning, school officials and teachers need adequate tools to maintain good discipline in schools. However, the General Assembly also recognizes that removal of students from school, while sometimes necessary, can exacerbate behavioral problems, diminish academic achievement, and hasten school dropout. School discipline must balance these interests to provide a safe and productive learning environment, to continually teach students to respect themselves, others, and property, and to conduct themselves in a manner that fosters their own learning and the learning of those around them."

From 2011-12 to 2015-16, North Carolina reduced the number of suspensions by 16% and reduced the proportion of students receiving suspensions from 17.8% to 14.5% (Council of State Governments Justice Center, 2017). Despite this reduction in suspension rates, North Carolina maintains a high rate of suspensions relative to other states. Both before and after this law change, we observe that school principals behave very differently from one another in their disciplinary responses to the same types of incidents (See Figure 1).

DATA

To perform an analysis of how principal severity affects outcomes for students involved in the disciplinary system (and for those uninvolved), we use longitudinal administrative records of students in public middle schools from the North Carolina Education Research Data Center (NCERDC) between 2008 and 2016, matched with later educational outcomes and adult conviction records from the North Carolina Department of Public Safety.

The disciplinary database from NCERDC includes a record for each referral of a student disciplinary offense, regardless of whether that referral results in an eventual suspension. This makes our dataset unique from many other studies which have only been able to observe those disciplinary incidents that result in suspension. North Carolina schools make a key distinction between "reportable offenses," which are mandated by statute to report to the central administration, and "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, resulting in 35 total offense types.[9] These 35 offense types represent 92.1% of all disciplinary incidents of students in grades six through eight that occurred during the time period of study.[10] Although many types of disciplinary consequences are observed in the data, we focus on whether student offenses result in an instance of exclusionary discipline ("removal"), which may include: out-of-school suspension, expulsion, or transfer to an Alternative Learning Program or alternative school. Typically, these out-of-school suspensions and other forms of removal are considered more serious and more likely to have lasting impacts on the student and his or her peers (Skiba et al., 2014).

Table 1 presents the frequency of each type of disciplinary offense in middle schools across the 2008-2016 school years, and also the average likelihood of removal for each offense type. The most common disciplinary events include disruptive behavior, insubordination, inappropriate language or disrespect, aggressive behavior, and fighting. The least common include bomb threats, assault resulting in serious injury, possession of a firearm, sexual offense, and robbery without a weapon. Offenses vary widely in terms of their likelihood of resulting in

[9]We remove the most serious reportable offenses, which occur extremely rarely and 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.
[10] None of the excluded offense types alone represent more than 1% of all disciplinary events.

out-of-school suspension, transfer to an Alternative Learning Program, or expulsion. For instance, 89.1% of assaults with a weapon lead to some type of student removal, whereas only 2.8% of excessive tardiness complaints lead to removal. For most categories of student offense, though, it appears that schools and administrators vary widely in their likelihood of assigning the consequence of out-of-school suspension or other form of removal.

To these disciplinary referral events we link administrative records on school principals present at the time of the referrals, also made available through NCERDC. Of use, we can track principals as they transfer across schools or change positions within schools. Principal records contain information on gender, race/ethnicity, years of experience, and years of tenure. We also estimate within-school and across-school principal value-added scores in reading and math to provide (Branch, Hanushek, & Rivkin, 2012; Grissom, Kalogrides, & Loeb, 2015). Such information may illuminate how principal characteristics and qualifications relate to their observed disciplinary practices.

After estimating principal severity within the referrals dataset (using a method described in the following section), we link principal severity measures to a variety of short-term outcomes of all students enrolled in that principal's school, regardless of their level of involvement in the disciplinary system. These outcomes include: an indicator of any disciplinary incident, an indicator of any removal from a disciplinary incident, a count of number of days absent, an indicator of grade retention in the following year, and standardized reading and math test scores. Reading and math scores were each normalized to have a mean of zero and standard deviation of one by grade level and year. We also 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.[11]

This study builds upon prior evidence exploring the short-term impacts of disciplinary practices, but it also extends to exploring any lasting impacts on students into young adulthood. To test this empirically, we consider high school graduation (noted above) and also match to adult criminal offense and conviction records through 2016. We obtain this data using a web scraper to iteratively capture historical and current records from the North Carolina Department of Public Safety (DPS) publicly-available offender search feature of their website. In this way, we can determine whether each middle school student within our sample ever ends up appearing in the North Carolina criminal justice system. Because the courts consider all individuals as adults at the early age of 16 in North Carolina, our longitudinal data provides sufficient time for students in the original sample to grow into eligibility for adult criminal conviction. For simplicity, we use as the dependent measure an indicator of whether or not the student in question receives a criminal conviction in North Carolina by age 20.[12] Of the students that entered middle school in time to be observed all the way through age 20, 3.3% were successfully matched to the DPS offender database.[13]

[11] There are multiple possible counterfactuals to on-time high school graduation: delayed high school graduation or GED, graduation from private schools in North Carolina, graduation outside the state of North Carolina, or dropout. We cannot differentiate between these alternative outcomes.

[12] By restricting the age range in which individuals can show up in criminal justice records and restricting to only the state of North Carolina, we are going to underestimate the true conviction rates of middle school students in North Carolina public schools. This could be problematic if, for example, students with more severe principals were systematically more or less likely to move out of state than other students attending that same school. We have no reason to believe this is the case. The same general argument holds for our analysis of high school graduation rates.

[13] One limitation is that the nine-year panel of data only allows matching to the long-term outcomes of high school graduation and criminal conviction for a subset of the cohorts in the full sample. Specifically, we can use within-school, across-cohort variation from five cohorts for tracking through to on-time high school graduation and from three cohorts for tracking through to criminal conviction by age 20.

In addition to the eight outcome variables described above and our independent variable of estimated principal conditional severity, we also develop an expansive set of control variables. These include measures at the individual level such as race/ethnicity, gender, limited English proficiency, economic disadvantage (through the proxy of free or reduced-price lunch eligibility), and incidence of disciplinary offenses. They also include time-varying measures at the school level such as student enrollment, number of full-time equivalent teachers, percent of students by race/ethnicity, and percent of students by free and reduced-price lunch eligibility. In alternate specifications (See Appendix Tables A3 and A4), we also control for principal characteristics including race/ethnicity, gender, years of experience, and value-added scores in reading and math.

METHODS

Step 1. Estimating Principal Disciplinary Severity

This section describes how we measured a given middle school principal's use of exclusionary discipline. 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. These data included information about the nature of the disciplinary event and about the student's disciplinary history at the time of the event. 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 is a disciplinary event that is reported through the school and the school district to the state Department of Public Instruction. These disciplinary events lead to a decision, usually by the principal or someone who works for him/her, for some type of punishment, including removal from school. Our goal is to isolate the component of that disciplinary consequence that is due to principal decision-making – and not due to the nature of the student offense, to the student's prior record, or to district- or state-wide policies. We then port our principal specific measure of removal likelihood back to the student (rather than disciplinary event) level in step two of our analysis. This second step will allow us to evaluate 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.[14] 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, the 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).[15] These models have shed light on the potentially deleterious consequences of policies

[14] There have been some attempts to use value-added models to study the impact of principals on standard outcomes like test scores (Grissom, Kalogrides and Loeb, 2015). There is also a movement 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 et al.2014). Our approach is different because it is focused on disciplinary events as the unit of analysis, rather than the student.

[15] 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).

that allowed judge's considerable discretion in sentencing decisions. More optimistically, the models allow for the identification of judge behavior that appears to be "optimal".

Our sample for estimating principal disciplinary severity encompasses 35 high-frequency types of reported disciplinary referrals for all 6th to 8th grade students. Because we acknowledge that principals may respond more harshly to some offense types than to others, and that each principal faces a different disciplinary environment in terms of offense types and frequencies, we split the sample of referrals into 35 subsamples, one for each offense type recorded. Within these subsamples, our estimation objective becomes straightforward: How much more, or less, likely is 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, principal severity is estimated as the principal-specific random coefficient ($\widehat{\mu_{1p}}$) on each distinct disciplinary incident type (*k=1, . . . , 35)* in the sample of referrals from disciplinary event-level estimating equation of the general form:

$$r_{ijpt} = \beta_0 + H_{it}\beta_1 + G_{it}\beta_2 + \mu_{0p} + O_j\mu_{1p} + \epsilon_{ijpt} \quad (1)$$

$$\begin{bmatrix}\mu_{0p}\\ \mu_{1p}\end{bmatrix} \sim N(0, \Omega_\mu);\ \epsilon_{ijpt} \sim N(0, \sigma_\epsilon^2)$$

The dependent variable (r_{ipt}) is an indicator of school removal for the disciplinary event *j* (for student *i*) supervised by principal *p* in year *t*. The only control variables in equation 1 are a vector describing the student's prior history of disciplinary events (H_{it}) at time of event *t* and a vector of grade level fixed effects (G_{it}). For students with multiple disciplinary events, the set of disciplinary history variables changes over time as the student's history accrues. The principal random coefficient on each element of the vector of offense types (O_j) is a time constant measure of how "severe" the principal is for that particular offense type, and the error term (ϵ_{ijpt}) is an idiosyncratic disciplinary incident effect that varies across disciplinary events and over time. If

we observe a principal for four years, we will have, for that principal, up to 35 unique estimates of the removal likelihood for each disciplinary type *k* that is observed within those four years at that school.

From model 1, we capture the raw estimates of principal severity for each offense type *k* ($\hat{\mu}_{pk}$), which each equal the estimated principal random intercept $\hat{\mu}_{0p}$ plus the estimated offense-specific random coefficient $\hat{\mu}_{1pk}$. In an approach similar to that used by Grissom et al. (2015), we form empirical Bayes estimates of these principal-offense severity scores so as to place greater weight on scores that are estimated with better precision. Each principal severity estimate for offense type *k* ($\hat{\mu}_{pk}$) is the sum of the "true" principal severity (μ_{pk}) plus some measurement error, which we assume is not correlated with the unobserved explanatory variable: $\hat{\mu}_{pk} = \mu_{pk} + \upsilon_{pk}$. The empirical Bayes estimate is therefore a weighted average of the estimated removal likelihood for principal *p* offense type *k* and the average removal likelihood for offense type *k* across all principals in the population: $E(\mu_{pk}|\hat{\mu}_{pk}) = (1 - \lambda_{pk})\overline{\mu_{pk}} + \lambda_{pk}\hat{\mu}_{pk}$. The weight is a function of the precision of each principal estimate – the more precise the estimate of the principal removal likelihood for type *k*, the more weight we place on the estimated individual principal estimate, rather than the average principal estimate across the full sample. To be more precise, $\lambda_{pk} = \frac{(\widehat{\sigma_\mu})^2}{(\widehat{\sigma_\varepsilon})^2/n_{pk}+(\widehat{\sigma_\mu})^2}$, where $(\hat{\sigma}_\mu)^2$ equals the estimated variance of random coefficient $\hat{\mu}_{pk}$, $(\hat{\sigma}_\varepsilon)^2$ equals the estimated variance of the residual ϵ_{ijpt}, and n_{pk} equals the number of referrals of type *k* under principal *k*. We can view the weight as the proportion of the total variation in the principal effect that is attributed to true differences between principals.

Our measure of principal severity seeks to capture that discretionary propensity for exclusionary discipline for each unique principal. In the construction of our measure of principal

severity, we did not account for the school and student effects -- we will control for student and school effects in the second stage models focused on student-level outcomes. As a result, we urge caution in using results of this model to judge a specific principal's use of discretion.[16]

We now have a set of estimated and shrunk removal likelihood scores for each principal for each disciplinary type *k*. There are 35 different disciplinary offense types in our data, ranging from disruptive behavior (n=631,231 events) to robbery with no weapon (n=64). The median category has 21,728 events. This gives us 35 different disciplinary removal scores for each principal. We want to create a single measure that reflects an individual principal's removal likelihood for a randomly drawn disciplinary event. At one extreme, a very severe principal suspends or expels every student for every offense (severity = 1). At the other extreme, a very lenient principal never suspends or expels a single student for any offense (severity = 0). To create one universal measure of removal probability for each principal, we need to combine the estimated (and shrunk) measures across the 35 offense types.

For the offense types, we chose to weight each principal score by the sample proportion of disciplinary events in each type. Please note that this is the *sample* proportion of disciplinary events, not the proportion of events faced by principal *p* in each offense type. As a result, each principal will face the same weight for each offense type, regardless of the relative frequency or rarity of that offense type at that particular school. This weighting creates a final removal likelihood that does not punish principals who face more serious offenses on average. In other

[16] We have compared our baseline specification (1.1) with value-added models for both principals and teachers (For teacher models, see Koedel, Mihaly, & Rockoff, 2015). The details of this specification varies widely across papers and applications. In teacher value added models or principal value added models, it is not unusual for the school or student fixed effects to be included, so that the teacher fixed effect captures only the change in the score from a baseline measure at either the school or student level. This is particularly valuable if the teacher value added score is to be used as a stand-alone measure of teacher quality.

words, this weighting scheme attempts to create a unified measure of removal likelihood for each principal *p* under the assumption that each principal faced the same offense mix.

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

with n_k = number of disciplinary events in the sample for offense type *k*

and n = total number of disciplinary events in the sample

After weighting by offense type, we have a single severity score for each principal. This severity 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 grade level.

Step 2. Estimating the Effect of Principal Discipline Severity on Student Outcomes

In this next step, we study the effects principal discipline severity on a variety of student outcomes. This study focuses on five short-term outcomes, generically referred to as Y. The first two measures focus directly on student discipline, including an indicator variable for whether or not the student committed any reported disciplinary offense that year, and an indicator variable for whether or not the student received out-of-school suspension, expulsion, or transfer that year (1=yes, 0=no). The next two measures represent academic achievement, an average of end-of-year reading and math test scores (standardized z-scores). The final two measures focus on other short-term academic outcomes: number of days absent and an indicator of grade retention.

Establishing a hypothesis for why our measure of principal discipline severity is linked to a student's likelihood of being removed from school is straightforward. We expect that students with a more "severe" principal will, on average, have a higher chance of removal and more removal events, and therefore more absenteeism. In the same way, students who are more likely to be suspended will on average do less well on academic tests than students who were in class.

Because suspension is relatively rare, we expect that these effects will be relatively small on average for all students, but should be larger for subsets of students more at risk for disciplinary offenses. Equation 3 starts with a simple model linking the student outcome to the principal disciplinary severity measure:

$$Y_{it} = \beta_0 + \beta_1 \hat{\mu}_p + \varepsilon_{it} \quad (3)$$

The problem with this model is that factors related to the school and student types could both be correlated with the student outcome Y and the principal's removal likelihood. As mentioned above, we did not account for these factors when we estimated the principal severity measure. Therefore, we need a model that controls for as much of this variation as possible. Our preferred model, specified in equation 4, controls for time-fixed school factors (through a school fixed effect α_s), time varying school factors X_{1st}, as well as time fixed student factors like sex, race, and socioeconomic status (X_{2i}) in addition to grade (γ_g) and year fixed effects (τ_t) .

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

In the results section, we will experiment with other models, including a model that controls for student fixed effects. In addition, we will interact our measure of principal likelihood with a measure that indicates whether or not the student has had a disciplinary removal this year. We are particularly interested in knowing if the principal's choices only affect students involved in the disciplinary system or if there are spillover effects to all students in the school.

With the inclusion of school fixed effects, the β_1 coefficient on principal severity can only be identified through variation in severity from within-school turnover of principals over time. For our estimates of β_1 to be unbiased, we must assume that no unobservable changes in the school environment are responsible for correlations between principal severity and student outcomes. For example, we may be concerned that principals with higher removal likelihoods

run an overall "tighter" ship that encourages more student participation, less absenteeism, and better learning outcomes. On the other hand, a principal who is less likely to remove a student might also be more likely to create a more welcoming culture that encourages more school attendance and better learning outcomes. Either scenario would lead the model to mistakenly link the principal's removal likelihood to general student outcomes. We provide robustness tests in the results section to explore the potential of these mechanisms to drive the results.

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 empirical relationship between principal severity and long-term student outcomes, we must slightly modify the model in equation 4. We collapse the data from disciplinary offense-level to the student level and then track 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 indicator outcomes is:

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

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

RESULTS

Descriptive Analysis of Principal Severity

Before discussing how principal severity affects students, we examine the nature of the estimated principal severity measure itself. What exactly does this measure tell us about a principal? How does it correlate with other principal characteristics and indicators of effectiveness? Figure 1 illustrates the distribution of principal severity in our sample. A value of one would imply that the principal assigns removal regardless of the type of disciplinary offense or disciplinary history of the student. A value of zero would imply that the principal never assigns removal. The average principal in our sample removes students for 28 percent of 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. Principals at the 10th percentile of severity would remove 12.8% of students for the average offense, whereas principals at the 90th percentile would remove 49.3%.

A common question in the literature on school discipline is whether students are treated differently based solely on their race. We re-estimate principal severity as described in the methods section for white student referrals and black student referrals separately.[17] For each school that has at least 10% of referrals coming from white students and 10% of referrals coming from black students, we calculate principal black-white "bias" as the principal's likelihood of removing a black student minus their likelihood of removing a white student, for the same offense and offense history. Figure 1 also provides a histogram of this principal bias measure. A black student is three percentage points more likely to be removed than a white student for the same offense by the average principal. We use this measure to examine the consequences of having a more biased principal for students of different racial groups.

[17] We chose to not examine other forms of racial bias such as white-Hispanic, black-Hispanic, white-other, etc. due to the small sample of schools that have racially diverse enough student populations to estimate severity differences by race. This could provide an interesting avenue for future research.

A growing literature in education has examined the importance of other principal characteristics, including value-added in student test scores (e.g., Branch, Hanushek, & Rivkin, 2012; Grissom, Kalogrides, & Loeb, 2015). In Appendix Table A2, we present correlation of our principal severity and principal bias estimates with value-added measures of principal effectiveness estimated on our data[18], as well as with years of principal experience. Overall, our measure of principal severity appears only weakly related with principal value-added, with a correlation of 0.028 in reading and 0.027 in math. More severe principals tend to be slightly more experienced, on average, with a correlation of 0.05 between severity and years of experience. These weak connections with other measures of principal effectiveness support the notion that principal disciplinary practices represents a unique dimension of behavior that deserves attention in its own right. They also support our belief that more severe principals are not more likely to be systematically "better" or "worse" principals in other dimensions.

The more severe principals tend to work at high-poverty schools – the average principal in an urban, high-poverty school uses removal 40% of the time for a typical offense. Principals of schools in urban areas are also more likely than principals in rural or suburban areas to assign student removal. Relatedly, we observe the highest levels of principal racial disproportionality in urban schools. Principals in low-poverty schools are actually more likely to exhibit racial bias than principals in high-poverty schools. These descriptive findings are presented in Appendix Figures A1 and A2.

[18] To estimate principal value-added, we run a regression of student test scores in year t on student test scores in year t-1 and a series of control variables, and capture principal fixed effects. We also perform this same method in a regression with school fixed effects. The value-added measures presented in this paper are using across-school value-added, but the main points remain consistent when we use within-school value-added measures instead.

Appendix Figure A3 shows that the average principal severity dropped steadily from 34% removal likelihood in 2008 to 27% removal likelihood in 2012 before flattening out.[19] Although there may be many reasons for the changing composition of principals, attention to the issue of exclusionary discipline certainly increased between 2008 and 2012 as civil rights lawsuits and state policy changes put pressure on schools to reduce their suspensions (Losen, 2011).

Effects of Principal Severity on Short-Term Outcomes

Table 3 documents the impacts of principal severity on students' disciplinary and academic behaviors in the short-term, in a model containing school fixed effects and a variety of controls. Each coefficient on principal severity represents the effect of substituting a principal who never removes students (severity = 0) with a principal who always removes students (severity = 1). Although there are principals in our sample who exhibit such behavior (see Figure 1), these coefficients should nonetheless be interpreted as the results of extreme changes in principal severity.

In column 1, we observe that a 100-percentage point increase in principal severity reduces students' likelihood of having any disciplinary referral that year by 25 percentage points. This implies that strict principals have a "deterrent" effect on student misbehavior. (A possible, but less likely, alternative explanation would be that principal severity causes teachers to report fewer students for the same level of offenses to "protect" the students from harsh consequences). Conditional on whether a student has a disciplinary referral that year, the same rise in principal severity increases the likelihood of student removal by 11.6 percentage points (column 2). A

[19] We estimate a single severity score for each principal across their entire tenure. The average severity of the total principal workforce overtime reflects changes in the principal workforce (i.e. as younger principals replace older principals.)

severe principal generates fewer student offenses, and 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.

Columns 3 through 5 of Table 3 present effects of principal severity on student academic outcomes. An increase in principal removal likelihood from 0% to 100% increases the average number of student absences by approximately one-half of a day. This result is only marginally statistically significant. The estimates in column 4 and 5 indicate null effects in aggregate of principal severity on student test scores and grade retention. Given that students under severe 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 make sense. We anticipate, however, that principal severity affects students involved in the disciplinary system quite differently from those not involved.

To investigate this question, we show results in Table 4 from a model that interacts principal severity with an indicator of whether or not the student was removed in that school year. This model is particularly useful for thinking about whether principal severity has any spillover effects on students for students who are not suspended. In columns 1 through 3, we see that principal severity has no significant effect on absences, test scores or grade retention for students not suspended during the school year. In other words, there are no general spillover effects, either positive or negative.

The coefficient on 'Any Removal,' however, reminds us that removal is strongly correlated with negative outcomes. A student removed from a school with a principal possessing the minimum removal likelihood will miss 3 more days in absences than a student who is not suspended in the same school. She will also have a test score that is almost 0.5 standard

deviations lower than a similar student without a suspension. Finally, in column 3, we see that those students that are removed are also 1.2 percentage points more likely to be retained.

The interaction term with Principal Severity asks whether it matters which type of principal removes the student. In column 1, we find that being removed by a very harsh principal more than doubles the absenteeism rate, perhaps suggesting that students in harsh disciplinary environments choose to simply skip school, or that those principals induce higher-than-average suspension lengths. Perhaps not surprisingly then, colum 3 shows that students removed by the harshest principal are more than 3 times more likely to be retained than a student removed by a less harsh principal. This could act through absenteeism or length of suspension, or reflect a more general correspondence between harshness in discipline and retention decisions. Finally, column 2 shows that those students who are suspended by the harshest principal have smaller test score reductions due to suspension than students suspended by a less harsh principal. This may well demonstrate a selection effect – harsh principals exclude more students, including students with slightly better test scores. More broadly, principal severity may selectively affect the group of students who eventually take end-of-grade standardized tests through its impacts on absences, suspension, transfer, and expulsion.

To better describe the practical implications of these findings, we translate the magnitude of the coefficient estimates to more reasonable margins. For example, substituting a 10^{th} percentile severity principal with a 90^{th} percentile severity principal would: (1) reduce incidence of disciplinary referrals by 9.1 percentage points; (2) increase removal conditional on offense by 4.2 percentage points; and (3) increase absences by 0.2 days per student; for the full population of students. For students who are removed from school due to principal discretion, substituting a 10^{th} percentile severity principal with a 90^{th} percentile severity principal would: (1) increase

absences by 1.7 days; (2) increase average test scores by 0.03 standard deviations; and (3) increase grade retention by 0.8%. The magnitudes of these impacts are fairly modest by most standards. Whether such consequences of having a severe principal would shape long-term student trajectories, or simply fade out, is the question of our next analysis.

Effects of Principal Severity on Long-Term Outcomes

Individual-level analyses test the impacts of middle school principal severity on high school graduation and criminal conviction in young adulthood (Table 5). As described in the methods section, we replace principal severity in year *t* with average principal severity experienced by the student from grades 6 through 8. Column 1 of Table 5 shows that an increase in average principal removal likelihood from 0% to 100% reduces the likelihood of high school graduation by a surprisingly large 32.2 percentage points. Replacing a 10th percentile severity principal with a 90th percentile severity principal would still reduce on-time graduation likelihood by 11.8 percentage points. In column 2, we observe that these adverse effects on graduation are similar across students who never had an out-of-school suspension or removal during middle school, and students who did. Below we interrogate the robustness of this result to different model specifications, and although the impact is always significant and negative, it may be more smaller than presented here (See Appendix Table A4).

Column 3 of Table 5 presents the effects of middle school principal severity on likelihood of having a criminal conviction by age 20. Replacing a principal having removal likelihood of 0% with one having removal likelihood of 100% leads to a 12.9 percentage point increase in the likelihood of those students having been convicted by age 20. The more realistic replacement of a 10th percentile severity principal with a 90th percentile severity principal would raise the

criminal conviction rate by 4.7 percentage points. In column 4, the coefficient estimates show that adverse impacts are 8.0 percentage points larger for students who are removed from middle school due to principal discretion than for students who are never removed. This direct translation of consequences within a school disciplinary system to consequences within the adult criminal justice system gives some credence to the school-to-prison pipeline metaphor (Skiba Arredondo, & Williams, 2014).

Robustness Tests

Our preferred model specification accounts for any time-invariant institutional or contextual factors related to the school that could confound the relation between principal severity on student outcomes. However, even with school fixed effects, the introduction of a more severe principal could introduce a host of changes for a school that may also influence student outcomes. For this reason, we perform an alternative estimation of each of the main results tables described above, controlling for a series of other principal characteristics. In particular, we add seven principal-specific controls: principal experience, value-added in reading, value-added in math, gender, and a series of race/ethnicity indicators. As can be seen in Appendix Table A3, the statistically significant decrease in disciplinary offenses but increase in disciplinary removals from principal severity remains with very similar magnitudes. The effect of principal severity on absences becomes slightly larger and more significant at 0.71 days.

In terms of long-term outcomes, the negative effect of principal severity on high school graduation and positive effect on criminal conviction remains highly significant, although the magnitude of the effects decrease somewhat. Once controlling for a multitude of characteristics of the principal, replacing a 10^{th} percentile severity principal with a 90^{th} percentile severity

principal ultimately produces a 7.9 percentage point decrease in high school graduation, and 4.8 percentage point increase in criminal conviction in young adulthood.

Effects of Principal Racial Bias on Student Outcomes

Finally, we seek to determine if the differential treatment in the disciplinary process of students by principals influences student outcomes. To do so, we run regressions of short-term student outcomes on principal black-white removal bias in the form presented in Table 3 (with school fixed effects and controls). We interact principal black-white removal bias with student race to examine specific effects on white, black, Hispanic, and other race students. Table 6 presents regression estimates, and Figure 2 presents the corresponding marginal effects by race from these regressions. First, we see that increased principal removal bias leads to increased rates of removal conditional on offense for black and Hispanic students. Next, we see that principal removal bias decreases absenteeism and increases test scores for white students, and substantially increases absenteeism and decreases test scores for black and Hispanic students. Finally, such bias appears to have permanent ramifications as it causes reductions in grade retention rates for white students and growth in grade retention rates for black students. Thankfully, these differential impacts are calculated from theoretically replacing a principal who treats black and white students identically to a principal who suspends black students 100% of the time and white students 0% of the time, a situation which never actually exists in our data. Still, the average principal in our sample removes black students for the same offenses and offense histories 3% more often than white students, which implies systematic disadvantage for later educational outcomes.

CONCLUDING DISCUSSION

Principal Joe Clark gained fame in real life (and through the portrayal in film by Morgan Freeman) because people believed that he made a dangerous school safer, and that he improved the life chances of his students. However, not everyone appreciated the harshness of his methods, particularly for those students that he expelled and suspended. This paper looks at both sides of that story using data on principals in middle schools in North Carolina. We find that in fact that principals who use harsh practices do make schools safer, at least as measured by the frequency of school disciplinary incidents. Specifically, we find that a middle school principal who is in the 90^{th} percentile in our distribution of removal likelihood reduces the percentage of students with disciplinary reports by 9.1 percentage points in a given year relative to a principal who is in the 10^{th} percentile in our distribution of removal likelihood. This is a big difference. In our study, 24% of students have a disciplinary offense in a given year, and so 9.1 percentage points represents a 38% reduction in the probability of disciplinary offenses among the student population. It is possible that this reduction does not represent real safety improvements – perhaps teachers stop reporting students if they fear that students will be suspended. Further research is needed to explore this possibility. However, in our data, a more severe principal does appear to have a dramatic impact on the behavior of students.

Of course, this benefit does not come for free. The harsher principal also increases exclusionary removal, conditional on offense, by 4.2 percentage points, an increase of 35% over the mean removal probability for a given student. Although students who are removed by any principal already have bad outcomes, dealing with a harsher principal appears to further increase student absenteeism and grade retention, although the increases here are modest.

The results are not modest for long-term educational attainment or involvement in the criminal justice system. We found that having a principal in the 90th percentile of severity decreases the likelihood of on-time high school graduation by 16.5% of the average graduation rate, relative to having a principal in the 10th percentile. This repercussion exists both for students who are suspended in middle school, and for students who are never suspended in middle school, suggesting that a harsh principal has a broad influence on students beyond simply enforcing removals. Similarly, we found that having a principal in the 90th percentile of severity doubles the likelihood of criminal conviction by age 20 for a student who is removed in middle schools, relative to what happens for a student who is removed by a less harsh principal. Again, both suspended students and not-suspended students experienced these long-term impacts, although with larger impacts for the suspended students. Future research should study directly how principal behavior and disciplinary decision-making can influence long-term trajectories, either through removal from school or through alternative mechanisms.

These consequences of principal harshness are particularly problematic given that there is evidence of principal bias in our sample. Harsher principals are not necessarily more biased, but we did find that some principals are in fact more likely to remove minorities, with direct consequences for the test score achievement, attendance records, and grade retention of minority students. Given that minority students already have higher rates of referrals than white students, and are therefore more susceptible to experiencing the negative consequences of exclusionary discipline, any differential treatment by principals on top of that baseline disadvantage will worsen the problem.

Our study provides normative evidence that principal choices about disciplinary consequences come with a sharp tradeoff between school security and negative outcomes for

students removed from school. These negative outcomes may have long term consequences, including involvement in the adult criminal justice system. These negative consequences are not outweighed by any positive spillover effects for other non-suspended students, at least not for the outcomes studied in this paper. Given this evidence, future work needs to be done to better calibrate the use of discretion by principal. At the very least, policy makers should work to ensure that the power of the principal to remove students be used in a fair and evidence-based manner.

REFERENCES

Barnes, J. S., & Motz, R. T. (2018). Reducing racial inequalities in adulthood arrest by reducing inequalities in school discipline: Evidence from the school-to-prison pipeline. *Developmental Psychology*, *54*(12), 2328-2340.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (No. w17803). National Bureau of Economic Research.

Berg, J. K., & Cornell, D. (2016). Authoritative school climate, aggression toward teachers, and teacher distress in middle school. *School psychology quarterly*, *31*(1), 122.

Bhuller, M., Dahl, G. B., Loken, K. V., & Mogstad, M. (2018). Intergenerational effects of incarceration. In *AEA Papers and Proceedings* (Vol. 108, pp. 234-40).

Chu, E. M., & Ready, D. D. (2018). Exclusion and Urban Public High Schools: Short-and Long-Term Consequences of School Suspensions. *American Journal of Education*, *124*(4), 479-509.

Cornell, D., & Huang, F. (2016). Authoritative school climate and high school student risk behavior: A cross-sectional multi-level analysis of student self-reports. *Journal of Youth and Adolescence*, *45*(11), 2246-2259. doi:10.1007/s10964-016-0424-3

Council of State Governments Justice Center. (2017). Realizing the full vision of school discipline reform: A framework for statewide change.

Curran, F. C. (2019). The law, policy, and portrayal of zero tolerance school discipline: Examining prevalence and characteristics across levels of governance and school districts. *Educational Policy*, *33*(2), 319-349.

Curran, C. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student Bbhavior. *Educational Evaluation and Policy Analysis*, 647-668.

DeMatthews, D., Carrola, P., & Mungal, A. (2015). Despite best intentions: A critical analysis of social justice leadership and decision making. *Administrative Issues Journal Education Practice and Research*, *5*(2). doi:10.5929/2015.5.2.4

DeMatthews, D. E., Carey, R. L., Olivarez, A., & Saeedi, K. M. (2017). Guilty as charged? Principals' perspectives on disciplinary practices and the racial discipline gap. *Educational Administration Quarterly*, *53*(4), 519-555.

Duke Law News. (2011). North Carolina has new school discipline law. Retrieved from: https://law.duke.edu/news/6608/.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). Breaking schools' rules: A state wide study of how school discipline relates to students' success and juvenile justice involvement.

Findlay, N. M. (2015). Discretion in Student Discipline. *Educational Administration Quarterly*, *51*(3), 472-507. doi:10.1177/0013161x14523617

Gershoff, E. T., & Font, S. A. (2016). Corporal punishment in US public schools: Prevalence, disparities in use, and status in state and federal policy. *Social policy report*, *30*.

Goodman, J. W. (2005). Leandro v. State and the constitutional limitation on school suspensions and expulsions in North Carolina. *North Carolina Law Review*, *83*, 1507-1525.

Gottfredson, G. D., & Gottfredson, D. C. (2001). What schools do to prevent problem behavior and promote safe environments. *Journal of Educational and Psychological Consultation*, *12*(4), 313-344. doi:10.1207/s1532768xjepc1204_02

Gottfredson, G. D., Gottfredson, D. C., Payne, A. A., & Gottfredson, N. C. (2005). School climate predictors of school disorder: Results from a national study of delinquency prevention in schools. *Journal of Research in Crime and Delinquency*, *42*(4), 412-444.

Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for black and white high school students. *American Educational Research Journal*, *48*(4), 904-934. doi:10.3102/0002831211398531

Gregory, A., Cornell, D., Fan, X., Sheras, P., Shih, T. H., & Huang, F. (2010). Authoritative school discipline: High school practices associated with lower bullying and victimization. *Journal of Educational Psychology*, *102*(2), 483.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3-28.

Harding, D. J., Morenoff, J. D., Nguyen, A. P., & Bushway, S. D. (2018). Imprisonment and labor market outcomes: Evidence from a natural experiment. *American Journal of Sociology*, *124*(1), 49-110.

Heilbrun, A., Cornell, D., & Konold, T. (2018). Authoritative school climate and suspension rates in middle schools: Implications for reducing the racial disparity in school discipline. *Journal of School Violence*, *17*(3), 324-338. doi:10.1080/15388220.2017.1368395

Heilbrun, A., Cornell, D., & Lovegrove, P. (2015). Principal attitudes regarding zero tolerance and racial disparities in school suspension. *Psychology in the Schools*, *52*(5), 489-499. doi:10.1002/pits.21838

Hoffman, S. (2014). Zero benefit: Estimating the effect of zero tolerance discipline polices on racial disparities in school discipline. *Educational Policy*, *28*(1), 69-95.

Huang, F. L., & Cornell, D. (2018). The relationship of school climate with out-of-school suspensions. *Children and Youth Services Review*, *94*, 378-389. doi:10.1016/j.childyouth.2018.08.013

Hwang, N. (2018). Suspensions and achievement: Varying links by type, frequency, and subgroup. *Educational Researcher*, 47(6), 363-374.

Kinsler, J. (2011). Understanding the black–white school discipline gap. *Economics of Education Review*, *30*(6), 1370-1383. doi:10.1016/j.econedurev.2011.07.004

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

Kupchik, A., & Catlaw, T. J. (2015). Discipline and participation. *Youth & Society*, *47*(1), 95-124. doi:10.1177/0044118x14544675

Lacoe, J., & Steinberg, M. P. (2018a). Do suspensions affect student outcomes. *Educational Evaluation and Policy Analysis*, *41*(1), 34-62. doi:10.3102/0162373718794897

Lacoe, J., & Steinberg, M. P. (2018b). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes. *Peabody Journal of Education*, *93*(2), 207-227. doi:10.1080/0161956x.2018.1435047

Lau, C., Wong, M., & Dudovitz, R. (2018). School disciplinary style and adolescent health. *Journal of Adolescent Health*, *62*(2), 136-142.

Laub, J. H. (2014). Understanding inequality and the justice system response: Charting a new way forward.

Lindsay, C. A., & Hart, C. M. D. (2017). Exposure to same-race teachers and student disciplinary outcomes for black students in North Carolina. *Educational Evaluation and Policy Analysis*, *39*(3), 485-510. doi:10.3102/0162373717693109

Losen, D. J. (2011). Discipline policies, successful schools, and racial justice. National Education Policy Center. Available at: https://nepc.colorado.edu/sites/default/files/NEPC-SchoolDiscipline.pdf.

Losen, D. J. & Martinez, T. E. (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.

Losen, D. J., & Skiba, R. J. (2010). Suspended education: Urban middle schools in crisis. The UCLA Center for Civil Rights Remedies at The Civil Rights Project.

Marchbanks, M. P., Peguero, A. A., Varela, K. S., Blake, J. J., & Eason, J. M. (2018). School strictness and disproportionate minority contact. *Youth Violence and Juvenile Justice*, *16*(2), 241-259. doi:10.1177/1541204016680403

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

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

Mowen, T., & Brent, J. (2016). School discipline as a turning point: The cumulative effect of suspension on arrest. *Journal of Research in Crime and Delinquency*, *53*(5), 628-653.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Unpublished Working Paper*, *18*.

Mukaria, G. (2002). Disciplinary challenges: How do principals address this dilemma? *Urban Education*, *37*, 432-452.

Nicholas, P. T., Ayana, A., & Chance, W. L. (2014). Zero tolerance, school shootings, and the post-Brown Quest for Equity in Discipline Policy: An Examination of How Urban Minorities Are punished for white suburban violence. *The Journal of Negro Education*, *83*(3), 352. doi:10.7709/jnegroeducation.83.3.0352

Nicholson-Crotty, S., Birchmeier, Z., & Valentine, D. (2009). Exploring the impact of school discipline on racial disproportion in the juvenile justice system. *Social Science Quarterly*, 1003-1018.

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

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

North Carolina Department of Public Instruction. (2000). Case studies of best practices: Alternative learning schools and programs 1998-99.

North Carolina General Statutes. (2013). Chapter 115C - Elementary and Secondary Education. Article 27 - Discipline.

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

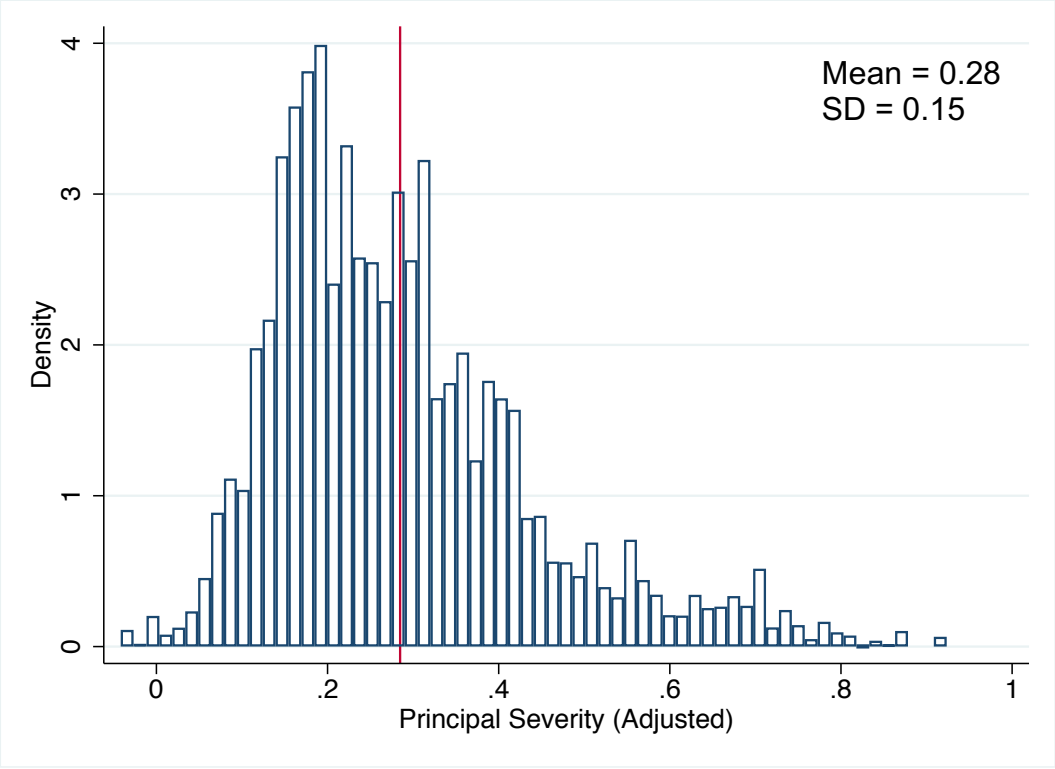
Public Schools of North Carolina, State Board of Education, Department of Public Instruction (2008), Report to the Joint Legislative Education Oversight Committee Consolidated Data Report, 2007-08. Available at http://www.ncpublicschools.org/docs/research/discipline/reports/consolidated/2007-08/consolidated-report.pdf

Public Schools of North Carolina, State Board Of Education (2018), Department of Public Instruction. Report to the North Carolina General Assembly Consolidated Data Report, 2016-17. Retrieved from http://www.ncpublicschools.org/docs/research/discipline/reports/consolidated/2016-17/consolidated-report.pdf

Pudelski, S. (2014). School discipline in the eyes of school superintendents. Retrieved from https://www.childrensdefense.org/wp-content/uploads/2018/08/school-discipline-in-the-eyes.pdf February 18, 2019.
Rausch, M. K., & Skiba, R. J. (2005). The academic cost of discipline: The relationship between suspension/expulsion and school achievement. In *Annual Meeting of the American Educational Research Association, Montreal, Canada*.
Rosenbaum, J. (2018). Educational and criminal justice outcomes 12 years after school suspension. *Youth & Society*.
Skiba, R. J., & Knesting, K. (2001). Zero tolerance, zero evidence: An analysis of school disciplinary practice. *New Directions for Youth Development*, *92*, 17-43.
Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More Than a Metaphor: The contribution of xxclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education*, *47*(4), 546-564. doi:10.1080/10665684.2014.958965
Skiba, R. J., Chung, C.-G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality. *American Educational Research Journal*, *51*(4), 640-670. doi:10.3102/0002831214541670
Skiba, R. J., & Edl, H. (2004). The disciplinary practices survey: How do Indiana's principals feel about discipline.
Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*(1), 85-108.
Skiba, R. J., & Peterson, R. (1999). The dark side of zero tolerance: Can punishment lead to safe schools. *The Phi Delta Kappan*, *80*(5), 372-382.
Steinberg, M. P., & Lacoe, J. (2017). What do we know about school discipline reform? *Education Next*, *17*(1).
Sullivan, M. K. (2011). Long-term suspensions and the right to an education: An alternative approach, 293 (2011). *North Carolina Law Review*, *293*.
Sykes, B. L., Piquero, A. R., Gioviano, J. P., & Pittman, N. (2015). The school-to-prison pipeline in America, 1972–2012.
U.S. Department of Education. (2018). North Carolina compilation of school discipline laws and regulations.
U.S. Department of Education. (2014). Guiding principles: A resource guide for improving school climate and discipline.
U.S. Department of Education. (1998). Violence and discipline problems in U.S. public schools: 1996-97. National Center for Education Statistics. Statistical Analysis Report NCES 98-030.
U.S. Government Accountability Office. (2018). Discipline disparities for black students, boys, and students with disabilities. Report GAO-18-258.
Wald, J., & Losen, D. J. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development*, *2003*(99), 9-15.
Welsh, R. O., & Little, S. (2018). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research*, *88*(5), 752-794.
Wettach, J. R. (2011). School discipline law affective North Carolina public school students. Retrieved
Wolf, K. C., & Kupchik, A. (2017). School Suspensions and Adverse Experiences in Adulthood. *Justice Quarterly*, *34*(3), 407-430.

TABLES AND FIGURES

Figure 1. Histograms of Estimated Principal Severity and Estimated Principal Bias



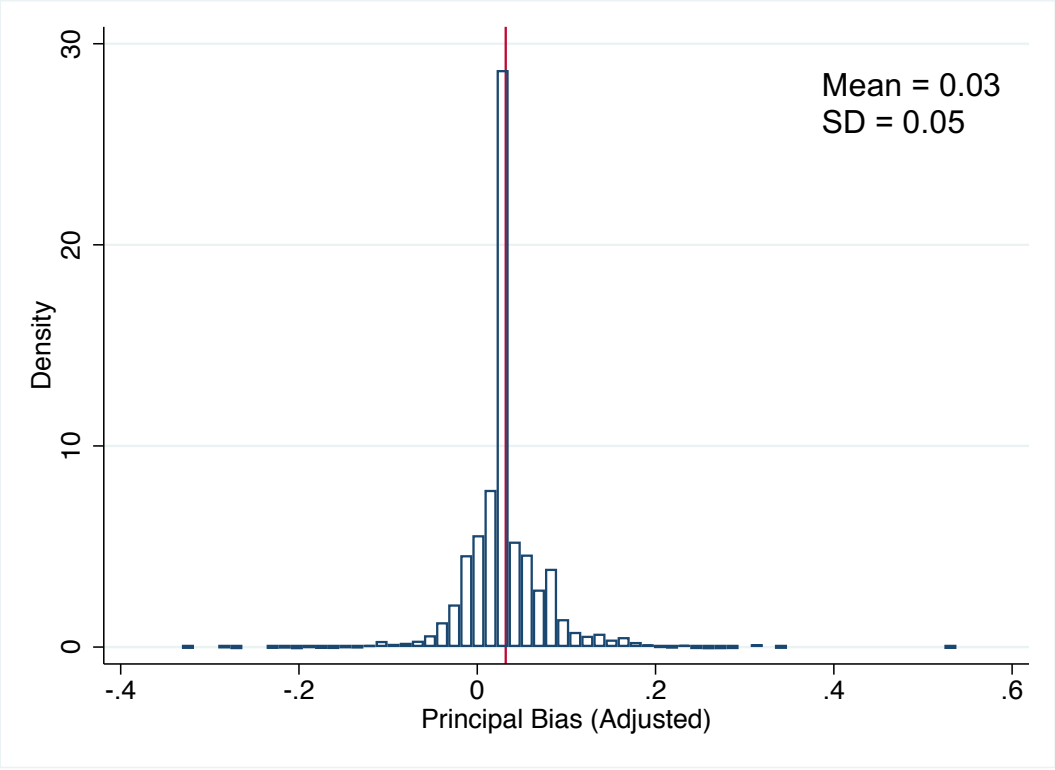
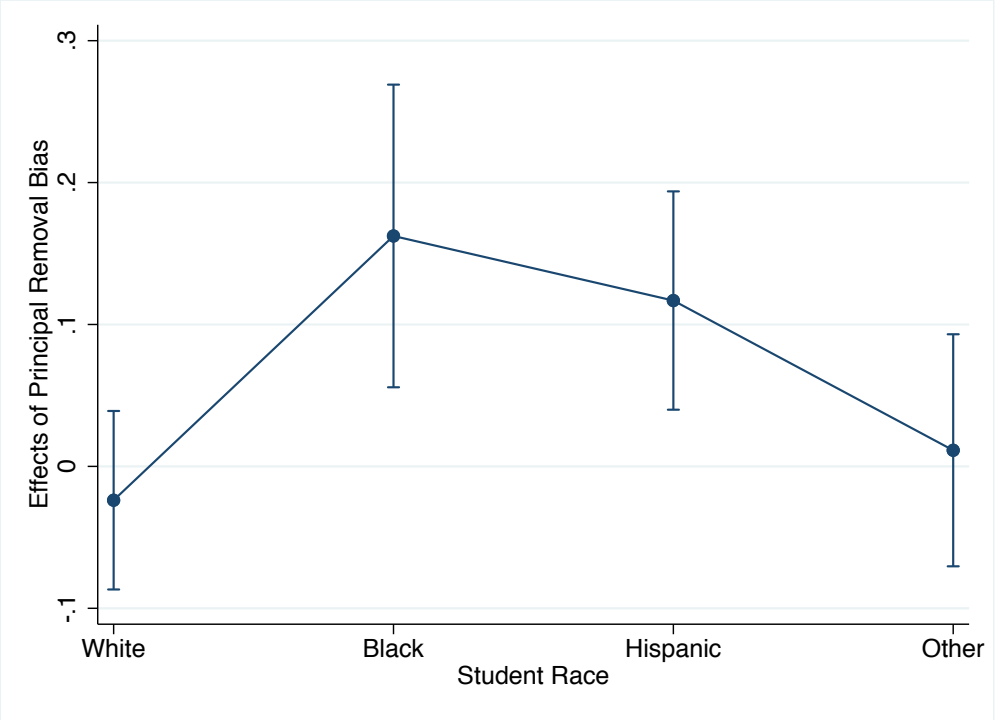
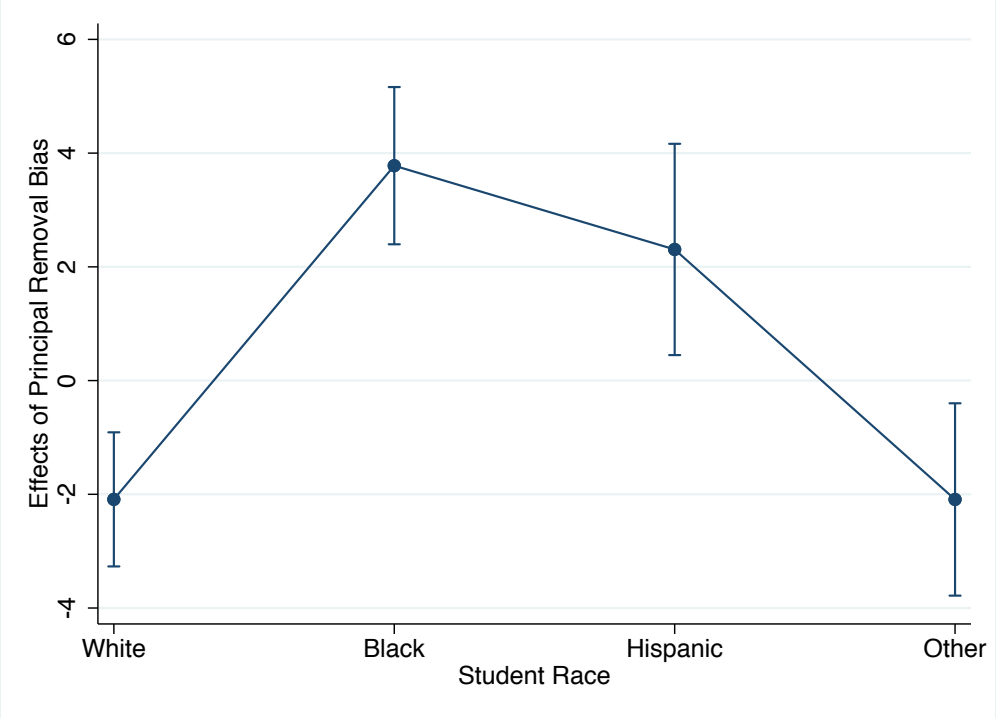


Figure 2. Effects of Principal White-Black Removal Bias by Student Race/Ethnicity

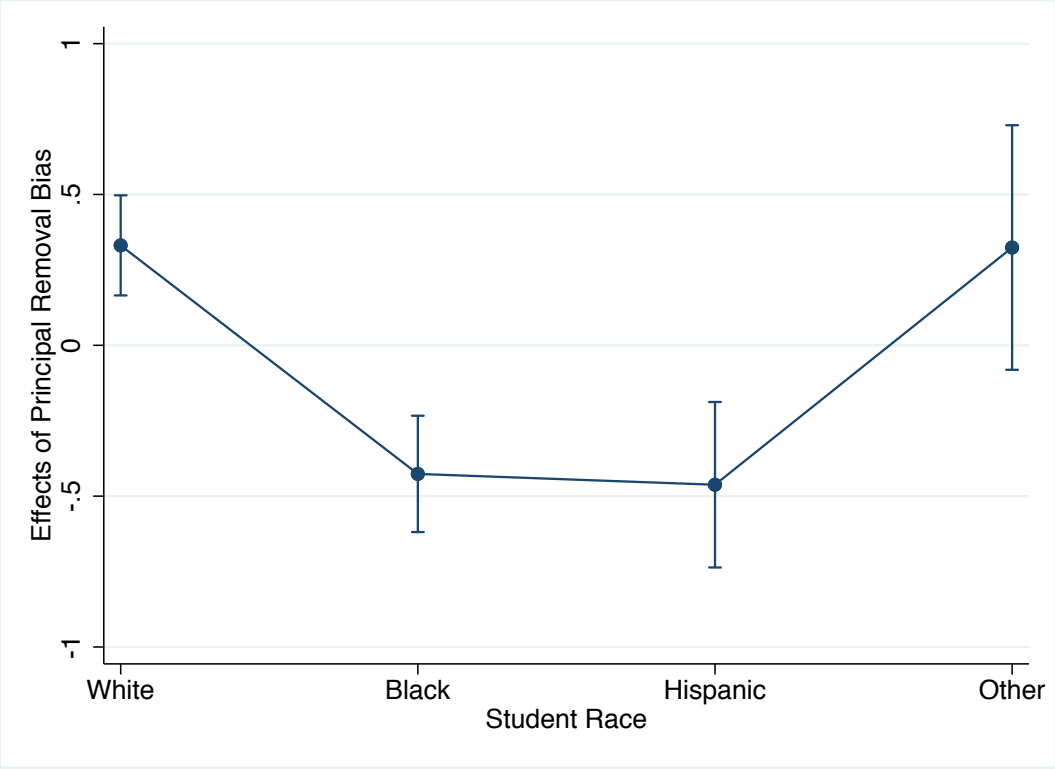
Effects on Student Removal



Effects on Student Absences



Effects on Student Test Scores (SDs)



Effects on Student Grade Retention

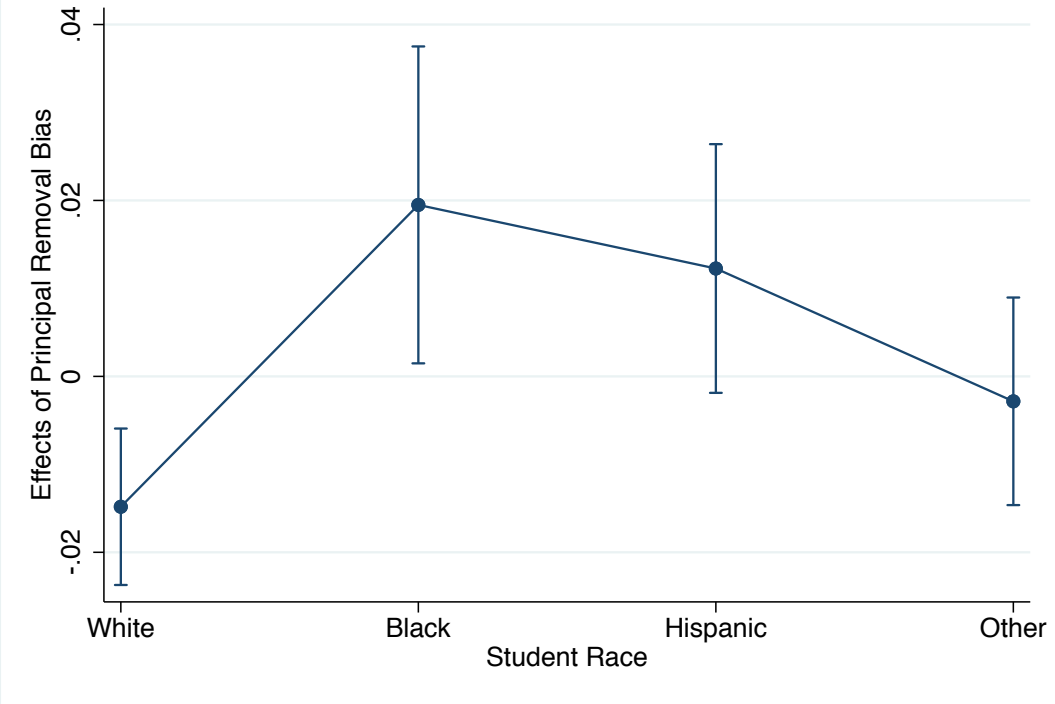


Table 1. Event Frequency and Probability of Removal by Disciplinary Offense Type

Offense Type	Code	Frequency	Pr(Removal)
Excessive tardiness	UB	34,811	0.028
Bus misbehavior	UB	184,140	0.035
Cutting class	UB	44,559	0.129
Cell phone use	UB	19,964	0.129
Skipping school	UB	42,678	0.179
Dress code violation	UB	30,690	0.201
Being in an unauthorized area	UB	29,566	0.203
Other school defined offense	UB	86,592	0.218
Disruptive behavior	UB	631,231	0.252
Insubordination	UB	278,488	0.268
Disrespect of faculty/staff	UB	109,461	0.299
Late to class	UB	67,918	0.307
Bullying	UB	68,203	0.342
Inappropriate language/disrespect	UB	201,577	0.347
Property damage	UB	17,687	0.399
Aggressive behavior	UB	194,624	0.467
Disorderly conduct	UB	32,439	0.489
Inappropriate items on school property	UB	15,966	0.546
Theft	UB	33,946	0.574
Communicating threats	UB	21,728	0.753
Sexual offense	RO	145	0.766
Robbery without a dangerous weapon	RO	64	0.766
Assault on student not resulting in serious injury	UB	17,085	0.780
Sexual assault (not involving rape or sexual offense)	RO	613	0.819
Burning of a school building	RO	419	0.823
Assault resulting in a serious injury	RO	235	0.838
Bomb threat	RO	257	0.848
Possession of a weapon (not firearm)	RO	10,529	0.858
Possession of alcoholic beverage	RO	2,637	0.864
Assault on school personnel	RO	3,077	0.865
Unlawfully setting a fire	UB	561	0.866
Possession of controlled substance in violation of law	RO	10,196	0.868
Fighting	UB	184,240	0.871
Possession of a firearm or powerful explosive	RO	185	0.881
Assault involving use of a weapon	RO	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. RO = reportable offense; UB = unacceptable behavior, as classified by NC DPI.

Table 2. Descriptive Statistics of Student-Level Analytical Dataset

	Observations	Mean	Std Dev	Minimum	Maximum
Outcome variables					
Any student offense	2,822,073	0.24	0.43	0.00	1.00
Removal indicator	2,822,073	0.12	0.33	0.00	1.00
Days absent	2,500,637	6.21	6.90	0.00	170.00
Test scores (SD)	2,815,337	0.00	1.00	-3.00	3.00
Grade retention	2,390,653	0.01	0.09	0.00	1.00
HS graduation*	302,247	0.71	0.45	0.00	1.00
Criminal conviction*	99,621	0.03	0.18	0.00	1.00
Principal variables					
Principal severity	2,822,073	0.28	0.15	-0.04	0.92
Principal bias	2,822,073	0.03	0.05	-0.33	0.54
Principal V-A (read)	2,764,467	0.00	0.08	-1.08	1.59
Principal V-A (math)	2,765,170	0.00	0.09	-0.93	0.55
Principal experience	2,708,707	22.04	7.67	0.00	48.00
Principal female	2,821,336	0.51	0.50	0.00	1.00
Principal black	2,822,073	0.25	0.43	0.00	1.00
Principal Hispanic	2,822,073	0.00	0.04	0.00	1.00
Principal other race	2,822,073	0.02	0.13	0.00	1.00
Student variables					
Student black	2,822,073	0.22	0.41	0.00	1.00
Student Hispanic	2,822,073	0.10	0.30	0.00	1.00
Student other race	2,822,073	0.06	0.24	0.00	1.00
Student female	2,821,860	0.49	0.50	0.00	1.00
Student LEP	2,822,073	0.06	0.24	0.00	1.00
Student free lunch	2,821,953	0.57	0.50	0.00	1.00
School variables					
Urban area	2,822,073	0.26	0.44	0.00	1.00
Suburban area	2,822,073	0.18	0.39	0.00	1.00
Town area	2,822,073	0.11	0.31	0.00	1.00
Title 1 eligible	2,794,314	0.66	0.47	0.00	1.00
Magnet school	2,822,073	0.06	0.24	0.00	1.00
FTE teachers	2,822,073	48.77	15.55	0.00	106.00
Percent free lunch	2,822,073	46.08	26.05	0.00	100.00
Percent reduced lunch	2,822,073	7.73	6.45	0.00	100.00
Enrollment	2,822,073	735.89	296.18	10.00	1781.00
School pct black	2,822,073	21.76	19.90	0.00	100.00
School pct Hispanic	2,822,073	9.69	8.42	0.00	68.68
School pct other race	2,822,073	6.34	7.34	0.00	96.74

*These long-term outcome measures are available in a collapsed student-level dataset (rather than student-year-level) and only for certain cohorts of students

Table 3. Effects of Principal Severity on Short-Term Student Outcomes

Variables	Any Offense	Any Removal	Days Absent	Test Scores (SD)	Grade Retention
Principal Severity	-0.2502**	0.1163**	0.4891+	0.0005	0.0036
	(0.019)	(0.010)	(0.259)	(0.027)	(0.003)
Any Offense		0.4952**			
		(0.008)			
Student Black	0.1342**	0.0239**	-1.7236**	-0.4342**	0.0005
	(0.003)	(0.001)	(0.052)	(0.009)	(0.000)
Student Hispanic	-0.0276**	-0.0075**	-1.5585**	0.0700**	-0.0031**
	(0.002)	(0.001)	(0.051)	(0.009)	(0.000)
Student Other	0.0018	0.0010	-1.2195**	0.0636**	-0.0006*
	(0.003)	(0.001)	(0.052)	(0.013)	(0.000)
Student Female	-0.1478**	-0.0131**	-0.2756**	0.0902**	-0.0057**
	(0.002)	(0.001)	(0.019)	(0.002)	(0.000)
Student LEP	-0.0255**	-0.0029**	-0.7783**	-0.6693**	0.0022**
	(0.003)	(0.001)	(0.045)	(0.009)	(0.000)
Student ED	0.1507**	0.0152**	2.4882**	-0.5554**	0.0083**
	(0.002)	(0.001)	(0.036)	(0.008)	(0.000)
Number Teachers	-0.0004	-0.0006*	-0.0114	0.0003	0.0001
	(0.000)	(0.000)	(0.008)	(0.001)	(0.000)
Percent FL	0.0005**	0.0001	0.0000	-0.0006**	0.0000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Percent RPL	0.0003	-0.0003*	-0.0094**	0.0007*	0.0000
	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)
Enrollment	-0.0000	0.0000**	0.0013**	-0.0001	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percent Black	-0.0010**	-0.0000	0.0147**	0.0027**	0.0001*
	(0.000)	(0.000)	(0.005)	(0.000)	(0.000)
Percent Hispanic	0.0003	-0.0002	0.0162+	-0.0003	-0.0000
	(0.000)	(0.000)	(0.009)	(0.001)	(0.000)
Percent Other	-0.0005	-0.0004+	0.0040	-0.0001	0.0000
	(0.000)	(0.000)	(0.010)	(0.001)	(0.000)
Year FE	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES
Observations	2,821,802	2,821,802	2,500,610	2,815,093	2,390,638
R-Squared	0.148	0.476	0.061	0.284	0.017

** p<0.01, * p<0.05, + p<0.1.
Robust standard errors in parentheses, clustered by school. LEP = Limited English Proficient; ED = Economically Disadvantaged; FL = Free Lunch; RPL = Reduced Lunch

Table 4. Effects of Principal Severity on Short-Term Outcomes, by Student Removal

Variables	Days Absent	Test Scores (SD)	Grade Retention
Principal Severity	-0.0803	-0.0151	0.0008
	(0.245)	(0.028)	(0.002)
Severity x Any Removal	4.6450**	0.0885**	0.0229**
	(0.543)	(0.028)	(0.006)
Any Removal	3.1463**	-0.4782**	0.0118**
	(0.157)	(0.010)	(0.002)
Student Black	-2.1586**	-0.3937**	-0.0013**
	(0.045)	(0.008)	(0.000)
Student Hispanic	-1.4624**	0.0605**	-0.0027**
	(0.047)	(0.008)	(0.000)
Student Other Race	-1.2304**	0.0645**	-0.0007*
	(0.047)	(0.013)	(0.000)
Student Female	0.1263**	0.0514**	-0.0041**
	(0.013)	(0.002)	(0.000)
Student Limited English	-0.7196**	-0.6762**	0.0025**
	(0.040)	(0.009)	(0.000)
Student Econ Disadvantage	2.0810**	-0.5150**	0.0066**
	(0.029)	(0.008)	(0.000)
Number of Teachers	-0.0074	-0.0000	0.0001
	(0.008)	(0.001)	(0.000)
Percent Free Lunch	-0.0009	-0.0005*	0.0000
	(0.001)	(0.000)	(0.000)
Percent Reduced Lunch	-0.0080**	0.0007*	0.0000
	(0.003)	(0.000)	(0.000)
Student Enrollment	0.0011**	-0.0000	-0.0000
	(0.000)	(0.000)	(0.000)
Percent Black	0.0148**	0.0024**	0.0001*
	(0.005)	(0.000)	(0.000)
Percent Hispanic	0.0186*	-0.0004	-0.0000
	(0.008)	(0.001)	(0.000)
Percent Other Race	0.0088	-0.0004	0.0000
	(0.010)	(0.001)	(0.000)
Year FE	YES	YES	YES
Grade FE	YES	YES	YES
School FE	YES	YES	YES
Observations	2,500,610	2,815,093	2,390,638
R-Squared	0.105	0.303	0.022

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

Robust standard errors in parentheses, clustered by school. LEP = Limited English Proficient; ED = Economically Disadvantaged; FL = Free Lunch; RPL = Reduced Price

Table 5. Effects of Principal Severity on Long-Term Student Outcomes

Variables	High School Graduation	High School Graduation	Criminal Conviction by Age 20	Criminal Conviction by Age 20
Principal Severity	-0.3222**	-0.3226**	0.1287**	0.0856**
	(0.053)	(0.054)	(0.026)	(0.024)
Severity x Any Removal		0.0203		0.0800**
		(0.023)		(0.014)
Any Removal		-0.1838**		0.0364**
		(0.008)		(0.005)
Student Black	0.0691**	0.0951**	0.0162**	0.0061**
	(0.004)	(0.004)	(0.002)	(0.002)
Student Hispanic	0.0216**	0.0177**	-0.0153**	-0.0143**
	(0.005)	(0.005)	(0.002)	(0.002)
Student Other Race	0.0061	0.0071	-0.0005	-0.0005
	(0.004)	(0.004)	(0.002)	(0.002)
Student Female	0.0492**	0.0220**	-0.0391**	-0.0285**
	(0.002)	(0.002)	(0.001)	(0.001)
Student Limited English	-0.0382**	-0.0402**	-0.0115**	-0.0111**
	(0.005)	(0.005)	(0.002)	(0.002)
Student Econ Disadvantage	-0.1451**	-0.1170**	0.0293**	0.0187**
	(0.004)	(0.003)	(0.001)	(0.001)
Number of Teachers	-0.0166**	-0.0174**	-0.0020**	-0.0016**
	(0.002)	(0.002)	(0.001)	(0.001)
Percent Free Lunch	0.0047**	0.0050**	0.0003	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)
Percent Reduced Lunch	0.0001	0.0001	-0.0003	-0.0002
	(0.001)	(0.001)	(0.000)	(0.000)
Student Enrollment	0.0011**	0.0011**	-0.0000	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)
Percent Black	-0.0030**	-0.0027**	0.0004	0.0004
	(0.001)	(0.001)	(0.000)	(0.000)
Percent Hispanic	0.0046**	0.0044**	-0.0008	-0.0006
	(0.001)	(0.001)	(0.001)	(0.001)
Percent Other	0.0035**	0.0035**	0.0005	0.0005
	(0.001)	(0.001)	(0.001)	(0.001)
Cohort FE	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Observations	301,478	301,478	99,381	99,381
R-Squared	0.080	0.103	0.053	0.074

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

Robust standard errors in parentheses, clustered by school.

Table 6. Effects of Principal Black-White Removal Bias on Student Outcomes, by Race

Variables	Any Offense	Any Removal	Days Absent	Test Scores (SD)	Grade Retention
Principal Bias	0.0715+	-0.0402+	-2.0883**	0.1808*	-0.0134**
	(0.041)	(0.021)	(0.554)	(0.083)	(0.004)
Bias x Black	-0.0505	0.1847**	6.2559**	-0.5946**	0.0346**
	(0.057)	(0.037)	(0.705)	(0.150)	(0.009)
Bias x Hispanic	0.0839	0.0742**	4.8816**	-0.6179**	0.0267**
	(0.051)	(0.022)	(0.905)	(0.186)	(0.006)
Bias x Other	-0.0262	0.0168	0.3616	0.0965	0.0114*
	(0.041)	(0.021)	(0.706)	(0.183)	(0.005)
Any Offense		0.4940**			
		(0.007)			
Student Black	0.1360**	0.0177**	-1.9354**	-0.4140**	-0.0007+
	(0.003)	(0.002)	(0.049)	(0.009)	(0.000)
Student Hispanic	-0.0302**	-0.0100**	-1.7126**	0.0897**	-0.0039**
	(0.003)	(0.001)	(0.056)	(0.009)	(0.000)
Student Other	0.0027	0.0005	-1.2291**	0.0603**	-0.0010**
	(0.003)	(0.001)	(0.057)	(0.014)	(0.000)
Student Female	-0.1478**	-0.0132**	-0.2757**	0.0902**	-0.0057**
	(0.002)	(0.001)	(0.019)	(0.002)	(0.000)
Student LEP	-0.0257**	-0.0028**	-0.7755**	-0.6696**	0.0022**
	(0.003)	(0.001)	(0.045)	(0.009)	(0.000)
Student ED	0.1507**	0.0152**	2.4821**	-0.5549**	0.0083**
	(0.002)	(0.001)	(0.036)	(0.008)	(0.000)
Number Teachers	-0.0005	-0.0006*	-0.0113	0.0003	0.0001
	(0.000)	(0.000)	(0.008)	(0.001)	(0.000)
Percent FL	0.0005**	0.0000	0.0000	-0.0006**	0.0000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Percent RPL	0.0002	-0.0002*	-0.0092**	0.0007*	0.0000
	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)
Enrollment	-0.0000	0.0000**	0.0013**	-0.0001	-0.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percent Black	-0.0010**	-0.0001	0.0145**	0.0027**	0.0001*
	(0.000)	(0.000)	(0.005)	(0.000)	(0.000)
Percent Hispanic	0.0002	-0.0001	0.0168+	-0.0004	-0.0000
	(0.000)	(0.000)	(0.009)	(0.001)	(0.000)
Percent Other	-0.0000	-0.0006*	0.0021	-0.0001	-0.0000
	(0.001)	(0.000)	(0.010)	(0.001)	(0.000)
Year FE	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES

Observations	2,821,802	2,821,802	2,500,610	2,815,093	2,390,638
R-Squared	0.146	0.476	0.061	0.284	0.018

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

Robust standard errors in parentheses, clustered by school. LEP = Limited English Proficient; ED = Economically Disadvantaged; FL = Free Lunch; RPL = Reduced Lunch

Appendix

Figure A1. Average Principal Severity by School Type

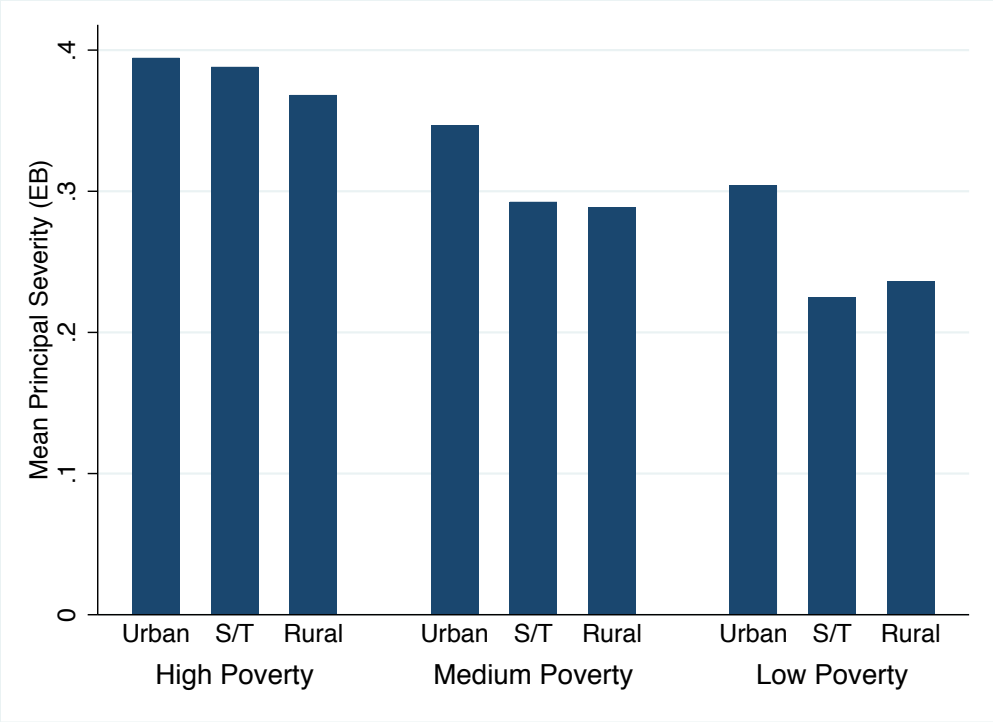


Figure A2. Average Principal White-Black Bias by School Type

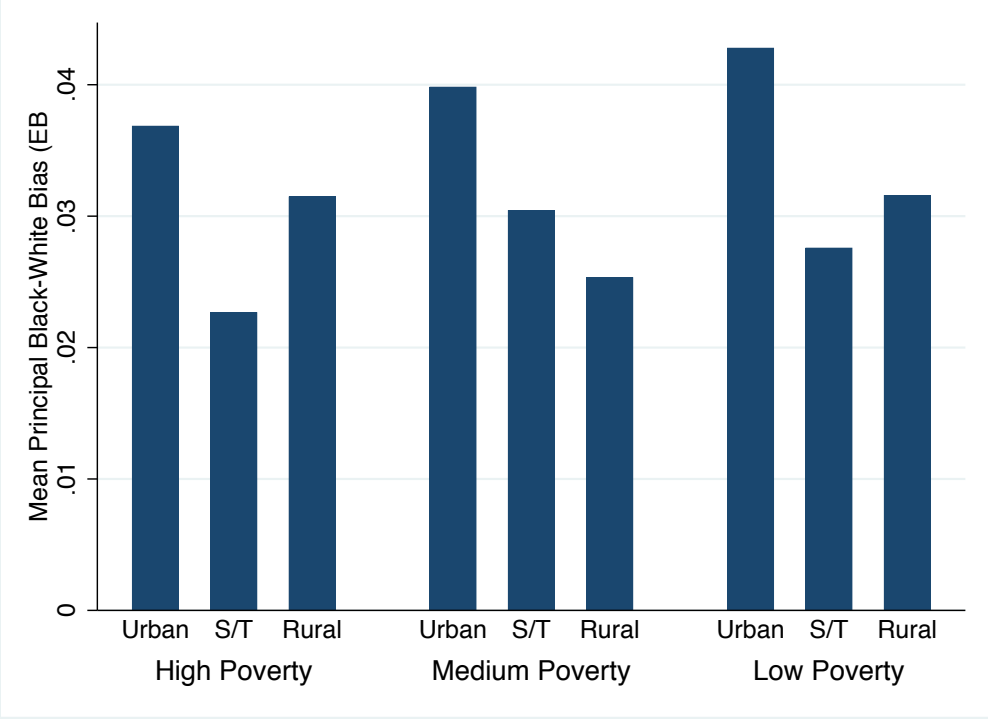
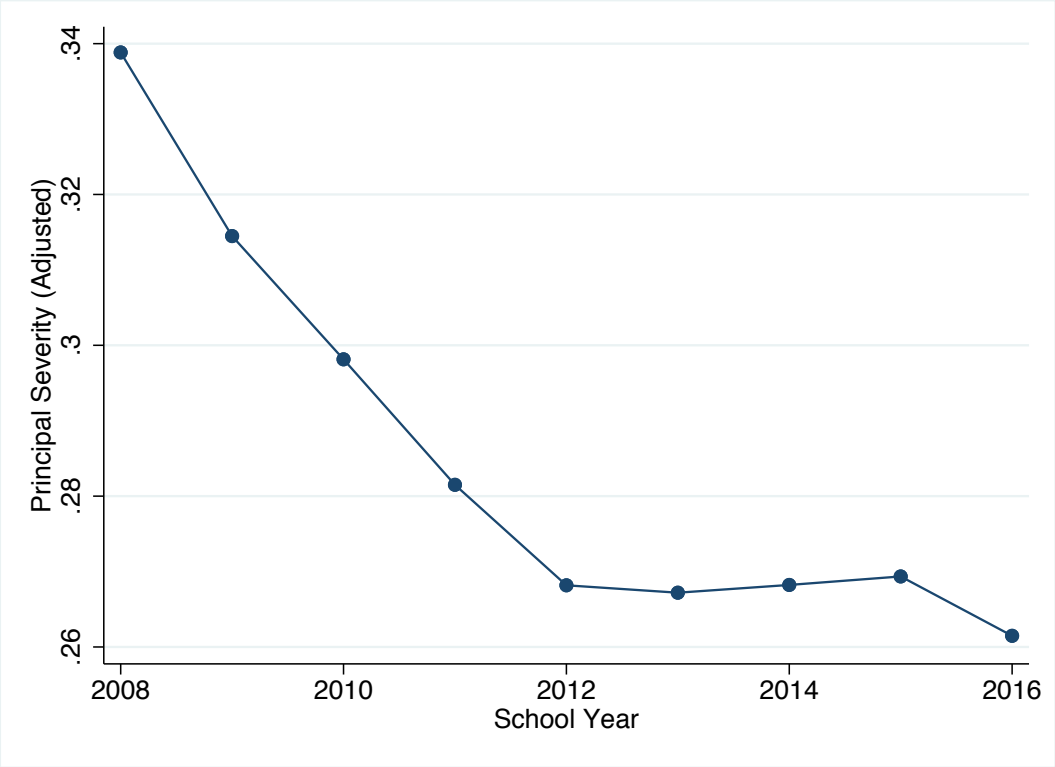


Figure A3. Average Severity of Current Middle School Principals by Year



*****Lucy: add table of random effects model**

Table A1. Number of Principals Serving at Each School

Number of Principals	Frequency (Sch-Year Observations)	Percent of Sch-Year Observations
1	227,643	8.07
2	916,507	32.48
3	921,701	32.66
4	534,798	18.95
5	195,759	6.94
6	20,758	0.74
7	4,907	0.17
Total	822,073	100

Table A2. Correlation Matrix of Principal Value-Added, Severity, Bias, and Experience

	V-A ELA	V-A Math	Severity	Bias	Experience
V-A ELA	1				
V-A Math	0.0662	1			
Severity	0.0279	0.0265	1		
Bias	-0.0056	-0.0643	0.1483	1	
Experience	0.0002	-0.005	0.0518	0.027	1

Table A3. Effects of Principal Severity on Short-Term Outcomes, Controlling for Principal Characteristics

Variables	(1) Any Offense	(2) Any Removal	(3) Days Absent	(4) Test Scores (SD)	(5) Grade Retention
Principal Severity	-0.2437**	0.1341**	0.7129*	0.0331	0.0042
	(0.020)	(0.010)	(0.292)	(0.027)	(0.003)
Any Offense		0.4864**			
		(0.008)			
Principal Experience	0.0003	-0.0003**	-0.0042	0.0003	0.0000
	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)
Principal V-A Read	0.0521+	-0.0081	-0.0658	0.0526	0.0001
	(0.027)	(0.012)	(0.355)	(0.041)	(0.003)
Principal V-A Math	-0.0012	-0.0080	-0.0921	0.6162**	0.0098*
	(0.028)	(0.012)	(0.389)	(0.050)	(0.005)
Principal Female	0.0023	0.0021	-0.0158	-0.0039	-0.0001
	(0.004)	(0.002)	(0.066)	(0.006)	(0.001)
Principal Black	0.0028	-0.0031	0.0231	-0.0139+	-0.0008
	(0.005)	(0.002)	(0.100)	(0.007)	(0.001)
Principal Hispanic	0.0226+	0.0269**	0.1098	-0.0507+	-0.0059**
	(0.012)	(0.004)	(0.097)	(0.030)	(0.001)
Principal Other	0.0141	-0.0015	-0.2678	-0.0037	-0.0032
	(0.013)	(0.005)	(0.172)	(0.015)	(0.004)
Covariates	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES
Observations	2,651,842	2,651,842	2,332,712	2,645,586	2,233,579
R-Squared	0.149	0.467	0.063	0.284	0.017

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

Robust standard errors in parentheses, clustered by school. Coefficients on student and school control variables not shown (available upon request).

Table A4. Effects of Principal Severity on Long-Term Outcomes, Controlling for Principal Characteristics

Variables	(1) High School Graduation	(2) High School Graduation	(3) Criminal Conviction by Age 20	(4) Criminal Conviction by Age 20
Principal Severity (EB)	-0.2170**	-0.2213**	0.1308**	0.0919**
	(0.050)	(0.050)	(0.030)	(0.028)
Severity x Any Removal		0.0102		0.0732**
		(0.023)		(0.014)
Any Removal		-0.1825**		0.0373**
		(0.008)		(0.005)
Principal Experience	-0.0018**	-0.0018**	-0.0002	-0.0002
	(0.001)	(0.001)	(0.000)	(0.000)
Principal V-A Read	0.4815**	0.4000**	-0.2077**	-0.1698**
	(0.090)	(0.089)	(0.049)	(0.047)
Principal V-A Math	0.0038	-0.0506	-0.1449**	-0.1198**
	(0.086)	(0.085)	(0.044)	(0.042)
Principal Female	-0.0025	-0.0022	-0.0015	-0.0016
	(0.013)	(0.013)	(0.007)	(0.007)
Principal Black	0.0151	0.0170	-0.0092	-0.0097
	(0.015)	(0.015)	(0.008)	(0.007)
Principal Hispanic	0.1511	0.1922*	0.0005	-0.0175
	(0.108)	(0.095)	(0.024)	(0.024)
Principal Other	-0.0490	-0.0491	-0.0180	-0.0219
	(0.041)	(0.042)	(0.029)	(0.028)
Covariates	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Observations	291,162	291,162	94,361	94,361
R-Squared	0.079	0.103	0.056	0.075

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

Robust standard errors in parentheses, clustered by school. Coefficients on student and school control variables not shown (available upon request).