

The Collateral Damage of In-School Suspensions: A Counterfactual Analysis of High-Suspension Schools, Math Achievement and College Attendance

Urban Education

1–37

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0042085920902256

journals.sagepub.com/home/ue



Jason Jabbari¹  and Odis Johnson Jr.¹

Abstract

Even the least severe forms of exclusionary discipline are associated with detrimental effects for students that attend schools that overuse them. With a nationally representative longitudinal study of high school students, we utilize propensity score weighting to limit selection bias associated with schools that issue high numbers of in-school suspensions. Accounting for school social order and individual suspensions, we find that high-suspension schools are negatively associated with students' math achievement and college attendance. We also find that when we account for high and low-suspension schools, attending an urban schools is associated with an increase in both math achievement and college attendance.

Keywords

discipline policies, African American students, urban education, mathematics, college

¹Washington University in St. Louis, MO, USA

Corresponding Author:

Jason Jabbari, 1 Brookings Dr, St. Louis, MO 63130, Campus Box 1196, USA.

Email: jabbari.jason@wustl.edu

Introduction

Since the onset of zero tolerance policies in the early 1990s, U.S. schools have increased their mechanisms of surveillance (e.g., school resource officers), as well as their menu of punishments (Kafka, 2011). At the same time, many schools have adopted an authoritarian approach to discipline, relying heavily on exclusionary practices, such as suspensions. These approaches may be especially prevalent in urban communities, which “have often been characterized by social isolation, heightened police surveillance, perceived family dysfunction, high rates of unemployment and poverty, high rates of violent crime, and overcrowded and underfunded schools” (Peguero et al., 2018, p. 7). As a result of these exclusionary practices, many students have been pushed closer toward the criminal justice system. A statewide study of Texas middle and high school students (Fabelo et al., 2011) found that 31% of suspended students repeated a grade, while 10% of suspended students dropped out. Demonstrating that the school-to-prison pipeline is “more than a metaphor,” Fabelo and his colleagues (2011) also found that nearly half of students with multiple (11+) suspensions were in contact with the juvenile justice system. At the same time, exclusionary practices have also moved students further away from academic achievement: using a nationally representative sample of students, Jabbari and Johnson (2020) found that being suspended in high school reduced the chances of taking advanced math courses, as well as attending college.

Moreover, the use of these exclusionary practices, as well as the impacts associated with them, have been found to vary across schools. When accounting for student and school risk factors, 23% of Texas high schools had discipline rates that were higher than what was projected, while 27% of schools had discipline rates that were lower than what was projected (Fabelo et al., 2011, p. xii). Even when accounting for a robust array of individual and school-level characteristics related advanced math course-taking and college attendance, Jabbari and Johnson (2020) still observed that a significant amount of the variation in the relationships among suspensions and these measures of achievement and attainment occurred between schools. Furthermore, given the segregated nature of U.S. schools and the discriminatory nature of suspensions (see Ibrahim & Johnson, 2019), separate studies on Arkansas (Anderson & Ritter, 2017) and another Midwestern state (Skiba et al., 2014) have demonstrated that much of the racial/ethnic disproportionality in discipline occurs *between* schools, rather than *within* them. Thus, breaking down the school-to-prison pipeline and ensuring more equitable outcomes in the future will require a greater understanding of high-suspension schools. Nevertheless, as the majority of students may not

directly receive a suspension in a given year—even in a *high-suspension school*, some stakeholders might still question the need to reduce high rates of school suspensions. As recent school shootings have increased concerns for student safety (see Johnson et al., 2019), stakeholders might also question the need to reduce high rates of *out-of-school* suspensions, which may be reserved for serious and potentially violent behaviors. As a result, we will explore how *non-suspended* students are impacted by attending schools that issue large amounts of in-school suspensions. To our knowledge, there is no research that has explored the indirect effects of in-school suspensions.

As schools that issue large amounts of in-school suspensions might represent environments with high levels of coercive control (see Kupchik et al., 2015), we explore the impacts that high-suspension schools have on learning that appears antithetical to coercive control. In doing so, we extend Jabbari and Johnson's (2020) previous work on in-schools suspensions by focusing on math, a subject in which mastery can often entail high levels of problem-solving and teamwork (see Adams & Hamm, 2010). As students often use their math knowledge to access post-secondary opportunities, we also consider the impact of high-suspension schools on college attendance.

In addition, while high rates of suspensions are often framed as an urban problem (see Losen & Skiba, 2010), recent research has demonstrated that the high-suspension schools—and the negative effects associated with them—extend to suburban and rural areas as well (Peguero et al., 2018). Therefore, while we conceptualize urban education as occurring in both major (urban intensive) and large cities (urban emergent), we recognize that the hyper-disciplining of vulnerable student populations is an urban characteristic that can apply to a variety of geographic contexts (see Milner, 2012). Finally, as we approach urban education from a policy and reform perspective, we attempt to isolate the inside-of-school factors—namely the relationship between suspensions and achievement—by controlling for outside-of-school factors (see Milner & Lomotey, 2014). We do so by employing a counterfactual framework that addresses selection bias associated with attending high-suspension schools. We pose the following questions:

- I. *What are the short-term (math achievement) and long-term (college attendance) impacts associated with attending a high-suspension school?*
- II. *What is the relationship between math achievement and college attendance in the context of high-suspension schools?*
- III. *How do student and school background characteristics, such as race/ethnicity, gender, social class, urbanicity, and school social order, relate to high suspension schools?*

We find that when controlling for selection into schools, students attending high-suspension high schools were associated with lower math achievement scores during their junior year of high school and were less likely to attend college—even when accounting for student-level suspensions and school-level social order. Moreover, we find that when we add junior year math achievement to the model predicting college attendance, the impact of high-suspension schools no longer remains significant, suggesting that high math achievement might operate as a protective barrier in schools that issue high numbers of suspensions. Furthermore, we find significant relationships among student and school background characteristics, high-suspension schools, math achievement, and college attendance. Most notably, we find that when we account for high- and low-suspension schools, attending urban schools is associated with an increase in both math achievement and college attendance.

Theoretical Framework: Social Control

Social control has been theorized to reduce anti-social behavior, maintain social order, and enhance the safety and wellbeing of societies and institutions *through the use of discipline* (see Durkheim et al., 1961). As an instrument of social control, discipline can be enacted within groups or communities to achieve internal regulation (informal social control) or externally through the actions of state agents (formal social control) (see Kirk, 2009). Within educational institutions, discipline takes on the added purpose of socializing youth toward adult roles and responsibilities, as well as ensuring the process of learning (Durkheim et al., 1961).

In the school discipline literature researchers often focus on circumstances where social control has become counterproductive (see Irby, 2014). While counterproductive social control can manifest itself in both overly strict (i.e., high social control schools) and overly lenient (i.e., low social control schools) environments, much of the research on school discipline focuses on the former: “In such situations, punishment becomes an end in itself, not an occasional means to an end of normative social order” (Perry & Morris, 2014, p. 5). To this end, previous research on school discipline has implicitly used social control to trace the historical transformation of overt racism and the legacy of slavery into modern school practices (see Duncan, 2000; Wacquant, 2001). More recently, research on school discipline has explicitly used social control to (a) describe criminogenic environments (see Kupchik, 2010), (b) demonstrate how students perceive these environments (Portillos et al., 2012) and how these perceptions negatively impact student outcomes (Peguero et al., 2015), (c) detail the criminalization of student

behavior (see Basile et al., 2019) and how this criminalization increases student discipline gaps (see Shabazian, 2020), and (d) theorize how discipline may have “collateral consequences” for non-disciplined students (see Perry & Morris, 2014). Similar to Perry and Morris (2014), we use social control to theorize how overly punitive environments may have collateral damages for all students.

To provide a comprehensive overview of social control, we consider how high and low-social control schools can impact student outcomes in both positive *and* negative ways. In an ethnographic study of a high-social control urban high school, Nolan (2011) observed that the threat of undeserved punishment increased feelings of anxiety for both well and poorly behaved students alike. Increased anxiety may, in turn, have a negative impact on learning. On the other hand, high-social control schools may operate in accordance with Wilson and Kelling’s (1982) “Broken Window Theory,” where minor signs of disorder can lead to serious crimes. Thus, by having a low tolerance for minor offenses, high-social control schools may be able to create an environment of order and conformity; in doing so, these schools may be able to avoid more serious offenses in the future. In an ethnographic study of six high-achieving, high-social control urban schools, Whitman (2008) observed that by incessantly “sweating the small stuff,” these schools were able to increase feelings of safety for students. In contrast to anxiety, increased safety may have a positive impact on learning.

Conversely, low-social control schools may allow students’ misbehavior to go unchecked, which may embolden their classmates to misbehave as well (Wilson & Kelling, 1982). The learning environments in these schools may become increasingly disruptive and disengaging. As an alternative form of institutional racism, low-social control schools may embody “the soft bigotry of low-expectations.” Finally, in accordance with a *laissez-faire* approach to discipline (see Rogers & Freiberg, 1969), low-social control schools may grant students greater freedom of expression, which may promote self-discipline, psychological well-being, and ultimately an engaging, student-centered learning environment.

Literature Review: High-Suspension Schools

Previous research has demonstrated that suspensions are inequitably distributed according to social and demographic characteristics, such as race and ethnicity, and that these inequitable distributions—often referred to as “punishment gaps”—can eventually lead to “achievement gaps” (Morris & Perry, 2016). While schools are often segregated along these social and demographic lines, only recently has research begun to explore the impacts of this

inequitable distribution of suspensions at the school-level (see Peguero et al., 2018).

In a recent study with middle and high schools students in Texas *both* overly strict *and* overly lenient schools were associated with higher grade retention across urban, suburban, and rural school contexts (Peguero et al., 2018). While these results suggest the need for balance, achieving a perfect match between discipline and behavior may be difficult for many schools. Thus, school leaders and teachers may want to know which side of the spectrum they should err on. Given the recent increase in the rate of school discipline (see Losen & Martinez, 2013), despite decreasing rates of misbehavior (see Robers et al., 2014), it appears that many schools are currently erring on the side of being overly strict. Therefore, we focus our literature review on overly strict schools—using suspensions as a common measure of school strictness.

Determinants of High-Suspension Schools

Initially, we consider the possibility that high-suspension schools may arise in response to higher levels of social disorder. While student misbehavior has been declining in recent decades (see Robers et al., 2014), an uneven distribution of students with behavioral problems across schools might lead some schools to have relatively high rates of social disorder. For example, in a study of Kentucky middle schools, Christie et al. (2004) found that the number of school violations increased the rate of out-of-school suspensions. Therefore, high-suspension schools may be needed to address higher levels of social disorder.

Alternatively, high-suspension schools may arise from differences in detection practices rather than differential in rates of misbehavior (see Ditton, 1979). For example, as evident in some of the “No-Excuses” schools (see Goodman, 2013), entire school systems may choose to suspend students for minor transgressions. Hence, rather than an actual escalation in student misbehavior, schools can become high-suspensions schools when the threshold for which a suspension is triggered gets lowered to include less serious offenses.

Similarly, high-suspension schools may also arise from differences in adherence to discipline policies and practices. For example, in a qualitative study of Louisiana principals in schools that predominantly serve Black students, Mukuria (2002) found that principals of high-suspension schools strictly adhered to the discipline policies of their district, while principals of low-suspension schools modified the discipline policies of their district to fit the needs of their students. Furthermore, using statewide data from a Midwestern state, Skiba and his colleagues (2014) found that schools with

principals that don't favor preventive discipline practices were more likely to suspend students out of school. Moreover, using a statewide sample of public high schools from Virginia, Gregory et al. (2011) found that schools with a less supportive school climate, as well as a less *challenging* school environment (referred to as "Academic Press" by the authors), were more likely to suspend students.

In addition, high-suspension schools may arise from academic achievement. Christie et al.'s (2004) study on middle school students also found that standardized test achievement decreased the rate of out-of-school suspensions. Here, schools may use suspensions to remove low-achieving students to increase test scores and meet accountability demands. For example, using administrator records from Florida at the onset of a high-stakes testing reform, Figlio (2006) found that schools assign longer suspensions (both in-school and out-of-school) to low-achieving students during the testing window and within testing grades.

Finally, high-suspension schools may arise from racial/ethnic bias. For example, in a nationally representative sample, Welch and Payne (2010) demonstrated that an increase in the percent of Black students within a school was directly related to an increase in suspensions (both in-school and out-of-school)—even when accounting for school levels of misbehavior and disorder. Furthermore, Skiba and his colleagues (2014) also found that the relationship between the percentage of Black students within a school and out-of-school suspensions remained significant when school-level poverty, achievement, and administrator attitudes were accounted for.

Impacts of High-Suspension Schools

In terms of the consequences of high-suspension schools, Lee et al. (2011) used a statewide sample of public high schools from Virginia to demonstrate that when controlling for both school demographics and student attitudes, schools with higher suspension rates were associated with higher dropout rates for both Black and White students. In addition, using a sample of middle and high schools students from a large urban school district in Kentucky, Perry and Morris (2014), demonstrated that higher rates of out-of-school suspension had a negative impact on math and reading achievement for non-suspended students—even when controlling for school-level behavior. Alternatively, using a cross-sectional sample of middle-school students in North Carolina, Kinsler (2013) found that the number of days students were suspended out of school deterred their future infractions, which ultimately increased the math achievement of their peers. Essentially, while being suspended entails a loss of instructional time for the suspended student, Kinsler's

(2013) findings suggest that repeated exposure to a disruptive student may entail a greater “loss” of instructional time for his or her peers.

Nevertheless, many relevant questions have been left unanswered by previous research. For example, since much of the existing research focuses on out-of-school suspensions, we do not know the collateral consequences of less-severe exclusionary practices, such as in-school suspension. These more common exclusionary practices are often used for more subjective offenses, such as “disorderly conduct” and “willful defiance” (Watanabe, 2013), as well as offenses relating to personal expression, such as hairstyles and dress codes (see Morris, 2005). Thus, high rates of in-school suspension may further represent an environment of coercive control and what Annamma (2018) describes as a “pedagogy of pathologization.” Furthermore, a focus on outcomes at a single point in time within existing research has not revealed the duration of collateral consequences, nor how these consequences might be mediated over time. Moreover, we do not know whether the magnitude of these consequences changes when methods are used that limit selection bias associated with suspensions. In addition, we do not know how student and school background characteristics, such as race/ethnicity, gender, social class, urbanicity, and school social order, relate to high-suspension schools. Finally, since much of the existing research relies on localized samples, we do not know under what circumstances the effects of suspensions might apply more broadly to schools throughout the nation and across geographic contexts.

In extending the previous literature, we (a) establish the impacts of less-severe exclusionary policies through measures of in-school suspension, (b) explore both the short-term (math achievement) and long-term (college attendance) impacts associated with high and low-suspension schools and demonstrate how these impacts are related, (c) limit bias associated with attending high and low-suspension schools by using a counterfactual model based on propensity scores, (d) explore how student and school background characteristics relate to high-suspension schools, math achievement, and college attendance, and (e) rely on students from a nationally representative longitudinal sample.

Methods

Research Overview

In this article, we are primarily interested in the “collateral” damages of suspensions. Specifically, we are interested in the indirect impact of attending a high-suspension school (as opposed to a low-suspension school) for students

that are *not* directly suspended. Estimating the collateral damages of suspensions will require three main elements. First, as students' attendance in high-suspension schools is not random, we will control for selection bias through a counterfactual approach to ensure that it is not the students, themselves, that are causing schools to issue a high number of suspensions. Second, to isolate the impact of high-suspension schools on non-suspended students we will include a student-level measure of suspensions in our outcome models. Finally, as schools with higher levels of social disorder may issue greater numbers of suspensions, we will include a school-level measure of social order in our outcome models. In doing so, we are able to estimate the collateral damages of suspensions that exist beyond the behaviors of both students and schools.

Research Design

Attendance in high-suspension schools is not random. Thus, estimating the impacts of attending high-suspension schools *without* adjusting for students' non-random attendance into these schools can yield biased results. We therefore employ a counterfactual framework based on propensity scores to adjust for students' non-random attendance in these schools. In a counterfactual framework treatment and control participants have potential outcomes in both states: the state in which they are observed in and the state in which they are not observed in (Rubin, 2005). In our counterfactual framework, students who attend high-suspension schools are viewed as being assigned to the "treatment" group, while students who attend low-suspension schools are viewed as being assigned to the "control" group. Within this counterfactual framework, propensity scores define the conditional probability of being "assigned" to a high- or low-suspension school based on a set of observed characteristics (see Rosenbaum & Rubin, 1983). Here, propensity scores can be seen as balancing property: "conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects" (Austin, 2011).

Specifically, we use propensity score weighting to balance students in the treatment and control groups, which utilizes the inverse probability for receiving the treatment (that the subject actually received) to weight these observations from a given sample (Austin, 2011). This allows for average treatment effects (ATE) to be estimated, which in this study is the difference in the potential outcomes associated with high-suspension schools for all students. In estimating our propensity scores, we include a set of observed variables that are related to the treatment (high-suspension schools), the underlying treatment mechanisms (suspensions), and ultimately, the outcomes associated with

the treatment (math achievement and college attendance). The inclusion of these variables will not only limit potential biases in our treatment assignment, but will also balance students' pre-dispositional characteristics that are related to the underlying treatment mechanisms, as well the outcomes in our analyses. Furthermore, to meet the temporal assumption that the treatment occurred before the outcome, we primarily include variables that occur before treatment assignment (i.e., high school attendance). Moreover, in following Guo and Fraser's (2014) recommendations for propensity score analysis, we utilize individual-level variables to estimate propensity scores for our group-level treatment. In doing so, the ATE weights for participants in the treatment group (high-suspension school attendees) are defined as $w_i = 1/p(x_i)$, while the ATE weights for participants in the control group (low-suspension school attendees) are defined as $w_i = 1/(1 - p(x_i))$ (Guo & Fraser, 2014).

When considering the nature of our treatment and data, PSW has two distinct advantages when compared to other propensity score strategies. First, PSW maintains all participants, which is especially desirable in an analysis where some participants will be automatically lost by using a treatment that only includes students from high- and low-suspension schools. Second, PSW allows for greater generalizability with the ability to easily multiply propensity score weights with survey weights in complex data, such as ours. Nevertheless, while counterfactual frameworks can allow researchers to make inferences that approach causality, because the initial measure of our math achievement outcome occurred *during* the treatment and because there is not an exact pre-treatment measure of our college attendance outcome, our counterfactual framework will allow for *associational* claims that are less prone to selection bias.

Data

Description. Our analyses use restricted-use data from the High School Longitudinal Study of 2009 (HSLs). We use the HSLs because it is the most recent nationally representative longitudinal study of high school students in the United States. In the HSLs's stratified random sampling design, 944 schools were selected in the first stage, and an average of 27 ninth-graders were selected from these schools in the second stage for a total of 25,206 eligible students (Ingels et al., 2011). Our analyses used two series of longitudinal waves. The first analysis, which tests the impact of attending a high-suspension school on math achievement, spans across the first wave and (9th grade) and the second wave (11th grade). The second analysis, which tests the impact of attending a high-suspension high school on college attendance, spans across the first wave (9th grade), second wave (11th grade), and fourth

wave (freshman year of college). Each analysis used information from used student, parent, and administrator questionnaires.

Imputation. We used multiple imputation with chained equations (MICE) to impute five sets of missing values. Not including key demographic and dependent variables, which we did not impute, most independent variables had less than 5% of their responses missing. The only exception was the measure of school social order, which was missing 23% of the responses in the original sample. Here, we successfully imputed 87% of the missing responses for this measure.

Weighting. Out of the 8,744 original students from high- and low-suspension schools in our sample, attrition across these waves (for different time points) and within waves (for different questionnaire types) resulted in sample sizes ranging from 7,680 to 7,920 students in the final analyses. Nevertheless, the National Center for Education Statistics (NCES) did provide analytic weights that account for these instances of non-response, which limit potential biases that can arise from attrition. In our math achievement analysis, we used the W2W1STU weight, which is recommended for student analyses that span across the base year and first follow-up waves. In our college attendance analysis, we used the W3W1W2STUTR weight, which is recommended for student analyses that span across the base year, first follow up, and college transcript waves.

Measures

Treatment variable. The treatment variable in this study is attending a high-suspension school, as opposed to attending a low-suspension high school (1 = *high-suspension school*; 0 = *low-suspension school*). This treatment variable was derived from a student-level, self-reported measure of in-school suspension collected during the first follow-up wave. This measure occurred on the following scale: 0 = *not suspended in the previous 6 months*, 1 = *suspended 1–2 times in the previous 6 months*, 2 = *suspended 3–6 times in the previous 6 months*, 3 = *suspended 7–9 times in the previous 6 months*, and 4 = *suspended 10 or more times in the previous 6 months*. Using the original survey weight (W1STUDENT), which provides a representative estimate of both schools and the students attending them at the start of the treatment, we calculated a survey-weighted mean of suspensions for each individual school in the survey. This created a school-level measure of in-school suspensions. Based on this measure, schools were then broken down into five quintiles that mirrored the weighted distribution of suspensions.

Table 1. Suspension Quintiles.

School suspension quintiles	Student suspensions					Number of students/ schools
	0 Suspensions	1–2 Susp.	3–6 Susp.	7–9 Susp.	10+ Susp.	
Lowest quintile	4,672	36	4	2	1	4,715/233
Middle/low quintile	3,250	176	10	0	5	3,441/145
Middle quintile	3,668	347	48	4	6	4,073/180
Middle/high quintile	3,437	497	105	24	16	4,079/194
Highest quintile	2,984	752	147	48	71	4,029/192

Because we are interested in the high and low extremes of school suspensions, we utilized the highest and lowest quintiles to create our treatment variable. The highest quintile (192 schools with 4,029 students) was operationalized as high-suspension schools, while the lowest quintile (233 schools with 4,715 students) was operationalized as low-suspension schools. Descriptive information on the treatment variable can be found in Table 1.

Variables in the propensity score estimation model. Stemming from the literature on high-suspension schools, which demonstrates that the overuse of suspensions often manifests itself in schools that predominantly serve students of color, we include the following demographic variables as treatment covariates in the propensity score estimation model: Black race/ethnicity (1 = *yes*; 0 = *no*), Hispanic race/ethnicity (1 = *yes*; 0 = *no*), socioeconomic status (SES) quintile (created by the NCES and derived from parent education, parent occupation, and family income, SES quintiles range from 1 to 5 with 1 representing the lowest quintile and 5 representing the highest quintile), and urban school location (1 = *yes*; 0 = *no*), which entails attending a school that is both inside an urbanized area *and* inside a principal city. In addition, to balance covariates that are also related to suspensions, such as gender (see Skiba et al., 2002) and household structure (see Manning & Lamb, 2003), the propensity score estimation model included being female (1 = *yes*; 0 = *no*) and having two parents/guardians at home (1 = *two-parent/guardian household*; 0 = *single-parent/guardian household*). Two separate scales depicting how often parents were previously contacted about their child's misbehavior and poor academic performance during their child's eighth grade year were also included (1 = *never*; 2 = *once or twice*; 3 = *three or four times*; 4 = *more than four times*). Finally, to balance covariates that are related to the outcomes in our analyses, pre-treatment math achievement and college attendance variables were included in the propensity score estimation model. This included a scale of advanced math course-taking during students' eighth

Table 2. Comparison of Treatment Selection Variables before Propensity Score Weighting.

Variable	High-suspension school	Low-suspension school	Standardized difference	p value
Race: Black	0.33	0.14	0.45	.00
Race: Hispanic	0.22	0.21	0.03	.46
Gender: Female	0.48	0.52	-0.09	.01
SES quintile	2.64	3.53	-0.62	.00
Urban school location	0.29	0.35	-0.14	.00
Two parent household	0.67	0.83	-0.35	.00
High parental college expectations	0.63	0.79	-0.34	.00
Eighth grade behavior	1.52	1.25	0.37	.00
Eighth grade performance	1.45	1.31	0.20	.00
Eighth grade math course	3.27	3.48	-0.11	.00
Eighth grade math grade	2.25	1.92	0.33	.00
Observations	4,150	4,710		
Effective sample size	2,052.88	1,949.67		

Note. Above results from Multiple Imputation set #1. Due to space limitations results from other Multiple Imputation sets were not included. However, it is worth noting that their results were nearly identical. SES = socioeconomic status.

grade year—ranging from 1 (“Math 8”) to 9 (“Other advanced math course such as pre-calculus or calculus”) and a scale of grades received in these math courses during students’ eighth grade year (1 = “A”; 2 = “B”; 3 = “C”; 4 = “D”; 5 = “below D”). A measure of parental expectations for their child’s college attainment was also included (1 = *child will receive a bachelor’s degree*; 0 = *child will not receive a bachelor’s degree/doesn’t know if child will receive a bachelor’s degree*).

As seen in Table 2, in the absence of propensity score weighting, high-suspension schools had more Black students than low-suspension schools (33% compared to 14%); less female students (48% compared to 52%), more students with lower SES quintiles (2.64 compared to 3.53); less students attending schools in urban areas (29% compared 35%); less students from two-parent/guardian households (67% compared to 83%); more students who were frequently contacted about their negative behavior and performance; more students who had taken less advanced math courses in eighth grade; more students that received lower math grades in their eighth grade math

courses; and less students with parents that expect them to go to college (63% compared 79%).

Outcome variables. The short-term outcome variable consisted of a norm-referenced math achievement test score taken during the spring of 11th grade (ranging from 22.24 to 84.91). This test was developed by the NCES to reflect growth in math achievement and preparedness for college science, technology, engineering and mathematics (STEM) programs; it primarily focused on algebraic reasoning and contained more difficult items than a similar test taken in the fall of ninth grade (Ingels et al., 2011). The long-term outcome variable consisted of full-time college attendance recorded during the fall of a student's freshman year of college (1 = *yes*; 0 = *no*).

Covariates in the outcome models. To isolate the impact of the treatment, final math achievement test scores were controlled for by students' initial math achievement test scores (ranging from 24.02 to 82.19) collected during the first wave. Similarly, full-time college attendance was controlled for by a students' initial expectations for graduating college (1 = *student will not receive a bachelor's degree/student doesn't know if he or she will receive a bachelor's degree*; 0 = *student expects to receive a bachelor's degree*). Furthermore, to estimate the collateral effects of high-suspension schools (i.e., the effects for students that are *not* suspended) our analyses includes students' individual suspension rates. Moreover, to further operationalize the treatment as a function of school practices and *not* a function of student behaviors, each analysis included a school measure of social order. This continuous scale of social order (ranging from -4.22 to 1.97 with higher values representing higher levels of social order) was provided by the NCES and created through principal component factor analysis from administrator frequency ratings of the following activities at his or her school: physical conflicts, robberies, vandalism, drug use, alcohol use, drug sales, weapon possessions, physical abuse of teachers, racial tensions, bullying, verbal abuse of teachers, in-class misbehavior, disrespect toward teachers, and gang activities. The Cronbach's alpha of this scale was 0.88.

In addition, as high-suspension schools might have the largest impact on individuals who attend them most regularly, the number of student absences (0 = *no absences*; 1 = *1 or 2 absences*; 2 = *3–6 times absences*; 3 = *7–9 absences*; 4 = *10 or more absences*) and classes skipped (0 = *no classes skipped*; 1 = *1 or 2 classes skipped*; 2 = *3–6 classes skipped*; 3 = *7–6 classes skipped*; 4 = *10 or more classes skipped*) were included in each analysis. Finally, indicators for Black, Hispanic, female, SES quintile, and urban school

location were included as model covariates to ensure robustness of the treatment impacts (see Bang & Robins, 2005), as well as to provide insight into how these factors impact the outcomes *after* we account for the extremes of the distribution of suspensions across schools. Also, to allow for meaningful inferences of the intercepts in the outcome analyses freshman and junior-year math test scores were mean-centered, while SES-quintile was rescaled to include zero (-2, -1, 0, 1, 2).

When considering the descriptive differences between the treatment and control groups (Table 3), high-suspension schools have students with lower freshman and junior-year math test scores, lower college attendance rates and expectation levels, lower levels of school social order, and higher rates of absences and classes skipped. However, when students are balanced on pre-treatment characteristics through propensity score weighting, these differences partially dissipate. This was also the case when comparing the correlation tables that use survey weights (Table 4) with the correlation tables that use propensity score weights (Table 5).

Propensity Score Estimation

Modeling. Nonparametric modeling techniques, such as generalized boosted modeling (GBM), have the ability to reduce the chance of misspecification errors in the estimation of propensity scores (see McCaffrey et al., 2004). GBM utilizes automated, data adaptive modeling algorithms to “predict treatment assignment from a large number of pretreatment covariates while also allowing for flexible, non-linear relationships between the covariates and the propensity score” (p. 3). Specifically, we utilized the TWANG—Toolkit for Weighting and Analysis of Non-equivalent Groups—package (Ridgeway et al., 2014) in STATA to estimate our propensity score weights through GBM. Finally, as recommended by DuGoff et al. (2014) for inferences on populations (as opposed to samples), we used TWANG to multiply the propensity score weights by the appropriate survey weights.

Balancing. Using TWANG’s default settings, we assessed the mean effect sizes for covariate balance. Results of the propensity score estimation models demonstrate that all treatment covariates were properly balanced (Table 6). In addition, propensity scores for both treatment and control groups shared an adequate region of common support (Figure 1), which ensures that participants with similar treatment covariates have a positive theoretical probability of being in either the treatment or control group (Rosenbaum & Rubin, 1983).

Table 3. Descriptive Statistics.

Variable	Survey-weighted				Propensity-weighted			
	Treatment group		Control group		Treatment group		Control group	
	M	SD	M	SD	M	SD	M	SD
Freshman year math score	-2.26	9.37	3.34	9.48	0.06	9.74	1.33	9.68
Junior year math score	-3.07	9.22	3.67	9.61	-0.88	9.66	1.85	9.59
College attendance	0.53	0.50	0.78	0.42	0.61	0.49	0.70	0.46
Low college expectation	0.45	0.50	0.34	0.47	0.39	0.49	0.38	0.49
In-school suspension	0.40	0.79	0.00	N/A	0.32	0.73	0.00	N/A
School social order	-0.19	1.02	0.48	0.98	-0.18	1.01	0.41	0.99
Absences	1.56	1.06	1.39	1.01	1.52	1.05	1.42	1.02
Classes skipped	0.38	0.86	0.20	0.60	0.34	0.82	0.22	0.61
Race: Black	0.31	0.46	0.12	0.32	0.22	0.41	0.19	0.40
Race: Hispanic	0.19	0.39	0.18	0.38	0.19	0.39	0.18	0.38
Gender: female	0.50	0.50	0.54	0.50	0.51	0.50	0.52	0.50
SES quintile	-0.26	1.35	0.60	1.39	0.18	1.43	0.19	1.45
Urban school location	0.26	0.44	0.33	0.47	0.30	0.46	0.29	0.45
Observations	4,290		3,620		4,290		3,620	

Note. Unweighted population statistics, such as the number of observations, have been rounded to the nearest 10 to comply with our restricted use data license agreement.
SES = socioeconomic status.

Results

Each analysis used STATA’s SVY program (StataCorp, 2013), which is designed for the analysis of complex surveys. After demonstrating the significant impacts of our treatment in unconditional models (Table 7), we created a series of conditional models that allow us to demonstrate the impacts of our treatment *in the presence* of control variables. Our conditional models also allow us to demonstrate the impact of controlling for treatment assignment, as well as the impact of including the treatment.

First, we ran “propensity” models, which controlled for treatment assignment with the inclusion of propensity score weights. Next, we juxtaposed these models with “standard” models, which did not control for treatment assignment with propensity score weights (original survey weights were used instead). We then juxtaposed these models with “null” models—that did not include the treatment—to demonstrate the difference between attending a high-suspension school and directly receiving a suspension (across high and

Table 4. Survey-Weighted Correlation Table.

	High- suspension school	Freshman year math score	Junior year math score	College attendance	Low college expectation	In-school suspension	School social order	Absences	Classes skipped	Race: Black	Race: Hispanic	Gender: female	SES quintile	Urban school location
High- suspension school	1.00													
Freshman year math score	-0.28	1.00												
Junior year math score	-0.34	0.75	1.00											
College attendance	-0.26	0.36	0.41	1.00										
Low college expectation	0.12	-0.31	-0.30	-0.31	1.00									
In-school suspension	0.33	-0.24	-0.27	-0.26	0.15	1.00								
School social order	-0.32	0.17	0.18	0.15	-0.09	-0.11	1.00							
Absences	0.08	-0.07	-0.13	-0.17	0.05	0.21	-0.04	1.00						
Classes skipped	0.11	-0.15	-0.16	-0.20	0.09	0.39	-0.09	0.29	1.00					
Race: Black	0.24	-0.25	-0.24	-0.10	-0.01	0.14	-0.13	-0.05	0.04	1.00				
Race: Hispanic	0.02	-0.08	-0.09	-0.12	0.08	0.00	-0.14	-0.01	0.08	-0.10	1.00			
Gender: female	-0.04	0.01	-0.02	0.07	-0.07	-0.12	0.00	0.04	-0.04	0.02	-0.01	1.00		
SES quintile	-0.30	0.41	0.42	0.39	-0.24	-0.19	0.20	-0.06	-0.12	-0.18	-0.29	0.02	1.00	
Urban school Location	-0.08	0.03	0.05	0.02	-0.02	0.01	-0.08	-0.03	0.08	0.06	0.15	0.03	-0.01	1.00

Note. Above results from Multiple Imputation Set #1. Due to space limitations results from other multiple imputation sets were not included. However, it is worth noting that their results were nearly identical. Bolded values are significant beyond the $p < .05$ level. SES = socioeconomic status.

Table 5. Propensity-Weighted Correlation Table.

	High- suspension school	Freshman year Math Score	Junior year math Score	College attendance	Low college expectation	In-school suspension	School social order	Absences	Classes skipped	Race: Black	Race: Hispanic	Gender: female	SES quintile	Urban school location
High-suspension school	1.00													
Freshman year math score	-0.07	1.00												
Junior year math score	-0.14	0.73	1.00											
College attendance	-0.09	0.35	0.40	1.00										
Low college expectation	0.01	-0.30	-0.29	-0.31	1.00									
In-school suspension	0.29	-0.17	-0.22	-0.20	0.13	1.00								
School social order	-0.28	0.09	0.11	-0.06	-0.06	-0.09	1.00							
Absences	0.05	-0.06	-0.12	-0.16	0.06	0.20	-0.05	1.00						
Classes skipped	0.08	-0.13	-0.14	-0.18	0.09	0.38	-0.07	0.28	1.00					
Race: Black	0.03	-0.22	-0.22	-0.10	0.00	0.07	-0.08	-0.05	0.02	1.00				
Race: Hispanic	0.01	-0.08	-0.10	-0.11	0.07	0.00	-0.13	0.00	0.07	-0.10	1.00			
Gender: female	-0.01	0.02	-0.02	0.07	-0.06	-0.11	0.00	0.05	-0.03	0.04	-0.01	1.00		
SES quintile	0.00	0.41	0.40	0.38	-0.26	-0.12	0.11	-0.05	-0.09	-0.17	-0.27	0.02	1.00	
Urban school location	0.01	0.05	0.06	0.03	-0.03	0.02	-0.13	-0.01	0.10	0.06	0.14	0.03	0.00	1.00

Note. Above results from Multiple Imputation Set #1. Due to space limitations results from other Multiple Imputation sets were not included. However, it is worth noting that their results were nearly identical. Bolded values are significant beyond the $p < .05$ level. SES = socioeconomic status.

Table 6. Comparison of Treatment Selection Variables after Propensity Score Weighting.

Variable	High-suspension school	Low-suspension school	Standardized difference	p value
Race: Black	0.24	0.22	0.03	.45
Race: Hispanic	0.21	0.21	0.00	.90
Gender: Female	0.50	0.50	0.00	1.00
SES quintile	3.06	3.11	-0.04	.26
Urban school location	0.31	0.31	0.01	.87
Two parent household	0.75	0.76	-0.03	.49
High parental college Expectations	0.72	0.73	-0.03	.46
Eighth grade behavior	1.38	1.35	0.04	.34
Eight grade performance	1.37	1.35	0.02	.60
Eighth grade math Course	3.36	3.37	-0.01	.87
Eighth grade math grade	2.08	2.06	0.02	.53
Observations	4,150	4,710		
Effective sample size	1,918.53	1,467.35		

Note. Above results from Multiple Imputation Set #1. Due to space limitations results from other Multiple Imputation sets were not included. However, it is worth noting that their results were nearly identical. SES = socioeconomic status.

low-suspension schools). Finally, to test the relationships between the short- and long-term outcomes, we created two additional sets of models for predicting college attendance: one model set included an additional control measure of freshman year math achievement test scores, and another model set included both freshman *and* junior year math achievement test scores.

Math Achievement Models

When controlling for attendance into high-suspension schools with propensity score weights (Model 1), attending a high-suspension school was associated with a 1.41 point decrease in junior year math achievement (Table 8). This was larger than the impact of directly being suspended, which was associated with a 0.98 point decrease in math achievement (for a one-unit increase in suspensions). Other negative predictors of math achievement were absences, identifying as Black, and identifying as female. Conversely,

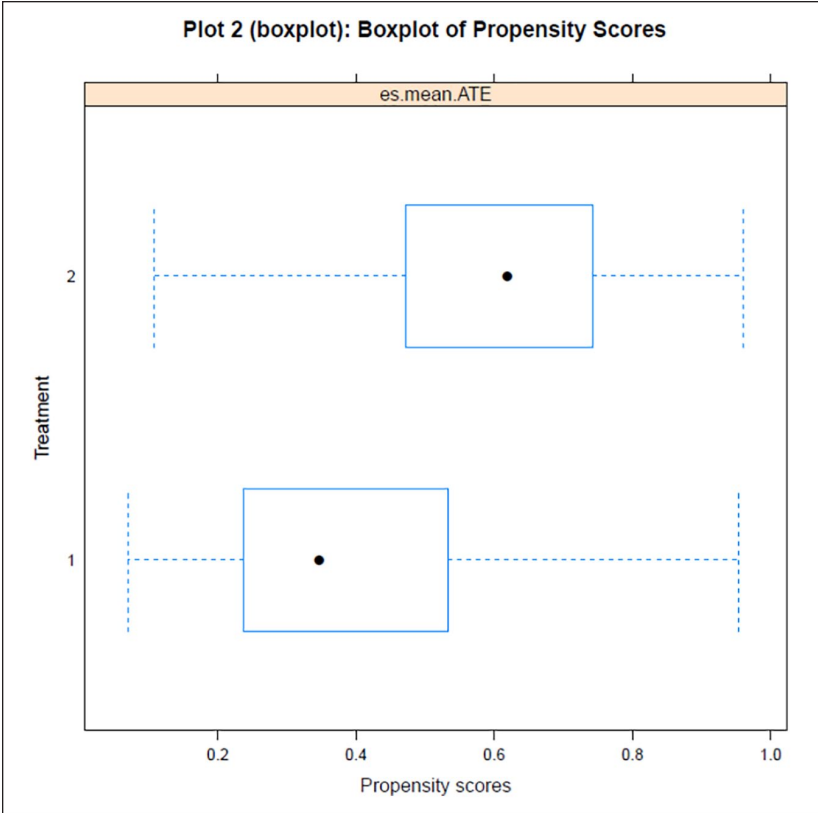


Figure 1. Boxplot of propensity scores.

Table 7. Unconditional Outcome Models.

	Math achievement: Propensity model	Math achievement: Standard model	College attendance: Propensity model	College attendance: Standard model
High-suspension school	-2.50 (0.64)***	-6.67 (0.58)***	0.67 (0.09)**	0.33 (0.04)***
Intercept	0.77 (0.48)	2.81 (0.43)***	2.36 (0.25)***	3.45 (0.32)***
Observations	7,830	7,830	7,920	7,920

Note. For Math Achievement Models, coefficients are provided, which are followed by robust standard errors in parentheses. For College Attendance Models, odds ratios are provided, which also are followed by robust standard errors in parentheses.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 8. Continuous Regressions of the Impact of High-Suspension Schools on Junior year Math Achievement.

	Model 1 (propensity model)	Model 2 (standard model)	Model 3 (null model)
High-suspension school	-1.41 (0.38)***	-1.73 (0.37)***	(omitted)
In-school suspension	-0.98 (0.25)***	-0.78 (0.25)**	-1.16 (0.24)***
Freshman year math score	0.65 (0.02)***	0.66 (0.01)***	0.66 (0.01)***
School social order	0.11 (0.17)	0.14 (0.16)	0.33 (0.15)*
Absences	-0.64 (0.12)***	-0.62 (0.11)***	-0.63 (0.12)***
Classes skipped	0.08 (0.16)	0.04 (0.17)	0.08 (0.17)
Race: Black	-1.21 (0.36)**	-1.06 (0.30)**	-1.29 (0.31)***
Race: Hispanic	-0.39 (0.48)	-0.23 (0.40)	-0.19 (0.40)
Gender: female	-0.53 (0.26)*	-0.48 (0.23)*	-0.47 (0.24)
SES quintile	0.61 (0.11)***	0.62 (0.11)***	0.72 (0.10)***
Urban school location	0.78 (0.36)**	0.69 (0.33)*	0.87 (0.34)*
Intercept	1.98 (0.41)***	1.97 (0.36)***	1.10 (0.30)***
Observations	7,680	7,680	7,680

Note. Coefficients followed by robust standard errors in parentheses. SES = socioeconomic status.

* $p < .05$. ** $p < .01$. *** $p < .001$.

freshman year math achievement scores (the primary control for the outcome), SES quintile, and attending an urban school were all positive predictors junior year math achievement.

When we did not control for selection into high-suspension schools with propensity score weights (Model 2), the impact of the treatment increased and was now associated with a 1.73 point decrease in junior year math achievement. With the exception of freshman year math achievement and SES quintile, which remained practically unchanged, the impact of all other significant predictors slightly decreased.

Finally, when we removed the treatment from the model (Model 3), the impact of directly receiving a suspension increased; for a one-unit increase in ISS, junior year math achievement scores were associated with a 1.16 point decrease. While the impact of freshman year math achievement, absences, and identifying as female remained practically unchanged (although gender lost statistical significance), the impact of identifying as Black, SES quintile, and attending an urban school slightly increased. In addition, school social

Table 9. Propensity Models: Propensity Score Weighted Logistic Regressions of the Impact of High-Suspension Schools on College Attendance.

	Model 4	Model 5	Model 6
High-suspension school	0.77 (0.09)*	0.79 (0.09)*	0.85 (0.11)
In-school suspension	0.76 (0.09)*	0.81 (0.09)	0.85 (0.09)
Low college expectation	0.36 (0.04)***	0.43 (0.04)***	0.45 (0.05)***
School social order	1.11 (0.08)	1.10 (0.08)	1.10 (0.08)
Absences	0.76 (0.04)***	0.75 (0.04)***	0.77 (0.04)***
Classes skipped	0.76 (0.06)**	0.78 (0.07)**	0.77 (0.07)**
Race: Black	0.72 (0.10)*	0.85 (0.12)	0.92 (0.13)
Race: Hispanic	0.88 (0.12)	0.87 (0.12)	0.91 (0.13)
Gender: Female	1.31 (0.12)**	1.34 (0.13)**	1.42 (0.14)***
SES quintile	1.67 (0.06)***	1.54 (0.06)***	1.51 (0.06)***
Urban school location	1.35 (0.18)*	1.28 (0.16)	1.22 (0.16)
Freshman year math score	(Omitted)	1.05 (0.01)***	1.01 (0.01)
Junior year math score	(Omitted)	(Omitted)	1.06 (0.01)***
Intercept	5.18 (0.69)***	4.82 (0.69)***	4.31 (0.62)***
Observations	7,920	7,920	7,920

Note. Odds ratios followed by robust standard errors in parentheses. SES = socioeconomic status.

* $p < .05$. ** $p < .01$. *** $p < .001$.

order, which was previously non-significant, was now associated with a significant increase in math achievement.

College Attendance Models

Propensity model results. In the first propensity model predicting college attendance—in which neither freshman nor junior year math achievement was included (Model 4), attending a high-suspension school was associated with a decrease in the odds of attending college to a ratio 0.77 to 1 (Table 9). This was nearly identical to the impact of directly being suspended, which was associated with a decrease in the relative odds of attending college to a ratio of 0.76 to 1 (for a one-unit increase in suspensions). In addition to having low-college expectations (the primary control for the outcome), absences, skipped classes, and identifying as Black were also negatively related to college attendance. Conversely, identifying as female, SES quintile, and attending in an urban school were positively related to college attendance.

In addition, when freshman year math achievement scores were added in Model 5, which turned out to be a significant predictor of college attendance

(a one point increase in math achievement scores was associated with an increase in the relative odds of college attendance to a ratio of 1.05 to 1), the impact of attending a high school slightly decreased, while the impact of directly being suspended no longer remained significant. When considering changes in the other predictors, identifying as Black and attending an urban school no longer remained significant predictors of college attendance.

Finally, when junior year math achievement scores were added in Model 6, which also turned out to be a significant predictor of college attendance (a one point increase in math achievement scores were associated with an increase in the relative odds of college attendance to a ratio of 1.06 to 1), additional changes occurred among the other predictors. Specifically, the impact of attending a high-suspension school, as well as the impact of freshman year math achievement, no longer remained significant predictors of the outcome.

Standard model results. Similar to the standard model predicting math achievement (Model 2), the impact of attending a high-suspension school increased when we did not control for selection into the treatment with propensity score weights (Table 10). As seen in Model 7, attending a high-suspension school was now associated with a decrease in the odds of college attendance to a ratio of 0.62 to 1. While the impact of identifying as female and SES quintile slightly decreased when compared to Model 7's equivalent propensity model (Model 4), the impact of identifying as Black and attending an urban school no longer remained significant in Model 7. All other model predictors remained similar to Model 7's equivalent propensity model (Model 4). When we added freshman year math achievement in Model 8, which again turned out to be a significant predictor of college attendance, the impact of attending a high-suspension school and directly being suspended slightly decreased. However, unlike its equivalent propensity model (Model 5), the direct impact of being suspended remained significant in Model 8. This was also the case when junior year math achievement scores were added in Model 9.

Null model results. In the first null model predicting college attendance—in which neither freshman nor junior year math achievement was included (Model 10), directly receiving a suspension was now associated with a decrease in the relative odds of attending college to a ratio of 0.61 to 1 (for a one-unit increase in suspensions) (Table 11). Unlike Model 10's equivalent standard model (Model 7), school social order and attending an urban school were positively related to college attendance. In addition, when we added freshman year math achievement scores in Model 11, which again turned out to be a significant predictor of college attendance, the impact of directly being suspended slightly decreased, but remained significant. Conversely,

Table 10. Standard Models: Non-Propensity Score Weighted Logistic Regressions of the Impact of High-Suspension Schools on College Attendance.

	Model 7	Model 8	Model 9
High-suspension school	0.62 (0.07)***	0.66 (0.08)**	0.72 (0.09)*
In-school suspension	0.69 (0.08)**	0.74 (0.08)**	0.78 (0.08)*
Low college expectation	0.36 (0.03)***	0.43 (0.04)***	0.46 (0.04)***
School social order	1.10 (0.07)	1.08 (0.07)	1.08 (0.07)
Absences	0.77 (0.04)***	0.76 (0.04)***	0.78 (0.04)***
Classes skipped	0.76 (0.06)**	0.78 (0.06)**	0.77 (0.06)*
Race: Black	0.87 (0.10)	1.02 (0.13)	1.10 (0.14)
Race: Hispanic	0.86 (0.12)	0.86 (0.12)	0.88 (0.12)
Gender: female	1.25 (0.11)*	1.29 (0.12)**	1.36 (0.12)**
SES quintile	1.62 (0.06)***	1.52 (0.05)***	1.48 (0.05)***
Urban school location	1.24 (0.15)	1.21 (0.15)	1.17 (0.14)
Freshman year math score	(Omitted)	1.05 (0.01)***	1.01 (0.01)
Junior year math score	(Omitted)	(Omitted)	1.06 (0.01)***
Intercept	5.88 (0.79)***	5.18 (0.70)***	4.63 (0.65)***
Observations	7,920	7,920	7,920

Note. Odds ratios followed by robust standard errors in parentheses. SES = socioeconomic status.

* $p < .05$. ** $p < .01$. *** $p < .001$.

the impact of attending an urban school no longer remained significant in Model 11. Finally, when we added junior year math achievement scores in Model 12, the impact of directly being suspended, again, slightly decreased (but remained significant), while the impact of school social order no longer remained significant in Model 12.

Sensitivity Analysis

To check the extent to which these analyses were sensitive to unobserved—and potentially confounding—treatment assignment covariates, analyses were replicated with all observed covariates deliberately removed from the propensity score estimation models on separate occasions (Table 12). When these variables were removed, outcomes were nearly identical to the original analyses. The only exception was SES; when this variable was removed from the propensity score estimation model high-suspension schools had a slightly larger impact on math achievement and college attendance. Nevertheless, when considering that this variable was a composite of multiple indicators for social class, this small change is to be expected. Moreover, as it is unlikely

Table 11. Null Models: Non-Propensity Score Weighted Logistic Regressions of Treatment Covariates on College Attendance.

	Model 10	Model 11	Model 12
High-suspension school	(Omitted)	(Omitted)	(Omitted)
In-school suspension	0.61 (0.07)***	0.67 (0.08)**	0.72 (0.08)**
Low college expectation	0.36 (0.03)***	0.43 (0.04)***	0.46 (0.04)***
School social order	1.16 (0.07)*	1.13 (0.07)*	1.12 (0.07)
Absences	0.77 (0.04)***	0.76 (0.04)***	0.78 (0.04)***
Classes skipped	0.77 (0.06)**	0.79 (0.06)**	0.78 (0.06)**
Race: Black	0.80 (0.09)	0.95 (0.11)	1.04 (0.13)
Race: Hispanic	0.88 (0.13)	0.88 (0.12)	0.89 (0.13)
Gender: female	1.25 (0.11)*	1.29 (0.12)**	1.36 (0.12)**
SES quintile	1.66 (0.06)***	1.54 (0.05)***	1.50 (0.05)***
Urban school location	1.30 (0.16)*	1.26 (0.15)	1.21 (0.15)
Freshman year math score	(Omitted)	1.05 (0.01)***	1.01 (0.01)
Junior year math score	(Omitted)	(Omitted)	1.07 (0.01)***
Intercept	4.60 (0.55)***	4.19 (0.52)***	3.91 (0.49)***
Observations	7,920	7,920	7,920

Note. Odds ratios followed by robust standard errors in parentheses. SES = socioeconomic status.

* $p < .05$. ** $p < .01$. *** $p < .001$.

that another variable containing a similar set of information as SES exists outside of the variables already included in our propensity score estimation model, the potential for an unobserved confounder of this type is low. Overall, our sensitivity analysis provides further support for the robustness of our estimation of treatment effects.

Findings

To summarize these results, we found (a) that high-suspension schools are associated with lower math achievement scores and lower college attendance rates; (b) that the relationship between high-suspension schools and college attendance is significantly impacted by junior year math achievement; and (c) that there are significant relationships among high-suspension schools and student/school background characteristics that demonstrate persistent inequities, as well as novel opportunities, in regards math achievement and college attendance. In addition, we found that results are upwardly biased in models that do not use propensity score weights and that the indirect effects of attending a high-suspension school are similar and—in some cases—larger than the direct effects of being suspended.

Table 12. Sensitivity Results.

Removed treatment covariate	Model type	Outcome	Sensitivity results	Original results
Race: Black	Unconditional	Math achievement (coefficient)	-2.70 (0.63)***	-2.50 (0.64)***
Race: Black	Unconditional	College attendance (odds ratio)	0.67 (0.09)**	0.67 (0.09)**
Race: Hispanic	Unconditional	Math achievement (coefficient)	-2.49 (0.64)***	-2.50 (0.64)***
Race: Hispanic	Unconditional	College attendance (odds ratio)	0.67 (0.09)**	0.67 (0.09)**
Gender: Female	Unconditional	Math achievement (coefficient)	-2.47 (0.64)***	-2.50 (0.64)***
Gender: Female	Unconditional	College attendance (odds ratio)	0.67 (0.09)**	0.67 (0.09)**
SES quintile	Unconditional	Math Achievement (coefficient)	-3.14 (0.62)***	-2.50 (0.64)***
SES quintile	Unconditional	College attendance (odds ratio)	0.57 (0.07)***	0.67 (0.09)**
Urban school location	Unconditional	Math achievement (coefficient)	-2.57 (0.64)***	-2.50 (0.64)***
Urban school location	Unconditional	College attendance (odds ratio)	0.67 (0.09)**	0.67 (0.09)**
Two parent household	Unconditional	Math achievement (coefficient)	-2.51 (0.64)***	-2.50 (0.64)***
Two parent household	Unconditional	College attendance (odds ratio)	0.67 (0.09)**	0.67 (0.09)**
High parental college expectations	Unconditional	Math achievement (coefficient)	-2.55 (0.64)***	-2.50 (0.64)***
High parental college expectations	Unconditional	College attendance (odds ratio)	0.67 (0.09)**	0.67 (0.09)**
Eighth grade performance	Unconditional	Math achievement (coefficient)	-2.47 (0.64)***	-2.50 (0.64)***
Eighth grade performance	Unconditional	College attendance (odds ratio)	0.68 (0.09)**	0.67 (0.09)**
Eighth grade behavior	Unconditional	Math achievement (coefficient)	-2.54 (0.64)***	-2.50 (0.64)***
Eighth grade behavior	Unconditional	College attendance (odds ratio)	0.64 (0.08)**	0.67 (0.09)**
Eighth grade math course	Unconditional	Math achievement (coefficient)	-2.76 (0.64)***	-2.50 (0.64)***
Eighth grade math course	Unconditional	College attendance (odds ratio)	0.66 (0.09)**	0.67 (0.09)**
Eighth grade math grade	Unconditional	Math achievement (coefficient)	-2.56 (0.64)***	-2.50 (0.64)***
Eighth grade math grade	Unconditional	College attendance (odds ratio)	0.68 (0.09)**	0.67 (0.09)**

Note. For Math Achievement Models, coefficients are provided, which are followed by robust standard errors in parentheses. For College Attendance Models, odds ratios are provided, which also are followed by robust standard errors in parentheses. SES = socioeconomic status. * $p < .05$. ** $p < .01$. *** $p < .001$.

Main Findings

The impacts of high-suspension schools. When controlling for selection, students that attend high-suspension high schools are associated with lower math achievement test scores in high school (Model 1) and are less likely to attend college full-time (Model 4)—even when accounting for individual-level suspensions and school-level social order. These impacts are not only statistically significant, but also practically significant. While the practical significance of the impact of high-suspension schools on math achievement is most apparent in its relationship with college attendance (explained in the section below), it is important to note that students who attend high-suspension schools have only a 43.5% chance of attending college full-time—compared to a 56.5% chance for students attending low-suspension schools.

The relationship among math achievement and college attendance. In regards to the relationship among math achievement and college attendance, it is first important to note that while the *direct* effect associated with receiving a suspension lost significance when *freshman* year math achievement scores were accounted for in the selection model for college attendance (Model 5), the *indirect* effect associated with attending a high-suspension high school did not lose significance until *junior year* math achievement scores were accounted for in the selection model for college attendance (Model 6). Thus, *early* math achievement accounts for part of the direct effect of being suspended, while *later* math achievement accounts for part of the indirect effect of attending a high-suspension school.

The importance of junior year math achievement in rendering the impacts of high-suspension schools insignificant allows for two plausible interpretations. First, based on the negative impact that high-suspension schools have on junior year math achievement (Model 1), it can be inferred that attending a high-suspension school lowers some students' junior year math achievement to the extent that the actual impact of the school no longer remains a significant predictor of college attendance. This implies that the long-term effects of suspensions on college attendance may be channeled through the short-term effects of suspensions on math achievement. Second, it can be also be inferred that higher junior year math achievement may act as a protective factor for other students—shielding them from the negative effects of suspensions. For these students, higher math achievement may serve as a way to achieve mobility in high-suspension schools. Of course, these interpretations are not mutually exclusive. In fact, based on the range of junior year math achievement scores within schools, it is likely that both phenomena are occurring at the same time.

Student and school background characteristics. Based on the literature that demonstrates racial inequity in discipline at the school-level (see Anderson & Ritter, 2017), we were unsurprised to find that Black and low-income students were more likely to attend high-suspension schools *prior* to applying propensity score weights. Interestingly, while both Black and Hispanic students are negatively correlated with school social order, which is negatively correlated with high-suspension schools, only Black students are also correlated with high-suspension schools. Here, the disciplinary responses to school disorder may be more extreme and exclusionary in predominantly Black schools, as opposed to predominantly Hispanic schools. Nevertheless, given their negative correlations with this study's outcomes, reforms for increasing math achievement and college attendance with Hispanic students should also be prioritized.

Moreover, as urban schools are negatively correlated with school social order, which again, is negatively correlated with high-suspension schools, we might assume that students who attend urban schools are also more likely to attend high-suspension schools. However, given Peguero and his colleagues (2018) recent research, which demonstrates that even though more schools tend to be overly strict in urban areas, more schools also tend to be overly lenient in urban areas, we were not surprised to find that slightly less students in urban areas attend high-suspension schools. Here, urban schools may respond differently to social order in terms of school discipline (e.g., through more lenient practices). When considering that Black students are more likely to attend schools in urban areas *and* more likely to attend high-suspension schools, *yet* urban schools are less likely to be high-suspending, we can infer that disciplinary responses may be more extreme and exclusionary in predominantly Black schools in urban areas. We can also infer that racial inequity in discipline for Black students at the school-level may be more likely to occur in less urban areas (e.g., suburbs).

Furthermore, when observing the differences across propensity, standard, and null models, we noticed a significant changes among student and school background characteristics. For example, when we did not control for attendance in high-suspension schools in the standard model (Model 7), identifying as a Black student was no longer negatively associated with college attendance, while attending an urban school was no longer positively associated with college attendance. Here, racial/ethnic inequity and urban opportunities related to college attendance, may be masked in studies that do not control for selection bias. In regards to racial/ethnic inequity in college attendance, these findings demonstrate that even if suspension rates were equalized across high and low suspension schools (and students were balanced accordingly), Black students would still face other significant obstacles in their pursuit of post-secondary educational opportunities. One of these

obstacles may revolve around early math achievement: when freshman year math achievement scores were included in the propensity model for college attendance (Model 5), disparities in college attendance no longer remained for Black students. Moreover, we can conclude that if school suspension rates were equalized across high and low suspension schools (and students were balanced accordingly), students attending urban schools would be better supported in their pursuit of college attendance. Again, one of these supports may revolve around early math achievement: when freshman year math achievement scores were included in the propensity model for college attendance (Model 5), advantages in college attendance no longer remained for students attending urban schools. As the freshman year math achievement test focuses on algebraic reasoning, these findings underscore the importance of algebra preparation in the first year of high school for Black students and students attending high-suspension schools in urban areas.

In addition, when we did not account for the treatment in the null model (Model 10), urban school location became associated with an increase in college attendance, while school social order became associated with an increase in math achievement and college attendance. Thus, high- and low-suspension schools can be seen as driving part of the positive impact of urban school location and school social order, which is expected when considering that there are less high-suspension schools in urban areas and that high-suspension schools have lower levels of social order. Here, it may not be the level of school disorder that impacts student achievement the most, but rather schools' responses to disorder—especially those responses that involve suspensions.

Finally, even after we account for the extremes of the distribution of suspensions across schools (and balanced students accordingly), significant inequalities persisted. Black students and low-SES students were associated with lower math achievement and were less likely to attend college; female students were associated with lower math achievement; and male students less likely to attend college. Thus, while policies aimed at decreasing suspensions should rightfully be pursued, more must be done to ensure reductions in math achievement inequality and college attendance disparities among different racial/ethnic, gender, and social class groups.

Additional Findings

Selection bias. When we do not control for attendance into high-suspension schools, the impacts of this treatment on both math achievement and college attendance are upwardly biased. As a result, the effects appearing in much of the research on school-level suspensions may be overstated by proportions that should not be ignored. Rather, counterfactual and other strategies that are

able to adjust for non-random attendance into high-suspension schools are needed to limit the biases associated with their effects. In addition, it is important to note that when freshman year and junior year math scores were added in the standard college attendance model (Models 8 and 9), the direct effects associated with receiving a suspension (Model 8) and the indirect effects associated with attending a high-suspension school (Model 9) remained a significant predictor of college attendance, which was not the case in equivalent propensity models (Models 5 and 6). Therefore, we can infer that students who naturally attend high-suspension schools may be more susceptible to the negative effects of them and that these negative effects are strong enough to withstand the impacts of math achievement.

Indirect effects. In the standard models of math achievement (Model 2) and college attendance (Model 7), the indirect effect associated with attending high-suspension school was larger than the direct effect associated with receiving a suspension, which—when compared to their equivalent null models (Models 3 and 10)—had weakened with the inclusion of the treatment. While the indirect effects of suspensions can be seen as absorbing a small portion of the direct effects of suspensions in the standard models, it is important to note that the indirect effects associated with ISS in the standard models were also *larger* than direct effects associated with ISS in the null models. Furthermore, even though the indirect effects associated with attending a high-suspension school slightly weakened in the propensity models for math achievement (Model 1) and college attendance (Model 4), these indirect effects still remained similar or larger (in the case of math achievement) than the direct effects associated with receiving a suspension in these models. Thus, for many students it may be worse to attend a high-suspension school and *not* be suspended than to *not* attend a high school and be suspended.

Discussion

In-school suspension was initially conceived as a less-severe alternative to out-of-school suspension. It was originally designed to remove disruptive students from classrooms to provide a secluded setting where the behavior of offending students could be reformed, while also ensuring the learning of their classmates (Sheets, 1996). This would ideally result in a reduction in recidivism and an increase in academic achievement—both for suspended students and their classmates. However, recent research by Cholewa et al. (2018) has demonstrated that—for suspended students—the intents of in-school suspension do not match its reality: using a nationally representative study of high school students, directly receiving an in-school suspension was found to be

significantly related to a decrease in students' grade point average (GPA) and an increase in dropout status. In our present study, we have demonstrated that the intents of in-school suspension do not match its reality for *non-suspended* students either. Rather, in-school suspension was associated with detrimental short and long-term effects for students that do not directly receive them, but—*by no fault of their own*—merely attend schools that overuse them. While a low-suspension school may be prone to some collateral damages of their own (see Peguero et al., 2018), we have demonstrated that it is far worse to attend a high-suspension school, as these schools decrease students' math achievement scores and, ultimately, their college attendance rates.

Thus, the notion that a greater reliance on suspensions would increase the achievement of non-suspended students by decreasing their exposure to disruptive students has not been supported in this study. Instead, a greater reliance on suspensions provides an additional mechanism by which educational opportunities, such as those related to STEM and college attendance, are stratified. Furthermore, as Black and low-income students are more likely to attend high-suspension schools, this mechanism of stratification not only exacerbates inequities between schools, but also between racial/ethnic and social class groups. As more calls are made for moratoriums on out-of-school suspension (e.g., the "Dignity in Schools" movement), while national trends demonstrate that the number of students receiving in-school suspension have recently *surpassed* the number of students receiving out-of-school suspensions (U.S. Department of Education, Office for Civil Rights, 2014), we fear that a decrease in out-of-school suspensions may be unintentionally coupled with an increase in in-school suspensions. Thus, in-school suspension—the original policy alternative to out-of-school suspensions—might require a policy alternative of its own.

In seeking an alternative to in-school suspension, recent research has suggested that restorative justice practices might offer a viable solution. Rather than separating offending individuals from their classroom communities, restorative justice practices seek to reintegrate these individuals by providing opportunities where relationships can be restored (Gonzalez, 2012). Through conferences, mediations, and talking circles, offenders are able to repair previous harms with their victims and make amends with their classroom communities (Gonzalez, 2012). Schools that adopt restorative justice practices are able to nurture caring relationships, increase students' sense of belonging and engagement, and provide students with opportunities to learn from their mistakes. Thus, in creating an environment of respect, dignity, and mutuality, restorative justice practices appear antithetical to high-social control environments. At the same time, restorative justice practices can be seen as achieving some of the intended outcomes of social control (i.e., decreased rates of transgressions) without the unintended consequences associated with exclusionary

discipline. Therefore, it is unsurprising that schools adopting restorative practices see a drastic reduction in offenses and suspension rates, as well as an increase in academic achievement (Eisenberg, 2016).

While the positive effects of restorative justice practices can extend to all students (Anyon et al., 2016), schools with higher proportions of minority and low-income students have been less likely to implement these practices (Payne & Welch, 2015). Thus, restorative justice efforts should be prioritized in high-suspension schools, which often serve a greater proportion of minority and low-income students. Restorative justice practices should also be prioritized in urban schools, as our results demonstrate that—when we account for high and low-suspension schools—urban schools increase both math achievement and college attendance for their students. However, as restorative justice approaches not only call for changes in a school's discipline practices, but also changes in a school's discipline philosophies, implementation with fidelity can require in-depth professional development, ongoing coaching and mentoring, additional instructional tools, and the development of new leadership teams (see RAND Corporation, 2018). Therefore, in schools that are less likely to implement restorative justice practices, critical and culturally responsive engagement strategies with students, parents, and the surrounding communities should be used to better design, implement, and evaluate restorative justice practices (see Ingraham et al., 2016). Moreover, as Lustick (2017) has pointed out in her recent ethnographic study of restorative justice coordinators in New York City, in order for restorative justice to move beyond maintaining order and toward restoring justice, teachers and administrators must first confront racial injustices in their schools: “justice cannot be restored if it does not exist in the first place” (p. 25). For more serious infractions, such as those that involve criminal justice courts, trauma-informed practices may also represent a promising alternative to in school-suspension (see Baroni et al., 2016).

Nevertheless, it is important to note that *even if we were to equalize suspension rates across high and low-suspension schools* by implementing restorative justice practices (or other suspension alternatives), our findings suggest that Black students, male students, and low-income students would still be significantly related to lowered math achievement and college attendance rates. Thus, while reducing high-suspension schools may be a good first step in achieving more equitable outcomes in education, more must be done to fill in the racial/ethnic, gender, and social class gaps in both math achievement and college attendance. As our findings suggest that algebraic reasoning may present an alternative access point for interventions that seek to curb the collateral damages of high-suspension schools (especially for Black students) “Double-Dosage Algebra”—an intensive instructional policy aimed at increasing the

amount of time spent learning algebra in 9th grade (Cortes et al., 2015)—should also be considered. Recent research on Double Dosage Algebra has found it to increase students' math credits and test scores, as well as students' high school graduation and college enrollment rates (Cortes et al., 2015).

Finally, as suspensions represent an essential piece of the school-to-prison pipeline, while high school math achievement and college attendance represent essential pieces of the STEM (science, technology, engineering, and math) pipeline (see Veenstra et al., 2009), we believe that reducing the amount of high-suspension schools has the potential to both *drain* the school-to-prison pipeline and *fill* the STEM pipeline. When considering the social and economic costs of an overpopulated prison population and an underdeveloped STEM workforce, we believe that reducing high-suspension schools will not only benefit students who attend these schools, but will also benefit the larger U.S. society and economy.

Authors' Note

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Work on this paper has been funded by the National Science Foundation (#1619843 & #1800199).

ORCID iD

Jason Jabbari  <https://orcid.org/0000-0002-0196-4966>

References

- Adams, D., & Hamm, M. (2010). *Demystify math, science, and technology: Creativity, innovation, and problem-solving*. R&L Education.
- Anderson, K. P., & Ritter, G. W. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. State. *Education Policy Analysis Archives*, 25, Article 49.
- Annamma, S. A. (2018). *The pedagogy of pathologization: Dis/abled girls of color in the school-prison nexus*. Routledge.
- Anyon, Y., Gregory, A., Stone, S., Farrar, J., Jenson, J., McQueen, J., Downing, B., Greer, E., & Simmons, J. (2016). Restorative interventions and school disci-

- pline sanctions in a large urban school district. *American Educational Research Journal*, 53(6), 1663–1697.
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3), 399–424.
- Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4), 962–973.
- Baroni, B. A., Day, A. G., Somers, C. L., Crosby, S., & Pennefather, M. (2016). The adoption of the Monarch Room as an alternative to suspension and expulsion in addressing school discipline issues among court-involved youth. *Urban Education*. http://works.bepress.com/angelique_day/15/
- Basile, V., York, A., & Black, R. (2019). Who Is the One Being Disrespectful? Understanding and Deconstructing the Criminalization of Elementary School Boys of Color. *Urban Education*. Advance online publication. doi:10.1177/0042085919842627
- Cholewa, B., Hull, M. F., Babcock, C. R., & Smith, A. D. (2018). Predictors and academic outcomes associated with in-school suspension. *School Psychology Quarterly*, 33(2), 191–199.
- Christie, C. A., Nelson, C. M., & Jolivet, K. (2004). School characteristics related to the use of suspension. *Education and Treatment of Children*, 27, 509–526.
- Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment long-run impacts of double-dose algebra. *Journal of Human Resources*, 50(1), 108–158.
- Ditton, J. R. (1979). *Controlology: Beyond the new criminology*. Palgrave Macmillan.
- DuGoff, E., Schuler, M., & Stuart, E. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*, 49(1), 284–303.
- Duncan, G. A. (2000). Urban pedagogies and the ceiling of adolescents of color. *Social Justice*, 27(3), 29–42.
- Durkheim, E., Schnurer, H., & Wilson, E. (1961). *Moral education. A study in the theory and application of the sociology of education*. Free Press.
- Eisenberg, D. (2016). Against school suspensions. *University of Maryland Law Journal of Race, Religion, Gender and Class*, 16, 163.
- Fabelo, T., Thompson, M., Plotkin, M., Carmichael, D., Marchbanks, M., & Booth, E. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. Council of State Governments Justice Center.
- Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90(4–5), 837–851.
- Gonzalez, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the school to prison pipeline. *Journal of Law & Education*, 41, Article 281.
- Goodman, J. F. (2013). Charter management organizations and the regulated environment: Is it worth the price? *Educational Researcher*, 42(2), 89–96.
- 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.

- Guo, S., & Fraser, W. (2014). *Propensity score analysis*. SAGE.
- Ibrahim, H., & Johnson, O. (2019). School discipline, race-gender and STEM readiness: A hierarchical analysis of the impact of school discipline on math achievement in high school. *The Urban Review*. Advance online publication. doi:10.1007/s11256-019-00513-6
- Ingels, S., Pratt, D., Herget, D., Burns, L., Dever, J., Ottem, R., Rodgers, J., Jin, Y., & Leinwand, S. (2011). *High school longitudinal study of 2009: Base-year data file documentation (NCES 2011-328)*. National Center for Education Statistics.
- Ingraham, C. L., Hokoda, A., Moehlenbruck, D., Karafin, M., Manzo, C., & Ramirez, D. (2016). Consultation and collaboration to develop and implement restorative practices in a culturally and linguistically diverse elementary school. *Journal of Educational and Psychological Consultation*, 26(4), 354–384.
- Irby, D. J. (2014). Trouble at school: Understanding school discipline systems as nets of social control. *Equity & Excellence in Education*, 47(4), 513–530.
- Jabbari, J., & Johnson, O., Jr. (2020). Veering off track in U.S. high schools? Redirecting student trajectories by disrupting punishment and math course-taking tracks. *Children and Youth Services Review*. Advance online publication. doi:10.1016/j.childyouth.2019.104734
- Johnson, O., Jr., Jabbari, J., Williams, M., & Marcucci, O. (2019). Disparate impacts: Balancing the need for safe schools with racial equity in discipline. *Policy Insights from the Behavioral and Brain Sciences*, 6(2), 162–169.
- Kafka, J. (2011). *The history of “zero tolerance” in American public schooling*. Palgrave Macmillan.
- Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, 54(1), 355–383.
- Kirk, D. S. (2009). Unraveling the contextual effects on student suspension and juvenile arrest: The independent and interdependent influences of school, neighborhood, and family social controls. *Criminology*, 47(2), 479–520.
- Kupchik, A. (2010). *Homeroom security: School discipline in an age of fear*. NYU Press.
- Kupchik, A., Green, D. A., & Mowen, T. J. (2015). School punishment in the US and England: Divergent frames and responses. *Youth Justice*, 15(1), 3–22.
- Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for black and white students. *Education and Treatment of Children*, 34, 167–192.
- Losen, D., & Martinez, T. (2013). *Out of school and off track: The overuse of suspensions in American middle and high schools*. The Center for Civil Rights Remedies.
- Losen, D., & Skiba, J. (2010). *Suspended education: Urban middle schools in crisis*. The Civil Rights Project / Proyecto Derechos Civiles.
- Lustick, H. (2017). “Restorative Justice” or Restoring Order? Restorative School Discipline Practices in Urban Public Schools. *Urban Education*. Advance online publication. doi:10.1177/0042085917741725
- Manning, W. D., & Lamb, K. A. (2003). Adolescent well-being in cohabiting, married, and single-parent families. *Journal of Marriage and Family*, 65(4), 876–893.

- McCaffrey, D., Ridgeway, G., & Morral, A. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9(4), 403–425.
- Milner, H. R., IV. (2012). But what is urban education? *Urban Education*, 47, 556–561.
- Milner, H. R., IV, & Lomotey, K. (Eds.). (2014). *Handbook of urban education*. Routledge.
- Morris, E. W. (2005). “Tuck in that shirt!” Race, class, gender, and discipline in an urban school. *Sociological Perspectives*, 48(1), 25–48.
- Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, 63(1), 68–86.
- Mukuria, G. (2002). Disciplinary challenges: How do principals address this dilemma? *Urban Education*, 37(3), 432–452.
- Nolan, K. (2011). *Police in the hallways: Discipline in an urban school*. University of Minnesota Press.
- Payne, A., & Welch, K. (2015). Restorative justice in schools: The influence of race on restorative discipline. *Youth & Society*, 47(4), 539–564.
- Peguro, A. A., Portillos, E. L., & González, J. C. (2015). School securitization and Latina/o educational progress. *Urban Education*, 50(7), 812–838.
- Peguro, A. A., Varela, K., Marchbanks, M., III, Blake, J., & Eason, J. (2018). School punishment and education: Racial/ethnic disparities with grade retention and the role of urbanicity. *Urban Education*. Advance online publication. doi:10.1177/0042085918801433
- Perry, B., & Morris, E. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, 79(6), 1067–1087.
- Portillos, E. L., González, J. C., & Peguro, A. A. (2012). Crime control strategies in school: Chicanas/os’ perceptions and criminalization. *The Urban Review*, 44(2), 171–188.
- RAND Corporation. (2018). *Can restorative practices improve school climate and curb suspensions? An evaluation of the impact of restorative practices in a mid-sized urban school district*.
- Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., & Griffin, B. (2014). *Toolkit for weighting and analysis of nonequivalent groups* (R Package). RAND Corporation.
- Robers, S., Kemp, J., Rathbun, A., Morgan, R., & Snyder, T. (2014). *Indicators of school crime and safety: 2013*. National Center for Educational Statistics, U.S. Department of Education.
- Rogers, C. R., & Freiberg, H. J. (1969). *Freedom to learn*. Pearson.
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rubin, D. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469), 322–331.
- Shabazian, A. N. (2020). Voices that matter: Chief Administrative Officers’ role in the student discipline gap. *Urban Education*, 55, 66–94.
- Sheets, J. (1996). Designing an effective in-school suspension program to change student behavior. *NASSP Bulletin*, 80(579), 86–90.

- Skiba, R., Chung, C., Trachok, M., Baker, T., Sheya, A., & Hughes, R. (2014). Parsing disciplinary disproportionality contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal*, 51(4), 640–670.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, 34(4), 317–342.
- StataCorp. (2013). *Stata: Release 13. Statistical Software*. College Station, TX: StataCorp LP.
- U.S. Department of Education, Office for Civil Rights. (2014). *2013-14 discipline estimations by discipline type*.
- Veenstra, C. P., Dey, E. L., & Herrin, G. D. (2009). A model for freshman engineering retention. *Advances in Engineering Education*, 1(3). <https://advances.asee.org/wp-content/uploads/vol01/issue03/papers/ae-vol01-issue03-p07.pdf>
- Wacquant, L. (2001). Deadly symbiosis: When ghetto and prison meet and mesh. *Punishment & Society*, 3(1), 95–133.
- Watanabe, A. (2013, May 14). L.A. Unified bans suspension for “willful defiance.” *Los Angeles Times*. <https://www.latimes.com/local/la-xpm-2013-may-14-la-me-lausd-suspension-20130515-story.html>
- Welch, K., & Payne, A. (2010). Racial threat and punitive school discipline. *Social Problems*, 57(1), 25–48.
- Whitman, D. (2008). *Sweating the small stuff: Inner-city schools and the new paternalism*. Thomas B. Fordham Institute.
- Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic Monthly*, 249(3), 29–38.

Author Biographies

Jason Jabbari is a senior analyst at the Social Policy Institute at Washington University in St. Louis. A former classroom teacher, school leader, and basketball coach, Jason has a MEd in School Leadership from Harvard University, a PhD in Education from Washington University in St. Louis, and an advanced certificate in Quantitative Data Analysis. Jason uses sociological theories and quantitative methodologies to explore education programs, policies, and practices related to equitable outcomes in urban communities. His most recent work has been published in *Policy Insights from the Behavioral and Brain Sciences* and *Children and Youth Services Review*.

Odin Johnson Jr. is a professor in the Departments of Sociology and Education, Director of the NSF Institute in Critical Quantitative, Computational, and Mixed Methodologies, and Associate Director of the Center for the Study of Race, Ethnicity, and Equity at Washington University in St. Louis. His research examines how neighborhoods, schools, and public policies relate to social inequality, youth development and the status of African American populations. Among other journals, his previous work has appeared in *Review of Educational Research*, *Social Science and Medicine*, and the *Annals of the American Academy of Political and Social Research*.