Suspending Suspensions: The Education Production Consequences of School Suspension Policies

Nolan G. Pope* George W. Zuo†

June 9, 2020

Abstract

Managing student behavior is integral to the education production process. Over the past half-century, a common but controversial approach has been to suspend disruptive students from the classroom, creating potential tradeoffs between disciplined students and their peers. We study these tradeoffs by modeling and estimating how changes in school suspension policies impact student performance and teacher turnover. We use administrative data from the Los Angeles Unified School District, where suspension rates fell by over 90 percent since 2003. We instrument for school suspension rates by interacting districtwide suspension rate changes with initial school suspension rate levels. Our results indicate that a reduction in suspension rates decreases math and English test scores, decreases GPAs, and increases absences. Teacher turnover also increases, particularly for inexperienced teachers. The overall negative impact of reducing suspension rates is driven by small but diffuse spillovers produced by more lenient disciplinary environments. These spillovers are only partially offset by large and concentrated benefits for the small number of students who are no longer suspended.

We thank Nathan Barrymore, Peter Blair, Mrinmoyee Chatterjee, Joshua Goodman, Judith Hellerstein, Melissa Kearney, Richard Patterson, Nathan Petek, John Shea, Cody Tuttle, and Sergio Urzua for helpful comments and discussions. Zuo graciously acknowledges financial support from the National Science Foundation Graduate Research Fellowship Program during the writing of this paper.

*University of Maryland, 3114 Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742. Email: npope@umd.edu

†University of Maryland, 3114 Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742. Email: gzuo@umd.edu

1 Introduction

Managing student behavior is an integral part of education production. Yet historically, most disciplinary policies have been formed with little empirical evidence of their costs and benefits. This is highlighted by the turbulent history of school suspensions in American education. Starting in the early 1990s, many states and school districts began implementing "zero tolerance" disciplinary policies in response to growing concerns of violence and disorder in schools.[1] These policies relied on suspending students for even minor forms of misbehavior in order to deter additional misconduct and limit negative spillovers on learning in the education production process (Lazear, 2001). Consequently, national suspension rates more than doubled from 3.5% to 8.4% between the mid-1970s and the turn of the century (see Figure 1).[2]

Despite their wide prevalence, suspensions have long been controversial. While advocates argue that suspensions are necessary for maintaining a productive and disruption-free learning environment, critics counter that suspensions needlessly remove students from the classroom and disproportionately impact minority students. These criticisms culminated in a joint initiative between the Department of Education and Department of Justice in 2011, leading many states and school districts to reform their disciplinary policies to be less strict and exclusionary (Lacoe and Steinberg, 2017). The reforms fueled a 40 percent reduction in suspensions over the past decade (see Figure 1). Still, approximately 3.5 million students are suspended each year, amounting to nearly 18 million days of lost instruction (Losen et al., 2015).

Despite the wide variety of policies accompanying the rise and decline of suspensions over the last 40 years, there remains little empirical evidence on the causal impact of suspension policies on students and teachers.[3] In this paper, we model and estimate how reducing school suspension rates affects student performance and teacher turnover. We focus on the Los Angeles Unified School District (LAUSD), where suspension rates fell from eight percent in 2003 to less than one percent in 2015. We use LAUSD administrative student-level data to estimate the effect of changes in school suspension rates on math and English test scores, GPAs, absences, and teacher turnover.

[1]These policies were designed based on the prevailing theory of "broken windows", which argued that even traces of disorder (e.g. a single broken window) could breed an atmosphere for more serious offenses to occur (Teske, 2011).

[2]Although we focus on out-of-school suspensions in this paper, many schools also employ in-school suspensions, which is the practice of removing students from regular classroom activities but keeping them within the supervision of a school administrator during the day.

[3]A notable exception is concurrent work by Bacher-Hicks, Billings and Deming (2019), who find that attending a high-suspension school leads to lower graduation rates and higher rates of future crime.

Estimating the causal impact of changes in school suspension rates is complicated by the fact that school suspension rates are likely endogenous to changes in school quality and student composition. Our empirical approach instruments for each school's suspension rate using year-to-year changes in the districtwide suspension rate interacted with school-specific suspension rates fixed to an initial pre-period. The instrument relies on year-to-year changes in districtwide suspension rates being exogenous to school-specific determinants of student performance. Initial suspension rates scale each school's exposure to these districtwide changes.

We find that a 10 percentage point decrease in school suspension rates decreases concurrent math and English test scores by 0.04 and 0.06 standard deviations, respectively. These effects, while modest, are equivalent to reducing teacher quality by 0.29 and 0.64 standard deviations for math and English (Chetty, Friedman and Rockoff, 2014). Suspension rates also impact students beyond test scores. A 10 percentage point decline in suspension rates decreases GPAs by 0.07 standard deviations and increases the fraction of days absent (excluding suspension days) by 1.1 percentage points (15.1%). This implies that declining suspension rates in Los Angeles during our sample period were detrimental to the average student across a variety of outcomes. The estimates are robust to a variety of sensitivity tests.

Declining suspension rates may also affect teachers' well-being. When limited in their capacity to suspend students, teachers may face more misbehavior and find the teaching environment to be more difficult and unpleasant. We quantify this by estimating the impact of school suspension rates on teacher turnover. We find that a 10 percentage point decrease in suspension rates increases teacher turnover by 2 percentage points (9.9%). The effect is particularly concentrated on inexperienced teachers. Teachers with less than three years of experience are more than three times as likely to leave their school in response to declining suspension rates. A back-of-the-envelope calculation suggests that inexperienced teachers would need to be paid $2,967 more per year to offset a 10 percentage point decrease in suspension rates (Clotfelter et al., 2008). In addition, we find that high school teachers are most responsive to declines in suspensions.

All schools – whether implicitly or explicitly – must determine the strictness of their suspension policies. We model the choice of suspension rates and show that this choice produces tradeoffs between students that average effects do not fully capture. When a misbehaving student is suspended, the impact of the suspension manifests in both a *direct* effect on that student as well as *indirect* spillover effects on his peers.[4,5] Changing school suspension policies will therefore impact students through both these channels. We use two approaches to separately quantify direct versus indirect effects. First, we leverage the fact that changes in suspension rates affect low-misbehavior students primarily through indirect effects. We construct a proxy for students' propensities to be suspended and estimate the effects of changes in suspension rates separately for low-, medium-, and high-suspension students. Differences across these estimates provide insight into the direction and magnitude of direct and indirect effects. Second, we explicitly model direct and indirect effects within the same estimating equation.

Estimates from both approaches suggest that the indirect effects of increasing suspension rates are small, positive, and diffuse, whereas the direct effects are large, negative, and concentrated. We show that indirect effects dominate in the aggregate because all students are exposed to indirect effects, whereas relatively few students experience direct effects. On the other hand, we find that students who become suspended due to higher suspension rates experience net declines in their test scores, since the direct cost of being suspended is approximately two times larger than the indirect spillover benefit of higher suspension rates. This implies that lowering suspension rates in the LAUSD decreased average test scores while increasing test scores for students at the margin of being suspended. Our findings illustrate that suspension policies exhibit a tradeoff between efficiency and equality (Okun, 2015). Higher suspension rates improve educational efficiency by increasing average test scores, but increase inequality by decreasing the test scores of students on the margin of suspension, who are disproportionately represented by low-performing, minority, and male students (Barrett et al., 2017).

Despite these tradeoffs, schools must ultimately make a decision on the strictness of their disciplinary policies. These choices are difficult given the lack of causal evidence surrounding suspension policies, as well as ongoing changes in public opinion surrounding school discipline. Our findings help quantify these tradeoffs and provide schools with evidence and a framework for understanding

[4] This direct effect may be negative due to suspended students being removed from the classroom and the potential stigma from suspension. It could alternatively be positive if suspensions improve future behavior and engagement in the classroom. Due to limited causal evidence in the literature, there is no clear consensus on either the direction or size of these effects (Anderson, Ritter and Zamarro, 2017; Lacoe and Steinberg, 2018*a*). A recent meta-analysis by Noltemeyer, Ward and Mcloughlin (2015) summarizes the correlational research across 34 studies, finding a negative correlation between achievement and being suspended.

[5] This indirect effect may be positive by improving the learning environment through removing disruptive students from the classroom. Alternatively, there could be negative indirect effects if strict discipline creates a stressful learning environment. Identifying the indirect effects of suspensions is also challenging, and existing estimates are generally descriptive (Lacoe and Steinberg, 2018*a*,*b*). Beyond suspensions however, there is a large literature on other academic spillovers (Carrell, Hoekstra and Kuka, 2018; Imberman, Kugler and Sacerdote, 2012; Sacerdote, 2011).

how students and teachers are affected by different suspension policies. While the short-term academic outcomes we study are not necessarily the sole objective for administrators, they provide a useful baseline for understanding the costs and benefits of changing school suspension policies.

2 Suspension Policies and Education Production

We use the following stylized framework to illustrate a school's decision to set optimal suspension rates based on education production and school-specific costs. The framework highlights the relationship between suspension policies and the public good nature of education, which creates tradeoffs in education production between disciplined students and their peers. We begin by building on the framework introduced by Lazear (2001). Suppose that for any given point in time, learning occurs if all n students in a classroom are behaving. Given the probability s of being suspended at any given point in time, students behave with probability $p(s)$. Learning is therefore produced with probability $p(s)^n$ when all students in the classroom are behaving.[6] We assume that $p'(s) > 0$, implying that students are more likely to behave as the likelihood of being suspended for misbehaving increases.

Each school chooses s to maximize the following profit function:

$$\Pi = V[np(s)^n - C(s)] - K(s) \tag{1}$$

where V is the value of a unit of learning and $C(s)$ represents learning that is lost by students who are suspended. $K(s)$ captures other school-specific costs of suspension policies not directly related to learning production. For example, $K(s)$ might be greater for schools with high s due to complaints from unhappy parents or scrutiny from district administrators. Schools with preferences for educational or racial equality could also incur a greater $K(s)$ for any given s. Taking these costs into account, schools choose s so that the marginal benefit from increasing suspension rates equals the marginal cost:

$$V\left[n^2 p(s)^{n-1}\frac{dp}{ds}\right] = V\left[\frac{dC(s)}{ds}\right] + \frac{dK(s)}{ds} \tag{2}$$

The marginal benefit on the left-hand side captures the increase in value from a marginal increase in s, which manifests through better classroom behavior and therefore more learning. The marginal cost on the right-hand side encompasses both the marginal learning costs incurred by suspended students

[6] To give a sense of what s represents in this context, consider a high-suspension LAUSD middle school, where 11% of students are suspended and each suspended student is suspended an average of 1.62 times for 2.32 days (see Table 1). In a school year with 180 days, the probability s that a given student is suspended on a given day is $(11.0\% \times 1.62 \times 2.32)/180 = 0.24\%$.

and the other marginal costs incurred by the school.

One way to explicitly model the cost to suspended students $C(s)$ is to assume that the costs are comprised of 1) forgone instruction time and 2) additional student-specific costs such as the socio-emotional effects of being suspended and the academic effects of disrupted learning continuity. This particular cost function can be written as follows:

$$C(s) = sn\left(p(s)^n + A\right) \tag{3}$$

Within the parentheses, the first term represents forgone learning from missing classroom time, and A is a constant capturing the additional learning costs. The individual cost is multiplied by sn to calculate the total cost to all suspended students in the classroom. Substituting this cost function into the first order condition in Equation (2) yields:

$$\underbrace{n\,V\left[np(s)^{n-1}\frac{dp}{ds}\right]}_{\text{Indirect Effect}} = \underbrace{sn\,V\left[np(s)^{n-1}\frac{dp}{ds} + A\right]}_{\text{Direct Effect}} + \frac{dK(s)}{ds} \tag{4}$$

Equation (4) implies that in the absence of other school-specific costs $K(s)$, the first order condition equates the marginal benefit of more learning (through less misbehavior) with the marginal learning costs incurred by suspended students. We refer to the marginal benefit of more learning for an individual student as the *indirect* or spillover effect of suspension policies. Because all students in the classroom are affected by spillovers, the total indirect effect for the classroom is n times larger than the individual indirect effect. Conversely, we refer to the costs directly incurred by suspended students as the *direct effect* of suspension policies. In contrast to indirect effects, direct effects only impact suspended students, implying that the total direct effect is only sn times larger than the individual direct effect. The fact that n is much larger than sn implies that indirect effects are smaller but more diffuse, compared to direct effects that are large but impact far fewer students.

Total indirect and direct effects may be unequal for two reasons, both of which tend to produce greater total indirect effects relative to direct effects. First, other school-specific costs of suspension policies $K(s)$ likely increase as suspension rates rise. $K'(s) > 0$ mechanically implies – via Equation (4) – that total indirect effects will exceed total direct effects even when suspension rates are at optimal levels. Second, schools may simply lack information about the marginal costs and benefits of education production with respect to suspension policies when choosing optimal suspension rates. Direct effects may be more salient to administrators since the effects of being suspended are concentrated and linked to specific students. As a result of this salience, schools may tend to overestimate total direct effects relative to total indirect effects and set suspension rates below optimal levels.

When total indirect effects exceed total direct effects, aggregate learning production can be increased by raising suspension rates. We quantify this empirically by estimating the effect of suspension rates on average test scores. A positive effect implies that the positive learning spillovers from a marginal increase in suspension rates exceed the learning lost by suspended students at the margin. However, the existence of a wedge between total direct and indirect effects can still be optimal if driven by other school-specific costs of suspension policies $K(s)$. If the wedge is instead driven by factors such as salience and lack of information, increasing s should increase the efficiency of education production.

This framework also has implications for how suspension policies affect equality in the classroom. As previously mentioned, increasing suspension rates produces positive spillovers for all students in the classroom, but the net benefit could be substantially smaller (or even negative) for students who become suspended as a result of the policy change. If these suspended students are lower-achieving students on average, then the benefits of stricter suspension policies will be greater for higher-achieving students. Thus, the increase in educational efficiency is accompanied by a corresponding decrease in equality.

3 Suspensions and School Discipline in Los Angeles

The use of suspensions peaked nationwide in the early 2000's (see Figure 1), when widespread "zero tolerance" policies enabled schools to suspend students for relatively minor non-violent infractions. These policies were designed based on the prevailing theory of "broken windows" and became commonly employed in school districts following the Gun-Free Schools Act of 1994, which mandated expulsion for infractions involving a firearm (Curran, 2016). The liberalized use of suspensions has recently come under additional scrutiny due to concerns over the punitive and regressive nature of school suspensions as well as allegations of discriminatory practices toward youth of color (U.S. Department of Education, 2014).[7] As of May 2015, 23 states had implemented laws to either limit the use of exclusionary discipline practices or implement non-punitive discipline strategies (Lacoe and

[7]Recent research has also begun to produce more rigorous evidence on implicit racial biases in the use of suspensions. For example, Barrett et al. (2017) show that in interracial fights, black students receive slightly longer suspensions, even after controlling for discipline histories and background characteristics. Their findings speak to a concern in the education community about inequities in disciplinary practices across racial lines.

Steinberg, 2017).

In line with this broader trend, the LAUSD implemented suspension policy reforms to reduce suspensions (Hashim, Strunk and Dhaliwal, 2018). Discipline reform in the district occurred in two main phases. Starting in the 2006-07 school year, the district implemented School-Wide Positive Behavior Supports (SWPBS), a program designed to address racial disparities in suspension rates. Schools were required to develop multi-tiered discipline plans to manage student behavior and avoid suspensions. SWPBS gave schools autonomy to develop their own discipline plans, placing the burden of lowering suspension rates upon local school administrators. Administrators were required to develop disciplinary policies tailored to their school's educational context while simultaneously providing instructors with resources and training to facilitate the transition to the new policies.

The second phase began in the summer of 2013, when the LAUSD implemented a ban on suspensions for "willful defiance", a discretionary catch-all for student misbehavior that enabled students to be suspended for a wide variety of non-violent offenses such as refusing to take off a hat or turn off a cellphone (Watanabe, May 14, 2013). The decision came following scrutiny by the U.S. Department of Education over continued racial disparities in LAUSD's school discipline policies (Blume, 2012). The suspension ban also led to the adoption of the School Discipline Policy and School Climate Bill of Rights, which emphasized the use of "restorative justice" methods as an alternative to suspensions beginning in the 2014-15 school year.

4 Data

In this paper, we use student-level administrative data from the Los Angeles Unified School District. The LAUSD is the second-largest school district in the United States and enrolls over 600,000 students annually. Within the district, 74 percent of students are Hispanic, 10 percent are white, and 8 percent are black. The administrative data we use include a panel of students in grades 2 through 11 beginning in the 2002-03 school year and ending in the 2014-15 school year. For consistency, we reference school years by year of graduation (e.g. 2003 represents the 2002-03 school year).

The student data include our main variables of interest: standardized test scores, grade point averages, absences, and suspensions. Standardized test scores for both math and English Language Arts come from the California Standards Test which is administered to all students in grades 2 through 11. The California Standards Test was discontinued in 2014; consequently, test scores in the data only extend through 2013. We normalize math test scores, English test scores, and GPA to mean zero and standard deviation one at the grade-year level.[8] Due to privacy restrictions, we do not observe student demographic characteristics, including gender, race, and socioeconomic status. We also use data on teachers in the district, and can observe whether a teacher leaves their school in any given year.

We summarize available student and school characteristics in Table 1. In this table, we divide students by grade category (elementary, middle, and high school) as well as whether their school's 2003 suspension rate was above or below the median for that grade category. In general, low-suspension schools have higher test scores than high-suspension schools, although this pattern is reversed for high schools. We also observe that high-suspension middle and high schools tend to have substantially larger student populations. As students progress from elementary to middle and high school, both absences and suspensions increase.

Data on suspensions include the number of times a student was suspended as well as the number of days suspended.[9] School and districtwide suspension rates are calculated by dividing the number of students who were suspended at least once by the total number of students enrolled. Districtwide suspension rates in the LAUSD decreased substantially during the sample time period. Nearly 8 percent of LAUSD students were suspended in 2003, compared to less than 1 percent in 2015. Figure 2 shows that suspension rates also vary substantially by grade. Suspension rates increase with each grade beginning in second grade, peak at eighth grade, and then decrease with each grade through the end of high school. Less than 5 percent of elementary school students were suspended in 2003, compared to over 15 percent of eighth grade students. Figure 3 illustrates how suspension rates evolved over time in elementary versus middle/high schools, separating schools into four quartiles based on 2003 suspension rates. Initial dispersion in 2003 is much greater in middle and high schools, as is the magnitude of the subsequent decline. Schools across the four quartiles nearly converge by the end of the sample period.

We also characterize the typical suspension in the district. Figure 4 describes the distribution of suspensions across the intensive margin of days suspended. Conditional on being suspended, approximately 50 percent of students are only suspended for a single day and 30 percent of students are suspended for two or three days. Suspension lengths tend to be longer for middle and high school students. Although not available at the student level, aggregate district data on suspension offenses is available through the California Department of Education beginning in 2012.[10] The top

[8]See Petek and Pope (2018) for a detailed description of these GPAs.

[9]The data unfortunately do not include offense-specific information.

[10]These statistics can be accessed via DataQuest at https://data1.cde.ca.gov/dataquest/.

three suspension offenses in LAUSD as of 2012 are violence (49%), defiance (26%), and drug-related offenses (14%). Statewide, the top three offenses are defiance (47%), violence (38%), and drug-related offenses (8%).

5 Empirical Strategy

Our main objective is to estimate the causal effect of school suspension rates on students' academic outcomes. However, suspension rates are likely to be endogenously chosen by schools based on many characteristics, such as the types of students enrolled and their propensity to misbehave. Suspension rates are often negatively correlated with many aspects of school quality, which likely biases standard OLS estimates downward.

With this in mind, we begin with the following estimating equation:

$$y_{isgt} = \alpha + \rho SuspendRate_{sgt} + \beta X_{isgt-1} + \theta S_{sgt-1} + \phi P_{isgt} + \lambda_{sg} + \epsilon_{isgt} \tag{5}$$

where y_{isgt} represents a measure of academic achievement for student i in school s, grade g, and year t. We focus on standardized math and English test scores, standardized GPAs, and the fraction of non-suspended days absent. The variable $SuspendRate_{sgt}$ is the fraction of students suspended in a school-grade-year multiplied by 10. We multiply suspension rates by 10 so that the main parameter of interest, ρ, represents the effect of increasing suspension rates by 10 percentage points (instead of 100 percentage points), approximately reflecting the overall suspension rate decline that we observe in the LAUSD during this time period. We calculate suspension rates at the school-grade level rather than the school level, given the substantial across-grade variation documented in Figure 2.[11] For student i, X_{isgt-1} is a vector that includes lagged math and English test scores, lagged GPA, lagged fraction of days absent, a lagged indicator of being suspended, and an indicator for being an English language learner. S_{sgt-1} is a vector of average math and English test scores in grade g and school s from the previous year. P_{isgt} includes average lagged math and English test scores of all students in the same school and grade in time t excluding student i. Lastly, λ_{sg} is a school-grade fixed effect. All standard errors are clustered at the school-grade level.

The control variables in Equation (5) account for certain observable variables that are correlated with suspension rates. However, estimates of ρ will remain biased if unobservable time-varying, within-school-grade characteristics are correlated with suspension rates. For example, changes in

[11] We find similar results in column (4) of Tables 7 and A.2 when we calculate suspension rates at the school level.

school academic policies could simultaneously affect suspension rates and test scores. To address these threats to identification, we instrument for suspension rates using annual growth in the districtwide suspension rate interacted with school-specific suspension rates fixed to an initial pre-period. The instrument relies on year-to-year changes in districtwide suspension rates being exogenous to school-specific determinants of student performance. Initial suspension rates scale each school's exposure to these districtwide changes. The instrument we use is as follows:

$$\widetilde{SuspendRate}_{sgt} = SuspendRate_{sg2003} \times G_{gt}^{-s} \tag{6}$$

where $SuspendRate_{sg2003}$ is the initial suspension rate of school s and grade g as of 2003, and $G_{gt}^{-s} = \frac{SuspendRate_{gt}^{-s}}{SuspendRate_{gt-1}^{-s}}$ is the year-to-year growth in district suspension rates in grade g between years $t-1$ and t, leaving out the contribution of school s. Therefore, the instrument simply multiplies initial 2003 suspension rates at school s and grade g by the leave-own-out districtwide growth in suspension rates for grade g.

The intuition underlying the instrument follows from its decomposition. We first decompose school suspension rates via the following identity:

$$SuspendRate_{sgt} = SuspendRate_{sgt-1} \times \frac{SuspendRate_{sgt}}{SuspendRate_{sgt-1}} = SuspendRate_{sgt-1} \times G_{sgt} \tag{7}$$

where G_{sgt} is the year-to-year growth of suspension rates at school s and grade g between $t-1$ and t. We first replace G_{sgt} with a leave-own-out districtwide growth rate G_{gt}^{-s}. The exclusion of school s addresses finite sample bias from using own-school information in the first stage (Goldsmith-Pinkham, Sorkin and Swift, 2018). However, using $SuspendRate_{sgt-1} \times G_{gt}^{-s}$ as an instrument would be problematic since both lagged suspension rates and lagged test scores would be on the right-hand side in the first-stage equation. This may introduce bias if lagged test scores are themselves an outcome of lagged suspension rates. We address this by fixing lagged suspension rates to 2003 (the earliest year available in the data) and restricting the estimation sample to begin in 2005. We could alternatively address this by using twice-lagged suspension rates in the construction of the instrument.[12] However, if $SuspendRate_{sgt}$ is correlated with ϵ_{isgt} and ϵ_{isgt} is serially correlated, twice-lagged suspension rates could still be endogenous.[13]

[12] Column (5) of Tables 7 and A.2 reports our main results when we use twice lagged suspension rates in the construction of the instrument. Our results remain very similar with this alternative construction of the instrument.

[13] As seen in columns (2) and (3) of Tables 7 and A.2, specifications that increase the number of years between the instru-

Given this instrument, the first stage equation is:

$$SuspendRate_{sgt} = \kappa + \eta \widetilde{SuspendRate}_{sgt} + \delta X_{isgt-1} + \gamma S_{sgt-1} + \xi P_{isgt} + \mu_{sg} + \nu_{isgt} \tag{8}$$

The instrument generates relevance from both initial suspension rates and district suspension rate growth. District suspension rate growth produces aggregate across-time variation, which is then scaled by each school-grade's initial suspension rate. Figure 3 shows that initial suspension rates are highly predictive of a school's exposure to changes in district suspension rates, with middle schools in the top quartile initially suspending 20 percent of students and bottom-quartile middle schools suspending 4 percent (for elementary school, the suspension rates are 6 percent and 0 percent, respectively). The eventual convergence of suspension rates by 2015 suggests that schools with low (high) initial suspension rates have low (high) exposure to districtwide changes. In addition, we visually inspect the first stage by plotting a binned scatterplot of suspension rates versus the suspension rate instrument Figure A.1. We observe a strong relationship between the instrument and actual suspension rates which is supported by a first-stage F-statistic exceeding 1,000.

The exclusion restriction requires that the instrument be uncorrelated with the structural error term ϵ_{isgt}, which contains within-school-grade, across-time variation in test scores not accounted for by lagged individual, school, or peer achievement. The instrument relies on the fact that changes in district suspension rates evolve externally from the idiosyncratic and endogenous decision-making of any given school. Initial suspension rates, while not directly contributing to across-time variation, scale the district suspension rate growth for each school. The exclusion restriction is violated if the interaction of these two variables is correlated with the regression residual. The exclusion restriction is supported by the fact that district growth evolves externally and captures changes in suspension rates (as opposed to levels), and that pre-period suspension rates are pre-determined. However, the variation produced by this interaction is complex, and we supplement our analysis with a combination of sensitivity tests, including a specification that only uses district suspension rate growth as the instrument.

mented year and the year used to measure the prior suspension rate of a school-grade produce similar results.

6 Results

6.1 Test Scores

We first estimate the impact of a school's suspension rate on test scores. Table 2 reports the OLS and IV estimates of school suspension rates and test scores using Equation (5). Panel A presents the effect on math test scores and Panel B on English test scores. Columns (1) through (5) provide OLS estimates while adding in control variables. In column (1), the relationship between suspension rates and test scores in the raw, uncontrolled OLS specification is large and negative. The estimates show that a 10 percentage point increase in suspension rates is associated with a -0.164 and -0.146 standard deviation decline in math and English test scores, respectively. This negative relationship is unsurprising since students with low test scores are more likely to be suspended. As such, schools with a high fraction of low-performing students are more likely to have high suspension rates. By adding school-grade fixed effects and controlling for the time-invariant characteristics of the school-grade, the coefficient in column (2) moves by an order of magnitude towards zero. Individual lagged achievement is also negatively correlated with suspension rates and its inclusion additionally increases the coefficient between columns (2) and (3). Subsequent inclusion of lagged school and peer test score controls do not meaningfully change the estimates. The estimates from the fully-controlled OLS specification in column (5) are positive and relatively small (0.005 and 0.019 standard deviations for math and English), with the estimate on English test scores being statistically significant.

In column (6), we use the instrument to address the endogeneity problem. The IV estimates for math and English test scores are both positive and statistically significant at the one-percent level. We find that a 10 percentage point increase in a school's suspension rate increases math and English test scores by 0.040 and 0.064 standard deviations. The IV estimates are larger than the corresponding OLS estimates in column (5), suggesting that unobservable, time-varying confounders negatively bias the fully-controlled OLS estimates. However, the IV estimates appear moderately sized given that they reflect the effect of a 10 percentage point change in suspension rates. To contextualize the effect size, the impact of increasing suspension rates by 10 percentage points is equivalent to having a 29% and 64% standard deviation higher-quality teacher as measured by value-added for math and English, respectively (Chetty, Friedman and Rockoff, 2014).[14] It should also be noted that the estimates we report are for one year of exposure and do not reflect the cumulative effect of being exposed to higher

[14] Chetty, Friedman and Rockoff (2014) find that a 1 standard deviation improvement in teacher value-added increases test scores by 0.14 standard deviations in math and 0.10 standard deviations in English.

or lower suspension rates in multiple grades.

The IV estimates represent the effect of suspension rates on average test scores. A utilitarian administrator—weighting all students' test scores equally—could use this parameter to determine optimal suspension rates at their school. However, Equation (4) implies that the effect on average test scores combines both the direct and indirect effect of suspension rates on student test scores. The positive estimates in Table 2 suggest that when suspension rates declined in Los Angeles, the average decline in test scores due to indirect learning spillovers exceeded the average increase in test scores due to direct benefits from suspending fewer students. Even if the direct effect of being suspended is large and negative, these impacts are concentrated on a limited number of students and are outweighed by the accumulation of the small but diffuse benefits from learning spillovers. We return to this discussion of indirect and direct effects in Section 7.

6.2 Absences and GPA

Suspension policies may also influence students' desire to attend school as well as other aspects of in-class achievement not captured by test scores. We therefore also estimate the effects of school suspension rates on GPA and absences. We standardize GPA by grade and year, and we measure absences as the fraction of non-suspended days a student is absent to prevent absences from being mechanically influenced by suspensions.

Table 3 presents OLS and IV estimates for these two outcomes. The raw OLS specification shows that higher suspension rates are negatively correlated with GPA and positively correlated with absences. Adding school-grade fixed effects and control variables changes the signs of both estimates. In the fully controlled OLS specification, a 10 percentage point increase in suspension rates increases GPAs by 0.011 standard deviations and decreases absences by 0.7 percentage points (9.6%); both estimates are statistically significant at the five percent level. The final column provides the IV estimates, which indicate that a 10 percentage point increase in suspension rates increases GPAs by 0.067 standard deviations and decreases absences by 1.1 percentage point (15.1%). The GPA estimate is similar in magnitude and direction as the previous test score estimates. The effect on absences suggests that harsher discipline policies could potentially deter a broad class of misbehavior that may include skipping class. Students may also feel less inclined to attend class when the classroom environment is more prone to disruption from misbehavior or when exposed to more bullying or violence.

6.3 Teacher Attrition

Suspension rate decreases may also affect teachers. Fewer disciplinary options could make classroom behavior management more taxing, especially for inexperienced teachers. Teaching in a classroom with more misbehavior is also generally less enjoyable. Ultimately, difficult and unpleasant working conditions could lead to increases in teacher turnover. Using classroom-level data linked to teachers, we estimate the effects of suspension rates on teacher turnover by using our IV approach while controlling for lagged school-grade test scores and school-grade fixed effects.[15] The outcome of interest is an indicator equal to one if a teacher leaves their school between years t and $t+1$.

In Table 4, we estimate that a 10 percentage point increase in suspension rates leads to a 2 percentage point (9.9%) decrease in teacher turnover. The baseline turnover rate in Los Angeles is quite high at 20.3%, an estimate which is consistent with previous research. Newton et al. (2011), for example, verify that the probability that an LAUSD elementary school teacher leaves their school is 21.6% after their first year and 19.5% after the second. Similarly, they find that 26.4% of high school teachers leave after their first year and 21.6% leave after the second. Panels B and C present separate estimates for inexperienced teachers with 0-2 years of experience (aligning with the years prior to when tenure decisions are made) and teachers with three or more years of experience.[16] The point estimate for inexperienced teachers is more than triple the size of the point estimate for experienced teachers. For inexperienced teachers, a 10 percentage point increase in suspension rates leads to a 8.8 percentage point (28%) increase in turnover. These effects on teacher turnover are quite large, particularly for inexperienced teachers.

Due to the large effect of suspension rates on teacher turnover, it appears that teachers value the ability to suspend students. Clotfelter et al. (2008) estimate that a $1,800 bonus payment reduces teacher turnover by 17%. Using their estimate as a benchmark, the school district would need to pay teachers $1,043 more per year in order to maintain stable attrition rates when suspension rates decrease by 10 percentage points. Inexperienced teachers would need to be paid $2,967 more to offset a 10 percentage point decrease in suspension rates.

[15]To assign teachers to a school-grade fixed effect, we choose the grade level with the greatest number of students that the teacher teaches.

[16]We do not directly observe teacher experience. Teachers who enter the data for the first time are assumed to have zero years of experience. Teachers who remain in the data from 2003 to 2005 mechanically have at minimum three years of experience. This assumption requires that we omit 2005 from our estimation sample.

6.4 Effects by Grade

There are several reasons why changes in suspension rates might exhibit differential impacts by grade. First, the large differences in suspension rates by grade, displayed in Figure 2, could cause the marginal effect of changing suspension rates to differ across grades. In addition, the disparate nature of how elementary, middle, and high schools are taught and organized may change how suspension policies affect achievement. Lastly, the nature of misbehavior could also differ across grades. Figure 4 suggests that the marginal suspension in middle and high school may be more serious (or treated more harshly) than the marginal suspension in elementary school. This could cause both direct effects and learning spillovers from suspensions to vary across grades.

We test for these differences by estimating our IV results separately for elementary, middle, and high school students. The identifying assumption for these estimates remains the same. Table 5 presents these estimates. The test score coefficients remain positive and significant for elementary, middle, and high school students. However, the effect appears much larger for elementary students (0.097 and 0.183 standard deviations for math and English). We note that since the baseline suspension rate for elementary students is three to four times lower than for middle and high school students, a 10 percentage point change in suspension rates represents a much larger policy change. The lower baseline suspension rate also suggests that an elementary student on the margin of being suspended could be more disruptive than the marginal student in other grades. The estimates for middle and high school students are similar in magnitude to our baseline estimates (0.034 and 0.060 standard deviations for math and English in middle school and 0.062 and 0.67 standard deviations in high school).

We also find that the effects on GPA are positive and significant across all types of students, with effect sizes of 0.121, 0.071, and 0.042 standard deviations for elementary, middle, and high school students. These effects suggest suspension policies likely impact students' grades similarly to test scores. In contrast to test scores and GPA, the effect of suspension rates on absences appears to increase with grade level. For elementary students, the effects are three times smaller than for high school students (but are similar in percent terms relative to the baseline mean). The smaller effect size could be due to the limited autonomy that elementary students have over the decision to attend class, a decision that parents typically make for their young children. In middle and high school, increasing suspension rates by 10 percentage points decreases absences by 1.0 percentage points (14.9%) and 2.7 (23.1%) percentage points. These effect sizes are quite large and are equivalent to 1.8 and 4.9 days of lost instruction for middle and high school students per school year. The results suggest that as students mature, their school attendance patterns may become more responsive to changes in suspension rates.

In addition to the effect on students, Table 6 shows that the effect on teacher attrition also appears to increase with grade level. While we find no significant impact on elementary school teachers, suspension rates have a large impact on teacher attrition for high school teachers and inexperienced middle school teachers. Inexperienced high school teachers are most impacted: a 10 percentage point increase in suspension rates decreases the likelihood of attrition by 11.7 percentage points (39%). Higher suspension rates also decrease attrition among inexperienced middle school teachers (-7.1 percentage points) and experienced high school teachers (-5.3 percentage points). These results suggest that the marginal misbehavior by older students is more costly to teachers than that of younger students. These differences could arise for several reasons. Alternative in-school disciplinary methods (i.e. non-suspension discipline) could be more effective for younger students than for older students. Misbehavior by older students could also affect teachers differently and may be more unpleasant to deal with. For example, high school students are more physically developed and approximately 9 percent of teachers are physically threatened each year.[17] Teachers who feel threatened but cannot safely respond without the use of suspensions may be more likely to leave.

6.5 Robustness Checks

Serial Correlation and Lagged Suspension Rates: To help verify our main results, we conduct several sensitivity tests. The main IV results are estimated on a sample beginning in 2005. However, serial correlation in the regression residual may cause lagged suspension rates in the instrument to be endogenous, even if they are fixed to an initial pre-period. We test the sensitivity of our results by increasing the lag between the initial suspension year and the first year in the sample. Specifically, we re-estimate the results on two alternate samples, one beginning in 2006 and the other beginning in 2007. These estimates are presented in columns (2) and (3) of Table 7. In both columns, effects on math and English test scores both increase slightly relative to our baseline estimates. The effects also remain precisely estimated and significant at the one percent level. We find that the same pattern holds for GPA and absences in Table A.2. These findings partially alleviate concerns about bias from serial correlation.

School-level variation: We also test whether the results are sensitive to the choice to use school-grade

[17] National Center for Education Statistics, Digest of Education Statistics, Table 228.70; Link available here.

level variation versus school-level variation. Schools may be unlikely to differentiate suspension policy (among other school policies) by grade. The regression residual may also exhibit school-specific serial correlation, implying that standard errors should be adjusted for clustering at the school level. Column (4) of Table 7 shows how the IV estimates change when using school-level suspension rates and school-level clustered standard errors. The point estimates and standard errors increase slightly; however, the effects are still significant at the one percent level. The results also hold for GPA and absences in Table A.2. The decision to use school-grade variation or school variation does not appear to meaningfully impact the results.

Twice-lagged suspension rates: We also consider a variation of our instrument in which we use twice-lagged suspension rates instead of fixing suspension rates to 2003. The reason for this stems from the accounting identity in Equation (7). Using twice-lagged suspension rates instead of fixed initial suspension rates may theoretically increase the instrument's predictive power. However, column (5) of Table 7 shows that the first-stage F-statistic *falls* when the instrument is constructed in this way. This may be because a school's exposure to districtwide suspension rate growth is more correlated with initial conditions than with recent suspension rates. We nevertheless find that the estimates for all four outcomes change little when constructing the instrument in this way. We also note that serial correlation could play a larger role when the instrument is constructed with twice-lagged suspension rates. However, the similarity of the two estimates provides additional assurance that the bias from serial correlation is relatively small.

District level instrument variation: Our main specification instruments for each school's suspension rate using year-to-year changes in the districtwide suspension rate interacted with school-specific suspension rates in 2003. We also estimate the effect of suspension rates using only the variation from year-to-year changes in the districtwide suspension rate without scaling each school's exposure by their initial pre-period suspension rate level. In this robustness test, the instrument used to estimate Equation (5) is simply G_{gt}^{-s}. We report these estimates in column (6) of Tables 7 and A.2. The effects are positive, significant, and 30-80 percent larger than our main results. Despite using only districtwide variation to construct the instrument, the results are broadly consistent with our previous findings.

7 Direct and Indirect Effects of Suspensions

7.1 Conceptual Framework

In this section, we discuss our approach to quantifying the direct and indirect effects underlying our main estimates. As outlined in Section 2, schools determine the strictness of their suspension policies by maximizing the profit function in Equation (1). The first order condition in Equation (4) implies that changes in suspension rates impact students through distinct direct and indirect effects.

The direct effect represents the individual impact of being suspended. The most immediate consequence of being suspended is reduced classroom time. On average, LAUSD students miss 2.1 days of school per suspension. As part of this forgone learning, suspensions may also disrupt learning continuity. Being suspended could also influence student motivation and engagement, although the direction of these effects is not clear *ex ante*. Due to lost instruction, increasing suspension rates will likely have negative direct effects on test scores. However, the direct effects could be positive if suspensions act as a catalyst for reforming students' future behavior. Direct effects only impact the marginal students that become suspended (not suspended) when suspension rates increase (decrease).

Indirect effects arise from the change in students' probability of misbehaving and the subsequent impact on learning in the classroom. Misbehavior disrupts class and diverts a teacher's time and energy away from instruction, potentially producing negative spillovers on learning. In 2016, 43 percent of teachers "agreed" or "strongly agreed" that student misbehavior interfered with their teaching during the year (Musu-Gillette et al., 2018). Suspensions provide teachers with one way to curtail disruptions and prevent misbehavior from escalating.[18] In contrast to direct effects, indirect effects impact all students.

Causal evidence on the direction and magnitude of direct effects is limited. A recent meta-analysis by Noltemeyer, Ward and Mcloughlin (2015) summarizes the correlational research across 34 studies in the education literature and finds a negative correlation between achievement and being suspended. In the causal literature, Lacoe and Steinberg (2018*a*) find negative direct effects in Philadelphia using an individual fixed effects approach while Anderson, Ritter and Zamarro (2017) find positive effects of being suspended using dynamic panel methods in Arkansas. Beyond test scores, concurrent work by Bacher-Hicks, Billings and Deming (2019) study students who are assigned to schools with varying suspension rates due to a school zone boundary change. They find that attending a

[18] However, not all suspensions are equally productive at reducing spillovers. Suspensions motivated by implicit racial biases or targeted towards minor infractions may produce small or no effects.

school with stricter suspension policies leads to lower graduation rates, lower rates of college attendance, and higher rates of future crime, while finding no impact on test scores. Causal identification of indirect effects is also challenging, and little work has been done to identify the spillover effects of suspension policies. However, beyond suspensions, there exists a robust literature on peer effects in schools (Sacerdote, 2011).[19]

7.2 Empirical Estimates

Our main results show that an increase in suspension rates increases average test scores, improves student GPAs, and decreases absenteeism. Since the observed effects combine indirect and direct effects, the positive estimates on test scores imply that the total indirect effect is larger than the total direct effect. However, the estimates provide little information on the underlying magnitudes of direct versus indirect effects. We use two approaches to disentangle and quantify these separate effects.

The first approach exploits the fact that the direct effect of being suspended only impacts students who are suspended. For students who rarely misbehave and are unlikely to ever be suspended, a change in suspension rates will have no effect on them through the direct effect. However, they are still affected by suspension rate changes through the indirect effect, which impacts all students. Under the strong assumption that indirect effects impact all students equally, we can potentially isolate the direct and indirect effect by comparing estimates between students that have a high and low likelihood of being suspended. Estimates for students unlikely to be suspended represent the indirect effect, whereas the difference in estimates between high- and low-suspension students, weighted by the differential likelihood of being suspended, represents the direct effect.

We first construct a proxy of an individual's propensity to be suspended. We estimate a linear probability model on an out-of-sample set of observations from 2004 to predict the probability that a student in future years was suspended.[20] The linear probability model we use is as follows and is estimated separately by grade:

$$Suspended_{isg,2004} = \beta_0 + \beta_1 X_{isg,2003} + \beta_2 S_{sg,2003} + \beta_3 P_{isg,2003} + \epsilon_{isg,2004} \tag{9}$$

[19] For two examples closely related to school discipline, Imberman, Kugler and Sacerdote (2012) study students with disciplinary problems who were displaced by Hurricanes Katrina and Rita. Local students in districts receiving these students experienced increases in disciplinary problems and absenteeism, although there was no impact on test scores. Carrell and Hoekstra (2010) study children from families matched to domestic violence cases, and find negative effects of such students on the performance of their peers.

[20] We use data from 2004 because it is the only year omitted from our analysis that contains lagged information about students. We note that suspension policies in 2004 were stricter than in the later years of our sample. This implies that the predictions will likely contain information about more minor forms of misbehavior that may not be captured by more recent suspensions.

where $X_{isg,2003}$ is a vector of lagged math and English test scores, GPA, fraction of non-suspended days absent, a suspension indicator, and days suspended.[21] $S_{sg,2003}$ is a vector of lagged school-grade math and English test scores, and $P_{isg,2003}$ includes lagged test scores for peers of student i.

We use the estimates from this model to produce a suspension propensity for each student in each future year. Figure A.2 provides a histogram of these predictions for elementary, middle, and high school students. Predictions in elementary school are clustered near zero, while predictions in middle and high school exhibit greater dispersion. Within each grade, we split students into terciles based on their predicted suspension propensities and we produce IV estimates for each of the three resulting subsamples.

We present the IV estimates for each tercile in Panel A of Table 8. The math estimates remain positive and significant for the bottom two terciles but are statically insignificant for the high suspension tercile (from low to high suspensions: 0.046, 0.090, and 0.002 standard deviations). The English estimates are similar in magnitude but are also significant for the high suspension tercile (from low to high suspensions: 0.050, 0.119, and 0.029 standard deviations). Both sets of estimates suggest that higher suspension rates are more beneficial to lower-suspension students than higher-suspension students.

Under the strong assumption that indirect effects impact all students equally and that low-suspension students are not suspended (the suspension rate for these students is 2%), the coefficients on low-suspension students represent the indirect effect of a 10 percentage point increase in suspension rates on test scores. As seen in Table 8, the indirect effects are 0.046 and 0.050 standard deviations for math and English. The direct effect of being suspended is represented by the difference in effect sizes between high- and low-suspension students divided by the difference in the fraction suspended. The English estimates suggest that the direct effect of being suspended is negative and very large, -0.49 standard deviations.[22] For math, the estimated direct effect is half as large, -0.23 standard deviations.

When comparing indirect and direct effects, the positive indirect effect per student on English is 0.050 standard deviations while the negative direct effect per student (including both suspended and non-suspended students) is only -0.014 standard deviations.[23] For math, the per-student indirect and direct effects are 0.046 and -0.029 standard deviations. The per-student net effect is therefore 0.017 and 0.036 standard deviations for math and English, respectively. This is about half as large as the per-

[21] Unfortunately, the data do not contain many demographics such as gender, race, and socioeconomic status. In addition, fraction of non-suspended days absent is not available for elementary students in 2003.

[22] We calculate this by subtracting the low-suspension estimate from the high-suspension estimate (0.002-0.046 standard deviations) and dividing by the difference in suspension rates between the two groups (0.11-0.02).

[23] The negative direct effect per student is calculated by taking the direct effect of being suspended (-0.23 standard deviations) and multiplying by the overall fraction of students suspended (0.06).

student effects presented in Table 2. Panel A of Table A.3 reports the results for GPA and absences. For GPA, the positive indirect effect per student is much larger than the negative direct effect per student. For absences, the indirect and direct effect are both beneficial for students.[24] Overall, this approach suggests that per-student indirect effects are larger than per-student direct effects, even though the individual direct effect of being suspended appears quite large.

Our second approach relaxes the assumption of constant indirect effects and allows the size of indirect effects to vary across low-, medium-, and high-suspension students. We do so by separating school suspension rates into a self and peer component. For a given student i, the self component is simply an indicator for whether the student was suspended. The peer component is the suspension rate of all other students in the same school and grade, excluding student i. The revised IV regression is as follows:

$$y_{isgt} = \alpha + \rho_1 SuspendRate_{sgt}^{-i} + \rho_2 Suspended_{isgt} + \beta_1 X_{isgt-1} + \beta_2 S_{sgt-1} + \beta_3 P_{isgt-1} + \lambda_{sg} + \epsilon_{isgt} \quad (10)$$

$SuspendRate_{sgt}^{-i}$ represents the suspension rate across all peers in grade g and school s, leaving out the suspension contribution from student i himself. $Suspended_{isgt}$ is a binary variable equal to 1 if student i is suspended in year t. Conceptually, ρ_1 and ρ_2 represent estimates of the indirect effect of suspension rates and the direct effect of being suspended, respectively. We use the instrument from our main analysis to instrument for the leave-own-out suspension rate. We then compare ρ_1, which is comprised solely of the indirect effect, to our previous estimates of ρ in Panel A, which represent the aggregate of both direct and indirect effects. Here, ρ_1 represents a causal estimate of the indirect effect per student whereas $\rho - \rho_1$ represents the direct effect per student. While we lack an instrument for estimating ρ_2, the coefficients may still be informative after controlling for lagged achievement controls and school-grade fixed effects. Comparing ρ_1 and ρ_2 also provides insight on the relative magnitude between indirect and direct effects for students at the margin of being suspended.

Panel B of Table 8 reports these estimates. The first row contains estimates of indirect effects, which vary across low-, medium-, and high-suspension students. For both math and English, the effect sizes are similar to the aggregate effect sizes in Panel A. This suggests that the overall effect of suspension rates is largely driven by indirect effects. The difference between the aggregate effect size

[24] For GPA, the estimated indirect effect per student is 0.043 standard deviations, whereas the estimated direct effect is 0.003 standard deviations. For fraction of days absent, the estimated indirect and direct effects per student are -0.005 and -0.007 percentage points.

in Panel A and the indirect effect size in Panel B is representative of the size of the direct effect per student. For both math and English, the direct effect per student is small for the average low-suspension student (-0.004 standard deviations for both math and English) but is larger for high-suspension students (-0.010 and -0.015 standard deviations for math and English). Since only 12 percent of high-suspension students are suspended, these results suggest that the individual direct effect of being suspended is an order of magnitude larger than the per-student direct effect.

The coefficients on $Suspended_{isgt}$, though not necessarily causal, may still shed light on the magnitude of the direct effect on suspended students. The effects are consistently large and negative. The negative effect on test scores of being suspended ranges from 0.076 to 0.147 standard deviations. Interestingly, the direct effect of being suspended is larger for low-suspension students (-0.147 and -0.138 standard deviations) than for high-suspension students (-0.076 and -0.102 standard deviations). Low-suspension students may find being suspended more traumatizing, and there may be diminishing negative effects to each additional suspension. Low-suspension students who are high achievers may also have more learning to lose by being suspended. These estimates of being suspended are roughly comparable to the effect of having a one standard deviation lower value-added teacher.

The sum of the indirect and direct effects in Panel B represents the effect of a 10 percentage point increase in suspension rates on a student who becomes suspended due to the suspension rate increase. The net effect is negative for all three categories of students in math (from low to high: -0.097, -0.010, and -0.064 standard deviations) and English (from low to high: -0.084, -0.004, and -0.058 standard deviations). While the average student benefits from higher suspension rates, the direct effect exceeds the indirect effect for students suspended as a result of the policy change. This implies that any decision to change suspension rates will generally create an efficiency versus equality tradeoff. Although higher suspension rates increase average student achievement due to lower misbehavior spillovers, higher suspension rates also produce inequality in education production because the cost of stricter suspension policies is disproportionately borne by students who are likely to be low achievers.

We also provide a breakdown of how indirect, direct, and aggregate effects differ for students in elementary, middle, and high school. The estimates are presented in Table 9. The previous patterns generally hold for all grade categories: a positive indirect effect per student accompanied by large, negative direct effects.

8 Conclusion

In this paper, we provide evidence on the multi-faceted consequences of changing school suspension policies. Our empirical approach instruments for each school's suspension rate using year-to-year changes in the districtwide suspension rate interacted with school-specific suspension rates fixed to an initial pre-period. We find that an increase in suspension rates leads to increases in math and English test scores, increases in GPAs, and decreases in absenteeism. In addition, we find a negative relationship between suspension rates and teacher turnover, suggesting that teachers value the ability to suspend students.

While we find that the suspension rate decrease in Los Angeles was detrimental to the average student, this result does not tell the entire story. The effect of changing suspension policies is comprised of direct effects on students who are suspended, as well as indirect spillover effects on all students schoolwide. This paper provides a framework to conceptualize and quantify these direct and indirect effects. We find that an increase in suspension rates leads to a positive indirect effect on students, which is likely due to reduced misbehavior in the classroom. However, we also find that the direct effect of being suspended is quite detrimental, but ultimately affects relatively few students. The positive net effects of higher suspension rates are therefore driven by the accumulation of small but diffuse benefits from indirect spillovers, which outweigh the large but concentrated direct effects of being suspended. Net achievement falls for students who become suspended as a result of higher suspension rates. Because such students are more likely to be low-achieving students, higher suspension rates may exacerbate educational inequality despite increasing average achievement.

The important role of student behavior in the learning process necessitates that school, district, and state administrators determine the best policies to manage their students' behavior. While other forms of behavior management are also important to consider, suspensions have historically played a key role in shaping these policies. Administrators (whether explicitly or implicitly) determine the level of their suspension rates through the strictness of their disciplinary policies. The estimates from this paper can help inform the tradeoffs underlying the decision-making process so that schools can better determine their optimal suspension policies.

References

Anderson, Kaitlin P., Gary W. Ritter, and Gema Zamarro. 2017. "Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?" EDRE Working Paper 2017-09.

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

Barrett, Nathan, Andrew McEachin, Jonathan N. Mills, and Jon Valant. 2017. "Disparities in Student Discipline by Race and Family Income." Education Research Alliance for New Orleans Technical Report.

Blume, Howard. 2012. "Suspension figures called "alarming"." *Los Angeles Times*.

Carrell, Scott E., and Mark L. Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics*, 2(1): 211–28.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review*, 108(11): 3377–3415.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina." *Journal of Public Economics*, 92(5): 1352 – 1370.

Curran, F. Chris. 2016. "Estimating the Effect of State Zero Tolerance Laws on Exclusionary Discipline, Racial Discipline Gaps, and Student Behavior." *Educational Evaluation and Policy Analysis*, 38(4): 647–668.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2018. "Bartik Instruments: What, When, Why, and How." National Bureau of Economic Research, Inc NBER Working Papers 24408.

Hashim, Ayesha K., Katharine O. Strunk, and Tasminda K. Dhaliwal. 2018. "Justice for All? Suspension Bans and Restorative Justice Programs in the Los Angeles Unified School District." *Peabody Journal of Education*, 93(2): 174–189.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review*, 102(5): 2048–82.

Lacoe, Johanna, and Matthew P. Steinberg. 2017. "What Do We Know About School Discipline Reform?" *Education Next*, 17(1).

Lacoe, Johanna, and Matthew P. Steinberg. 2018*a*. "Do Suspensions Affect Student Outcomes?" *Educational Evaluation and Policy Analysis*.

Lacoe, Johanna, and Matthew P. Steinberg. 2018*b*. "Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes." *Peabody Journal of Education*, 93(2): 207–227.

Lazear, Edward P. 2001. "Educational Production*." *The Quarterly Journal of Economics*, 116(3): 777–803.

Losen, Daniel, Cheri Hodson, Michael A. Keith III, Katrina Morrison, and Shakti Belway. 2015. "Are We Closing the School Discipline Gap?" The Center for Civil Rights Remedies.

Musu-Gillette, Lauren, Anlan Zhang, Jizhi Zhang, Jana Kemp, Melissa Diliberti, and Barbara A. Oudekerk. 2018. "Indicators of School Crime and Safety: 2017." National Center for Education Statistics, Institute of Education Sciences.

Newton, Xiaoxia A., Rosario Rivero, Bruce Fuller, and Luke Dauter. 2011. "Teacher Stability and Turnover in Los Angeles: The Influence of Teacher and School Characteristics." Policy Analysis for California Education Los Angeles School Infrastructure Project.

Noltemeyer, Amity L., Rose Marie Ward, and Caven Mcloughlin. 2015. "Relationship Between School Suspension and Student Outcomes: A Meta-Analysis." *School Psychology Review*, 44(2): 224–240.

Okun, Arthur M. 2015. *Equality and Efficiency: The Big Tradeoff.* Brookings Institution Press.

Petek, Nathan, and Nolan Pope. 2018. "The Multidimensional Impact of Teachers on Students." PhD diss. University of Chicago.

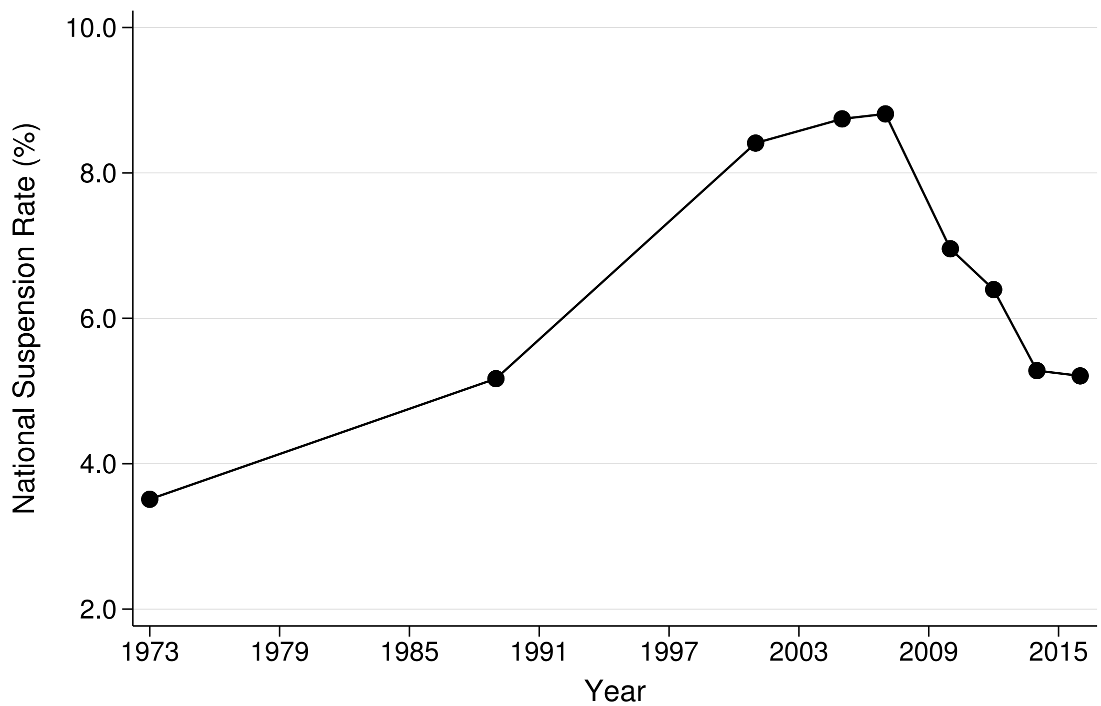
Sacerdote, Bruce. 2011. "Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In . Vol. 3 of *Handbook of the Economics of Education*, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 249 – 277. Elsevier.

Teske, Steven C. 2011. "A Study of Zero Tolerance Policies in Schools: A Multi-Integrated Systems Approach to Improve Outcomes for Adolescents." *Journal of Child and Adolescent Psychiatric Nursing*, 24(2): 88–97.

U.S. Department of Education. 2014. "U.S. Departments of Education and Justice Release School Discipline Guidance Package to Enhance School Climate and Improve School Discipline Policies/Practices."

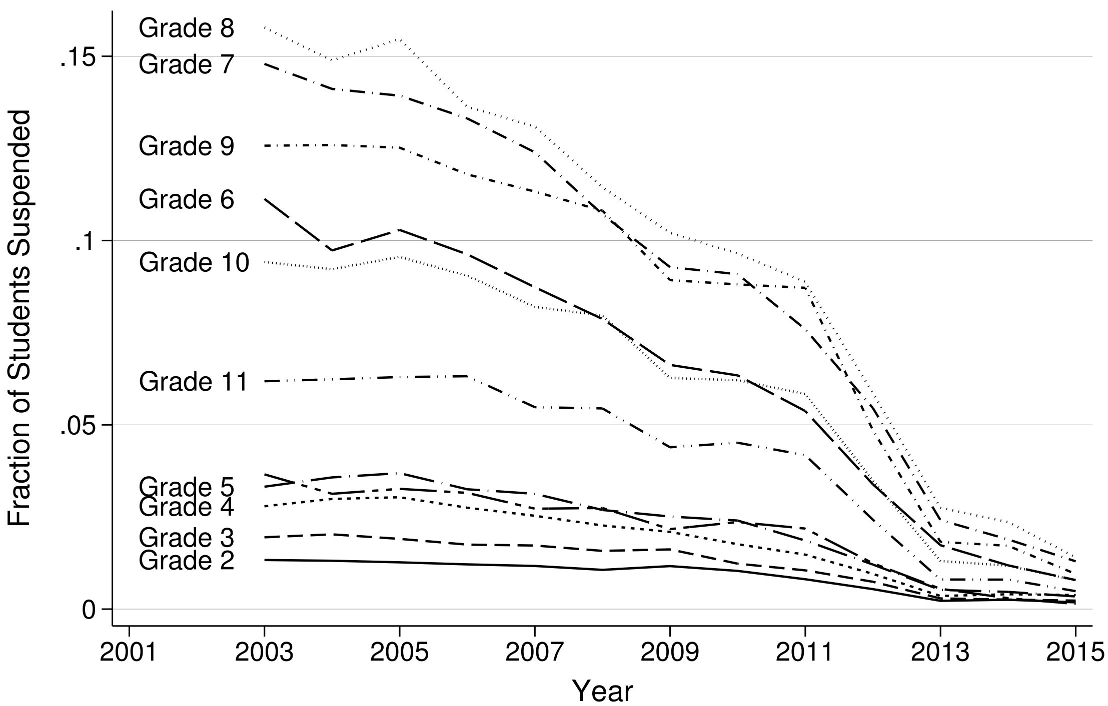
Watanabe, Teresa. May 14, 2013. "L.A. Unified bans suspension for willful defiance." *Los Angeles Times*.

Figure 1: National Suspension Rates



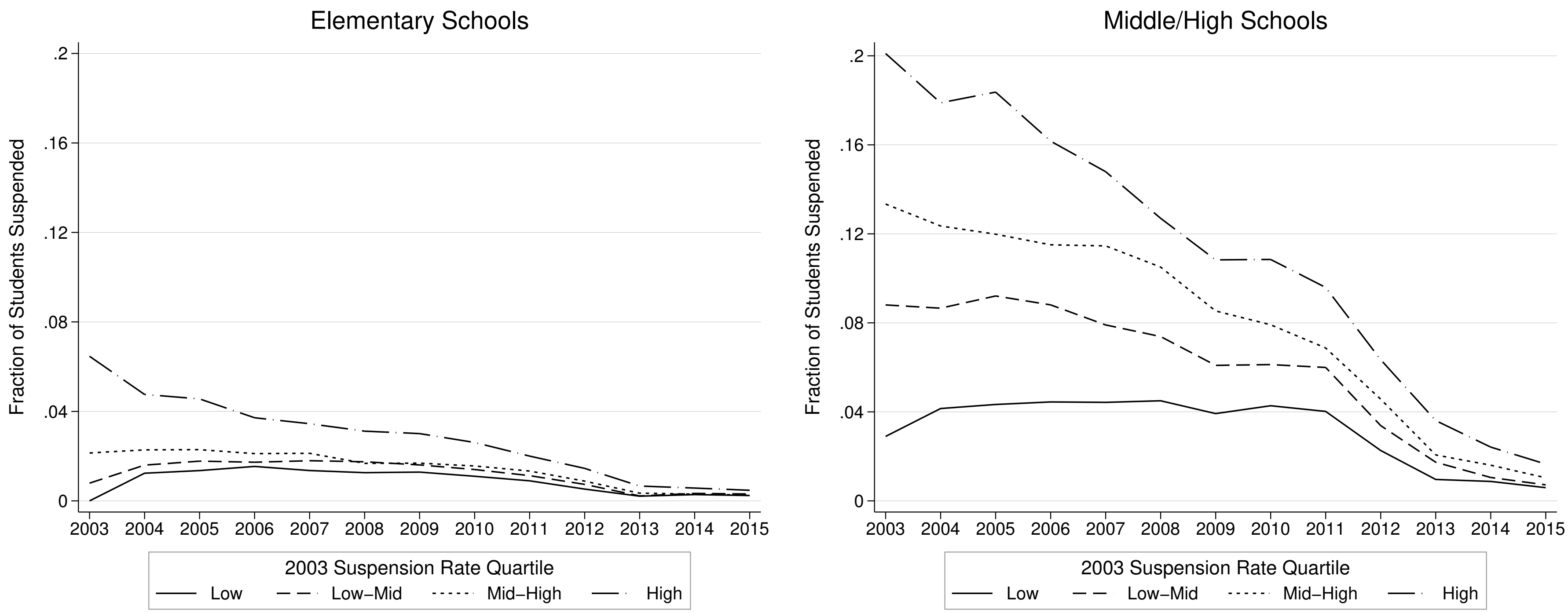
Note: This figure shows national suspension rates using data compiled by the Civil Rights Data Collection by the Office for Civil Rights. Suspension rates are calculated by dividing the total number of suspended students (including both students with and without disabilities) by the total number of enrolled students.

Figure 2: LAUSD Suspension Rates by Grade



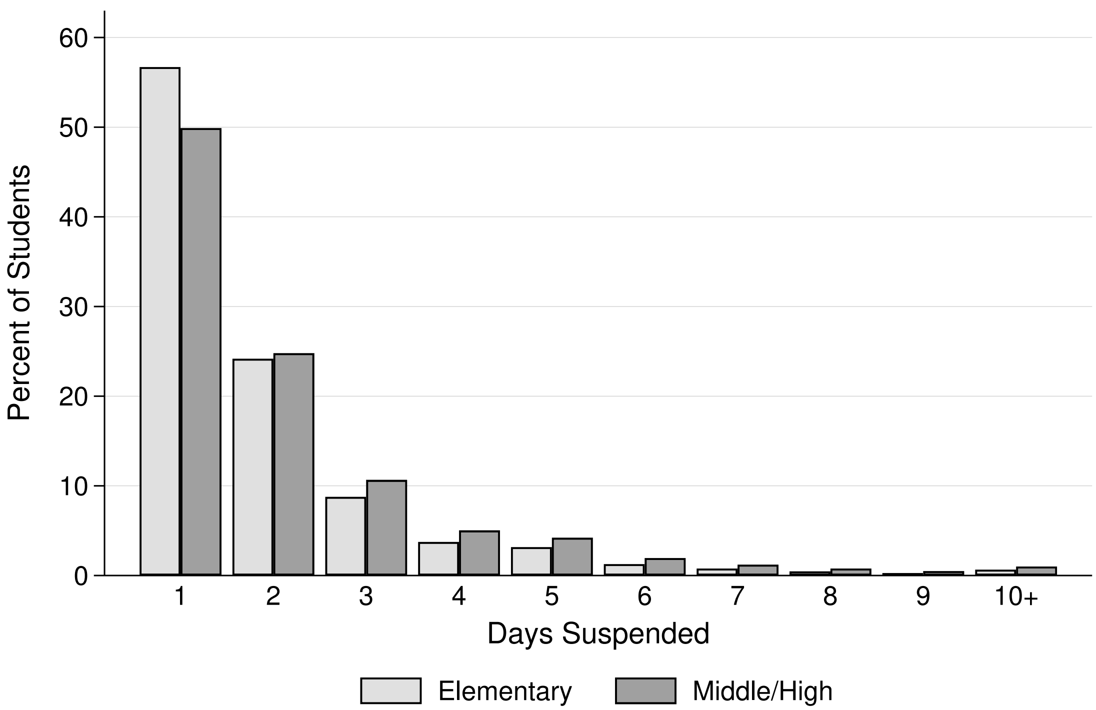
Note: This figure shows districtwide suspension rates for each grade in the LAUSD for each year. Suspension rates are calculated by dividing the total number of suspended students by the total number of enrolled students in each grade.

Figure 3: School-Grade Suspension Rates by 2003 Suspension Rate Quartile



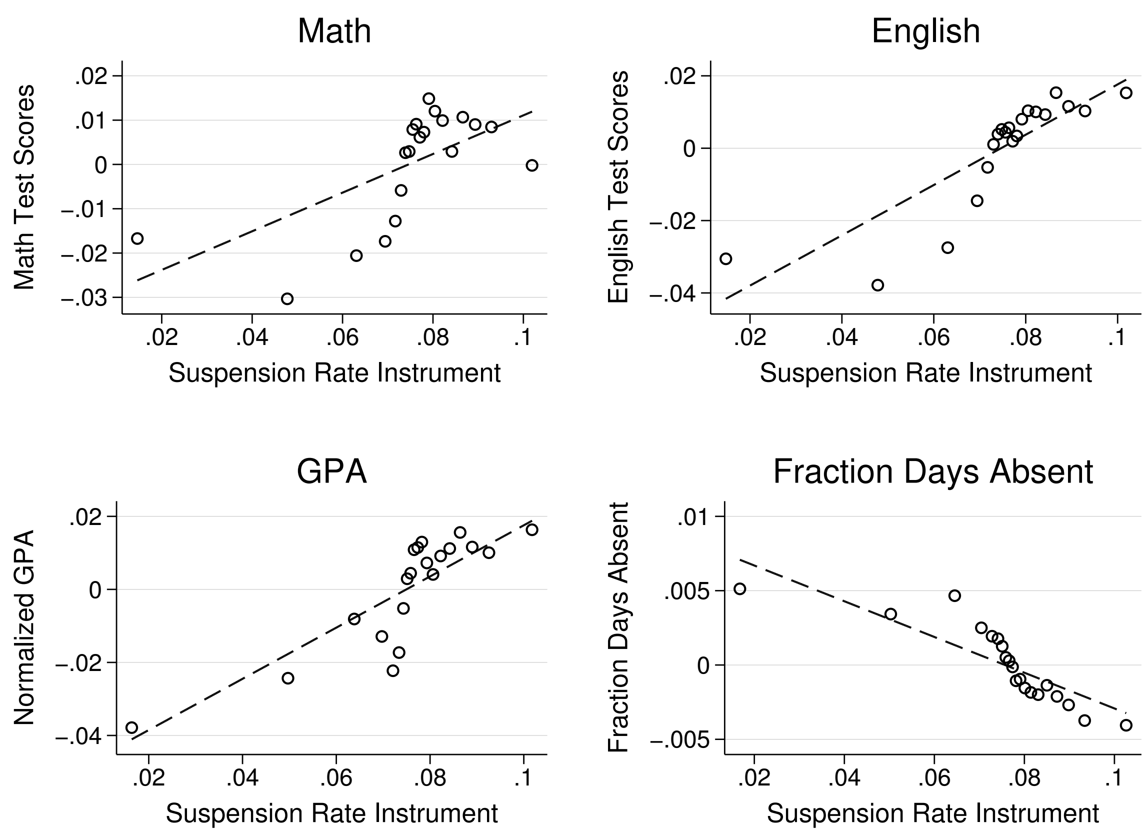
Note: This figure plots the trajectory of suspension rates for elementary and middle/high students in the LAUSD, dividing school-grades into one of four equally-sized quartiles based on initial 2003 suspension rates. Cutoffs for each of the four quartiles are as follows. For elementary school-grades: 0%, 1.3%, and 3.2%. For middle/high school-grades: 6.1%, 10.9%, and 15.5%. Average suspension rates are then calculated for each quartile, weighted by the number of students in each school-grade.

Figure 4: Distribution of Days Suspended



Note: This figure shows the distribution of the number of days suspended in the LAUSD, conditional on having been suspended at least once during the year. The average number of days suspended is 2.1. The data are aggregated across the years 2003-2015.

Figure 5: Reduced Form Relationship between Academic Outcomes and Suspension Rate Instrument



Note: This figure plots average math and English test scores, normalized GPA, and the fraction of non-suspended days absent against binned values of the instrument, after residualizing both axes with respect to lagged individual, school, and peer achievement as well as school-by-grade fixed effects (as shown in Equation (5)). The instrument is calculated based on Equation (6).

Table 1: LAUSD Summary Statistics

	Elementary School				Middle School				High School			
	Low Suspensions		High Suspensions		Low Suspensions		High Suspensions		Low Suspensions		High Suspensions	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Standardized Math Scores	0.05	1.02	-0.05	0.98	0.18	1.06	-0.08	0.96	-0.03	0.94	0.00	1.00
Standardized English Scores	0.05	1.02	-0.04	0.97	0.17	1.03	-0.08	0.98	-0.12	0.97	0.01	1.00
Standardized GPA	0.03	1.01	-0.03	0.99	0.10	1.00	-0.05	1.00	0.14	0.97	-0.01	1.00
Fraction Days Absent	0.04	0.05	0.04	0.05	0.05	0.07	0.06	0.08	0.13	0.21	0.10	0.13
English Language Learner	0.38	0.48	0.37	0.48	0.20	0.40	0.26	0.44	0.09	0.29	0.18	0.38
Suspended	0.01	0.10	0.02	0.15	0.06	0.24	0.11	0.31	0.02	0.13	0.07	0.25
Days Suspended (If Suspended)	1.82	1.44	1.93	1.64	2.05	1.79	2.32	2.08	1.82	1.40	2.04	1.57
# Times Suspended	1.23	0.65	1.33	0.80	1.45	0.94	1.62	1.17	1.29	0.72	1.36	0.80
School Size	433	213	440	213	850	879	1,570	734	474	746	2,008	1,562
Number of Schools	190		191		51		51		61		61	
Number of Observations	1,032,545		1,086,721		515,097		1,038,062		340,503		1,508,420	

Note: This table provides summary stats for student and school characteristics, split by elementary, middle, and high school students. Within each category, schools are divided into “low” and “high” suspension schools based on whether the school’s suspension rate in 2003 was above or below the median. The sample includes all students enrolled in grades 2-11 from 2003 to 2015. However, test scores are only available through 2013.

Table 2: Effects of School Suspension Rates on Test Scores

	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
A. Math Test Scores						
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.164***	0.003	0.007	0.006	0.005	0.040***
	(0.013)	(0.008)	(0.006)	(0.005)	(0.005)	(0.009)
N	2,335,653	2,335,653	2,335,653	2,335,653	2,335,653	2,335,653
F-Statistic (IV First Stage)						1,421
School-Grade Fixed Effects		Yes	Yes	Yes	Yes	Yes
Individual Lagged Achievement			Yes	Yes	Yes	Yes
Lagged Average School Test Scores				Yes	Yes	Yes
Lagged Peer Test Scores					Yes	Yes
	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
B. English Test Scores						
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.146***	0.009*	0.020***	0.020***	0.019***	0.064***
	(0.013)	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
N	2,208,372	2,208,372	2,208,372	2,208,372	2,208,372	2,208,372
F-Statistic (IV First Stage)						1,267
School-Grade Fixed Effects		Yes	Yes	Yes	Yes	Yes
Individual Lagged Achievement			Yes	Yes	Yes	Yes
Lagged Average School Test Scores				Yes	Yes	Yes
Lagged Peer Test Scores					Yes	Yes

Note: This table presents the effect of a 10 percentage point increase in suspension rates on normalized math and English test scores. The full OLS regression is estimated as in Equation (5). The IV estimates use instrumented suspension rates as calculated in Equation (6). The sample includes all students enrolled in grades 3-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: Effects of School Suspension Rates on GPA and Absences

	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
A. Normalized GPA						
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.108***	-0.011*	0.011**	0.011**	0.011**	0.067***
	(0.009)	(0.006)	(0.005)	(0.005)	(0.005)	(0.009)
N	2,701,775	2,701,775	2,701,775	2,701,775	2,701,775	2,701,775
F-Statistic (IV First Stage)						1,237
School-Grade Fixed Effects		Yes	Yes	Yes	Yes	Yes
Individual Lagged Achievement			Yes	Yes	Yes	Yes
Lagged Average School Test Scores				Yes	Yes	Yes
Lagged Peer Test Scores					Yes	Yes
	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
B. Fraction Days Absent (Non-Suspended)						
(Suspension Rate)$_{sgt}$ $\times$ 10	0.014***	0.002***	-0.007***	-0.007***	-0.007***	-0.011***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
N	2,744,787	2,744,787	2,744,787	2,744,787	2,744,787	2,744,787
Baseline Mean	0.073	0.073	0.073	0.073	0.073	0.073
F-Statistic (IV First Stage)						1,212
School-Grade Fixed Effects		Yes	Yes	Yes	Yes	Yes
Individual Lagged Achievement			Yes	Yes	Yes	Yes
Lagged Average School Test Scores				Yes	Yes	Yes
Lagged Peer Test Scores					Yes	Yes

Note: This table presents the effect of a 10 percentage point increase in suspension rates on normalized GPA and the fraction of non-suspended days absent. The full OLS regression is estimated as in Equation (5). The IV estimates use instrumented suspension rates as calculated in Equation (6). The sample includes all students enrolled in grades 3-11 from 2005 to 2014 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Effects of School Suspension Rates on Teacher Attrition

P(Teacher leaves school between t, $t+1$)	OLS			IV
	(1)	(2)	(3)	(4)
A. All Teachers				
(Suspension Rate)$_{sgt}$ $\times$ 10	0.021***	-0.015**	-0.015**	-0.020**
	(0.005)	(0.007)	(0.007)	(0.010)
N	97,762	97,762	97,762	97,762
Baseline Mean	0.203	0.203	0.203	0.203
F-Statistic (IV First Stage)				728
B. Teachers with 0-2 Years of Experience				
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.010	-0.056***	-0.056***	-0.088***
	(0.007)	(0.011)	(0.011)	(0.020)
N	31,534	31,534	31,534	31,534
Baseline Mean	0.314	0.314	0.314	0.314
F-Statistic (IV First Stage)				356
C. Teachers with 3+ Years of Experience				
(Suspension Rate)$_{sgt}$ $\times$ 10	0.018***	-0.013**	-0.013**	-0.026***
	(0.004)	(0.006)	(0.006)	(0.010)
N	66,136	66,136	66,136	66,136
Baseline Mean	0.150	0.150	0.150	0.150
F-Statistic (IV First Stage)				844
School-Grade Fixed Effects		Yes	Yes	Yes
Lagged School-Grade Test Scores			Yes	Yes

Note: This table presents IV estimates of a 10 percentage point increase in suspension rates on the probability that a teacher leaves his or her school after the current year. Each estimate is based on Equation (5) using the respective controls listed in the bottom panel. The instrument used for suspension rates is calculated in Equation (6). Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Effects of School Suspension Rates by Grade Category

	Math			English		
	Elementary	Middle	High	Elementary	Middle	High
$(\text{Suspension Rate})_{sgt} \times 10$	0.097***	0.034***	0.062***	0.183***	0.060***	0.067***
	(0.033)	(0.010)	(0.017)	(0.025)	(0.007)	(0.009)
N	895,411	858,321	581,921	895,775	859,608	622,463
Fraction Students Suspended	0.02	0.08	0.07	0.02	0.08	0.07
F-Statistic (IV First Stage)	338	681	661	337	682	658
	GPA			Fraction Days Absent		
	Elementary	Middle	High	Elementary	Middle	High
$(\text{Suspension Rate})_{sgt} \times 10$	0.121**	0.071***	0.042***	-0.007***	-0.010***	-0.027***
	(0.061)	(0.012)	(0.014)	(0.002)	(0.001)	(0.002)
N	983,863	956,972	760,940	993,361	983,342	768,084
Fraction Students Suspended	0.02	0.08	0.07	0.02	0.08	0.07
F-Statistic (IV First Stage)	225	605	583	228	584	575
Baseline Mean				0.040	0.067	0.117

Note: This table presents IV estimates for the effect of a 10 percentage point increase in suspension rates on normalized math and English test scores, normalized GPA, and the fraction of non-suspended days absent, separated by grade category. Elementary students include those in grades 3-5, middle school students include those in grades 6-8, and high school students include those in grades 9-11. Each estimate includes the full set of controls as described in Equation (5). The instrument used for suspension rates is calculated in Equation (6). Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Effects of School Suspension Rates on Teacher Attrition by Grade Category

	Elementary	Middle	High
A. All Teachers			
(Suspension Rate)$_{sgt}$ $\times$ 10	0.066	0.004	-0.050***
	(0.064)	(0.014)	(0.013)
N	31,346	35,447	30,969
Baseline Mean	0.206	0.208	0.198
F-Statistic (IV First Stage)	161	331	365
B. Teachers with 0-2 Years of Experience	Elementary	Middle	High
(Suspension Rate)$_{sgt}$ $\times$ 10	0.158	-0.071**	-0.117***
	(0.108)	(0.029)	(0.025)
N	9,329	12,423	9,782
Baseline Mean	0.342	0.312	0.299
F-Statistic (IV First Stage)	100	142	356
C. Teachers with 3+ Years of Experience	Elementary	Middle	High
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.083	0.002	-0.052***
	(0.065)	(0.013)	(0.014)
N	21,925	23,024	21,187
Baseline Mean	0.147	0.151	0.153
F-Statistic (IV First Stage)	146	476	345

Note: This table presents the IV results from Table 4, estimated separately for elementary, middle, and high school teachers. The instrument used for suspension rates is calculated in Equation (6). Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
A. Math Test Scores						
(Suspension Rate)$_{sgt}$ $\times$ 10	0.040***	0.043***	0.045***	0.057***	0.038***	0.074***
	(0.009)	(0.010)	(0.012)	(0.010)	(0.011)	(0.010)
N	2,335,653	2,035,173	1,742,084	2,335,663	2,330,308	2,335,653
F Stat (IV First Stage)	1,421	960	952	434	464	1,009
B. English Test Scores						
(Suspension Rate)$_{sgt}$ $\times$ 10	0.064***	0.066***	0.066***	0.076***	0.063***	0.107***
	(0.005)	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)
N	2,377,846	2,070,929	1,772,062	2,377,856	2,372,470	2,377,846
F Stat (IV First Stage)	1,429	966	958	435	472	1,000
Modification	None	Years: 2006-13	Years: 2007-13	Unit: Schools	Instrument: $t-2$	Instrument: District Growth

Note: This table shows the effect of a 10 percentage point increase in suspension rates on normalized math and English test scores, using the various alternative specifications discussed in Section 6.5. Column (1) provides baseline results. Columns (2) and (3) estimate the baseline results while omitting earlier years in the estimation sample to increase elapsed time between initial 2003 suspension rates. Column (4) uses school-level suspension rates (instead of school-grade) and adjusts standard errors for clustering at the school level. Column (5) uses an instrument derived from twice-lagged school-grade suspension rates instead of fixed initial 2003 suspension rates. Column (6) uses just the district suspension rate of growth to instrument for school suspension rates without interacting the district suspension rate of growth with the initial 2003 suspension rates. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 8: Direct and Indirect Effects of Suspension Rates

	Math			English		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
A. Aggregate Effects						
Aggregate Effect: (Suspension Rate)$_{sgt}$ $\times$ 10	0.046***	0.090***	0.002	0.050***	0.119***	0.029***
	(0.013)	(0.012)	(0.009)	(0.007)	(0.006)	(0.007)
N	778,421	778,476	778,552	792,481	792,546	792,619
Fraction suspended	0.02	0.04	0.11	0.02	0.04	0.11
F Stat (IV First Stage)	1,695	1,460	974	1,710	1,478	972
	Math			English		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
B. Indirect and Direct Effects						
Indirect Effect: (Suspension Rate)$_{sgt}^{-i}$ $\times$ 10	0.050***	0.097***	0.012	0.054***	0.129***	0.044***
	(0.013)	(0.012)	(0.009)	(0.007)	(0.007)	(0.007)
Direct Effect: Suspended$_{isgt}$	-0.147***	-0.107***	-0.076***	-0.138***	-0.133***	-0.102***
	(0.007)	(0.005)	(0.003)	(0.006)	(0.004)	(0.003)
N	778,421	778,476	778,552	792,481	792,546	792,619
Fraction suspended	0.02	0.04	0.11	0.02	0.04	0.11
F Stat (IV First Stage)	1,707	1,473	988	1,723	1,492	988

Note: Panel A shows the effect of a 10 percentage point increase in suspension rates on normalized test scores, estimated separately for students in three terciles based on their predicted probability of being suspended. Equation (9) in Section 7 shows how these probabilities are calculated. Terciles are assigned based on a student's rank within a given grade-year. The coefficients in this panel result from estimating Equation (5) separately for each tercile and instrumenting for suspension rates with the instrument from Equation (6). Panel B presents estimates using Equation (10), which includes an indicator for whether student i was suspended in year t and modifies the school-grade suspension rate to leave out student i's contribution. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

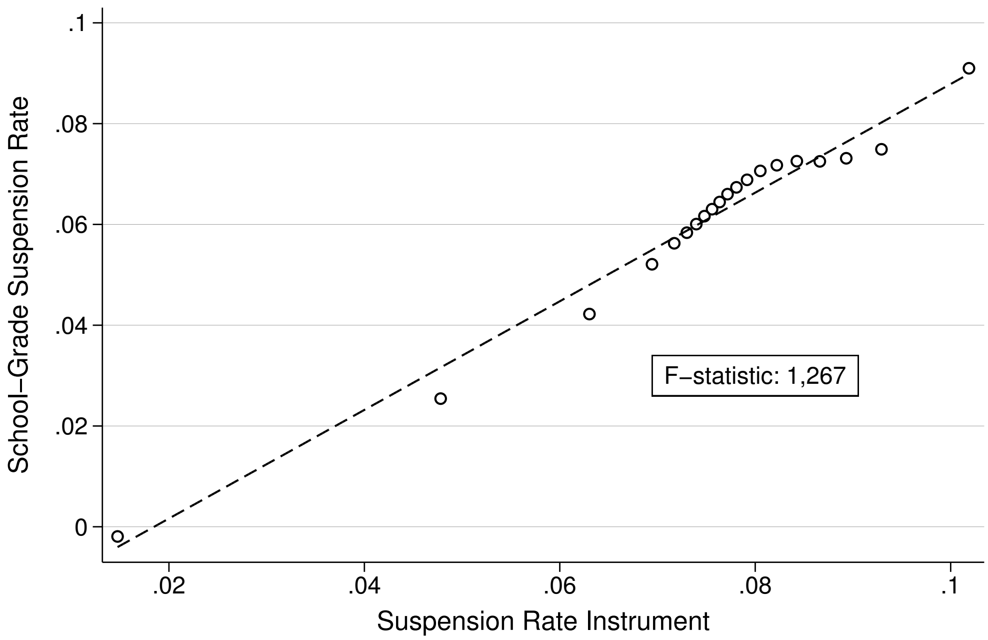
Table 9: Direct and Indirect Effects of Suspension Rates by Grade Category

Subject:	Math			English		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
A. Elementary School Students						
Aggregate Effect	0.023	0.204***	0.000	0.175***	0.306***	0.125***
	(0.052)	(0.046)	(0.031)	(0.039)	(0.034)	(0.027)
Indirect Effect	0.032	0.222***	0.019	0.184***	0.323***	0.143***
	(0.053)	(0.047)	(0.032)	(0.039)	(0.035)	(0.028)
Direct Effect	-0.174***	-0.204***	-0.132***	-0.158***	-0.181***	-0.116***
	(0.015)	(0.012)	(0.008)	(0.013)	(0.010)	(0.007)
N	298,399	298,420	298,444	298,521	298,540	298,565
Fraction suspended	0.01	0.01	0.04	0.01	0.01	0.04
B. Middle School Students						
Aggregate Effect	0.035**	0.079***	0.007	0.041***	0.113***	0.030***
	(0.016)	(0.014)	(0.010)	(0.009)	(0.008)	(0.008)
Indirect Effect	0.038**	0.086***	0.017*	0.045***	0.122***	0.045***
	(0.016)	(0.014)	(0.010)	(0.009)	(0.008)	(0.008)
Direct Effect	-0.123***	-0.100***	-0.074***	-0.117***	-0.124***	-0.099***
	(0.009)	(0.006)	(0.004)	(0.007)	(0.005)	(0.004)
N	286,084	286,105	286,121	286,514	286,532	286,550
Fraction suspended	0.02	0.06	0.18	0.02	0.06	0.17
C. High School Students						
Aggregate Effect	0.144***	0.115***	-0.005	0.099***	0.142***	-0.016
	(0.023)	(0.022)	(0.016)	(0.011)	(0.010)	(0.012)
Indirect Effect	0.147***	0.121***	0.004	0.102***	0.152***	-0.001
	(0.023)	(0.023)	(0.017)	(0.011)	(0.010)	(0.012)
Direct Effect	-0.158***	-0.089***	-0.062***	-0.165***	-0.142***	-0.105***
	(0.016)	(0.011)	(0.006)	(0.012)	(0.007)	(0.006)
N	193,938	193,951	193,987	207,446	207,474	207,504
Fraction suspended	0.02	0.05	0.13	0.02	0.05	0.13

Note: This table replicates Table 8, combining Panels A and B and estimating all effects separately for elementary, middle, and high school students. F-statistics are not shown but exceed 200 for all estimates. Standard errors are clustered at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

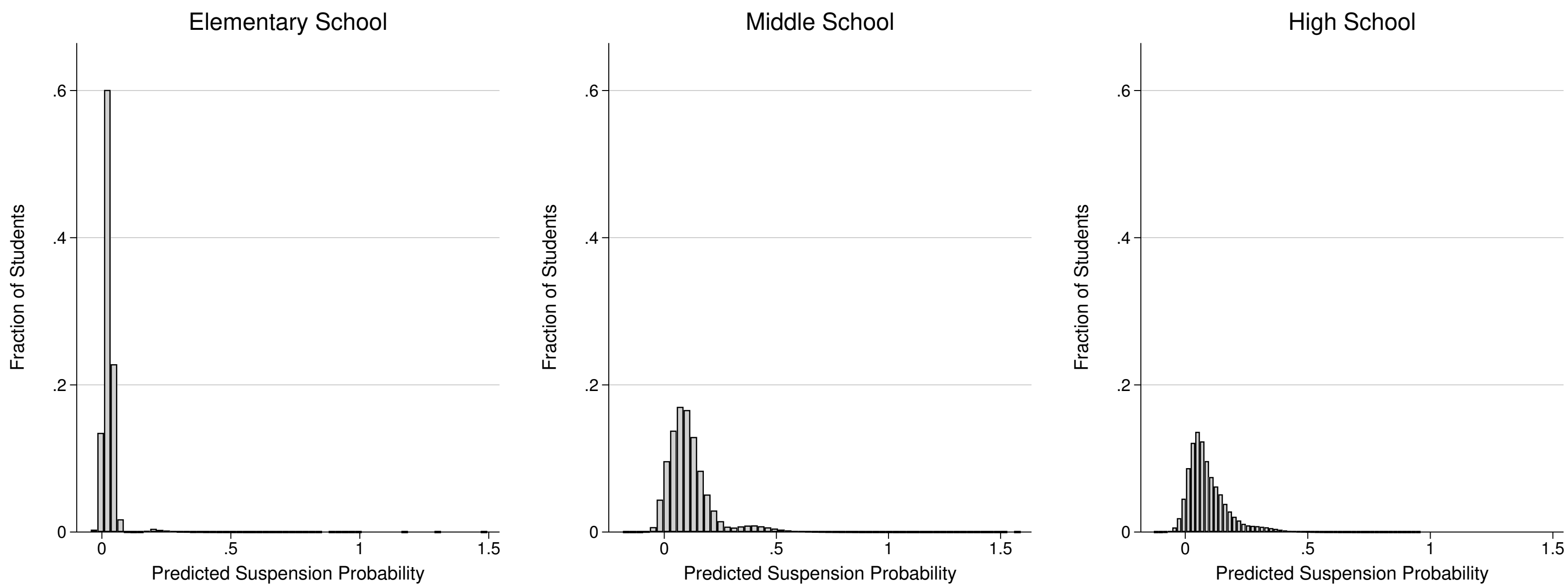
Appendix A: Appendix Tables and Figures

Figure A.1: IV First Stage



Note: This figure shows a binned scatterplot, with the suspension rate instrument calculated in Equation (6) on the x-axis and the actual suspension rate on the y-axis. Both axes are residualized with respect to lagged individual, school, and peer achievement as well as school-grade fixed effects (as shown in Equation (5)). The F-stat presented is for students with a English test score, which can be found in column (6) of Table 2.

Figure A.2: Distribution of Predicted Suspension Propensity



Note: This figure shows the distribution of predicted suspension propensities for elementary, middle, and high school students, calculated via Equation (9) using 2004 data to train the model. Predicted values are not strictly between zero and one because predictions are generated from a linear probability model.

Table A.1: Reduced Form and First Stage Regressions

	(1) First Stage	(2) RF - Math	(3) RF - English	(4) RF - GPA	(5) RF - Absences
$(\text{Suspension Rate})_{sgt} \times 10$	1.049***	0.044***	0.069***	0.070***	-0.012***
	(0.030)	(0.010)	(0.006)	(0.010)	(0.001)
Lagged Math	0.001	0.512***	0.135***	0.135***	-0.001***
	(0.001)	(0.004)	(0.002)	(0.003)	(0.000)
Lagged English	0.008***	0.203***	0.618***	0.157***	-0.002***
	(0.001)	(0.003)	(0.002)	(0.003)	(0.000)
Lagged GPA	0.012***	0.125***	0.119***	0.506***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.007)	(0.000)
Lagged Suspension Status	0.112***	0.010***	-0.041***	-0.172***	0.017***
	(0.006)	(0.003)	(0.002)	(0.005)	(0.001)
Lagged Fraction Days Absent	0.408***	-0.322***	0.041***	-1.449***	0.716***
	(0.030)	(0.020)	(0.012)	(0.036)	(0.011)
English Language Learner	0.052***	0.004	-0.125***	0.030***	-0.008***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.000)
Lagged School Math	0.044	0.312***	0.052***	-0.021	0.001
	(0.051)	(0.018)	(0.009)	(0.014)	(0.001)
Lagged School English	-0.016	-0.125***	0.062***	0.009	-0.008***
	(0.055)	(0.020)	(0.012)	(0.017)	(0.002)
Lagged Peer Math	-0.023	-0.171***	-0.129***	-0.092***	0.002*
	(0.050)	(0.017)	(0.010)	(0.016)	(0.001)
Lagged Peer English	-0.085	0.023	-0.018	-0.016	-0.007***
	(0.054)	(0.018)	(0.013)	(0.018)	(0.002)
N	2,804,135	2,335,653	2,377,846	2,701,775	2,744,787
Adjusted R-squared	0.665	0.626	0.721	0.563	0.389

Note: This table presents first stage and reduced form regression results. The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.2: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
A. Normalized GPA						
(Suspension Rate)$_{sgt}$ $\times$ 10	0.067***	0.074***	0.077***	0.138***	0.053***	0.089***
	(0.009)	(0.010)	(0.011)	(0.033)	(0.010)	(0.011)
N	2,701,775	2,381,670	2,066,776	2,701,780	2,696,304	2,701,775
F Stat (IV First Stage)	1,237	842	808	509	713	823
B. Fraction Days Absent (Non-Suspended)						
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.011***	-0.012***	-0.013***	-0.012***	-0.011***	-0.020***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	2,744,787	2,416,033	2,092,128	2,744,794	2,739,159	2,744,787
F Stat (IV First Stage)	1,212	838	816	457	721	822
Modification	None	Years: 2006-13	Years: 2007-13	Unit: Schools	Instrument: $t-2$	Instrument: District Growth

Note: This table shows the effect of a 10 percentage point increase in suspension rates on normalized GPA and the fraction of non-suspended days absent, using the various alternative specifications discussed in Section 6.5. Column (1) provides baseline results. Columns (2) and (3) estimate the baseline results while omitting earlier years in the estimation sample to increase elapsed time between initial 2003 suspension rates. Column (4) uses school-level suspension rates (instead of school-grade) and adjusts standard errors for clustering at the school level. Column (5) uses an instrument derived from twice-lagged school-grade suspension rates instead of fixed initial 2003 suspension rates. Column (6) uses just the district suspension rate of growth to instrument for school suspension rates without interacting the district suspension rate of growth with the initial 2003 suspension rates. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A.3: Direct and Indirect Effects of Suspension Rates

	GPA			Absences		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
A. Aggregate Effects						
Aggregate Effect: (Suspension Rate)$_{sgt} \times 10$	0.043***	0.083***	0.048***	-0.005***	-0.010***	-0.016***
	(0.009)	(0.012)	(0.012)	(0.009)	(0.007)	(0.006)
N	900,465	900,534	900,613	914,793	914,858	914,940
Fraction suspended	0.01	0.04	0.11	0.02	0.04	0.11
F Stat (IV First Stage)	1,469	1,292	837	1,459	1,276	810
	GPA			Absences		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
B. Indirect and Direct Effects						
Indirect Effect: (Suspension Rate)$^{-i}_{sgt} \times 10$	0.056***	0.113***	0.094***	-0.005***	-0.011***	-0.017***
	(0.009)	(0.012)	(0.013)	(0.001)	(0.001)	(0.001)
Direct Effect: Suspended$_{isgt}$	-0.420***	-0.400***	-0.308***	0.006***	0.007***	0.004***
	(0.010)	(0.009)	(0.007)	(0.000)	(0.000)	(0.001)
N	900,465	900,534	900,613	914,793	914,858	914,940
Fraction suspended	0.01	0.04	0.11	0.02	0.04	0.11
F Stat (IV First Stage)	1,484	1,303	848	1,473	1,287	821

Note: Panel A shows the effect of a 10 percentage point increase in suspension rates on normalized GPA and the fraction of non-suspended days absent, estimated separately for students in three terciles based on their predicted probability of being suspended. Equation (9) in Section 7 shows how these probabilities are calculated. Terciles are assigned based on a student's rank within a given grade-year. The coefficients in this panel result from estimating Equation (5) separately for each tercile and instrumenting for suspension rates with the instrument from Equation (6). Panel B presents estimates using Equation (10), which includes an indicator for whether student i was suspended in year t and modifies the school-grade suspension rate to leave out student i's contribution. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4: Direct and Indirect Effects of Suspension Rates by Grade Category

Subject:	GPA			Absences		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
A. Elementary School Students						
Aggregate Effect	0.146	0.143**	0.031	0.007**	0.010***	0.005
	(0.091)	(0.066)	(0.076)	(0.003)	(0.002)	(0.003)
Indirect Effect	0.163*	0.169**	0.068	0.007**	0.010***	0.005
	(0.092)	(0.067)	(0.078)	(0.003)	(0.002)	(0.003)
Direct Effect	-0.310***	-0.320***	-0.261***	0.002**	0.002***	0.001
	(0.019)	(0.016)	(0.017)	(0.001)	(0.001)	(0.001)
N	327,882	327,906	327,939	331,044	331,070	331,099
Fraction suspended	0.01	0.01	0.03	0.01	0.01	0.03
B. Middle School Students						
Aggregate Effect	0.034***	0.076***	0.049***	-0.004***	-0.008***	-0.014***
	(0.011)	(0.015)	(0.017)	(0.001)	(0.001)	(0.001)
Indirect Effect	0.050***	0.111***	0.105***	-0.004***	-0.009***	-0.015***
	(0.011)	(0.015)	(0.017)	(0.001)	(0.001)	(0.001)
Direct Effect	-0.494***	-0.475***	-0.381***	0.005***	0.006***	0.007***
	(0.016)	(0.014)	(0.011)	(0.001)	(0.001)	(0.001)
N	318,967	318,987	319,007	327,758	327,780	327,797
Fraction suspended	0.02	0.06	0.17	0.02	0.06	0.17
C. High School Students						
Aggregate Effect	0.030**	0.095***	0.031*	-0.013***	-0.026***	-0.035***
	(0.014)	(0.020)	(0.018)	(0.001)	(0.002)	(0.003)
Indirect Effect	0.039***	0.119***	0.065***	-0.013***	-0.027***	-0.036***
	(0.014)	(0.020)	(0.019)	(0.001)	(0.002)	(0.004)
Direct Effect	-0.352***	-0.323***	-0.216***	0.010***	0.011***	0.000
	(0.014)	(0.009)	(0.008)	(0.001)	(0.001)	(0.001)
N	253,616	253,641	253,667	255,991	256,008	256,044
Fraction suspended	0.02	0.05	0.14	0.02	0.05	0.14

Note: This table replicates Table A.3, combining Panels A and B and estimating all effects separately for elementary, middle, and high school students. F-statistics are not shown but exceed 200 for all estimates. Standard errors are clustered at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.