Suspending suspensions: The education production consequences of school suspension policies

Suspending Suspensions

Nolan G. Pope[1,*] and George W. Zuo[2,**]

Abstract: Managing student behaviour is integral to the education production process. We study the tradeoffs of school suspension policies by modelling and estimating how changes in school suspension policies causally impact student performance and teacher turnover. Our results indicate that the reduction in suspension rates in LAUSD decreased math and English test scores, decreased GPAs, and increased absences. Teacher turnover also increased, particularly for inexperienced teachers. We also document an efficiency-equity tradeoff: while achievement decreased for most students in the district, the highest-risk students experienced moderate gains in achievement.

Keywords: Education, Suspensions, School Discipline

Classification: I21, I24

[*] **Correspondence address:** npope@umd.edu

[**] gzuo@rand.org

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.

1 Introduction

Managing student behaviour is an integral part of education production. Yet historically, most disciplinary policies have been formed with little empirical evidence of their costs and benefits—a notion 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 misbehaviour in order to deter misconduct and limit negative learning spillovers in the education production process (Lazear, 2001). National suspension rates consequently increased from 3.5% to 8.4% between the mid-1970s and the turn of the century (see Figure 1).[2]

Despite their 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 punish 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 fuelled a reversal in national suspension policy, leading to a 40% 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 relatively little empirical evidence on the causal impact of suspension

[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.

policies on student performance and teacher turnover. In this paper, we model and estimate the effects of a large and persistent decline in suspension rates in the Los Angeles Unified School District (LAUSD), where suspension rates fell from 8% in 2003 to less than 1% in 2015. The decline was largely driven by a districtwide push for schools to revise their respective disciplinary policies to rely less on suspensions. Individual schools were required to both design and resource their revamped disciplinary policies. 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.

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. As we detail in Section 5, our approach shares similarities with the classical shift-share IV approach, but departs from it in several key ways.

We find that a 10 percentage point decrease in school suspension rates decreases concurrent math and English test scores by 0.040 and 0.065 standard deviations, respectively. These effects, while modest, are equivalent to reducing teacher quality by 0.29 and 0.65 standard deviations for math and English (Chetty *et al.*, 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. Our findings are supported by a falsification test that shows little correlation between future values of the instrument and current achievement, suggesting that the results are unlikely driven by correlations in the trajectory of schools with high initial suspension rates and districtwide suspension rate growth. The estimates are also robust to a variety of specifications which probe the sensitivity of the estimates to bias from serial correlation and the endogeneity of initial suspension rates.

Declining suspension rates may also affect teachers' well-being. When limited in their capacity to suspend students, teachers may face more misbehaviour 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 attrition rates among high school teachers – with older students whose misbehaviour may be more difficult to manage – are most responsive to declines in suspension rates.

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.[3,4] Changing school suspension policies will impact students through both these channels. Although we cannot separately identify direct and indirect effects with our empirical approach, we can infer the relative contribution of these effects from how suspension rates affect different groups of students. Those who are well-behaved will rarely experience direct effects and changes in suspension policies will have little to no impact on achievement through this channel. Students on the other side of the spectrum – particularly those whose behaviour is frequently on the margin of suspension – may find that changes in suspension policies dramatically affect their disciplinary outcomes. Using this intuition, we construct a proxy for students' propensities to be suspended and divide students into five quintiles. We then estimate the impacts of suspension rates separately for each quintile.

Our estimates show that the suspension rate decline in LA had very different impacts across the distribution of suspension propensity. We find that low-risk students in the lowest quintile experienced little to no impact on their achievement. By contrast, students in the middle three quintiles all experienced large decreases in achievement. Finally, students in the top quintile experienced a moderate *increase* in achievement. Alongside the fact that suspension rates only steeply increase between the fourth and fifth quintiles, these results suggest that direct effects are likely only meaningful for the highest-risk students. Our estimates imply that suspension policies entail a difficult tradeoff between efficiency (achievement falling for a majority of students) versus equity (achievement increasing for the highest-risk students). Explicitly quantifying direct and

[3] 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 behaviour 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 *et al.*, 2017; Lacoe and Steinberg, 2018a; Bacher-Hicks *et al.*, 2018; Sorensen *et al.*, 2021). A recent meta-analysis by Noltemeyer *et al.* (2015) summarises the correlational research across 34 studies, finding a negative correlation between achievement and being suspended.

[4] 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 even though existing estimates tend to be more descriptive in nature, they remain conflicting in their findings (Lacoe and Steinberg, 2018a,b; Craig and Martin, 2021; Sorensen *et al.*, 2021).

indirect effects, as well as further exploring the mechanisms which drive these effects, are important next steps as research on suspension policies continues to develop.

Quasi-experimental studies on the impacts of school suspension policies are limited. Bacher-Hicks *et al.* (2018) use quasi-random assignment to schools with varying suspension rates and find negative impacts on graduation rates and future criminal activity from attending schools with high suspension rates. Several recent studies focus on one-time bans on suspensions for low-level offenses, particularly in New York City (Craig and Martin, 2021) and Philadelphia (Lacoe and Steinberg, 2018a,b). We specifically compare our results to Craig and Martin (2021), who provide rigorous evidence that achievement *increased* in New York as a result of the suspension ban. We provide several explanations for why our findings differ. First, Craig and Martin (2021) focus on students grades 6 through 8. While we find that the suspensions decline in LA had negative effects for elementary, middle, and high school students, we do document that the magnitudes were smallest for middle school students. Second, suspension rates in LA and NYC experienced very different trajectories. In New York, suspension rates fell sharply within a year of the ban: suspensions for low-level offenses fell one percentage point to effectively zero, and suspensions for all infractions fell by 2.5 percentage points. In LA, suspension rates continued to fall for a decade but the total decline was roughly 10 percentage points. Finally, and perhaps most importantly, the suspensions decline in LA and NYC were handled very differently between the two districts. While the suspension ban in NYC was accompanied by substantial administrative and cultural districtwide changes, the change in the LAUSD during the mid-2000's was mandated at the district level but ultimately left implementation autonomy to individual schools (Hashim *et al.*, 2018). External reviews of school-level implementation in the LAUSD ultimately documented widespread non-compliance with the district's mandate (Chin *et al.*, 2010).

Given this, our research provides four important contributions to the literature on suspensions. First, we provide a model for how direct and indirect effects can lead to different impacts across the distribution of students, and we quantify the tradeoffs that suspension policies produce. Second, our IV strategy is novel to this literature and provides a way to address the endogeneity of suspension rates and separate the correlation with other school characteristics. Third, we quantify how suspension rates not only impact students, but teachers as well. Finally, the dramatic decline in LAUSD suspension rates is novel to a literature which has primarily focused on smaller shocks such as suspension bans for low-level offenses. Our findings ultimately provide educators and administrators with a framework for assessing the tradeoffs associated with suspension policy changes.

Our findings also have important implications for discipline in education more broadly as well as a rich peer effects literature on negative behaviour spillovers (Carrell *et al.*, 2018; Imberman *et al.*, 2012; Sacerdote, 2011). Our findings are particularly important given how little evidence exists on alternatives to suspensions. One such alternative that has gained momentum in recent years is "restorative justice", which uses a community-based approach to re-integrate misbehaving students into the classroom.[5] While evidence on restorative justice is still nascent, (Augustine *et al.*, 2018) use random assignment of restorative justice practices across schools in Pittsburgh and document small but negative effects of such practices on achievement. The lack of evidence on effective disciplinary alternatives casts a shadow over schools where misbehaviour is common. The implications of suspension policies also extend beyond the short-term outcomes studied in this paper. Researchers have linked suspensions to future incarceration (a phenomenon known as the "school-to-prison

[5]Restorative justice is a practice of bringing students, teachers, and families together in response to misbehaviour in order to resolve conflict and directly address reasons underlying the misconduct. Although suspension can still be used as a last-resort measure in response to repeated misbehaviour, restorative justice emphasises practices such as verbal confrontation, "restorative circles" (sessions in which all stakeholders convene together to openly discuss the misbehaviour and work together toward a joint solution), peer court, and one-on-one counselling.

pipeline"), which could lead to outsized negative economic and social effects for youth (Aizer and Doyle, 2015; Bacher-Hicks *et al.*, 2018; Wald and Losen, 2003). School districts must also contend with implicit racial bias in the implementation of suspension policies (Barrett *et al.*, 2017).

Given the tradeoffs that school suspensions entail, 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 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

[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 $(11.0\% \times 1.62 \times 2.32)/180 = 0.24\%$.

$p'(s) > 0$, implying that students are more likely to behave as the likelihood of being suspended for misbehaving increases.

Each school chooses s to maximise 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.

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:

$$C(s) = sn\,(p(s)^n + A) \tag{2}$$

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. Taking these costs into account, schools choose s so that the marginal benefit from increasing suspension rates equals the marginal cost via the following first-order condition:

$$\underbrace{n\,V\left[np(s)^{n-1}\frac{dp}{ds}\right]}_{\text{Indirect Effect}} = sn\,\underbrace{V\left[np(s)^{n-1}\frac{dp}{ds}+A\right]}_{\text{Direct Effect}} + \frac{dK(s)}{ds} \tag{3}$$

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

Equation (3) 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 misbehaviour) 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 (3) – 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; U.S. Department of Education (2022)), 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). Within the past decade, the prevalent use of suspensions has come under scrutiny due to concerns over the punitive and regressive nature of suspending students 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 Steinberg, 2017).

In line with this broader trend, the LAUSD implemented suspension policy reforms to reduce suspensions (Hashim *et al.*, 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), an initiative designed to address racial disparities in suspension rates. Schools were required to develop multi-tiered discipline plans to limit the use of suspensions to discipline students. 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

[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.

hands-off approach of SWPBS often led to "serious non-compliance" at the school-level implementation, especially with respect to training and professional development for school staff (Chin *et al.*, 2010).

The second phase of suspensions decline began in the summer of 2013, when the LAUSD implemented a ban on suspensions for "willful defiance", a discretionary catch-all for student misbehaviour that permitted students to be suspended for a wide variety of non-violent offenses such as refusing to take off a hat or turn off a cell phone (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 emphasised 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 (Los Angeles Unified School District, 2022). The LAUSD is the second-largest school district in the United States and enrols over 600,000 students annually. Within the district, 74% of students are Hispanic, 10% are white, and 8% 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: standardised test scores, grade point averages, absences, and suspensions. Standardised 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 normalise 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 observe whether a teacher leaves their school in any given year.

We summarise 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 out-of-school 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 enroled. Districtwide suspension rates in the LAUSD decreased substantially during the sample time period. Nearly 8% of LAUSD students were suspended in 2003, compared to less than 1% in 2015. Figure 2 shows that suspension rates also vary substantially by grade. Suspension rates increase with each grade beginning in second grade, peaking at eighth grade and subsequently decreasing with each grade through the end of high school. Less than 5% of elementary school students were suspended in 2003, compared to over 15% of eighth grade students. Figure 3

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

[9] The data unfortunately do not include offense-specific information, nor do they capture in-school suspensions.

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 characterise 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% of students are only suspended for a single day and 30% 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 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 factors, including student characteristics, principal preferences, school conditions, past performance, and more. Suspension rates are often negatively correlated with many aspects of school quality (as seen in Table 1), likely biasing standard OLS estimates downward.

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

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

$$y_{isgt} = \alpha + \rho SuspendRate_{sgt} + \beta X_{isgt-1} + \theta S_{sgt-1} + \phi P_{isgt} + \lambda_{sg} + \epsilon_{isgt} \tag{4}$$

where y_{isgt} represents a measure of academic achievement for student i in school s, grade g, and year t. We focus on standardised math and English test scores, standardised 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, restricting the identifying variation to changes within school-grade but across time.[12] All standard errors are clustered at the school level.

The control variables in Equation (4) account for certain observable variables that are correlated with suspension rates. With school-grade fixed effects, we focus on within-school-grade

[11]We find similar results in column (4) of Tables 8 and A.2 when we calculate suspension rates at the school level.

[12]One possible alternative to school-grade fixed effects is to rely on school-year fixed effects, which restricts the identifying variation to across-grade variation within the same school and year. However, this variation is potentially problematic given that suspension policies are typically determined and implemented at the school level with little differentiation between grades.

comparisons over time. However, estimates of ρ will remain biased if unobserved time-varying characteristics of school-grades are correlated with suspension rates. For example, changes in 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:

$$\widetilde{SuspendRate}_{sgt} = SuspendRate_{sg2003} \times G_{gt}^{-s} \tag{5}$$

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:

$$SuspendRate_{sgt} = SuspendRate_{sgt-1} \times \frac{SuspendRate_{sgt}}{SuspendRate_{sgt-1}} = SuspendRate_{sgt-1} \times G_{sgt} \tag{6}$$

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 *et al.*, 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.[13] However, if $SuspendRate_{sgt}$ is correlated with ϵ_{isgt} and ϵ_{isgt} is serially correlated, twice-lagged suspension rates could still be endogenous.[14] While this instrument somewhat resembles the classical shift-share ("Bartik") instrument in its derivation, its construction is more closely related to a simulated instrument (Currie and Gruber, 1996; Gruber and Saez, 2002) reflecting what suspension rates would have been at a given school had they followed the overall district trajectory.[15]

Given this instrument, the first stage equation is:

$$SuspendRate_{sgt} = \kappa + \eta \widetilde{SuspendRate}_{sgt} + \delta X_{isgt-1} + \gamma S_{sgt-1} + \xi P_{isgt} + \mu_{sg} + \nu_{isgt} \qquad (7)$$

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% of students and bottom-quartile middle schools

[13]Column (5) of Tables 8 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.

[14]As seen in columns (2) and (3) of Tables 8 and A.2, specifications that increase the number of years between the instrumented year and the year used to measure the prior suspension rate of a school-grade produce similar results.

[15]To see the distinction more clearly, consider that the instrument in Equation (5) has neither "shifts" nor "shares" in its construction and is based on an estimating equation in levels as opposed to changes.

suspending 4% (for elementary school, the suspension rates are 6% and 0%, 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 in Figure A.2. Figure A.3 shows analogous results for the subsamples of elementary, middle, and high school students. We observe a strong relationship between the instrument and actual suspension rates which is supported by a first-stage F-statistic exceeding 500.

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 any within-school-grade variation that is produced by 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 error.

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, pre-period suspension rates may endogenously determine the trajectory of test scores independent of suspension rate changes, and district suspension rate growth could be confounded by other concurrent changes in district policies. In particular, the exclusion restriction is violated if achievement changes differentially for high-initial suspension schools at the same trajectory as districtwide suspension rate growth. To address these concerns, we supplement our analysis with a combination of falsification and sensitivity tests in Tables 7 and 8 to probe the identifying assumption and its potential threats. The falsification test assesses whether future values of the instrument, conditional on the current value of the instrument, are correlated with current achievement. A strong correlation would indicate that the instrument may be affecting achievement in unobserved ways beyond its impact on current suspension rates. Our sensitivity checks look at the extent to which our estimates may be biased by serial correlation and the endogeneity of initial suspension rates.

One potential concern is that the district may have implemented other reforms during this time period which coincide with the decline in suspension rates and differentially affected schools with high suspension rates.[16] However, a confounding reform would need to not only differentially impact schools with high initial suspension rates, but also mirror the trajectory of *suspension rate growth* G_{gt} shown in Figure A.1, which is both nonlinear and varies across different grades in both timing and magnitude. Furthermore, reform in the LAUSD during this time period was notoriously difficult to achieve due to logistical and bureaucratic barriers (Mulholland Institute, 2005). Relatedly, the districtwide suspension reform that began in 2007 was largely relegated to individual schools to implement (see Section 3). Furthermore, to the extent that education reform tends to differentially *benefit* low-achieving schools (which typically have higher suspension rates), any confounding district reforms would likely *understate* the negative overall effect of the suspension rate decline that we estimate.

[16] A potential solution would be to leverage the steep drop in suspension rates beginning in 2012 (see Figure 2). However, the test score data extend only through 2013, when standardised testing reform began in the district.

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 (4). 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 (0.006 and 0.020 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. Figure 5 provides visual evidence of the reduced form relationship.[17] The IV estimates for math and English test

[17] Table A.1 provides first stage and reduced form estimates. We note that the reduced form scatterplots exhibit several high-leverage points that skew the overall slope of the plot; removing these points would in fact strengthen the relationships we observe.

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 0.065 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. Relative to a 10 percentage point change in suspension rates, these estimates are comparatively moderate in magnitude. To contextualise the effect size, the impact of the near-10 percentage point decline in suspension rates in the LAUSD had approximately the same impact as a year of exposure to a 0.29 and 0.65 standard deviation higher-quality teacher as measured by value-added for math and English, respectively (Chetty *et al.*, 2014).[18] We note that the estimates reflect one year of exposure and do not encompass the cumulative effect of being exposed to higher 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 (3) 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.

[18] Chetty *et al.* (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.

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 consequently estimate the effects of school suspension rates on GPA and absences. We standardise GPA by grade and year, and we measure absences as the fraction of non-suspended days that 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 suggests 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 0.010 standard deviations and decreases absences by 0.6 percentage points (8.2%); 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 0.066 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 misbehaviour that may include skipping class. Students may also feel less inclined to attend class when the classroom environment is more prone to disruption from misbehaviour or when exposed to more bullying or violence.

6.3 Teacher Attrition

Decreasing suspension rates may also affect teachers. Fewer disciplinary options could make classroom behaviour management more taxing, especially for inexperienced teachers. Teaching in a classroom with more misbehaviour 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.[19] 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 which is not statistically significant. The baseline turnover rate in Los Angeles is quite high at 20.2%, an estimate which is consistent with previous research. Newton *et al.* (2011) document 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.[20] 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 large and significant 8.3 percentage point (27%) increase in turnover. For more experienced teachers, the same increase in suspension rates leads to a 2.6 percentage point (17%) increase in turnover.

Due to the large effect of declining suspension rates on teacher turnover, it appears that teachers valued the ability to suspend students. Clotfelter *et al.* (2008) estimate that a $1,800 bonus payment

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

[20]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.

reduces teacher turnover by 17%. Using their estimate as a benchmark, the school district would need to pay teachers $1,048 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 $2,835 more to offset a 10 percentage point decrease in suspension rates.

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 organised may change how suspension policies affect achievement. Lastly, the nature of misbehaviour 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 (0.035 and 0.060 standard deviations for math and English in middle school and 0.061 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, 0.070, and 0.041 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 baseline means). 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 0.9 percentage points (13.4%) and 2.6 (22.2%) percentage points. These effect sizes are quite large and are equivalent to 1.6 and 4.7 days of lost instruction for middle and high school students per school year. The results suggest that as students advance through school, their 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 10.9 percentage points (37%). Higher suspension rates also decrease attrition among inexperienced middle school teachers (-7.3 percentage points) and experienced high school teachers (-4.8 percentage points). These results suggest that the marginal misbehaviour 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. Misbehaviour 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% of teachers are physically threatened each year.[21] Teachers who feel threatened but cannot safely respond without the use of suspensions may be more likely to leave.

6.5 Falsification Test

To help verify our main results, we conduct a falsification test assessing whether future values of the instrument affect present achievement, conditional on current values of the instrument. Future values of the instrument should have little to no impact on the current academic outcomes except through correlation with the current instrument. Therefore, if the instrument captures an exogenous shock to suspension rates in year t, there should be no relationship between the instrument and test scores in a future period. For test scores in year t, we estimate the following equation:

$$y_{isgt} = \alpha + \rho \widetilde{SuspendRate}^{resid}_{sgt+j} + \beta X_{isgt} + \theta S_{sgt} + \phi P_{isgt} + \lambda_{sg} + \epsilon_{isgt} \tag{8}$$

The equation takes the baseline estimating equation in (4) and replaces endogenous suspension rates with a version of the school-grade suspension rate instrument from j years in the future, $\widetilde{SuspendRate}_{sgt+j}$. We then residualize the future instrument with respect to the current value of the instrument in time t to remove all information contained in $\widetilde{SuspendRate}_{sgt+j}$ captured by the

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

current instrument, which by construction will be correlated with the current outcome. We denote this residualized instrument $\widetilde{SuspendRate}^{resid}_{sgt+j}$.[22]

Table 7 presents the effect of the residualized suspension rate instrument in years $t+1$ through $t+3$ on academic outcomes in year t. The future placebo estimates for math and English test scores are small relative to the effects in year t and are nearly all insignificant (one estimate is marginally significant but is in the opposite direction). In addition, there is no clear trend in effect sizes moving from $t+3$ to $t+1$. We find some significant effects for absences and GPA in the pre-periods; however, the significant effects are generally small or in the opposite direction relative to the baseline result in year t. In all instances, the effect size jumps noticeably from $t+1$ to t. These falsification tests suggest that achievement did not differentially change for high-suspension schools at the same trajectory as districtwide suspension rate growth, which may rule out a broad set of potential violations to the exclusion restriction.

6.6 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 8. In both columns, effects on

[22] One potential concern is that if the decline in suspension rates is linear, then the residualization process will ultimately remove most of the variation in future iterations of the instrument. The test would consequently be weakened by the high noise-to-signal ratio of the remaining variation. In Figure A.1, we provide a breakdown of district suspension rate growth by grade level. District suspension rate growth is consistently negative but varies quite substantially over time, particularly in the latter half of the sample period. This suggests that future iterations of the instrument may still contain meaningful information even after the residualization process.

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 level variation versus school-level variation. Schools may be unlikely to differentiate suspension policy (among other school policies) by grade. Column (4) of Table 8 shows how the IV estimates change when using school-level suspension rates. 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 (6). Using twice-lagged suspension rates instead of fixed initial suspension rates may theoretically increase the instrument's predictive power. However, column (5) of Table 8 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.

Days suspended as endogenous variable: We also analyse how our results change when using the average number of days suspended as the endogenous variable—a measure of suspension policy intensity not fully captured by suspension rates. For each school-grade cell, we calculate the average number of days suspended across all students. To construct the instrument, we calculate the year-to-year district growth in average days suspended. Column (6) of Table 8 shows the effect of increasing average days suspended by 0.1 days—nearly doubling the baseline value (0.12 days). The direction and significance of the estimates remain the same, and the magnitude is qualitatively similar.

7 Direct and Indirect Effects of Suspensions

7.1 Conceptual Framework

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 (3) 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 behaviour. 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. Misbehaviour disrupts class and diverts a teacher's time and energy away from instruction, potentially producing negative spillovers on learning. In 2016, 43% of teachers "agreed" or "strongly agreed" that student misbehaviour interfered with their teaching during the year (Musu-Gillette *et al.*, 2018). Suspensions provide teachers with one way to curtail disruptions and prevent misbehaviour from escalating.[23] 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 *et al.* (2015) summarises 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 (2018a) find negative direct effects in Philadelphia using an individual fixed effects approach while Anderson *et al.* (2017) find positive effects of being suspended using dynamic panel methods in Arkansas. Beyond test scores, concurrent work by Bacher-Hicks *et al.* (2018) study students who are assigned to schools with varying suspension rates due to a school zone boundary change. They find that attending a 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.

[23] 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.

However, beyond suspensions, there exists a robust literature on peer effects in schools (Sacerdote, 2011).[24]

7.2 Empirical Estimates

The positive relationship between suspension rates and achievement in our baseline results suggests that even if direct effects were highly beneficial to students at the margin of suspension, indirect effects ultimately drive the overall estimates. We now explore this narrative by quantifying how suspension rates affect achievement across the distribution of students based on their likelihood of suspension. This approach leverages the intuition that students who are well-behaved will rarely experience direct effects and changes in suspension policies will have little to no impact on achievement through this channel.

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.[25] The linear probability model we use is as follows and is estimated separately by grade:

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

[24]For two examples closely related to school discipline, Imberman *et al.* (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.

[25]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 misbehaviour 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.[26] $S_{sg,2003}$ 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.4 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 quintiles based on their predicted suspension propensities and we produce IV estimates for each of the five resulting subsamples.

We present the IV estimates for each quintile in Table 9. The math estimates are positive but small and insignificant in the first quintile (0.021 standard deviations). The estimates for the next three quintiles are all positive and statistically significant (in order: 0.082, 0.094, and 0.074 standard deviations). The estimates for the highest quintile for predicted suspension probability are negative and significant (-0.036 standard deviations). The English estimates are similar in magnitude but are also significant for the lowest suspension quintile (from low to high suspensions: 0.024, 0.091, 0.126, 0.117, and -0.022 standard deviations).

These estimates suggest that the suspension rate decline in LA had disparate impacts on three groups of students. First, students with low suspension propensity were not noticeably affected by the change in suspension rates. Because suspension rates for this group are low (1%), this estimate is nearly entirely comprised of indirect effects, which appear small. One reason why this group of students might be insulated from misbehaviour spillovers is that they could be

[26]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.

tracked into different classrooms with few misbehaving peers. Even a large schoolwide change in suspension rates might not affect students who primarily learn in classrooms where misbehaviour is rare. Another possibility is that students in this quintile are high performers who are capable of achieving and learning in spite of peer misbehaviour.

The second, third, and fourth quintiles tell a different story. Higher school suspension rates are associated with large increases in test scores for this group. Suspension rates only differ between first and second quintiles by one percentage point, suggesting that the difference between the two estimates is not likely driven by direct effects. Instead, it appears that students in the second quintile are highly sensitive to misbehaviour spillovers. The same appears to apply to the third and fourth quintiles, which only differ in suspension rates by four percentage points across the three quintiles. Given the large, positive estimates and the relatively low suspension rates among these students, it is unlikely that direct effects play a major role. The overall relationship between suspension rates and test scores in our baseline estimates from Section 6.1 are clearly driven by this majority of students.

Students in the highest quintile of predicted suspension probabilities appear moderately harmed by higher suspension rates. The point estimate is negative but less than half the magnitude of the effect sizes for the previous three quintiles. Roughly 14% of students in this quintile are suspended in any given year, a rate which is more than double that of the previous quintile. These results point to two possibilities: indirect effects may be small for this group (or possibly negative), and direct effects could be negative and large. Small indirect effects would imply that learning among students in this group is not responsive to changes in misbehaviour spillovers. Negative indirect effects could be possible if a less punitive school climate improves learning among this group of students (see for example, Craig and Martin (2021)). It is also possible that this estimate is driven entirely by large, negative direct effects given that this group of students also has the greatest exposure to direct effects. Unfortunately, our empirical approach does not enable us to separately identify the relative size of the two effects.

Overall, these estimates paint a narrative which suggest that the impacts of the LA suspension decline were unevenly distributed across students. Low-risk students were generally unaffected, high-risk students experienced moderate increases in achievement, and students within the inner three quintiles experienced large declines in achievement. Given the large, positive estimates in the inner quintiles and the near-zero suspension rates in the lowest quintile, direct effects are unlikely meaningful for all but the highest-risk students. The results also highlight that suspensions policies entail a tradeoff between efficiency and equity. When suspension rates fell in LA, overall achievement in the district decreased, driven by a majority of students within the inner three quintiles. At the same time, achievement among the highest-risk students rose moderately. This leaves school districts with the difficult task of choosing a suspension policy which could likely benefit one group at the detriment of the other.

We note that these results may be context-specific to the LAUSD decline in suspensions. Through a similar line of reasoning, Craig and Martin (2021) argue that the overall impact of direct effects in the NYC setting is small, though they do not attempt to estimate the size of direct effects per suspended student. However, the authors find that the overall impact of reducing suspension rates in NYC is *positive*, suggesting that the indirect effects of reducing suspension rates could be either positive or negative depending on the context. In NYC, the authors attribute the impacts to improvements in school culture and perceptions of safety. The overhaul of low-level discipline in NYC focused on improving student-teacher relationships and provided concrete, districtwide guidance for how to reduce suspension rates and replace them with alternatives. On the other hand, the LAUSD's approach to reducing suspension rates left autonomy to individual schools to determine how they would lower suspension rates, and how they would ultimately fund and resource those changes. These differences are perhaps most succinctly captured by the fact that Craig and Martin (2021) find no changes in teacher turnover, as opposed to the notable rates teacher turnover that we document in the LAUSD.

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. Our findings are supported by a falsification test that fails to find a clear correlation between future values of the instrument with current achievement and are robust to different specification choices.

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. Emprically, we find that the decline in suspension rates in LAUSD decreased achievement among a majority of students but increased achievement among the highest-risk students. These findings heavily suggest that the impacts of suspension rates are mostly driven by indirect effects, and that direct effects are only salient to the students at the highest risk of being suspended. We conclude that higher suspension rates may exacerbate educational inequality despite increasing average achievement.

The important role of student behaviour in the learning process necessitates that school, district, and state administrators determine the best policies to manage their students' behaviour. While other forms of behaviour 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.

Affiliations

[1]Department of Economics, University of Maryland, 3114 Tydings Hall, 7343 Preinkert Dr., College Park, MD 20742
[2]RAND Corporation, 4570 Fifth Avenue, Pittsburgh, PA 15213

References

Aizer, A. and Doyle, Joseph J., J. (2015). 'Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges', *The Quarterly Journal of Economics*, vol. 130(2), pp. 759–803, ISSN 0033-5533, doi:10.1093/qje/qjv003.

Anderson, K.P., Ritter, G.W. and Zamarro, G. (2017). 'Understanding a Vicious Cycle: Do Out-of-School Suspensions Impact Student Test Scores?', EDRE.

Augustine, C.H., Engberg, J., Grimm, G.E., Lee, E., Wang, E.L., Christianson, K. and Joseph, A.A. (2018). 'Can restorative practices improve school climate and curb suspensions? an evaluation of the impact of restorative practices in a mid-sized urban school district', *RAND Research Report*.

Bacher-Hicks, A., Billings, S.B. and Deming, D.J. (2018). 'The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime', National Bureau of Economic Research, Inc.

Barrett, N., McEachin, A., Mills, J.N. and Valant, J. (2017). 'Disparities in Student Discipline by Race and Family Income', Education Research Alliance for New Orleans.

Blume, H. (2012). 'Suspension figures called "alarming"', *Los Angeles Times*.

Carrell, S.E., Hoekstra, M. and Kuka, E. (2018). 'The long-run effects of disruptive peers', *American Economic Review*, vol. 108(11), pp. 3377–3415, doi:10.1257/aer.20160763.

Carrell, S.E. and Hoekstra, M.L. (2010). 'Externalities in the classroom: How children exposed to domestic violence affect everyone's kids', *American Economic Journal: Applied Economics*, vol. 2(1), pp. 211–28, doi:10.1257/app.2.1.211.

Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014). 'Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates', *American Economic Review*, vol. 104(9), pp. 2593–2632, doi:10.1257/aer.104.9.2593.

Chin, M., Cusick, R., Faer, L., Lawrence, A., McGowan, R., Ruelas, A. and Ventura, B. (2010). 'Refining Dignity in Our Schools: A Shadow Report on School-Wide Positive Behavior Support Implementation in South Los Angeles, 2007-2010', .

Clotfelter, C., Glennie, E., Ladd, H. and Vigdor, J. (2008). 'Would higher salaries keep teachers in high-poverty schools? evidence from a policy intervention in north carolina', *Journal of Public Economics*, vol. 92(5), pp. 1352 – 1370, ISSN 0047-2727, doi:https://doi.org/10.1016/j.jpubeco.2007.07.003.

Craig, A.C. and Martin, D.C. (2021). 'Discipline Reform, School Culture, and Student Achievement', .

Curran, F.C. (2016). 'Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior', *Educational Evaluation and Policy Analysis*, vol. 38(4), pp. 647–668, doi:10.3102/0162373716652728.

Currie, J. and Gruber, J. (1996). 'Health Insurance Eligibility, Utilization of Medical Care, and Child Health*', *The Quarterly Journal of Economics*, vol. 111(2), pp. 431–466, ISSN 0033-5533, doi:10.2307/2946684.

Goldsmith-Pinkham, P., Sorkin, I. and Swift, H. (2018). 'Bartik Instruments: What, When, Why, and How', National Bureau of Economic Research, Inc.

Gruber, J. and Saez, E. (2002). 'The elasticity of taxable income: evidence and implications', *Journal of Public Economics*, vol. 84(1), pp. 1 – 32, ISSN 0047-2727, doi:https://doi.org/10.1016/S0047-2727(01)00085-8.

Hashim, A.K., Strunk, K.O. and Dhaliwal, T.K. (2018). 'Justice for All? Suspension Bans and Restorative Justice Programs in the Los Angeles Unified School District', *Peabody Journal of Education*, vol. 93(2), pp. 174–189, doi:10.1080/0161956X.2018.1435040.

Imberman, S.A., Kugler, A.D. and Sacerdote, B.I. (2012). 'Katrina's children: Evidence on the structure of peer effects from hurricane evacuees', *American Economic Review*, vol. 102(5), pp. 2048–82, doi:10.1257/aer.102.5.2048.

Lacoe, J. and Steinberg, M.P. (2017). 'What Do We Know About School Discipline Reform?', *Education Next*, vol. 17(1).

Lacoe, J. and Steinberg, M.P. (2018a). 'Do Suspensions Affect Student Outcomes?', *Educational Evaluation and Policy Analysis*, doi:10.3102/0162373718794897.

Lacoe, J. and Steinberg, M.P. (2018b). 'Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes', *Peabody Journal of Education*, vol. 93(2), pp. 207–227, doi:10.1080/0161956X.2018.1435047.

Lazear, E.P. (2001). 'Educational Production*', *The Quarterly Journal of Economics*, vol. 116(3), pp. 777–803, ISSN 0033-5533, doi:10.1162/00335530152466232.

Los Angeles Unified School District (2022). 'Student records obtained from freedom of information act request [dataset]', (last accessed: 21 October, 2022).

Losen, D., Hodson, C., III, M.A.K., Morrison, K. and Belway, S. (2015). 'Are We Closing the School Discipline Gap?', The Center for Civil Rights Remedies.

Mulholland Institute (2005). 'Reforming Schools: Strategies for Restructuring the Los Angeles Unified School District', .

Musu-Gillette, L., Zhang, A., Zhang, J., Kemp, J., Diliberti, M. and Oudekerk, B.A. (2018). 'Indicators of School Crime and Safety: 2017', National Center for Education Statistics, Institute of Education Sciences.

Newton, X.A., Rivero, R., Fuller, B. and Dauter, L. (2011). 'Teacher Stability and Turnover in Los Angeles: The Influence of Teacher and School Characteristics', Policy Analysis for California Education.

Noltemeyer, A.L., Ward, R.M. and Mcloughlin, C. (2015). 'Relationship Between School Suspension and Student Outcomes: A Meta-Analysis', *School Psychology Review*, vol. 44(2), pp. 224–240, doi:10.17105/spr-14-0008.1.

Petek, N. and Pope, N. (2018). *The Multidimensional Impact of Teachers on Students*, Ph.D. thesis, University of Chicago.

Sacerdote, B. (2011). 'Chapter 4 - peer effects in education: How might they work, how big are they and how much do we know thus far?', pp. 249 – 277, vol. 3 of *Handbook of the Economics of Education*, Elsevier, doi:https://doi.org/10.1016/B978-0-444-53429-3.00004-1.

Sorensen, L., Bushway, S. and Gifford, E. (2021). 'Getting tough? the effects of discretionary principal discipline on student outcomes', *Education Finance and Policy*, pp. 1–74, doi:10.1162/edfp_a_00341.

Teske, S.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*, vol. 24(2), pp. 88–97, doi:10.1111/j.1744-6171.2011.00273.x.

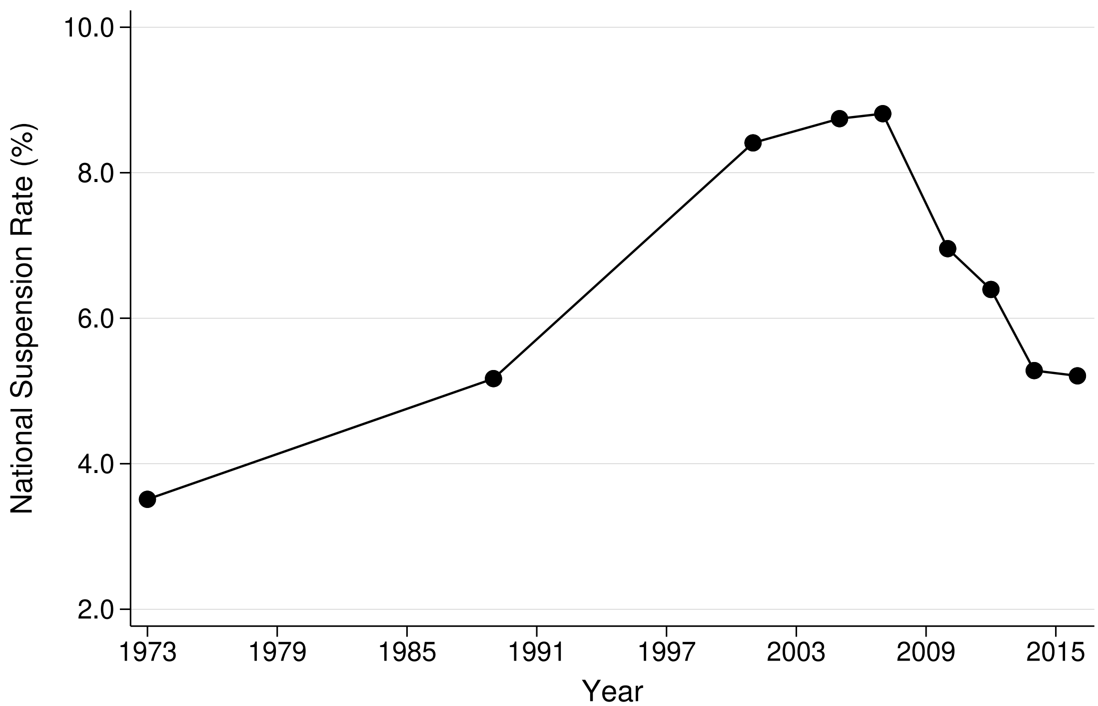
U.S. Department of Education (2014). 'U.S. Departments of Education and Justice Release School Discipline Guidance Package to Enhance School Climate and Improve School Discipline Policies/Practices', .

U.S. Department of Education (2022). Office for Civil Rights, Civil Rights Data Collection (1972–73, 1988–89, 2000–01, 2003–04, 2005–06, 2009–10, 2011–12, and 2015–16) [dataset], https://ocrdata.ed.gov/ (last accessed: 21 October, 2022).

Wald, J. and Losen, D.J. (2003). 'Defining and redirecting a school-to-prison pipeline', *New Directions for Youth Development*, vol. 2003(99), pp. 9–15, doi:10.1002/yd.51.

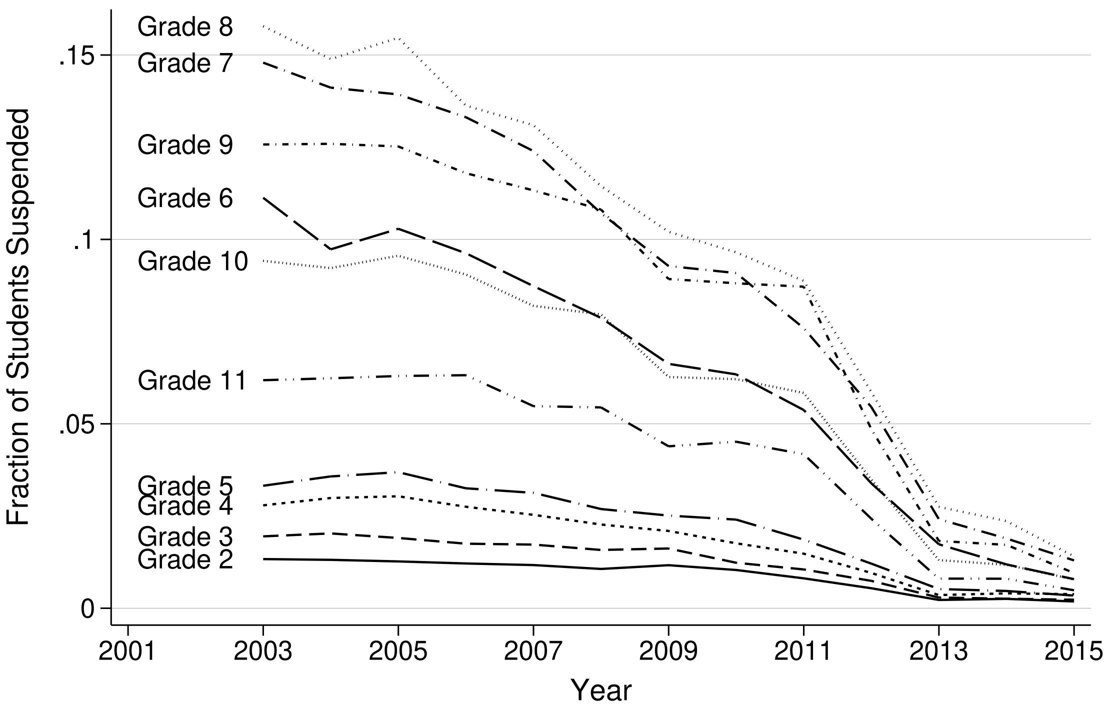
Watanabe, T. (May 14, 2013). 'L.A. Unified bans suspension for willful defiance', .

Figure 1: National Suspension Rates



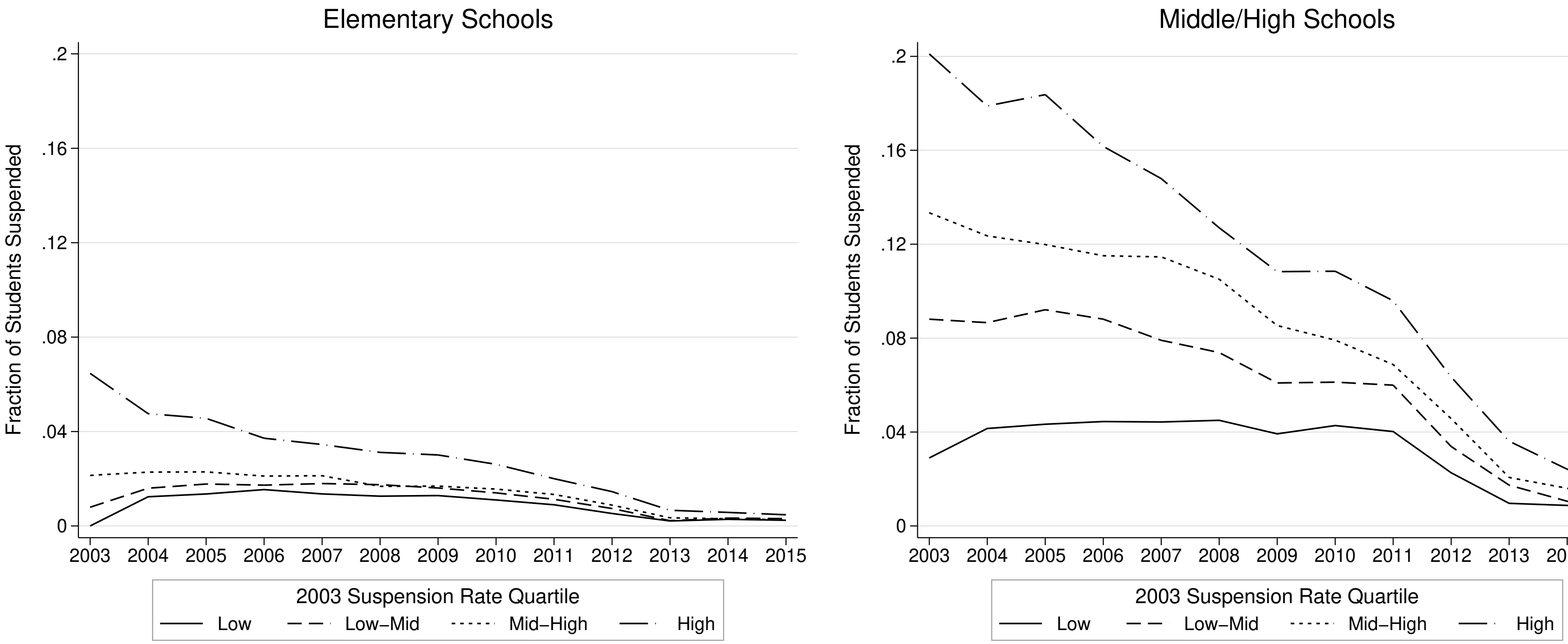
Note: This figure shows national suspension rates using data compiled by the Civil Rights Data Collection by the Office for Civil Rights (U.S. Department of Education, 2022). Suspension rates are calculated by dividing the total number of suspended students (including both students with and without disabilities) by the total number of enrolled students.

Figure 2: LAUSD Suspension Rates by Grade



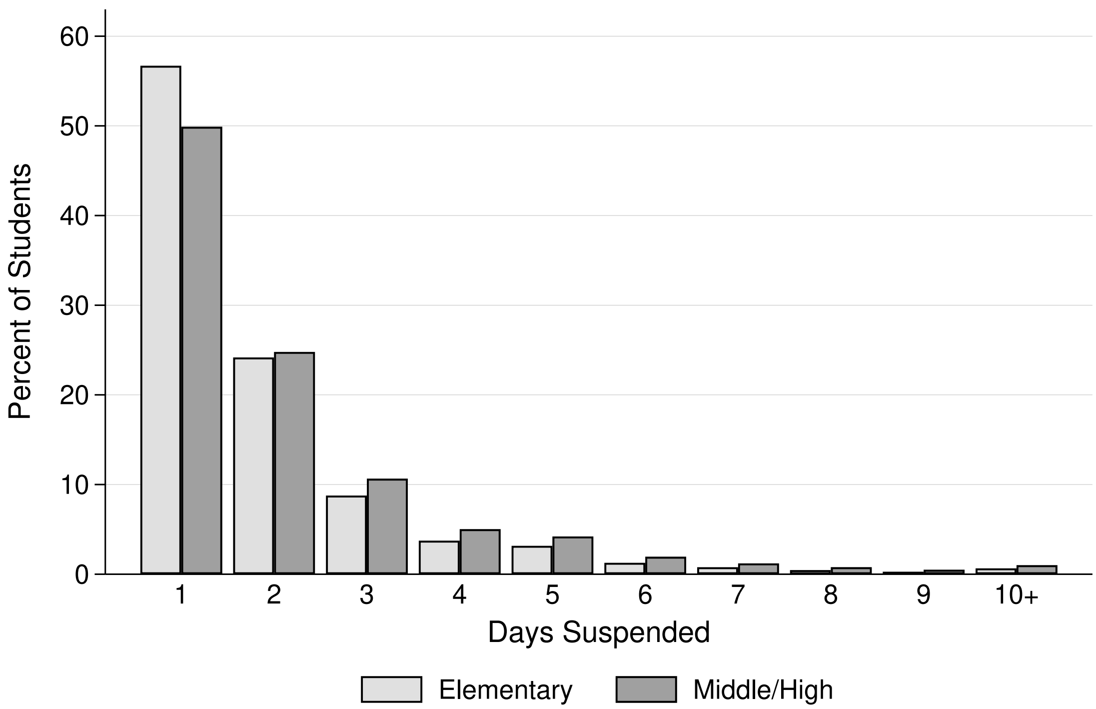
Note: This figure shows districtwide suspension rates for each grade in the LAUSD for each year. Suspension rates are calculated by dividing the total number of suspended students by the total number of enrolled students in each grade.

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



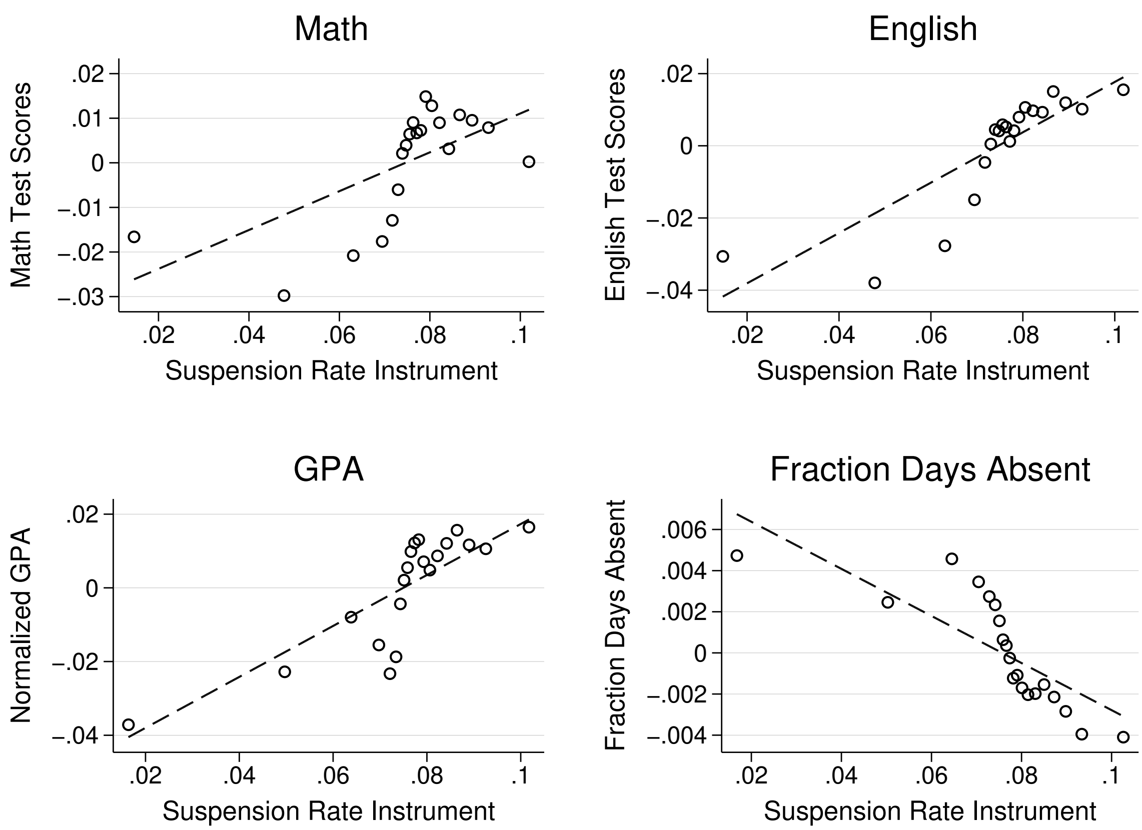
Note: This figure plots the trajectory of suspension rates for elementary and middle/high students in the LAUSD, dividing school-grades into one of four equally-sized quartiles based on initial 2003 suspension rates. Cutoffs for each of the four quartiles are as follows. For elementary school-grades: 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 (4)). The instrument is calculated based on Equation (5). The binned scatterplot (generated by the Stata command "binscatter") groups the data into twenty equally-sized groups based on the value of the suspension rate instrument and plots the mean x- and y- values within each bin.

Table 1: LAUSD Summary Statistics

	Low Suspensions		High Suspensions	
Elementary School	Mean	SD	Mean	SD
Standardized Math Scores	0.05	1.02	-0.05	0.98
Standardized English Scores	0.05	1.02	-0.04	0.97
Standardized GPA	0.03	1.01	-0.03	0.99
Fraction Days Absent	0.04	0.05	0.04	0.06
English Language Learner	0.38	0.48	0.37	0.48
Suspended	0.01	0.10	0.02	0.15
Days Suspended (If Suspended)	1.82	1.44	1.93	1.64
# Times Suspended	1.23	0.65	1.33	0.80
School Size	433	213	440	213
Number of Schools	190		191	
Number of Observations	1,032,545		1,086,721	
	Low Suspensions		High Suspensions	
Middle School	Mean	SD	Mean	SD
Standardized Math Scores	0.18	1.06	-0.08	0.96
Standardized English Scores	0.17	1.03	-0.08	0.98
Standardized GPA	0.10	1.00	-0.05	1.00
Fraction Days Absent	0.05	0.07	0.06	0.08
English Language Learner	0.20	0.40	0.26	0.44
Suspended	0.06	0.24	0.11	0.31
Days Suspended (If Suspended)	2.05	1.79	2.32	2.08
# Times Suspended	1.45	0.94	1.62	1.17
School Size	850	879	1,570	734
Number of Schools	51		51	
Number of Observations	515,097		1,038,062	
	Low Suspensions		High Suspensions	
High School	Mean	SD	Mean	SD
Standardized Math Scores	-0.02	0.94	0.00	1.00
Standardized English Scores	-0.12	0.97	0.01	1.00
Standardized GPA	0.13	0.97	-0.01	1.00
Fraction Days Absent	0.13	0.21	0.10	0.13
English Language Learner	0.09	0.29	0.18	0.38
Suspended	0.02	0.13	0.07	0.25
Days Suspended (If Suspended)	1.82	1.40	2.04	1.57
# Times Suspended	1.29	0.72	1.36	0.80
School Size	469	746	1,914	1,577
Number of Schools	62		60	
Number of Observations	338,769		1,510,154	

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***
	(0.019)	(0.011)	(0.007)	(0.005)	(0.006)	(0.010)
N	2,336,068	2,336,068	2,336,068	2,336,068	2,336,068	2,336,068
F-Statistic (IV First Stage)						618
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***	0.065***
	(0.020)	(0.007)	(0.004)	(0.004)	(0.004)	(0.007)
N	2,378,265	2,378,265	2,378,265	2,378,265	2,378,265	2,378,265
F-Statistic (IV First Stage)						615
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 (4). The IV estimates use instrumented suspension rates as calculated in Equation (5). 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 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	0.010	0.010*	0.010*	0.066***
	(0.014)	(0.008)	(0.006)	(0.006)	(0.006)	(0.010)
N	2,702,204	2,702,204	2,702,204	2,702,204	2,702,204	2,702,204
F-Statistic (IV First Stage)						547
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	0.015***	0.002**	-0.006***	-0.006***	-0.006***	-0.011***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	2,745,393	2,745,393	2,745,393	2,745,393	2,745,393	2,745,393
Baseline Mean	0.072	0.072	0.072	0.072	0.072	0.072
F-Statistic (IV First Stage)						518
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 (4). The IV estimates use instrumented suspension rates as calculated in Equation (5). 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 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	0.023***	-0.014	-0.014	-0.020
	(0.006)	(0.009)	(0.009)	(0.014)
N	101,103	101,103	101,103	101,103
Baseline Mean	0.202	0.202	0.202	0.202
F-Statistic (IV First Stage)				347
B. Teachers with 0-2 Years of Experience				
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.005	-0.054***	-0.053***	-0.083***
	(0.009)	(0.014)	(0.014)	(0.024)
N	32,933	32,933	32,933	32,933
Baseline Mean	0.310	0.310	0.310	0.310
F-Statistic (IV First Stage)				165
C. Teachers with 3+ Years of Experience				
(Suspension Rate)$_{sgt}$ $\times$ 10	0.019***	-0.012	-0.012	-0.026**
	(0.005)	(0.008)	(0.008)	(0.013)
N	68,038	68,038	68,038	68,038
Baseline Mean	0.149	0.149	0.149	0.149
F-Statistic (IV First Stage)				438
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 (4) using the respective controls listed in the bottom panel. The instrument used for suspension rates is calculated in Equation (5). Standard errors are adjusted for clustering at the school 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
(Suspension Rate)$_{sgt}$ $\times$ 10	0.097***	0.035***	0.061***	0.183***	0.060***	0.067***
	(0.034)	(0.011)	(0.019)	(0.026)	(0.008)	(0.011)
N	895,560	858,530	581,978	895,923	859,822	622,520
Fraction Students Suspended	0.02	0.08	0.07	0.02	0.08	0.07
F-Statistic (IV First Stage)	226	283	333	225	284	315
	GPA			Fraction Days Absent		
	Elementary	Middle	High	Elementary	Middle	High
(Suspension Rate)$_{sgt}$ $\times$ 10	0.121*	0.070***	0.041***	-0.007***	-0.009***	-0.026***
	(0.063)	(0.013)	(0.013)	(0.002)	(0.001)	(0.003)
N	984,003	957,188	761,013	993,659	983,571	768,163
Fraction Students Suspended	0.02	0.08	0.07	0.02	0.08	0.07
F-Statistic (IV First Stage)	151	267	277	153	243	274
Baseline Mean				0.041	0.068	0.120

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 (4). The instrument used for suspension rates is calculated in Equation (5). Standard errors are adjusted for clustering at the school 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	0.068	0.001	-0.046***
	(0.054)	(0.020)	(0.017)
N	31,471	36,335	33,297
Baseline Mean	0.204	0.208	0.197
F-Statistic (IV First Stage)	104	138	215
B. Teachers with 0-2 Years of Experience	Elementary	Middle	High
(Suspension Rate)$_{sgt}$ $\times$ 10	0.190**	-0.073**	-0.109***
	(0.094)	(0.035)	(0.029)
N	9,391	12,811	10,731
Baseline Mean	0.339	0.311	0.293
F-Statistic (IV First Stage)	76	62	210
C. Teachers with 3+ Years of Experience	Elementary	Middle	High
(Suspension Rate)$_{sgt}$ $\times$ 10	-0.070	0.000	-0.048**
	(0.063)	(0.016)	(0.019)
N	21,967	23,523	22,548
Baseline Mean	0.146	0.151	0.153
F-Statistic (IV First Stage)	89	208	214

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 (5). Standard errors are adjusted for clustering at the school level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7: Falsification Test: Effects of Suspension Rates on Past and Current Academic Outcomes

	Math in Year t				English in Year t			
	Future Instruments			Instrument	Future Instruments			Instrument
	$t+3$	$t+2$	$t+1$	t	$t+3$	$t+2$	$t+1$	t
Estimate	0.009	-0.008	-0.002	0.044***	0.009	-0.008*	-0.009	0.070***
	(0.010)	(0.008)	(0.009)	(0.011)	(0.006)	(0.004)	(0.006)	(0.007)
N	2,108,339	2,329,441	2,333,283	2,336,068	2,147,783	2,371,651	2,375,482	2,378,265
	Fraction Days Absent in Year t				GPA in Year t			
	Future Instruments			Instrument	Future Instruments			Instrument
	$t+3$	$t+2$	$t+1$	t	$t+3$	$t+2$	$t+1$	t
Estimate	-0.005***	0.004***	0.002***	-0.011***	0.012	-0.014*	0.002	0.069***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.008)	(0.007)	(0.006)	(0.011)
N	2,279,856	2,514,903	2,742,167	2,745,393	2,244,390	2,475,432	2,699,113	2,702,204

Note: This table presents IV estimates for the effect of a 10 percentage point increase in suspension rates in year t on normalized test scores in years $t-3$ through t. Each estimate includes the full set of controls outlined in Equation (4). Controls and fixed effects are indexed to the timing of the outcome variable (e.g. lagged math scores for the $t-2$ outcome are set to $t-3$). Equation (8) provides the structural equation; instrumented suspension rates are calculated based on Equation (5). Standard errors are adjusted for clustering at the school level and are reported in parentheses. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table 8: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
A. Math Test Scores						
(Suspension Rate)$_{sgt}$ $\times$ 10	0.040***	0.043***	0.045***	0.056***	0.038***	0.032***
	(0.010)	(0.011)	(0.013)	(0.010)	(0.014)	(0.006)
N	2,336,068	2,035,327	1,742,188	2,336,078	2,330,723	2,336,068
F Stat (IV First Stage)	618	418	433	434	210	292
B. English Test Scores						
(Suspension Rate)$_{sgt}$ $\times$ 10	0.065***	0.066***	0.066***	0.076***	0.063***	0.048***
	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.005)
N	2,378,265	2,071,082	1,772,165	2,378,275	2,372,889	2,378,265
F Stat (IV First Stage)	615	416	431	435	213	286
Modification	None	Years: 2006-13	Years: 2007-13	Unit: Schools	Instrument: $t-2$	Instrument: Days Suspended

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 6.6. 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. Column (5) uses an instrument derived from twice-lagged school-grade suspension rates instead of fixed initial 2003 suspension rates. Column (6) uses average days suspended as the endogenous variable as well as the corresponding days-based instrument. The estimate for this specification is interpreted as the change in test scores when average days suspended increases by 0.1 (relative to the baseline value of 0.12). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9: Direct and Indirect Effects of Suspension Rates

	Predicted Suspension Quintile (1 = Low Risk, 5 = High Risk)				
	1	2	3	4	5
A. Math Test Scores					
(Suspension Rate)$_{sgt}$ $\times$ 10	0.021	0.082***	0.094***	0.074***	-0.036***
	(0.017)	(0.015)	(0.014)	(0.011)	(0.009)
N	467,074	467,124	467,129	467,125	467,202
Fraction suspended	0.01	0.02	0.04	0.06	0.14
F Stat (IV First Stage)	826	729	657	590	342
	Predicted Suspension Quintile (1 = Low Risk, 5 = High Risk)				
	1	2	3	4	5
B. English Test Scores					
(Suspension Rate)$_{sgt}$ $\times$ 10	0.024***	0.091***	0.126***	0.117***	-0.022**
	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)
N	475,511	475,556	475,569	475,571	475,659
Fraction suspended	0.01	0.02	0.04	0.06	0.14
F Stat (IV First Stage)	827	727	662	589	336

Note: This table shows the effect of a 10 percentage point increase in suspension rates on normalized test scores, estimated separately for students in five quintiles based on their predicted probability of being suspended. Equation (9) in Section 7 shows how these probabilities are calculated. Quintiles are assigned based on a student's rank within a given grade-year. The coefficients in this panel result from estimating Equation (4) separately for each quintile and instrumenting for suspension rates with the instrument from Equation (5). Standard errors are adjusted for clustering at the school level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 10: Direct and Indirect Effects of Suspension Rates by Grade Category

Subject:	*Predicted Suspension Quintile (1 = Low Risk, 5 = High Risk)*				
Predicted Suspension Tercile:	1	2	3	4	5
A. Elementary School Students					
Math Effect	-0.034	0.157***	0.198***	0.117**	-0.047*
	(0.066)	(0.057)	(0.048)	(0.046)	(0.028)
N	179,025	179,048	179,057	179,044	179,085
Fraction suspended	0.01	0.01	0.01	0.02	0.05
English Effect	0.103*	0.277***	0.324***	0.266***	0.056**
	(0.061)	(0.040)	(0.037)	(0.041)	(0.024)
N	179,095	179,123	179,128	179,118	179,160
Fraction suspended	0.01	0.01	0.01	0.02	0.05
B. Middle School Students					
Math Effect	0.014	0.065***	0.081***	0.081***	-0.039***
	(0.020)	(0.016)	(0.015)	(0.013)	(0.010)
N	171,685	171,705	171,703	171,708	171,715
Fraction suspended	0.02	0.04	0.06	0.10	0.22
English Effect	0.020**	0.078***	0.116***	0.124***	-0.025**
	(0.010)	(0.009)	(0.010)	(0.011)	(0.012)
N	171,948	171,958	171,961	171,961	171,979
Fraction suspended	0.02	0.04	0.06	0.10	0.22
C. High School Students					
Math Effect	0.129***	0.142***	0.124***	0.067***	-0.045**
	(0.028)	(0.025)	(0.026)	(0.020)	(0.019)
N	116,364	116,371	116,369	116,373	116,402
Fraction suspended	0.01	0.03	0.05	0.08	0.16
English Effect	0.080***	0.132***	0.152***	0.099***	-0.081***
	(0.011)	(0.014)	(0.016)	(0.014)	(0.020)
N	124,468	124,475	124,480	124,492	124,520
Fraction suspended	0.01	0.03	0.05	0.08	0.16

Note: This table replicates Table 9, 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 level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.