Suspending Suspensions: The Education Production Consequences of School Suspension Policies*

Nolan Pope† George Zuo‡

February 7, 2020

Abstract

Managing student behavior is integral to the education production process. Many schools suspend disruptive students from the classroom, an increasingly controversial practice, which creates potential tradeoffs between disciplined students and their peers. In this paper, we study these tradeoffs by looking at how changes in school suspension policies impact student performance and teacher turnover. We use administrative data from the Los Angeles Unified School District, where suspension rates fell by over 90% from 2003 to 2013. Our empirical approach instruments for school suspension rates by interacting concurrent changes in district-wide suspension rates with school-specific suspension rates fixed to an initial pre-period. Our results indicate that a 10 percentage point reduction in suspension rates decreases math and English test scores by 0.03 to 0.04 standard deviations, decreases GPAs by 1.7%, and increases absences by 5.6%. Declining suspension rates also adversely impact teachers, leading to a 9.7% increase in teacher turnover. The overall negative effect of reducing suspension rates is driven by small but diffuse spillovers arising from a more lenient disciplinary environment, although this is partially offset by large and concentrated benefits for a small number of students who are no longer suspended at the margin.

*This version: February 7, 2020. We thank Nathan Barrymore, Peter Blair, Joshua Goodman, Judith Hellerstein, Melissa Kearney, Richard Patterson, Nathan Petek, John Shea, Cody Tuttle, and Sergio Urzua for helpful comments and discussions. Zuo graciously acknowledges financial support from the National Science Foundation Graduate Research Fellowship Program during the writing of this paper.

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

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

1 Introduction

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

Suspensions have long been controversial despite their wide prevalence. While advocates argue that suspensions are necessary for maintaining a productive and disruption-free learning environment for other students, critics counter that suspensions needlessly remove students from the classroom and disproportionately impact minority students. These criticisms culminated in a joint initiative between the Department of Education and Department of Justice in 2011, leading many states and school districts to reform their disciplinary policies and make them less strict and exclusionary (Lacoe and Steinberg, 2017). These reforms subsequently fueled a 40 percent reduction in suspensions over the past decade (see Figure 1). Despite this reduction, approximately 3 million students are still suspended each year, amounting to nearly 18 million days of lost instruction (Losen et al., 2015).[3]

Despite the wide variety of policies accompanying the rise and decline of suspensions over the last 40 years, there remains little empirical evidence on the causal impact of suspension policies on students and teachers.[4] In this paper, we estimate how reducing school suspension rates impacts student performance and teacher turnover. We focus on the Los Angeles Unified School District (LAUSD), where suspension rates in the district fell from eight percent in 2003 to less than one percent in 2015.[5] We use LAUSD administrative student-level data to estimate the effect of changes in school suspen-

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

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

[3] National Center for Education Statistics, Digest of Education Statistics Table 233.28. https://nces.ed.gov/programs/digest/d17/tables/dt17_233.28.asp.

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

[5] Throughout this paper, any reference to suspension rates refers to the number of students that were ever suspended during a given academic year, divided by the total number of enrolled students.

sion rates on math and English Language Arts (ELA) test scores, absences, GPA, and teacher turnover.

Estimating the 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 performance. Our empirical approach instruments for each school's suspension rate using year-to-year changes in the district-wide suspension rate interacted with school-specific suspension rates fixed to an initial pre-period. The instrument relies on year-to-year changes in district-wide suspension rates being exogenous to concurrent year-to-year changes in school-specific determinants of student performance. Initial suspension rates provide a way of scaling each school's exposure to these district-wide changes.

We find that a 10 percentage point decrease in school suspension rates decreases concurrent math and English test scores by 0.03 and 0.04 standard deviations, respectively. These effects, while modest, are highly significant and are equivalent to reducing teacher quality by 0.21 and 0.38 standard deviations for math and English (Chetty, Friedman and Rockoff, 2014). These effects also impact students beyond test scores. A 10 percentage point decrease in suspension rates increases absences (excluding suspension days) by 5.6% (0.3 percentage point) and decreases GPAs by 1.7% (0.045 grade points). This implies that declining suspension rates in Los Angeles during our sample period were detrimental to the average student across a variety of outcomes. The estimates are robust to a variety of sensitivity tests, as well as a falsification test showing that suspension rate changes have no effect on prior-year test scores.

Declining suspension rates may also affect teachers' well-being. With a limited ability to suspend students, teachers may be exposed to more misbehavior and may find behavior management more taxing. 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 9.7% (2 percentage points). 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. The effects are also disproportionately borne by high school teachers. These results suggest that the ability to suspend students is particularly important for inexperienced teachers, especially those working with older students. We also verify that these estimates remain robust to the aforementioned falsification test.

Teachers, principals, and schools face tradeoffs when determining the strictness of their suspension policies. When a misbehaving student is suspended, there is likely a *direct* effect on that student[6]

[6]This direct effect may be negative due to suspended students being removed from the classroom and the potential

as well as *indirect* spillover effects on his peers.[7] The effect of changing school suspension policies will therefore influence students through these two channels. We use two approaches to separately quantify direct versus indirect effects. First, we leverage the fact that changes in suspension rates affect low-misbehavior students primarily through indirect effects. We construct a proxy for students' propensities to misbehave and estimate the effects of changes in suspension rates separately for low-, medium-, and high-misbehavior students. Differences across these estimates provide insight into the direction and magnitude of direct and indirect effects. Second, we explicitly model direct and indirect effects within the same estimating equation.

Estimates from both approaches suggest that indirect effects of changing suspension rates are small, positive, and diffuse, whereas direct effects are large, negative, and concentrated. However, indirect effects dominate in the aggregate because all students are exposed to indirect effects, whereas relatively few students experience direct effects. We find that a student who is suspended due to a policy increase in suspension rates experiences a net decline in test scores since the direct cost of being suspended is approximately two times larger than the indirect spillover benefit to the student from less misbehavior. These results imply that lowering suspension rates decreases average test scores while increasing the test scores of students at the margin of being suspended. This illustrates that suspension policies exhibit a tradeoff between efficiency and equality (Okun, 2015). Higher suspension rates improve educational efficiency by increasing average test scores, but increase inequality by decreasing the test scores of students that misbehave, who are disproportionately represented by low-achievers, minorities, and male students.

Despite the tradeoffs that schools face, they 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

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

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

useful baseline for understanding the costs and benefits of changing school suspension policies.

2 Motivating Framework: Suspension Policies and Learning Production

We begin with a highly stylized static model illustrating a school's decision to set suspension rates based on learning production and other school-specific preferences. The model builds upon a parsimonious framework introduced by Lazear (2001): suppose that for any given moment in time, learning occurs if all n students in a given classroom are behaving. Given a suspension probability s, students behave with probability $p(s)$ and learning is produced with probability $p(s)^n$.[8] We assume that $p'(s) > 0$, implying that students are more likely to behave as the likelihood of being suspended for misbehaving increases.

Each school chooses s to maximize the following profit function:

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

Here, V is the value of a unit of learning. $C(s)$ represents learning that is lost by students who are suspended. $K(s)$ captures other costs of suspension policies not directly related to learning production. For example, schools with high s may face more frequent compliance warnings from the district and complaints from unhappy parents. Schools that value educational equity could also incur preference-based costs from instituting a higher s. Taking these costs into account, schools choose s so that the marginal benefit from increasing suspension rates equals the marginal cost:

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

The marginal benefit on the left-hand side captures the increase in learning production from better classroom behavior caused by a marginal increase in s. The marginal cost on the right-hand side encompasses both the marginal learning costs incurred by suspended students and the marginal non-learning costs incurred by the school.

One way to explicitly model the cost to suspended students $C(s)$ is to assume that the costs are comprised of 1) forgone instruction time and 2) residual learning costs, including required catch-up from missing classroom time as well as socioemotional effects of being suspended. This particular

[8] To give a sense of what s represents in this context, consider a classroom of 20 students in a high-suspension middle school, where 11% of students are suspended and each suspended student is suspended an average of 1.64 times (see Table 1). In a school year with 180 days, the daily suspension rate s can be approximated as $(20 \times 11.0\% \times 1.64)/180 = 2.0\%$. This implies that on any given day, such a classroom has a 2% probability of having a student suspended.

cost function can be written as follows:

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

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

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

In the absence of non-learning costs $K(s)$, the increase in learning production from a marginal increase in suspension rates equals the loss in learning production from an increase in suspended students. We refer to the change in learning production from better classroom behavior as *indirect effects* or *spillovers* of suspension policies. On the other hand, we refer to the costs directly incurred by suspended students as the *direct effects* of suspension policies.

These quantities may be unequal for two reasons, both of which tend to lean towards higher marginal indirect effects relative to marginal direct effects. First, non-learning costs of suspension $K(s)$ are likely to increase as suspension rates rise. $K'(s) > 0$ mechanically implies via Equation (4) that marginal indirect effects will exceed marginal direct effects. Second, schools may simply lack information about the marginal costs and benefits of learning production with respect to suspension rates. Given the disproportionate salience of direct effects, schools may tend to overestimate marginal direct effects relative to marginal indirect effects and set suspension rates below optimal levels.

When marginal indirect effects exceed marginal direct effects, aggregate learning production can be increased by raising suspension rates. Empirically, test scores should increase by marginally raising suspension rates if test scores are an increasing function of learning production. However, the existence of a wedge between marginal direct and indirect effects can still be optimal for certain schools if driven by non-learning costs $K(s)$. If instead the wedge is driven by factors such as salience and lack of information, changing s leads to an increase in overall welfare.

3 Suspensions and School Discipline in Los Angeles

The use of suspensions peaked nationwide in the early 2000's, when widespread "zero tolerance" policies enabled schools to suspend students for relatively minor non-violent infractions.[9] These policies were designed based on prevailing theories of "broken windows" and became commonly employed in school districts following the Gun-Free Schools Act of 1994, which mandated expulsion for infractions involving a firearm (Curran, 2016). The liberalized use of suspensions has recently come under additional scrutiny due to concerns over the punitive and regressive nature of school suspension as well as allegations of discriminatory practices toward youth of color (U.S. Department of Education, 2014).[10] 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, Strunk and Dhaliwal, 2018). Discipline reform in the district occurred in two main phases. Starting in the 2006-07 school year, the district implemented School-wide Positive Behavior Supports (SWPBS), a program designed to address racial disparities in suspension rates. Schools were required to develop multi-tiered discipline plans to manage student behavior and avoid suspensions. SWPBS gave schools autonomy to develop their own discipline plans, placing the burden of lowering suspension rates upon local school administrators. Administrators were required to develop disciplinary policies tailored to their school's educational context while simultaneously providing instructors with resources and training to adjust to the new policies.

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

[9] National Center for Education Statistics, Digest of Education Statistics Table 233.20. https://nces.ed.gov/programs/digest/d17/tables/dt17_233.20.asp.

[10] 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. However, such issues are beyond the scope of this paper.

restorative justice began with a cohort of 153 schools in the 2014-15 school years, and will continue across four additional cohorts of schools until full, district-wide implementation in 2020. Schools in each cohort will receive training and support from the district, including specially-hired restorative justice specialists and counselors.

4 Data

In this paper, we use student-level administrative data from the Los Angeles Unified School District (LAUSD). The LAUSD is currently the second-largest school district in the United States and enrolls over 600,000 students. Within the district, 74 percent of students are Hispanic, 10 percent are white, and 8 percent are black.[11] The administrative data we use include a panel of students in grades 2 through 12 beginning in the 2002-03 school year and ending in the 2014-15 school year. For consistency, we reference school years by year of graduation (e.g. 2003 represents the 2002-03 school year).

The student data include our main variables of interest: standardized test scores, grade point averages, absences, and suspensions. Standardized test scores for both math and English Language Arts scores come from the California Standards Test (CST) which is administered to all students in grades 2 through 11. The California Standards Test was discontinued in 2014 and consequently, test scores only extend through 2013. We normalize math and English CST scores to mean zero and standard deviation one for each grade level and school year. Grade point averages are calculated on a four-point scale.[12] Due to privacy restrictions, we do not observe student demographic characteristics, including gender, race, and socioeconomic status. We summarize available student and school characteristics in Table 1. In Table 1, 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 difference is most noticeable for middle school students. We also observe that high-suspension middle and high schools tend to have substantially larger student bodies. As students progress from elementary to middle and high school, GPAs fall, absences increase, and suspensions increase.

Data on suspensions include the number of times a student was suspended as well as the number of days suspended.[13] School and district-wide suspension rates are calculated by dividing the number

[11] Statistics can be accessed via the California Department of Education via DataQuest: https://dq.cde.ca.gov/dataquest/

[12] See Petek and Pope (2018) for a detailed description of how GPAs were computed, as well as availability of other non-achievement based grades.

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

of students who were suspended at least once by the total number of students enrolled. District-wide suspension rates in the LAUSD decreased substantially during the years available in the data. Nearly 8 percent of LAUSD students were suspended in 2003, compared to less than 1 percent in 2015. Figure 2 shows that suspension rates also vary substantially by grade. Suspension rates increase with each grade beginning in second grade, peak at eighth grade, and decrease with each grade through the end of high school. Less than 5 percent of elementary school students were suspended in 2003, compared to over 15 percent of eighth grade students. Figure 3 illustrates how suspension rates evolved over time in elementary versus middle/high schools, separating schools into four quartiles based on 2003 suspension rates. Initial dispersion in 2003 is much greater in middle and high schools, as is the magnitude of the subsequent decline. Schools across the four quartiles nearly converge by the end of the sample period. Finally, Figure 4 describes the distribution of suspensions across the intensive margin of days suspended. Conditional on being suspended, approximately 50 percent of students are only suspended for a single day and 30 percent of students are suspended for two to three days. In addition, suspension lengths tend to be longer for middle and high school students.

5 Empirical Strategy

Our objective is to estimate the effect of school suspension rates on standardized test scores and other academic outcomes. Suspension rates are negatively correlated with many aspects of school quality, biasing standard OLS estimates downward.[14] With this in mind, we begin with the following estimating equation:

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

y_{isgt} represents normalized math or English test scores for student i in school s, grade g, and year t. We calculate suspension rates at the school-grade level given the substantial across-grade variation documented in Figure 2. $X_{isg,t-1}$ is a vector of lagged math test scores, English test scores, and GPA for student i. $S_{sg,t-1}$ is a vector of average math and English test scores in grade g and school s from the previous year. P_{isgt} includes average prior-year math and English test scores of all students in the same school and grade as student i in time t. All standard errors are adjusted for clustering at the school-grade level. The parameter of interest is ρ, which represents the effect of changing suspension

[14] Consequently, much of the existing correlational research on suspension rates finds a negative association between suspension rates and achievement; see Noltemeyer, Ward and Mcloughlin (2015) for a meta-analysis summarizing this research.

rates from 0% to 100%. We multiply suspension rates by 10 so that ρ represents the effect of a 10 percentage point change, reflecting the overall suspension rate decline that we observe in the LAUSD data.

The fixed effects and control variables in Equation (5) account for certain observable omitted variables that are correlated with suspension rates. School-grade fixed effects absorb permanent differences in school-grade characteristics and restrict the identifying variation to across-time comparisons of students in the same school and grade. Lagged individual test scores and GPA modify the regression residual ϵ_{ist} so that it contains only variation from education production occurring in the current year. Lagged school-grade test scores capture observable fluctuations in school quality. Finally, peer lagged achievement accounts for observable changes in peer quality.

Despite these controls, estimates of ρ will remain biased if unobservable time-varying, within-school-grade characteristics 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 by utilizing variation in district suspension rate trends. District-wide changes in suspension rates are predictive of school-specific suspension rates without directly affecting individual student test scores. The instrument we use is as follows:

$$\widetilde{SuspendRate}_{sgt} = SuspendRate_{sg2003} \times G_g^{-s}(t-1,t) \tag{6}$$

where $SuspendRate_{sg2003}$ is the initial suspension rate of school s and grade g as of 2003, and $G_g^{-s}(t-1,t) = \frac{SuspendRate_{gt}^{-s}}{SuspendRate_{g,t-1}^{-s}}$ is the year-over-year growth in district suspension rates in grade g between years $t-1$ and t, leaving out the contribution of school s. The instrument predicts suspension rates by multiplying initial 2003 suspension rates at school s and grade g by the leave-own-out district-wide growth in suspensions.

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

$$SuspendRate_{sgt} = SuspendRate_{sgt-1} \times \frac{SuspendRate_{sgt}}{SuspendRate_{sgt-1}} \equiv SuspendRate_{sgt-1} \times G_{sg}(t-1,t) \tag{7}$$

where $G_{sg}(t-1,t)$ is the year-over-year growth of suspension rates at school s and grade g between $t-1$ and t. We first replace $G_{sg}(t-1,t)$ with a district-wide growth rate $G_g(t-1,t)$. We leave out the contribution from school s so that $G_g^{-s}(t-1,t)$ represents district grade-level suspension rate

growth that does not include endogenous own-school suspension rates. The exclusion of school s also addresses finite sample bias from using own-school information in the first stage (Goldsmith-Pinkham, Sorkin and Swift, 2018).

The instrument relies on the fact that changes in district suspension rates occur externally from the idiosyncratic decision-making of any given school. The instrument also leverages the fact that school-specific exposure to changes in the district suspension rate is linked to prior suspension rates (see Figure 3). However, the instrument constructed from lagged suspension rates in Equation (7) remains problematic since both lagged suspension rates and lagged test scores will be on the right-hand side in the first-stage equation. This introduces bias if lagged test scores are themselves an outcome of lagged suspension rates (Angrist and Pischke, 2008). We could address this by using twice-lagged suspension rates in the construction of the instrument. However, if $SuspendRate_{sgt}$ is correlated with ϵ_{isgt} and ϵ_{isgt} happens to be serially correlated, twice-lagged suspension rates could still be endogenous. This motivates fixing lagged suspension rates to 2003 (the earliest year available in the data) and restricting the estimation sample to begin in 2005.

Given this instrument, the first stage equation is:

$$SuspendRate_{sgt} = \gamma + \xi[\widetilde{SuspendRate}_{sgt}] + X_{isg,t-1}\delta + S_{sg,t-1}\gamma + P_{isgt}\xi + \mu_{sg} + \nu_{isgt} \tag{8}$$

The instrument generates relevance from both intial 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 inital suspension rate. Figure 3 shows that initial suspension rates are highly predictive school's exposure to changes in district suspension rates: a low initial suspension rate implies that the school has low exposure to district-wide changes. We visually inspect the first stage by plotting a binned scatterplot of suspension rates versus instrumented suspension rates in Figure A1.[15] We observe a strong linear relationship between the instrument and actual suspension rates, which is supported by a large first-stage F-statistic exceeding 1,000.

The exclusion restriction requires that the instrument be uncorrelated with the structural error term ϵ_{isgt}, which contains within-school-grade, across-time variation in test scores not accounted for lagged individual, school, and peer achievement. Initial suspension rates, while not directly contributing to across-time variation, scales the variation generated by district suspension rate growth. District

[15] Both axes have been residualized with respect to school-grade fixed effects and all control variables.

suspensions growth evolves "externally" from the variables in the model but does not strictly satisfy the exclusion restriction since it could still be correlated with the regression residual (Jaeger, Ruist and Stuhler, 2018; Deaton, 2010). The exclusion restriction is violated if the *interaction* of these two variables is correlated with the regression residual. Because the variation produced by this interaction is complex, we probe the exclusion restriction using a combination of falsification and sensitivity tests.

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 (Equation 5) and IV (Equation 6) estimates of school suspension rates and test scores. 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 OLS is large and negative. A 10 percentage point increase in suspension rates is associated with a -0.139 and -0.127 standard deviation decline in math and English test scores, respectively. This negative relationship is expected since students with low test scores are more likely to be suspended. As such, schools with a high fraction of low performing students will be more likely to have high suspension rates. By adding school-grade fixed effects and controlling for the time-invariant characteristics of the school-grade (including the average type of student that attends the school), the coefficient in column (2) moves by a order of magnitude towards zero. 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 (0.003 and 0.008 standard deviations for math and English), with only the estimate on English being statistically significant.

In column (6), we use the instrument to address the endogeneity problem. The IV estimates are both positive and statistically significant at the 1-percent level. We find that a 10 percentage point increase in a school's suspension rate increases math and English tests scores by 0.029 and 0.038 standard deviations. The IV estimate is larger than its corresponding OLS estimate in column (5), suggesting that unobservable, time-varying confounders negatively bias the fully-controlled OLS estimates. The IV estimate is still somewhat small given it reflects the effect of a large, 10 percentage point change in suspension rates. To contextualize the effect size, a one standard deviation increase in teacher

value-add in a given year increases math and English test scores by 0.14 and 0.10 standard deviations (Chetty, Friedman and Rockoff, 2014). The impact from one year of exposure to a better teacher more than doubles the impact from a decade of declining suspension rates in Los Angeles.

The IV estimates represent the effect of suspension rates on average test scores, net of direct and indirect effects. A utilitarian administrator weighing all students' test scores equally would use this parameter to determine optimal suspension rates. The positive estimates suggest that when suspension rates declined in Los Angeles, the average cost of greater exposure to learning spillovers exceeded the average benefit from lower suspension probabilities. Intuitively, even if direct effects are large, the effects are concentrated on such a small number of students that they become outweighed by the small but far more diffuse effect of learning spillovers. We return to this discussion in Section 7.

6.2 Absences and GPA

We can also estimate the effects of school suspension rates on absences and GPA. Days suspended are subtracted from days absent to prevent absence counts from being inflated by suspensions. Figure A2 shows that absences and GPA both exhibit substantial time trends. To avoid conflating the effect of suspension rates with aggregate time trends, we de-trend both outcome variables by subtracting the aggregate year-grade mean from each observation.

Table 3 presents OLS and IV estimates for these two outcomes. Suspension rates are positively correlated with absences and negatively correlated with GPA in the raw OLS. Adding school-grade fixed effects and control variables changes the signs of both estimates. In the fully controlled OLS, a 10 percentage point decrease in suspension rates decreases absences by 0.3 percentage point (5.6%) and increases GPAs by 0.045 grade points (1.7%); both estimates are statistically significant at the 1% level in the OLS estimates. The final column provides the IV estimates, which indicate that a 10 percentage point increase in suspension rates decreases absences by 0.6 percentage points (11%) and increases GPAs by 0.14 grade points (5.6%). The effects on GPA are consistent with the patterns we observe for test scores. The effects on absences suggest that harsher discipline policies could deter a broad class of misbehavior that includes skipping class. Students may also feel less inclined to attend class when the classroom environment is more prone to disruption from misbehavior. We note that absences and grades can be manipulated, complicating the interpretation of these findings.[16]

[16]For example, LAUSD has been alleged of using "informal" off-the-books suspensions (Watanabe, May 31, 2014), presumably as a workaround to remain compliant to district guidance on acceptable uses of suspensions. Students can also be suspended for truancy, which is a failure to report to school for an extended period of time.

6.3 Teacher Attrition

Spillovers from misbehavior can also affect teachers. Lower suspension rates could make classroom behavior management more taxing, especially for inexperienced teachers. Difficult working conditions could ultimately lead to increased 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.[17] The outcome of interest is the probability that a teacher leaves their school between years t and $t+1$. We do not differentiate between teachers who switch schools and teachers who leave the district altogether.

In Table 5, we show that a 10 percentage point increase in suspension rates is associated with a 2 percentage point decrease (9.7%) in teacher turnover. We note that the baseline turnover rate in Los Angeles is quite high (20.6%). Newton et al. (2011) verify that in Los Angeles, the probability that an elementary teacher leaves his or her school is 21.6% in the first year and 19.5% in the second. For high school teachers, 26.4% leave in the first year and 21.6% leave in the second. In Panels B and C, we provide separate estimates for inexperienced teachers with 0-2 years of experience (prior to when tenure decisions are made) and teachers with three or more years of experience.[18] The point estimate for inexperienced teachers is more than triple the size of the point estimate for experienced teachers (8.7 versus 2.6 percentage points). We conclude that lower suspension rates are detrimental for teacher retention, and that the effect is compounded for inexperienced teachers.

6.4 Effects by Grade

We also estimate our IV results separately for elementary, middle, and high school students. The identifying assumption is the same for these estimates because the identifying variation remains constrained to within-school-grade comparisons of suspension rates over time. However, suspension rates vary greatly across grades (see Figure 2). A 10 percentage point change for elementary schools (with a baseline suspension rate of 1.7%) will have a different impact than the same change for middle schools (with a baseline suspension rate of 9.0%). The nature of misbehavior also differs across grades; Figure 4 suggests that the marginal suspension in middle and high school is likely more serious than the marginal suspension in elementary school. This could lead to disparate impacts on

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

[18] 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 least three years of experience. This assumption requires that we omit 2005 from our estimation sample.

learning spillovers across grades.

Table 4 presents these estimates. The test score coefficients remain positive and significant for elementary and middle school students but are not significant for high school math test scores. The effect is quite large for elementary students (0.127 standard deviations for math; 0.150 standard deviations for English), which could be due to the low baseline suspension rate of elementary students. The effect on middle school students is similar in magnitude to our baseline estimates (0.028 and 0.034 standard deviations). The effect on high school students is not significant for math, and is small but significant for English (0.016 standard deviations). It is possible that the spillover effects of misbehavior become smaller as classes become more varied and dispersed in high school. However, we also find that the effects on GPA are positive and significant across all types of students, with effect sizes of 0.104, 0.150, and 0.131 grade points for elementary, middle, and high school students.

The effects of suspension rates on absences appear to increase with grade level. For elementary students, the effects are small and insignificant. We find this unsurprising given that elementary students have little autonomy 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 0.4 and 0.7 percentage point. Both of these estimates are significant at the 1 percent level. As students mature, their school attendance patterns become more responsive to changes in suspension rates.

Table 6 shows that the effects on teacher attrition also appear to increase with grade level. Inexperienced high school teachers are most impacted: a 10 percentage point increase in suspension rates decreases the likelihood of attrition by 11.6 percentage points (39%). Higher suspension rates also decrease attrition among inexperienced middle school teachers (-7.1 percentage points) and experienced high school teachers (-5.3 percentage points). These results suggest that behavior that We hypothesize that alternative disciplinary methods that can be used for younger students may be less effective for older students. High school students are also more physically imposing; approximately 9 percent of teachers are physically threatened each year.[19] Teachers who feel threatened but cannot respond to misbehavior with suspensions may be more likely to leave.

6.5 Placebo Test: Timing of Estimated Effects

We conduct a simple falsification test by estimating the effect of instrumented suspension rates in year t on test scores in earlier years. Suspension rates in year t should not have a causal effect on test scores

[19] National Center for Education Statistics, Digest of Education Statistics, Table 228.70; link available here

in earlier years. For test scores in period $t-1$, we estimate the following structural equation:

$$y_{isg,t-1} = \alpha + \rho SuspendRate_{sgt} + \mathbf{X'}_{isg,t-2}\beta + \mathbf{S'}_{sg,t-2}\theta + \mathbf{P'}_{isg,t-1}\phi + \lambda_{sg} + \epsilon_{isg,t-1} \quad (9)$$

The equation takes the baseline estimating equation in (5) and lags all variables except for suspension rates by one period. The fixed effect λ_{sg} is indexed to the school-grade attended in $t-1$ in order to ensure that test score comparisons are restricted to students within the same school and grade at the time of the measurement of y_{isg}^{t-1}. We then instrument for suspension rates in the same manner as in our baseline IV specification.

Table 7 presents the effect of suspension rates in year t on test scores in years $t-3$ through t. The results for math and English test scores are small and are generally not significant at conventional thresholds (although math is negative and significant in $t-3$) until time t. The effect size increases over time for math test scores, but does not display any noticeable pattern for English test scores. We find significant effects for absences and GPA in the pre-periods; however, these effects are small are likely the product of large sample sizes and high statistical precision. The effect size jumps noticeably from $t-1$ to t, which suggests that the significant pre-period effects are likely the product of noise estimated with low standard errors.

6.6 Robustness Checks

Serial Correlation and Lagged Suspension Rates—The main IV results are estimated on a sample beginning in 2005. This is because 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 year. 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 math and English test scores both increase slightly relative to our baseline estimates. The effects also remain precisely estimated and significant at the 1-percent level.

School-level variation—We also test whether the results are sensitive to the choice to use school-grade level variation instead of school-level variation. Schools may be unlikely to differentiate suspension policy (among other school policies) by grade. The regression residual may also exhibit school-specific serial correlation, implying that standard errors should be adjusted for clustering at the school level.

Column (4) of Table 8 shows how our preferred, fully-controlled IV estimates change when using school-level suspension rates and school-level clustered standard errors. We observe that both the point estimates and standard errors increase slightly, and that the effects are still significant at the 1 percent level. We conclude that the decision to use school-grade variation does not meaningfully impact the results.

Using twice-lagged suspension rates—We also consider a variation of our instrument in which we use twice-lagged suspension rates instead of fixing suspension rates to 2003. The reason for this stems from the accounting identity in Equation (7). Using twice-lagged suspension rates instead of fixed initial suspension rates may theoretically increase the instrument's predictive power. Interestingly, Column (5) of Table 8 shows that the first-stage F-statistic *falls* when the instrument is constructed in this way. This is because a school's exposure to district suspension rate growth appears more highly correlated with initial conditions than with recent suspension rates. We nevertheless find that the estimates change little when constructing the instrument in this way. We also note that serial correlation will more noticeably bias estimates derived from the instrument constructed with twice-lagged suspension rates. The similarity of the two estimates provides some additional assurance that bias from serial correlation is relatively trivial.

6.7 Correlates of Instrumented Suspension Rates

Our instrument captures the portion of within-school-grade variation in suspension rates that is tied to district suspensions growth, weighed by initial school-grade suspension rates. The exclusion restriction is violated if the instrument is correlated with omitted variables that affect within-school-grade test scores independent of suspension rates. While the exclusion restriction is inherently untestable, we can test whether the instrument is correlated with observable confounders. We do so by regressing several confounders η on instrumented suspension rates via the following school-grade regression:

$$\eta_{sgt} = \alpha + \widetilde{SuspendRate}_{sgt}\theta + \mathbf{S}'_{sg,t-1}\beta + \eta_{sg,t-1}\phi + \lambda_{sg} + \nu_{sgt} \tag{10}$$

where $\mathbf{S}$ again represents lagged school-grade math and English test scores. We consider four confounders available in our data. First, the number of students within a school-grade could affect class sizes, concentration effects,[20] and other school characteristics that influence test scores independently

[20]Concentration effects refer to a phenomenon where misbehavior and crime increase with the number/density of students in a given area; see Jacob and Lefgren (2003).

of suspension rates. Second, we consider whether the relationship between test scores and suspension rates are driven by selection among students who are English Language Learners (ELL), who may have different instructional and test-taking needs than traditional students. Approximately 30 percent of our sample is composed of ELL students. Although ELL students are not observably suspended at different rates (5.5% versus 5.4%), they have notably lower math and English test scores. Third, we consider the effect of suspension rates on school-switching and attrition from the LAUSD itself.[21] Parents may view a district-wide decrease in suspensions from $t-1$ to t as a sign of a deteriorating school learning environment, and opt to move their child in the following school year. The positive relationship between instrumented suspension rates and test scores could potentially be driven by high-achieving students strategically leaving schools with declining suspension rates.

Table A2 shows the correlation between each of these confounders and a 10 percentage point increase in instrumented suspension rates. We find that the instrument is not correlated with the number of students, the proportion of students leaving the district, or the proportion of students switching schools. However, a ten percentage point increase in instrumented suspension rates leads to a statistically signfiicant decrease in the proportion of ELL students by 1.9 percentage points. We find that adding ELL status as a covariate in Equation (5) results in a slight *increase* in the magnitudes of our estimates without affecting statistical significance. This is likely due to the fact that the correlation between suspension rates and ELL status, while significant, is not economically meaningful in magnitude.

7 Direct and Indirect Effects of Suspensions

7.1 Conceptual Framework

Schools determine the strictness of school suspension policies by weighing the benefits and costs to students. Suspension policies affect some students by changing the probability that they are suspended for misbehavior. The effects of being suspended are the *direct effects* of suspension policies. The most immediate consequence of being suspended is reduced classroom time: LAUSD students miss 2.1 days of school on average per suspension. Being suspended could also influence student motivation and engagement, although the the direction of these effects is not obvious *ex ante*. Causal

[21] We flag a student as having switched schools if the school attended in one year is not the same as the school attended in the prior year, not including students currently in the terminal grade of a certain grade category (typically, grades 5 and 8). A student leaves the district if they do not have a record in the following year and are not in the 12th grade.

evidence on the direction and magnitude of direct effects is limited.[22] Lacoe and Steinberg (2018*a*) find negative direct effects in Philadelphia using a individual fixed effects approach with a supplementary IV approach, while Anderson, Ritter and Zamarro (2017) find *positive* effects of being suspended using dynamic panel methods on in Arkansas. Beyond academic outcomes, Bacher-Hicks, Billings and Deming (2019) find that being suspended leads to lower graduation rates, lower rates of college attendance, and higher rates of future crime and arrest.

Suspension policies also affect students through *indirect effects* on misbehavior. Misbehavior disrupts class and diverts a teacher's time and energy away from instruction, producing negative spillovers on learning (Lazear, 2001). In 2016, 43 percent of teachers "agreed" or "strongly agreed" that student misbehavior interfered with their teaching during the year (Musu-Gillette et al., 2018). Suspensions provide teachers with a low-cost way to curtail disruptions and prevent misbehavior from escalating.[23] When a school mandates a decrease in the use of suspensions, teachers that previously relied on suspensions must adopt new approaches to managing misbehavior.

Causal identification of indirect effects of suspensions is challenging, and existing approaches leave much to be desired. Beyond suspensions, there exists a robust economics literature on peer effects in schools (Sacerdote, 2011). For example, Imberman, Kugler and Sacerdote (2012) study students with disciplinary problems who were displaced by Hurricanes Katrina and Rita. Local students in districts receiving these students experienced increases in disciplinary problems and absenteeism, although there was no impact on test scores. Carrell and Hoekstra (2010) study children from families matched to domestic violence cases, and find negative effects of such students on the performance of their peers. Deming (2011) finds that high-risk youth who received their top choice in a school choice lottery experienced large reductions in future arrests and days incarcerated, and partially attributes these effects to peer effects.

The concept of deterrence also plays a role in shaping direct and indirect effects. The seminal Beckerian theory of crime (Becker, 1968) posits that the decision to commit crime is influenced by the probability and expected severity of punishment. While we typically do not expect youth to make perfectly rational decisions, some empirical evidence does suggest that students are capable of adjusting their behavior in response to changes in disciplinary environments. Waddell (2012) studies marijuana possession in schools, and uses IV methods to show that second offenses for possession

[22] A recent meta-analysis by Noltemeyer, Ward and Mcloughlin (2015) summarizes the correlational research across 34 for studies, finding a negative correlation between achievement and being suspended.

[23] However, not all suspensions are equally productive at reducing spillovers. Suspensions motivated by implicit racial biases or targeted towards minor infractions may produce relatively smaller effects.

decrease as the severity of second-offense punishments increase. Within the suspensions context, a decline in suspension rates lowers the cost of misbehaving. More students misbehave, but fewer students are suspended conditional on misbehaving. Direct effects respond ambiguously as deterrence weakens whereas indirect spillover effects from misbehavior increase.

In summary, a decline in suspension rates has an ambiguous impact on the prevalence of direct effects (fewer students suspended but more students misbehave), whereas indirect effects increase in magnitude (more misbehavior occurs, less existing misbehavior leads to suspension). We explicitly model the interplay of direct and indirect effects in a simplified static model presented in Appendix A. The model begins with a student's discrete choice to misbehave, and illustrates how a change in suspension rates ultimately translates into direct and indirect effects. The model also suggests that the "optimal" suspension rate at a given school depends on each school's preferences for efficiency versus equality in student test scores. Specifically, schools set suspension rates to solve a social planner problem, which maximizes the weighted average achievement across all students. Schools that place greater weight on the achievement of marginal students are more likely to prefer reducing or eliminating suspensions. We return to this idea in our discussion of empirical findings.

7.2 Empirical Estimates

Changes in suspension rates affect test scores through direct and indirect effects. If direct effects are hypothetically negative and indirect effects are hypothetically positive, the signs of the net effects estimated in Section 6.1 suggest that the average student is more affected by indirect effects than direct effects. This is likely due to the fact that relatively few students are suspended, whereas learning spillovers affect all students. However, this comparison of direct versus indirect effects remains speculative without further analysis. In this section, we use two approaches to disentangle and quantify these effects.

The first approach exploits the fact that direct effects should only impact students who misbehave. Students who do not misbehave are not typically suspended, and a change in suspension rates should not affect their exposure to direct effects. Changing suspension rates *will* affect exposure to indirect spillover effects for all students. Under the (somewhat restrictive) assumption of constant indirect effects, we can isolate direct and indirect effects by comparing estimates between high- and low-misbehavior students. Estimates for low-misbehavior students represent average indirect effects, whereas the difference in estimates between high- and low-misbehavior students represents average

direct effects.

We first construct a crude proxy of an individual's propensity to misbehave since we do not directly observe misconduct in our data. We use the fact that a student in 2004 is approximately eight times more likely to be suspended than the same student in 2013. This suggests that suspensions in 2004 are more likely to capture minor forms of misconduct. We begin with a simple linear probability model to predict the probability that a student was suspended in 2004. The linear probability model we use is as follows:

$$Suspended_{isg,2004} = \beta_0 + \mathbf{X}'_{isg,2003}\beta_1 + \mathbf{S}'_{isg,2003}\beta_2 + \mathbf{P}'_{isg,2003}\beta_3 + \epsilon_{isg,2004} \tag{11}$$

Here, $\mathbf{X}$ is a vector of lagged math/English test scores, GPA, fraction of non-suspended days absent, grade retention status, suspension status, and days suspended.[24] $\mathbf{S}$ is a vector of lagged school-grade math and English test scores, and $\mathbf{P}$ includes the lagged test scores for peers of student i. Table A3 provides the estimated coefficients from 2004.

We then use the coefficients produced by this model to proxy for misbehavior in later years. An advantage of this approach is that the 2004 data are not included in our estimation sample, meaning that the predictions are trained using an out-of-sample set of observations. Figure A3 provides a histogram of these predictions, which tend to be concentrated near zero but are also display a fair amount of dispersion. The primary concern with this measure is that aggregate trends in several of the model inputs may impede our ability to compare predictions across time. The predictions are also unlikely comparable across grades. To account for this, we construct within-grade-year percentile ranks for each student based on their predicted misbehavior propensities. After assigning percentile ranks to all students, we split the distribution into equal thirds and produce IV estimates for each of the resulting subsamples.

We present these estimates in Panel A of Table 9. The math estimates remain positive and significant and are similar across all three groups (from low to high misbehavior: 0.049 SD, 0.044 SD, 0.046 SD).[25] The English estimates are quite high for low-misbehavior students and decline steeply for medium- and high-misbehavior students (from low to high misbehavior: 0.072 SD, 0.028 SD, 0.018 SD). Both sets of estimates suggest that higher suspension rates are beneficial to the average student, regardless of whether they are low-, medium-, or high-misbehavior students.

[24] Unfortunately, the data do not contain many core demographics such as gender, race, and socioeconomic status.

[25] We observe that the math estimates for all three groups are greater than our baseline estimate of 0.028 SDs. There are two main reasons for why this might occur. First, the prediction model uses covariates not found in the baseline model, meaning that students with missing entries for these covariates will be omitted from this analysis. Second, splitting the sample into three different subsamples changes the weights associated with the underlying LATE calculation.

Under the assumption of constant indirect effects, the coefficients on low-misbehavior students characterize the average impact of a 10 percentage point suspension rate through only spillovers. The difference in effect sizes between high- and low-misbehavior students characterizes the average impact of direct effects. The English estimates suggest that the direct effects of being suspended are negative and potentially quite large (difference = 0.054 SDs). To recover the effect of actually being suspended (as opposed to the average direct effect), we can scale the difference between low- and high-misbehavior students by the difference in suspension rates between the two groups (10 percentage points). This calculation suggests that the effect of being suspended is extremely large—roughly 0.5 standard deviations. On the other hand, our estimates for math test scores display negligible differences across the three groups, suggesting trivial direct effects in math. While it is possible that direct and indirect effects operate differently across math and English, it is also possible that the assumption of constant indirect effects is simply too restrictive.

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

$$y_{isgt} = \alpha + \rho_1 SuspendRate_{sgt}^{-i} + \rho_2 Suspended_{isgt} + \mathbf{X}'_{isg,t-1}\beta_1 + \mathbf{S}'_{isg,t-1}\beta_2 + \mathbf{P}'_{isg,t-1}\beta_3 + \lambda_{sg} + \epsilon_{isgt} \quad (12)$$

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

Panel B of Table 9 presents our estimates. The first row contains estimates of indirect effects,

which vary across low-, medium-, and high-misbehavior students. The effect sizes appear similar to the aggregate effect sizes in Panel A. This again suggests that the average effect of suspension rates is driven by indirect effects. The difference between the aggregate effect size in Panel A and the indirect effect size in Panel B is representative of the size of the average direct effect. Average direct effects are small for the average low-misbehavior student (-0.003 SD for math; -0.004 SD for English) but are much larger for high-misbehavior students (-0.013 SD for math; -0.016 SD for English). For English test scores of high-misbehavior students, average direct effects offset nearly half of average indirect effects. Only 12 percent of high-misbehavior students are suspended, suggesting that the magnitude of direct effects may be quite large.

The coefficients for direct effects, though not strictly causal, shed light onto these magnitudes. The effects are consistently large and negative. Interestingly, the effect of being suspended is worse for low-misbehavior students (-0.138 SD for math, -0.142 SD for English) than for high-misbehavior students (-0.074 SD for math, -0.092 SD for English). One potential explanation is that low-misbehavior students tend to have higher test scores to begin with and have a longer way to fall. The estimate of being suspended is roughly comparable to the effect of year-long exposure to a teacher with a value-add of one standard deviation below the mean.

The sum of the indirect and direct effects in Panel B represents the effect of a 10 percentage point increase in school suspension rates for a student at the margin of being suspended. The net effect is negative for all three categories of students in math (low: -0.086 SD, medium: -0.046 SD, high: -0.015 SD) and English (low: -0.066 SD, medium: -0.082 SD, high: -0.058 SD). We conclude that while the average student benefits from higher suspension rates, the effect of being suspended more than offsets the benefits from a reduction in learning spillovers. This implies that any decision to change suspension rates must contend with an equality versus efficiency tradeoff: lower suspension rates decrease average student achievement due to inefficiencies from misbehavior spillovers. However, being suspended has adverse effects on student achievement. Lower suspension rates produce educational inequality because the cost of stricter suspension policies is disproportionately borne by students who are likely to be low achievers in the first place.

We also provide a breakdown of how indirect, direct, and aggregate effects differ for students in elementary, middle, and high school. We calculate these estimates by splitting the sample into grade categories after generating predicted misbehavior rates and percentiles. The estimates are presented in Table 10. The previous patterns generally hold for elementary and middle school students: small,

positive indirect effects accompanied by large, negative direct effects. The exception appears to be for high school students. The indirect effects are smaller and less precisely estimated, while the direct effects are large enough that the average net effect of suspension rates is nearly zero for medium- and high-misbehavior students.

8 Conclusion

School suspensions are a controversial topic in education. In this paper, we provide new evidence on the multi-faceted consequences of changing school suspension rates. Our empirical approach instruments for endogenous school suspension rates by interacting lagged suspension rates with district-wide growth in suspension rates. We find that a 10 percentage point increase in suspension rates leads to small but statistically significant increases in math and English test scores (0.029 and 0.038 SDs, respectively). This implies that higher suspension rates were beneficial for the average student in Los Angeles.

The effect of changing suspension rates is comprised of *direct* effects on students who are suspended, as well as *indirect* spillover effects on peers of suspended students. This paper provides the first causal estimates of these direct and indirect effects. We find that an increase in suspension rates leads to small but positive indirect effects on students by decreasing (and potentially deterring) misbehavior in the classroom. We also find that the direct effect of being suspended is quite large and negative. In the aggregate, the small and positive effect of higher suspension rates is driven by indirect spillover effects because relatively few students are actually suspended. However, test scores fall for students who are suspended due to suspension rate increases. Such students are also more likely to be low-achieving students. Higher suspension rates increased educational inequality in Los Angeles by placing the burden of suspensions on low-achieving students.

Suspensions are simultaneously a go-to instrument for teachers to manage misbehavior and a regressive method of punishing students for misbehavior. Large-scale suspension rate declines have occurred in Los Angeles and in other school districts across the nation. In many cases, these declines coincide with initiatives such as School-wide Positive Behavior Supports (implemented in LAUSD in 2006), which provide schools with broad guidelines for how to decrease suspension rates. However, such initiatives give substantial autonomy to schools (Skiba et al., 2014), generating substantial heterogeneity in the effectiveness of suspension rate decreases. In more extreme cases, districts have implemented widespread bans on school suspensions (implemented in LAUSD in 2013). These bans

frequently come with limited supports for instructors and administrators, who must quickly adapt to managing classrooms without the use of suspensions (Lacoe and Steinberg, 2018*b*). Our results suggest that suspension rate declines produce negative effects on average achievement; however, it is possible that these negative effects can be offset with proper resources, training, and disciplinary alternatives for educators. The distribution of such resources should be amplified in schools where misbehavior is common.

Of course, finding alternatives to suspensions is not easy. In 2014, the LAUSD began piloting its first meaningful alternative to suspensions in the form of restorative justice. Unfortunately, early findings from a randomized controlled trial in Pittsburgh Public Schools do not generate optimism (Augustine et al., 2018). While restorative justice reduced disciplinary disparities and improved school climates for teachers, academic outcomes for students remained largely unchanged and even worsened in treated middle schools. These results, alongside the findings documented in this paper, reinforce the notion that there does not currently exist a silver bullet to solving the problem of school discipline. Without a viable alternative, administrators can only acknowledge that any suspension policy will produce winners and losers, and commit to a policy that is consistent with their preferences for efficiency versus equality in education production.

References

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

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Augustine, Catherine H., John Engberg, Geoffrey E. Grimm, Emma Lee, Elaine Lin Wang, Karen Christianson, and Andrea A. Joseph. 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, Andrew, Stephen B. Billings, and David J. Deming. 2019. "The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime." Working Paper.

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

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 76(2): 169–217.

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

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

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

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

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

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature*, 48(2): 424–55.

Deming, David J. 2011. "Better Schools, Less Crime?" *The Quarterly Journal of Economics*, 126(4): 2063–2115.

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

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

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

Jacob, Brian A., and Lars Lefgren. 2003. "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime." *American Economic Review*, 93(5): 1560–1577.

Jaeger, David A., Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and the Impact of Immigration." National Bureau of Economic Research, Inc NBER Working Papers 24285.

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

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

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

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

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

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

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

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

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

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

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

Skiba, Russell J., Choong-Geun Chung, Megan Trachok, Timberly L. Baker, Adam Sheya, and Robin L. Hughes. 2014. "Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion." *American Educational Research Journal*, 51(4): 640–670.

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

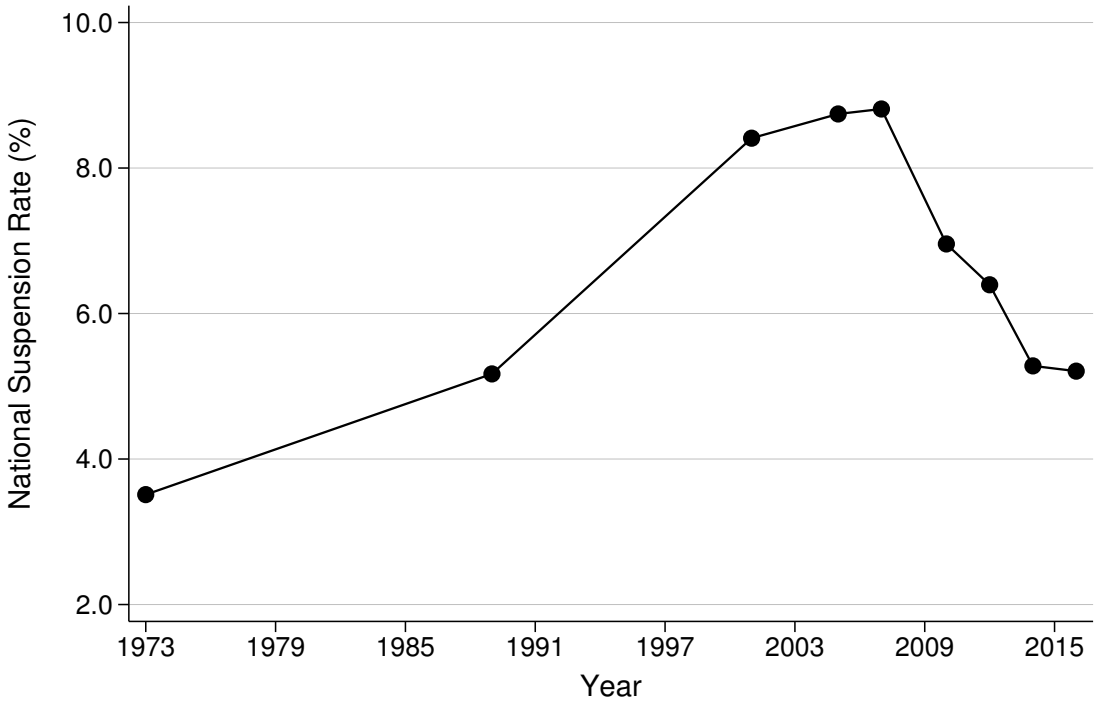
U.S. Department of Education. 2014. "U.S. Departments of Education and Justice Release School Discipline Guidance Package to Enhance School Climate and Improve School Discipline Policies/Practices."

Waddell, G.R. 2012. "Adolescent drug use and the deterrent effect of school-imposed penalties." *Economics of Education Review*, 31(6): 961 – 969.

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

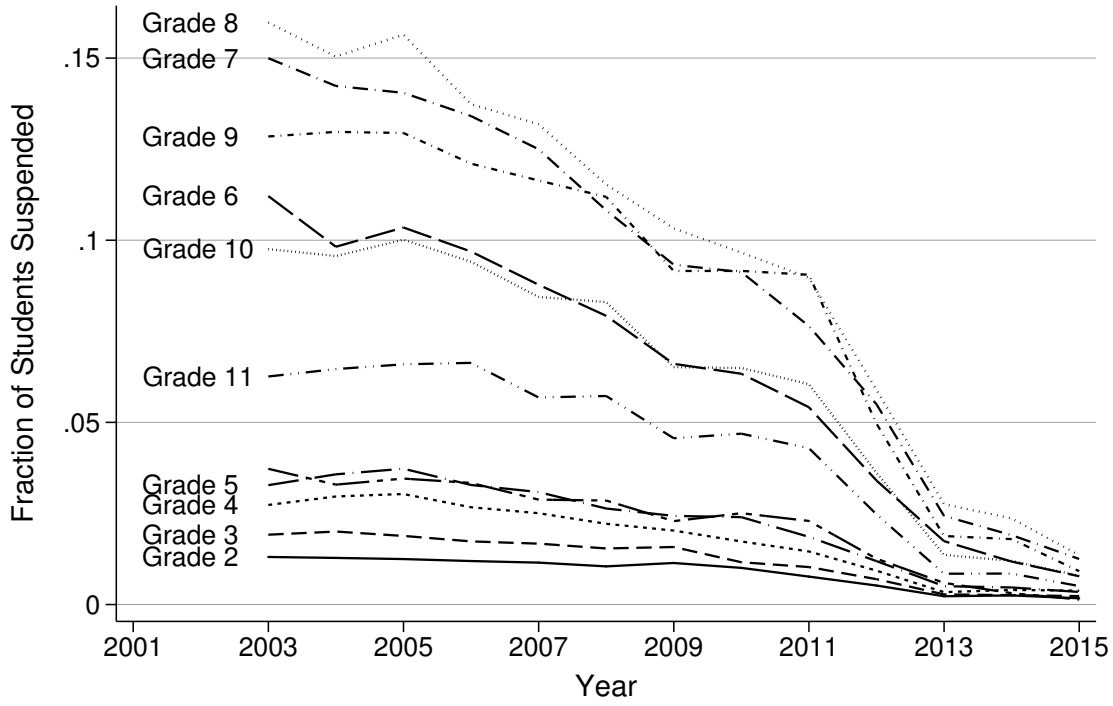
Watanabe, Teresa. May 31, 2014. "L.A. Unified suspension rates fall but some question figures' accuracy." *Los Angeles Times*.

Figure 1: National Suspension Rates, 1972-2015



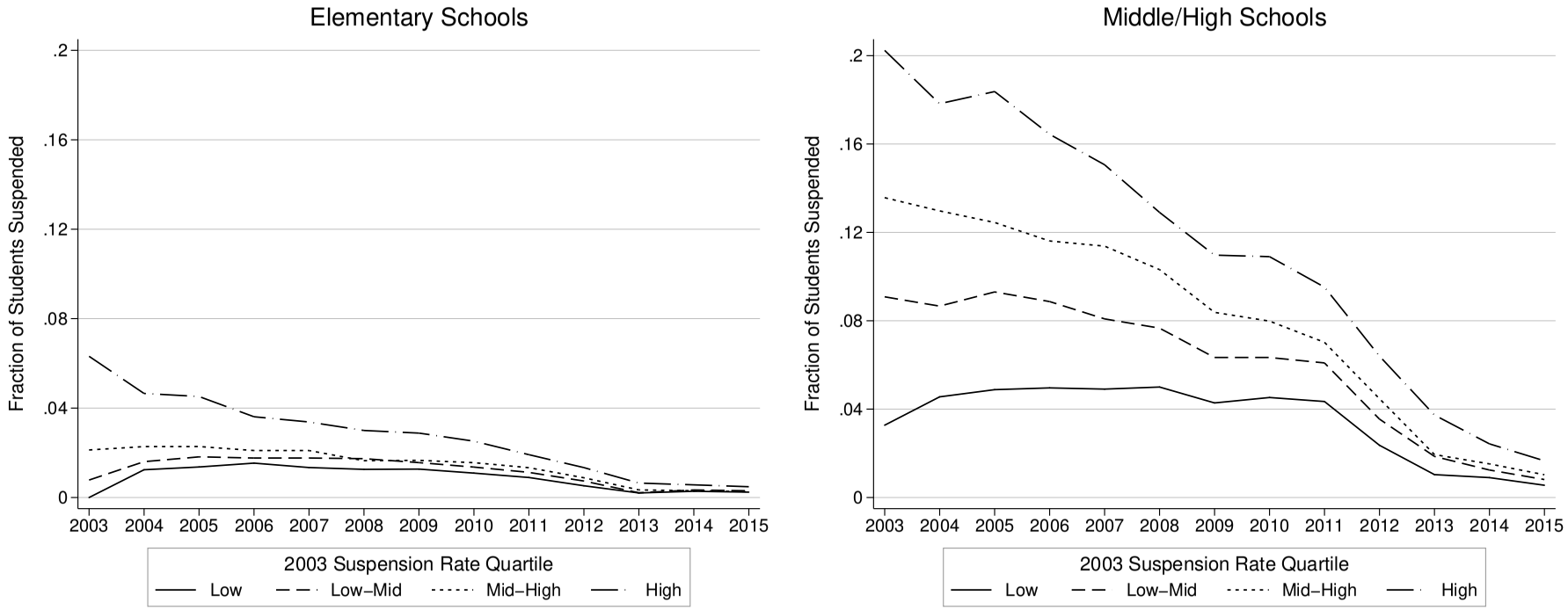
Note: This figure shows national suspension rates using data compiled by the Civil Rights Data Collection by the Office for Civil Rights. Suspension rates are calculated by dividing the number of total suspended students (summed across students with and without disabilities) by the total number of enrolled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

Figure 2: LAUSD Suspension Rates by Grade



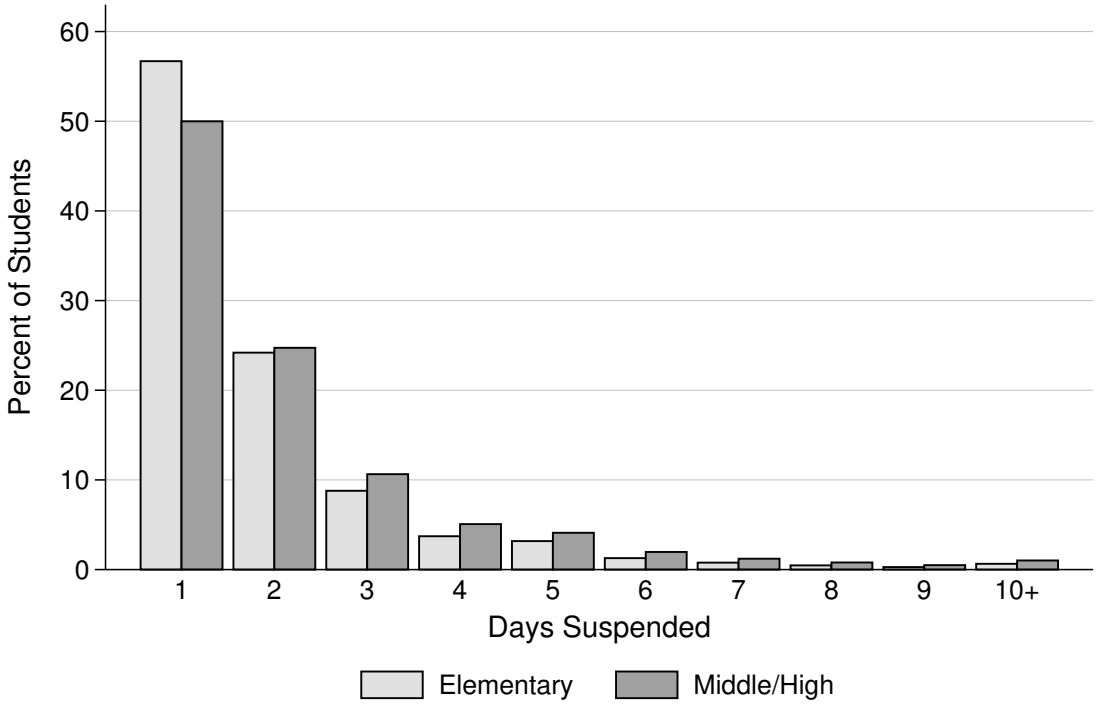
Note: This figure shows district-wide suspension rates for each grade in the LAUSD. Suspension rates are calculated by dividing the number of total suspended students by the total number of enrolled students in each grade.

Figure 3: Variation in School-Grade Suspension Rates by 2003 Suspension Rate Quartile and Grade Category



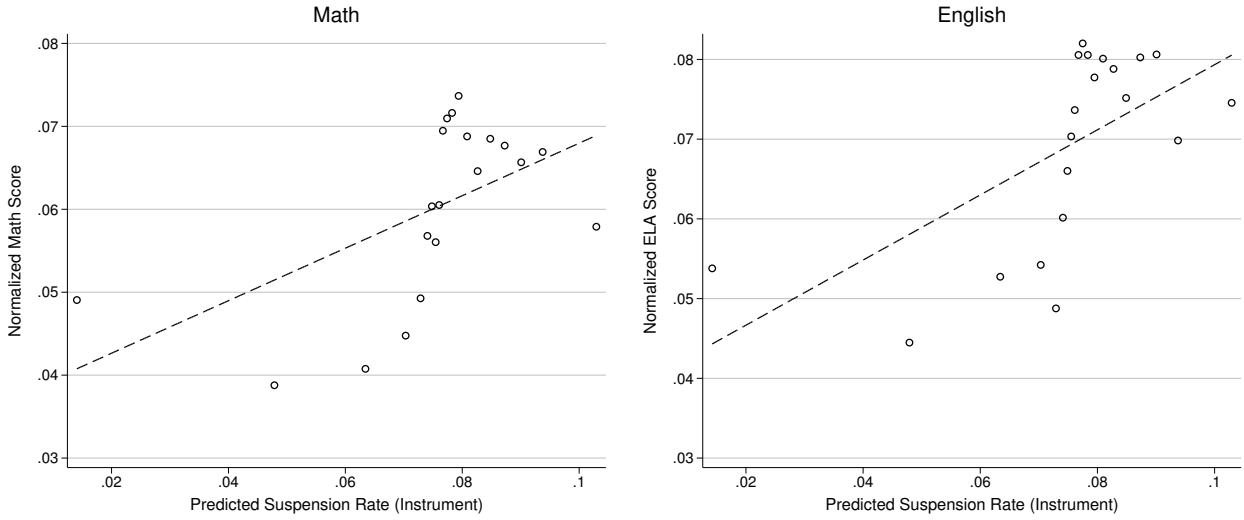
Note: This figure plots the trajectory of suspension rates for elementary and middle/high students in the LAUSD, dividing school-grades into one of four equally-sized quartiles based on initial 2003 suspension rates. Cutoffs for each of the four quartiles are as follows. For elementary school-grades: Q1 [0%], Q2 [0.2%, 1.3%], Q3 [1.3%, 3.2%], Q4[3.2%, 32.1%]. For middle/high school-grades: Q1 [0%, 6.7%], Q2 [6.7%, 11.4%], Q3 [11.4%, 15.5%], Q4[15.7%, 100%]. 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 by Grade Category



Note: This figure shows the distribution of 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. Suspension rates are calculated by dividing the number of total suspended students by the total number of enrolled students in each grade.

Figure 5: Reduced Form Relationship between Test Scores and Instrumented Suspensions



Note: This figure plots average math and English test scores 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. The instrument is calculated based on Equation (6). Math and English test scores have been standardized to have a mean of zero and a standard deviation of one; the axes in the figure reflect a different scaling due to the fact that control variables and fixed effects have been used to residualize test scores.

Table 1: LAUSD Summary Statistics

	Elementary Schools				Middle Schools				High Schools			
	Low Suspensions		High Suspensions		Low Suspensions		High Suspensions		Low Suspensions		High Suspensions	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Standardized Math Scores	0.05	1.02	-0.05	0.98	0.20	1.06	-0.11	0.95	0.04	1.03	0.01	1.00
Standardized English Scores	0.04	1.02	-0.05	0.97	0.19	1.04	-0.11	0.96	0.00	0.98	0.02	1.01
GPA - Total	2.89	0.44	2.86	0.44	2.56	0.93	2.36	0.93	2.10	1.04	2.06	1.03
GPA - Math	2.69	0.70	2.63	0.70	2.22	1.21	2.01	1.20	1.65	1.24	1.59	1.26
GPA - English	2.84	0.52	2.81	0.52	2.48	1.13	2.25	1.15	2.03	1.25	1.98	1.24
Fraction Days Absent	0.04	0.05	0.05	0.05	0.06	0.07	0.07	0.09	0.10	0.13	0.11	0.13
English Language Learner	0.39	0.49	0.38	0.49	0.20	0.40	0.27	0.44	0.18	0.38	0.20	0.40
Held Back	0.01	0.10	0.01	0.11	0.00	0.06	0.00	0.07	0.18	0.39	0.18	0.38
Suspended	0.01	0.10	0.02	0.15	0.06	0.25	0.11	0.31	0.05	0.23	0.09	0.28
Days Suspended (If Suspended)	1.83	1.46	1.93	1.64	2.06	1.79	2.35	2.11	1.98	1.47	2.07	1.63
# Times Suspended (If Suspended)	1.23	0.65	1.32	0.79	1.46	0.95	1.64	1.18	1.31	0.73	1.40	0.85
School Size	445	209	441	209	1,122	886	1,660	686	1,361	1,313	2,422	876
Number of Schools	177		177		43		43		27		27	
Number of Observations	1,021,742		1,014,242		587,931		928,211		442,330		850,259	

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. Roughly 30% of schools in 2015 did not have a suspension rate in 2003. The sample includes all students enrolled in grades 2-11 from 2003 to 2015. However, CST 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. Normalized Math CST Scores						
(Suspension Rate)$_{sgt} \times 10$	-0.139***	-0.011	0.005	0.004	0.003	0.029***
	(0.013)	(0.008)	(0.006)	(0.005)	(0.005)	(0.009)
N	2,906,313	2,906,303	2,255,891	2,251,225	2,250,647	2,250,647
F-Statistic (IV First Stage)						1345
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. Normalized English CST Scores						
(Suspension Rate)$_{sgt} \times 10$	-0.127***	-0.009*	0.009***	0.009***	0.008**	0.038***
	(0.013)	(0.005)	(0.003)	(0.003)	(0.003)	(0.005)
N	2,956,227	2,956,217	2,295,452	2,290,763	2,290,185	2,290,185
F-Statistic (IV First Stage)						1350
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 CST scores. The full OLS regression is estimated as in Equation (5). IV estimates use instrumented suspension rates as calculated in Equation (6). The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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

	OLS					IV
	(1)	(2)	(3)	(4)	(5)	(6)
A. Fraction Days Absent (Non-Suspended)						
(Suspension Rate)$_{sgt} \times 10$	0.016***	-0.001	-0.006***	-0.006***	-0.006***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	3,447,475	3,447,420	2,453,636	2,312,043	2,311,424	2,311,424
Baseline Mean	0.065	0.065	0.065	0.065	0.065	0.065
F-Statistic (IV First Stage)						1137
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. GPA						
(Suspension Rate)$_{sgt} \times 10$	-0.344***	-0.010*	0.068***	0.068***	0.068***	0.163***
	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)	(0.010)
N	3,684,345	3,684,305	2,750,065	2,599,093	2,598,429	2,598,429
	2.554	2.554	2.554	2.554	2.554	2.554
F-Statistic (IV First Stage)						1176
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 fraction of non-suspended days absent and total GPA. Both outcomes have been de-trended by subtracting out the year-grade-specific mean. The full OLS regression is estimated as in Equation (5). IV estimates use instrumented suspension rates as calculated in Equation (6). The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: Effects of School Suspension Rates by Grade Category

	Math			English		
	Elementary	Middle	High	Elementary	Middle	High
$(\text{Suspension Rate})_{sgt} \times 10$	0.134***	0.032***	-0.000	0.152***	0.033***	0.018**
	(0.035)	(0.010)	(0.018)	(0.025)	(0.007)	(0.008)
N	864,112	832,924	553,611	864,452	834,149	591,584
% Students Suspended	1.5%	8.1%	5.7%	1.5%	8.1%	5.7%
F-Statistic (IV First Stage)	305	664	593	305	664	590
	Fraction Days Absent			GPA		
	Elementary	Middle	High	Elementary	Middle	High
$(\text{Suspension Rate})_{sgt} \times 10$	0.001	-0.006***	-0.014***	0.121***	0.164***	0.160***
	(0.002)	(0.001)	(0.002)	(0.028)	(0.012)	(0.017)
N	800,015	837,674	673,735	948,751	928,758	720,920
% Students Suspended	1.5%	8.1%	5.7%	1.5%	8.1%	5.7%
F-Statistic (IV First Stage)	234	570	489	237	599	507
Baseline Mean	0.046	0.059	0.096	2.897	2.476	2.211

Note: This table presents IV estimates for the effect of a 10 percentage point increase in suspension rates on normalized math and English CST scores, separated by grade category. Elementary students include those in grades 2-5, 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 and school-grade fixed effects as in Equation (5). The instrument used for suspension rates is calculated in Equation (6). The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: 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				
$(\text{Suspension Rate})_{sgt} \times 10$	0.021***	-0.015**	-0.015**	-0.020**
	(0.004)	(0.007)	(0.007)	(0.010)
N	103,561	103,560	97,762	97,762
Baseline Mean	0.206	0.206	0.206	0.206
F-Statistic (IV First Stage)				728
B. Teachers with 0-2 Years of Experience				
$(\text{Suspension Rate})_{sgt} \times 10$	-0.012*	-0.056***	-0.056***	-0.088***
	(0.007)	(0.011)	(0.011)	(0.020)
N	33,025	32,944	31,534	31,534
Baseline Mean	0.286	0.286	0.286	0.286
F-Statistic (IV First Stage)				356
C. Teachers with 3+ Years of Experience				
$(\text{Suspension Rate})_{sgt} \times 10$	0.017***	-0.013**	-0.013**	-0.026***
	(0.004)	(0.006)	(0.006)	(0.010)
N	70,536	70,524	66,136	66,136
Baseline Mean	0.151	0.151	0.151	0.151
F-Statistic (IV First Stage)				844
School-Grade Fixed Effects		Yes	Yes	Yes
Lagged School-Grade Test Scores			Yes	Yes

Note: This table presents IV estimates of a 10 percentage point increase in suspension rates on the probability that a teacher leaves his or her school after the current year. Each estimate includes the full set of controls and school-grade fixed effects as in Equation (5). The instrument used for suspension rates is calculated in Equation (6). The sample includes all students enrolled in grades 2-11 from 2006 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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

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

Note: This table presents the results from Table 5, estimated separately for elementary, middle, and high school teachers. Teachers are assigned to a grade level (for school-grade fixed effects) based on the modal number of students across all grades taught. In the case of a tie, the higher grade was assigned. Each estimate includes the full set of controls and school-grade fixed effects as in Equation (5). The instrument used for suspension rates is calculated in Equation (6). The sample includes all students enrolled in grades 2-11 from 2006 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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

	Math				English			
	$t-3$	$t-2$	$t-1$	t	$t-3$	$t-2$	$t-1$	t
(Suspension Rate)$_{sgt} \times 10$	-0.017**	-0.005	0.015	0.029***	-0.008	-0.012*	0.008	0.038***
	(0.009)	(0.009)	(0.010)	(0.009)	(0.007)	(0.006)	(0.007)	(0.005)
N	928,249	1,338,726	1,692,576	2,250,647	930,882	1,349,715	1,719,817	2,290,185
F-Statistic (IV First Stage)	1973	1519	1062	1345	1980	1527	1073	1350
	Fraction Days Absent				GPA			
	$t-3$	$t-2$	$t-1$	t	$t-3$	$t-2$	$t-1$	t
(Suspension Rate)$_{sgt} \times 10$	0.004***	0.004***	0.002	-0.010***	-0.068***	-0.066***	-0.009	0.163***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.012)	(0.018)	(0.020)	(0.010)
N	811,132	1,184,950	1,647,039	2,311,424	937,762	1,362,321	1,888,781	2,598,429
F-Statistic (IV First Stage)	2018	1566	1344	1137	1985	1522	1284	1176
Baseline Mean	0.048	0.050	0.055	0.065	2.710	2.655	2.603	2.554
School-Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Individual Achievement	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Average School Test Scores	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Peer Test Scores	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table shows IV estimates for the effect of a 10% 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 and school-grade fixed effects. 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 (9) provides the structural equation; instrumented suspension rates are calculated based on Equation (6). The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 8: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)
A. Normalized Math CST Scores					
School-Grade Suspension Rate $\times 10$	0.029***	0.038***	0.041***	0.035***	0.032***
	(0.009)	(0.013)	(0.014)	(0.011)	(0.011)
N	2,250,647	2,245,325	2,250,653	1,951,412	1,668,031
F Stat (IV First Stage)	1344.636	462.605	412.177	296.652	266.519
B. Normalized English CST Scores					
School-Grade Suspension Rate $\times 10$	0.038***	0.042***	0.042***	0.040***	0.038***
	(0.005)	(0.008)	(0.009)	(0.008)	(0.007)
N	2,290,185	1,984,764	1,695,987	2,290,191	2,284,833
F Stat (IV First Stage)	1350.274	301.632	270.678	412.257	469.899
School-Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes
Lagged Individual Achievement	Yes	Yes	Yes	Yes	Yes
Lagged Average School Test Scores	Yes	Yes	Yes	Yes	Yes
Lagged Peer Test Scores	Yes	Yes	Yes	Yes	Yes
Modification	None	Years: 2006-13	Years: 2007-13	School	$t-2$ Instrument

Note: This table shows the effect of a 10% increase in suspension rates on academic achievement, 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 the first year of the sample. Column (4) uses school-level suspension rates (instead of school-grade) and adjusts standard errors for clustering at the school level. Column (5) uses an instrument derived from twice-lagged school-grade suspension rates instead of fixed initial 2003 suspension rates. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9: Direct and Indirect Effects of Suspension Rates

	Math			English		
Predicted Misbehavior Tercile:	Low	Medium	High	Low	Medium	High
A. Aggregate Effects						
Aggregate Effect: $(\text{Suspension Rate})_{sgt} \times 10$	0.049***	0.044***	0.046***	0.072***	0.028***	0.018***
	(0.013)	(0.011)	(0.008)	(0.007)	(0.006)	(0.007)
N	652,124	643,705	633,849	657,222	655,920	652,652
Fraction suspended	0.02	0.05	0.12	0.02	0.05	0.12
F Stat (IV First Stage)	1585.515	1341.150	868.861	1587.955	1356.083	877.525
School-Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Individual Achievement	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Average School Test Scores	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Peer Test Scores	Yes	Yes	Yes	Yes	Yes	Yes
	Math			English		
Predicted Misbehavior Tercile:	Low	Medium	High	Low	Medium	High
B. Indirect and Direct Effects						
Indirect Effect: $(\text{Suspension Rate})^{-i}_{sgt} \times 10$	0.052***	0.050***	0.059***	0.076***	0.035***	0.034***
	(0.013)	(0.011)	(0.009)	(0.007)	(0.006)	(0.007)
Direct Effect: Suspended_{isgt}	-0.138***	-0.096***	-0.074***	-0.142***	-0.117***	-0.092***
	(0.007)	(0.005)	(0.004)	(0.006)	(0.004)	(0.003)
N	652,124	643,705	633,849	657,222	655,920	652,652
Fraction suspended	0.02	0.05	0.12	0.02	0.05	0.12
F Stat (IV First Stage)	1601.244	1350.442	883.052	1603.501	1365.430	895.300
School-Grade Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Individual Achievement	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Average School Test Scores	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Peer Test Scores	Yes	Yes	Yes	Yes	Yes	Yes

Note: Panel A shows the effect of a 10% increase in suspension rates on normalized test scores, separately estimated students in three terciles based on their predicted probability of misbehavior. Equation (11) in Section 7 shows how these probabilities are calculated. Terciles are assigned based on a student's rank within a given grade-year. The coefficients in this panel are derived from estimating Equation (5) and instrumenting for suspension rates with Equation (6) separately for each tercile. Panel B shows the effects when suspension rates are decomposed into individual suspensions (direct effects) and peer suspensions (indirect effects), as estimated via Equation (12). Peer suspensions are calculated as the leave-own-out suspension rate of a student's school-grade cohort. The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. F-statistics are not shown but exceed 800 for all estimates. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

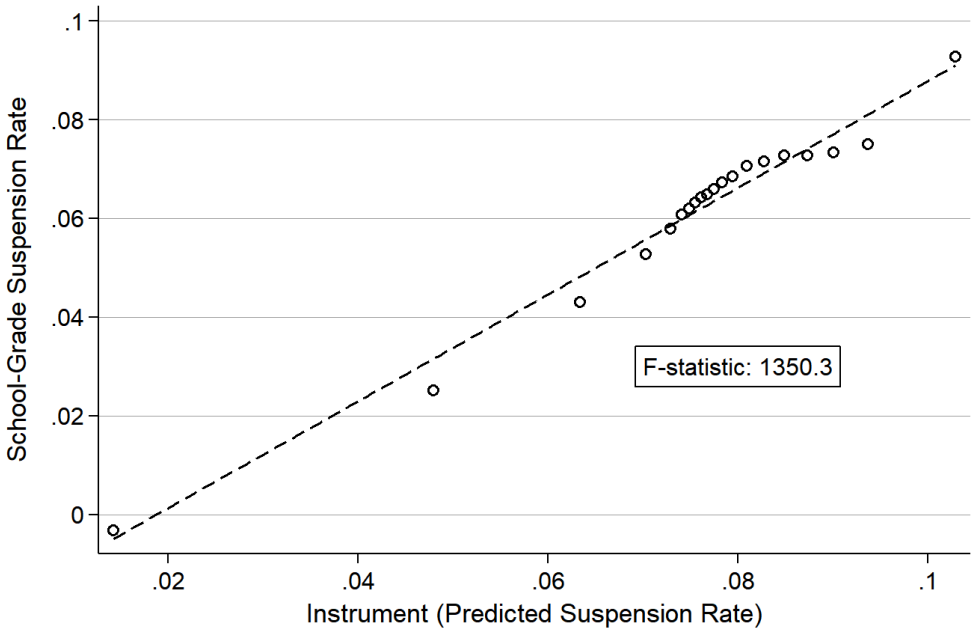
Table 10: Direct and Indirect Effects of Suspension Rates by Grade Category

Subject:	Math			English		
Predicted Suspension Tercile:	Low	Medium	High	Low	Medium	High
A. Elementary School Students						
Aggregate Effect	0.133**	0.204***	0.056*	0.262***	0.173***	0.018
	(0.054)	(0.045)	(0.030)	(0.044)	(0.032)	(0.023)
Indirect Effect	0.143***	0.219***	0.081***	0.271***	0.186***	0.031
	(0.054)	(0.046)	(0.031)	(0.045)	(0.032)	(0.024)
Direct Effect	-0.208***	-0.185***	-0.133***	-0.186***	-0.152***	-0.075***
	(0.017)	(0.012)	(0.008)	(0.015)	(0.010)	(0.007)
N	241,215	241,108	240,955	241,302	241,259	241,024
Fraction suspended	0.01	0.02	0.04	0.01	0.02	0.04
B. Middle School Students						
Aggregate Effect	0.028*	0.040***	0.060***	0.049***	0.036***	0.017**
	(0.015)	(0.013)	(0.009)	(0.009)	(0.007)	(0.008)
Indirect Effect	0.031**	0.046***	0.073***	0.052***	0.043***	0.032***
	(0.015)	(0.013)	(0.010)	(0.009)	(0.007)	(0.009)
Direct Effect	-0.120***	-0.090***	-0.072***	-0.121***	-0.111***	-0.086***
	(0.009)	(0.005)	(0.004)	(0.008)	(0.005)	(0.004)
N	239,282	238,721	237,524	239,508	239,129	237,944
Fraction suspended	0.03	0.07	0.19	0.03	0.07	0.19
C. High School Students						
Aggregate Effect	0.045**	0.005	0.029*	0.064***	-0.017*	0.011
	(0.022)	(0.022)	(0.016)	(0.011)	(0.010)	(0.012)
Indirect Effect	0.048**	0.010	0.039**	0.068***	-0.010	0.029**
	(0.022)	(0.022)	(0.017)	(0.011)	(0.010)	(0.012)
Direct Effect	-0.133***	-0.074***	-0.054***	-0.161***	-0.115***	-0.105***
	(0.016)	(0.011)	(0.007)	(0.011)	(0.006)	(0.007)
N	171,627	163,876	155,370	176,412	175,532	173,684
Fraction suspended	0.02	0.05	0.14	0.02	0.05	0.14

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-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

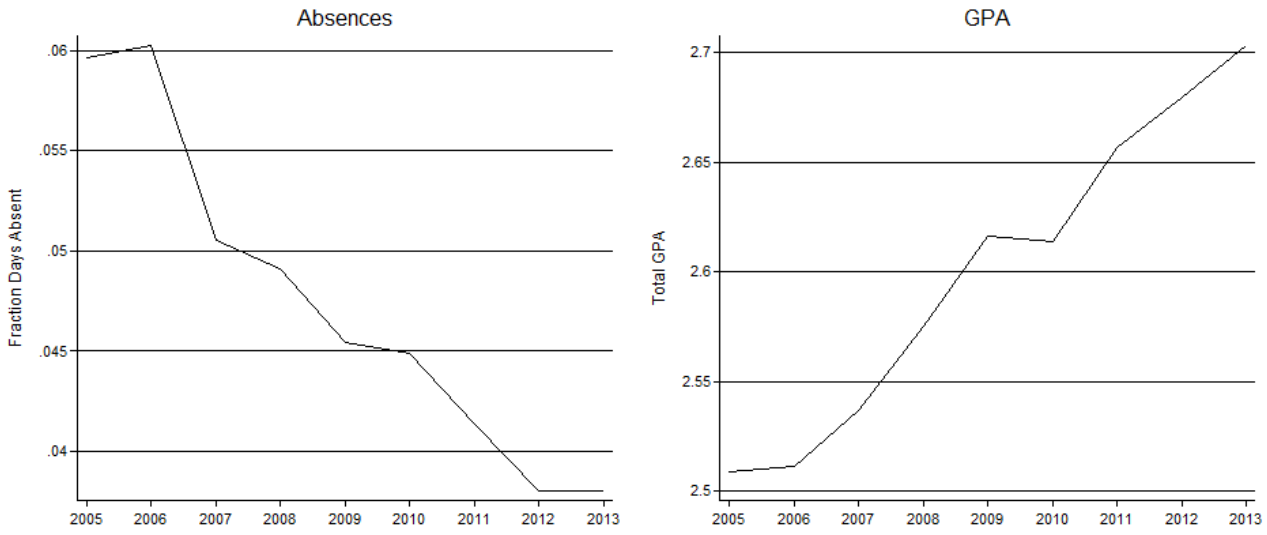
Appendix A: Appendix Tables and Figures

Figure A1: IV First Stage



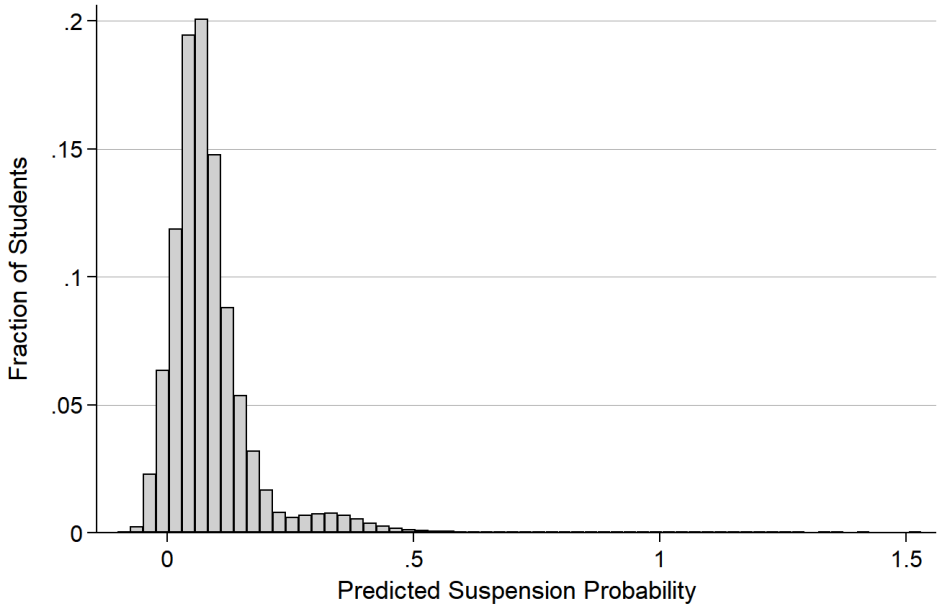
Note: This figure shows a binned scatterplot, with instrumented suspension rates on the x-axis and the actual suspension rate on the y-axis. Both axes are residualized with respect to lagged individual, school, and peer achievement as well as school-grade fixed effects. The F-stat presented is for the set of observations in our analysis sample with a valid entry for ELA test scores, which can be found in column (6) of Table 2.

Figure A2: Time Trends in Absences and GPA



Note: This figure shows the average fraction of days absent and GPA district-wide in LAUSD. Days absent do not include the number of days suspended. The sample includes students attending schools with a valid suspension rate in 2003 and with nonmissing math and ELA test scores. The figures suggest that these outcomes may need to be de-trended to avoid conflating the effect of suspension rates with aggregate time trends.

Figure A3: Distribution of Predicted Misbehavior Propensity



Note: This figure shows the distribution of predicted suspension propensities, calculated via Equation (11) using 2004 data to train the model. Predicted values are not strictly between zero and one due to the fact that predictions are generated from a linear probability model.

Table A1: Reduced Form and First Stage Regressions

	(1) First Stage	(2) RF - Math	(3) RF - English
School-Grade Suspension Rate $\times 10$	1.083***	0.032***	0.041***
	(0.030)	(0.010)	(0.006)
Lagged Math	0.004***	0.523***	0.142***
	(0.001)	(0.004)	(0.002)
Lagged English	0.004***	0.214***	0.652***
	(0.001)	(0.003)	(0.002)
Lagged GPA	-0.020***	0.139***	0.142***
	(0.002)	(0.004)	(0.003)
Lagged School Math	0.040	0.312***	0.060***
	(0.054)	(0.020)	(0.009)
Lagged School English	-0.088	-0.130***	0.048***
	(0.059)	(0.021)	(0.013)
Lagged Peer Math	-0.044	-0.192***	-0.140***
	(0.055)	(0.018)	(0.011)
Lagged Peer English	-0.143**	0.024	-0.037***
	(0.060)	(0.019)	(0.013)
N	2,482,718	2,250,647	2,290,185
Adjusted R-squared	0.721	0.622	0.715

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

Table A2: Correlates of Instrumented Suspension Rates

	(1)	(2)	(3)	(4)
	Grade Size	ELL	Left LAUSD	Switched Schools
Instrumented Suspensions	4.800	-0.019***	-0.001	-0.001
	(8.875)	(0.003)	(0.008)	(0.002)
N	14,050	14,050	14,050	14,050
Mean	186.495	0.319	0.247	0.065
School-Grade Fixed Effects	Yes	Yes	Yes	Yes
Lagged School Controls	Yes	Yes	Yes	Yes

Note: This table shows the relationship between instrumented suspension rates and four potential confounders. All coefficients are estimated via OLS using Equation (10). Grade size represents the number of students in a given school-grade-year cell. ELL represents an indicator for English Language Learner students. Left LAUSD is an indicator for whether a student left the school district at the end of the year (and was not in a graduating year). Switched schools is an indicator for whether a student changed schools within LAUSD in the following year (and was not in the terminal grade of a school). The sample includes students enrolled in grades 2-11 from 2005 to 2013. The sample includes all students enrolled in grades 2-11 from 2005 to 2013 whose school and grade had a non-missing suspension rate as of 2003. Standard errors are adjusted for clustering at the school-grade level and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A3: Coefficients on Predicted 2004 Suspensions

	OLS
Dependent Variable: Suspended in 2004	
Lagged Math	-0.000
	(0.001)
Lagged English	-0.013***
	(0.001)
Lagged GPA	-0.062***
	(0.001)
Fraction Days Absent Last Year	-0.001
	(0.013)
Held Back	-0.030***
	(0.003)
Suspended Last Year	0.134***
	(0.003)
Days Suspended Last Year	0.045***
	(0.001)
Lagged School Math	0.024***
	(0.006)
Lagged School English	-0.036***
	(0.009)
Lagged Peer Math	-0.042***
	(0.008)
Lagged Peer English	0.053***
	(0.010)
N	157,682
Adjusted R^2	0.121

Note: This table shows the estimates on coefficients used to predict student suspensions as of 2004, which are calculated via Equation (11). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.