**Banning Math Progress: The academic impact of California’s suspension bans**

*A growing number of states and school districts have banned suspensions – at least in some grades – for all but the most serious offenses. While the literature generally expects bans to improve academics, this paper complements and extends new studies that find unintended negative consequences of discipline reform. We exploit the exogenous introduction of bans to the California policy landscape during 2013-2015 to compare growth before (2007-2013) and after (2016-2018) Los Angeles, San Francisco, Oakland, and Pasadena adopted bans. We find a negative but sometimes non-significant impact on reading growth in grades four through eight as well as a consistently significant, negative impact on math growth (-0.09 SD) in grades four through seven. The impact occurred each of the three years after districts adopted bans. Cumulatively, the impact was equivalent to a decline from 50th percentile before a ban to the 39th percentile three years later.*

Schools are in the midst of a revolution in their utilization of suspensions. In the 1990s, most researchers and practitioners supported zero tolerance policies: mandatory administration of harsh punishment – often suspensions – for relatively minor rule violations. Following the logic of the broken window theory from criminology, proponents believed zero tolerance policies would improve the behavior of all students (Wilson & Kelling, 1982). School staff suspended more students, with the national suspension rate doubling from the 1980s to the early 2000s (Losen, 2011). However, the evidence shows that zero tolerance did not lead to improved behavior and disproportionately harmed African American students (Hoffman 2014, Curran 2016). Researchers also find that suspensions harm students: they grow less academically and are more likely to drop out and go to prison than similar students who do not get suspended (Arcia, 2006; Lacoe & Steinberg, 2018; Rumberger & Losen, 2017; Chu & Ready, 2018; Hwang, 2018). Even non-suspended students are harmed academically when more suspensions occur around them (Perry & Morris, 2014; Lacoe & Steinberg, 2018). Instead of a pillar of order and learning, suspensions are an obstacle to creating high-quality schools.

Most troubling, school staff disproportionately issue suspensions to disadvantaged subgroups of students: African American, Latino, Native American, students with disabilities, and lesbian, gay, bisexual, and transgender students (Wallace et al. 2008). While many factors account for these differences, experiments demonstrate that a portion is caused by bias (Okonofua & Eberhardt, 2015; Gilliam et al., 2016). Researchers also point to mechanisms such as racial threat to explain why schools serving a relatively large proportion of racial minorities tend to have higher rates of exclusionary discipline (Payne & Welch, 2010; Welch & Payne, 2012; Curran, 2017). Regardless of the exact set of causes, schools will not be able to close the achievement gap if they systematically exclude already disadvantaged students from the learning environment (Gregory et al., 2010).

One policy that attempts to address these discipline disparities is a suspension ban. A ban dictates that school staff are not allowed to suspend students (sometimes in specified grades) for relatively minor or subjective infractions. Since removing the subjectivity from suspensions should also remove the bias, the hope is that bans would significantly reduce the subgroup disparities in suspension rates.[[1]](#footnote-1) Even if disparities persist, bans should at least lessen the scale of the problem by reducing suspension rates for all subgroups (Hashim et al., 2018). Four California districts – Los Angeles, San Francisco, Oakland, and Pasadena – adopted bans between 2013 and 2015. Figures 1 and 2 show the average English and Math achievement for those districts before, during, and after that ban uptake period. In both subjects, districts with a ban declined slightly during the uptake period and experienced an even steeper decline at least the first year after that. This foreshadows our results, and does not bode well for the relationship between bans and academic growth.

*Figure 1: English Achievement for California Districts, 2007-2018*



*Figure 2: Math Achievement for California Districts, 2007-2018*



*Note: Achievement is averaged across district-grades 3-8 for English and 3-7 for Math in each year; achievement is standardized using the statewide average and the distribution of scores for all California districts; vertical lines indicate the ban uptake period (2013 to 2015) during which four California districts adopted bans.*

This paper is the first to test the academic impact of California’s suspension bans. We conduct fixed effect analyses that exploit the exogenous timing of the ban uptake period in California. The results for reading achievement are negative but not always significant; the results for math are consistently significant and negative. This negative math impact is significant even several years after the implementation of a ban. The cumulative impact on math is substantial, equivalent to a drop from the 50th percentile before a ban to the 39th percentile of all districts three years later.

This is not an argument in support of suspensions or other exclusionary discipline. In fact, we find that math performance dropped the most in schools that continued to issue the banned suspensions. We hypothesize that the negative effect we observe is caused by changes in the relationships between students, teachers, and district leaders. Schools should make more efforts to reduce suspensions and their disproportionalities. However, the evidence reveals a surprising trade-off: bans reduce suspension rates but harm math performance. Policymakers uncomfortable with this tradeoff will want to pursue other approaches to discipline reform.

The paper proceeds in four sections. The first explains the background necessary to understand suspension bans, how they spread in California, and what the literature expects their impact on academics to be. The next section details the data and analyses we utilize to evaluate the impact of bans. The third section examines the results, and the conclusion discusses the limitations and implications of those results.

**Background**

*The most problematic suspensions*

Data suggests that the reasons for suspensions fall along a spectrum from serious to superfluous. At the first extreme, some suspensions are a response to an extreme threat to safety in schools (Cornell and Mayers, 2010). While these suspensions may seem necessary in the moment, they often are preventable. We can predict future suspensions with simple information about previous student behavior (Raffaele Mendez, 2003). Additionally, teachers can utilize strategies that both make students less likely to misbehave and more likely to de-escalate after doing something wrong (McFarland, 2001). Therefore, prevention programs and staff training can prevent many students from getting to the point of threatening school safety; this is true even for the case of bullying (Orpinas and Horne, 2006; Swearer et al. 2009). Some of the more common approaches to schoolwide behavior management are the Behavior Education Program (Crone et al., 2010), function-based interventions (e.g. Liaupsin et al., 2006), and school-wide positive behavioral interventions and supports (Luiselli et al., 2005; Putnam et al., 2006; Skiba and Sprague, 2008). Freiberg and Lapointe (2006) find dozens of programs that have evidence of successfully improving student behavior.

At the other extreme, some suspensions are the consequence of relatively minor or subjective infractions. While these infractions may demand a consequence, it is more difficult to justify issuing suspensions for them. In California, for example, *Education Code* Section 48900(k) allows schools to suspend students if they have “[d]isrupted school activities or otherwise willfully defied the valid authority of supervisors, teachers, administrators, school officials, or other school personnel engaged in the performance of their duties.” Defiance is the most subjective basis on which California schools can issue suspensions; other rationales concern various forms of theft, violence, and possession of illegal substances. Many schools have made frequent use of this flexibility to remove defiant students from school. When California began collecting suspension data with these categories in 2011-12, it reported over 200,000 based on defiance, constituting 47% of all suspensions.[[2]](#footnote-2)

Bans directly target the most problematic suspensions. Bans do so by legally dictating that school staff are no longer allowed to issue suspensions on the basis of particularly minor or subjective rationales. This compels school staff to find other ways of addressing such infractions. Importantly, a ban is easy for policymakers to implement. Unlike the prevention and training programs mentioned above, a ban does not require any funding or time. The statewide policy team of a large non-profit organization summarized suspension reform as “an issue that framed correctly could have legs in Sacramento,” the California capitol, because it was a widespread problem that could be remedied through relatively small changes to state education law (Martinez et al., 2013, 7).[[3]](#footnote-3)

*The spread of suspension bans*

In 2012, California was one signature away from adopting a statewide suspension ban. A sudden push for legislative action on discipline reform resulted in seven bills passing California’s state house and senate (Martinez et al., 2013, p. 18). Governor Brown signed five of those seven bills, but vetoed the ban on defiance suspensions. Prior to 2012, no large districts or states had adopted a ban; between 2013 and 2015, four California districts implemented bans for grades Kindergarten through twelve.[[4]](#footnote-4) Los Angeles Unified School District was the first, in 2013. Three districts soon followed: San Francisco Unified in 2014 as well as Pasadena Unified and Oakland Unified in 2015 (Frey, 2015). We find no reports that any other California school districts adopted a ban between 2015 and 2018.

Of course, the four districts that adopted a ban are not a random or representative sample of California districts. Their leaders and interest groups decided to prioritize this particular, new discipline reform while other districts did not. A difference-in-differences analysis therefore would be inappropriate, because we expect districts who implemented a ban to be different from other districts in ways we cannot account for in a quantitative analysis.[[5]](#footnote-5)

However, the history detailed above shows that the timing of bans was not random. After California nearly adopted a statewide ban, district leaders suddenly perceived a ban as a viable policy option. During the next two years, districts decided to adopt a ban or not. This exogenous shock means that districts essentially did not control the timing of suspension bans; bans either occurred between 2013 and 2015, or they did not happen. Therefore, it is appropriate to conduct a fixed effect analysis that compares districts to themselves over time, before and after the 2013-2015 period of ban implementation. Such an analysis allows us to estimate the impact bans had on those four districts.

*The impact of suspension bans on discipline*

Evaluating bans is complicated by the fact that bans are part of a set of simultaneous changes. The four school districts that implemented suspension bans took different actions to address disruptive students. Los Angeles Unified began a seven year process of training staff at its schools about restorative justice practices.[[6]](#footnote-6) San Francisco Unified began a three-year, evidence-based training program for teachers that focused on de-escalation (Frey, 2014). Pasadena Unified did not report a major shift in training, while Oakland Unified spent over two million dollars to expand restorative justice programs (Frey, 2015). An analysis of those four bans is also in part an analysis of those additional disciplinary changes.

A second complication is that bans are implemented with different degrees of fidelity. In all four districts, state records show that some staff continued to issue suspensions that were officially banned. Additionally, there are news reports of schools giving off-the-record suspensions in an effort to appear compliant while not actually implementing the new policy (Watanabe, 2014). It is impossible to know exactly how many informal suspensions these districts gave. In theory, school staff could simply shift all banned suspensions from official and recorded to unofficial, off-the-record suspensions.

Nevertheless, all bans by definition share two fundamental features. First, bans alter the official authority and autonomy of school staff. By passing a ban, policymakers tell teachers and administrators they are not allowed to suspend students for relatively subjective reasons. Teachers and administrators can no longer use their judgement to decide to remove a student from school. A previously available tool for managing behavior is no longer allowed. This is true even if staff did not previously give the now-banned suspensions, and it is true even if staff ignore the policy and continue to give the banned suspensions. In all cases, staff know that policymakers have adjusted their authority to manage discipline in their schools.

Second, bans change the nature of student-teacher relationship. By passing a ban, policymakers tell students that they are not allowed to be suspended for relatively subjective reasons. This is true even if students did not previously receive the now-banned suspensions, and it is true even if staff ignore the policy and continue to give the banned suspensions. Students – or at least their parents and other advocates – know that policymakers have limited the authority of school staff to issue suspensions.

In summary, bans demand new relations between students, school staff, and district leaders. The fact that some schools refuse to comply with bans is evidence that some school staff oppose those new relationships. It is hard to predict exactly how this may impact the behavior of students and school staff. While we lack the evidence or space to determine the exact mechanisms by which bans impact academics, we posit that the mechanisms likely are related to this demand for new relationships among the main actors involved in public education.

*The impact of suspension bans on academics*

The majority of the literature expects a suspension ban to increase academic growth. While suspension rates are negatively correlated with academic growth (Rausch & Skiba, 2004; Losen et al., 2015), that in itself does not prove that suspensions cause academic growth to decline. Two lines of research support such a causal connection. One focuses on suspended students, while the other focuses on their non-suspended peers.

First, students who would have been suspended without the suspension ban should now have higher academic growth. While a relatively small portion of students are suspended, some research found that the negative impact on academic growth was enormous (Arcia, 2006). To conclude this, Arcia (2006) matches students who received suspensions with students with similar observed characteristics – such as academics, demographics, and prior behavior issues – who did not receive suspensions. However, research using more rigorous quasi-experimental methods such as panel data with student fixed effects finds that suspensions have a much more modest impact on suspended students (Lacoe & Steinberg, 2018; Hwang, 2018; Chu & Ready, 2018).

Second, the literature argues that even students who never would have been suspended experience higher growth because they suffer less from the negative spillover effects of suspensions. This second mechanism finds support in Perry and Morris’s work in *American Sociological Review* (2014). They use multilevel fixed-effect regressions to find that, controlling for the same non-suspended student over time, an increase in the number of classroom suspensions causes a decrease in academic performance. As in the other line of research, however, recent evidence points to much smaller effects on non-suspended students. Lacoe and Steinberg (2018) use the timing of discipline reform in Philadelphia to conduct an instrumental variable analysis on the spillover effects of suspensions on non-suspended students. Only suspensions for serious misconduct – not the types of suspensions targeted by a suspension ban – negatively impact the achievement of non-suspended students, and the effect is very small (Lacoe & Steinberg, 2018).

The literature almost exclusively debates the magnitude of impact, not the direction. The more suspensions harm academic growth for suspended and non-suspended students, the more we would expect a suspension ban – by decreasing suspension rates – to improve academic growth. Regardless of the true magnitude of the impact, both lines of research lead us to expect that a ban that reduces suspensions should cause academic growth to improve for both suspended and non-suspended students.

Nevertheless, recent analyses provide reasons to think bans and the discipline changes that follow may harm academic growth. An analysis of a suspension ban in Philadelphia found no change in academic performance for previously suspended students and lower performance for non-suspended students (Steinberg & Lacoe, 2018). Interestingly, that negative effect is driven by schools that did not comply with the new discipline policy. One interpretation is that non-compliant Philadelphia schools experienced negative unintended consequences of a ban while not experiencing the positive impact of a decreased suspension rates. Our analysis of California’s bans find a similar pattern, with more harm at schools that do not comply with the ban. However, we find that even schools that comply with the ban experience the negative unintended consequence of lower math growth.[[7]](#footnote-7)

After its ban, Los Angeles Unified implemented district-wide restorative justice training. Restorative justice attempts to reduce exclusionary discipline by improving relationships among students and staff. RAND conducted a randomized controlled trial of “SaferSanerSchools” Whole-School Change program from the International Institute for Restorative Practices (Augustine et al., 2018). At the end of the two year implementation, researchers found that the 22 treatment schools had significantly lower math performance in grades three through eight: negative 0.068 standard deviations, with a p-value less than 0.05 (Augustine et al, 2018, 56). This means that even if the ban in Los Angeles did not harm math achievement, the subsequent restorative justice training may have had that effect.

**Data and Analyses**

We create two datasets – one with district data and the other school data – and run parallel analyses on each. The advantages of each dataset complement the weaknesses of the other. The district dataset provides comprehensive average scores for all students in each district. However, there are concerns that changes in cohort size or the percent of students tested from year to year may explain an unknown amount of the changes in academics that we see. To directly address those concerns, the school dataset only includes observations where:

* Current and prior scores each reflect a participation rate of at least 95%
* There is a less than 5% change in cohort size from the prior year to the current year

Those inclusion rules greatly increase the probability that the sample contains current and prior test scores from consistent sets of students. The average grade in the resulting Math sample had 110 students with valid scores, representing 98.5% of enrolled students and only a 0.2% change from the number of students with valid scores in the prior year. This significantly alleviates the concern that schools in the sample used suspensions to prevent low performing students from testing. Of course, the tradeoff is that the dataset only contains scores for a portion schools.

Our dataset is based entirely on public files made available by the California Department of Education (CDE).[[8]](#footnote-8) The CDE reports grade-level assessment data and the number of valid scores in various demographics for each school starting in 2006-07, with the exception of 2014.[[9]](#footnote-9) We have reading scores for grades two through eleven until 2012-13 and in grades three through eight and eleven after that. We have math scores for grades two through seven until 2012-13 and in grades three through eight and eleven after that.[[10]](#footnote-10) Since analyses require both current and prior scores (explained in more detail below), the reading sample is able to include grades three through eleven from 2006-07 to 2012-13 as well as grades four through eight from 2014-15 to 2017-18. That is 45,456 grade-level observations clustered within 884 districts; using our inclusion rules with school data, we obtain 66,051 grade-level observations clustered within 8,031 schools. The math sample is able to include grades three through seven from 2006-07 to 2012-13 as well as grades four through eight from 2014-15 to 2017-18. That is 32,234 grade-level observations clustered within 779 districts; our school dataset obtains 59,422 grade-level observations within 7,018 schools.

The dependent variables are normalized reading and math performance for all students. Scores are normalized by subject separately for each grade and year. For the mean, we use the statewide average. To calculate the standard deviation, we use all available district data. Reassuringly, our scores have a 0.97 Pearson-R correlation with the nationally normed Cohort Standardized scale from the Stanford Education Data Archive, which is currently available at the district level for 2008 to 2013. The normalization process alleviates the concern that statewide trends over time or between grades may impact the analysis. We also include time (i.e. the first year of data, 2006-07, equals 1) and time squared as independent variables to remove any remaining time trends.

An important independent variable is the prior score – in the respective subject – for that cohort. The prior score is the normalized average score for the same cohort of students in the previous year. For example, if the current reading score is from fifth grade at District ABC, then the prior reading score is from fourth grade the previous year at District ABC. The one exception is caused by California not reporting reading or math scores in 2013-14, the year the state piloted a new assessment system. For the 2014-15 school year, if the current reading score is from fifth grade at District ABC, then the prior reading score is from third grade in 2012-13 at School ABC. Reassuringly, prior score and current score are highly correlated, and that relationship is just as strong for the 2014-15 school year (Pearson’s R = 0.9) as in other years. For methodological reasons detailed below, we exclude the 2014-15 from most of the analyses.

We follow Perry and Morris (2014) by including demographics at two levels: grade and unit (i.e. district or school). At the grade level, we include the portion of students with valid scores who are low-income, have disabilities, and are English Learners. At the school or district level, we include the three previous subgroups as well as the largest racial/ethnic groups: African American, Asian, Latino, White (the excluded category), and Other. Table 1 reports the descriptive statistics for the four main samples, used in Tables three through five.

**Table 1**

Averages for Independent Variables included in Main Analyses

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | English | | Math | |
|  | District Data | School Data | District Data | School Data |
| Observations | 45,456 | 66,051 | 32,234 | 59,422 |
| Time (2007 = 1) | 5.5 | 6.1 | 6.1 | 6.6 |
| Time squared | 41.3 | 53.0 | 50.1 | 59.5 |
| Current Achievement | 0.03 | 0.23 | 0.01 | 0.19 |
| Prior Achievement | 0.04 | 0.24 | 0.01 | 0.18 |
| % Participation Rate | 95.5% | 98.4% | 95.4% | 98.5% |
| # Students Tested | 723 | 144 | 653 | 110 |
| % Change in Cohort Students Tested | 1.7% | -0.3% | 0.2% | -0.2% |
| Grade-Level  % Low Income | 53.5% | 53.6% | 54.9% | 56.1% |
| % Students with Disabilities | 7.9% | 8.7% | 8.3% | 9.1% |
| % English Learner | 18.3% | 20.6% | 20.5% | 22.6% |
| Unit-Level  % African American | 4.0% | 5.2% | 3.6% | 5.1% |
| % Asian | 6.4% | 10.2% | 6.4% | 9.9% |
| % Latino | 44.9% | 48.2% | 45.7% | 50.3% |
| % White | 37.9% | 29.4% | 37.5% | 27.7% |
| % Other | 6.7% | 7.0% | 6.8% | 7.0% |
| % Low Income | 53.8% | 53.7% | 55.1% | 56.2% |
| % Students with Disabilities | 7.9% | 8.7% | 8.3% | 8.9% |
| % English Learner | 19.6% | 22.1% | 21.9% | 24.0% |
| # Observations when Binary = 1  Post-ban period (2015-16 to 2017-18) | 60 | 2,501 | 60 | 2,681 |
| * only in Los Angeles | 15 | 2,037 | 15 | 2,151 |
| * only in San Francisco | 15 | 248 | 15 | 261 |
| * only in Oakland | 15 | 159 | 15 | 210 |
| * only in Pasadena | 15 | 57 | 15 | 59 |
| * only in 2015-16 | 20 | 775 | 20 | 844 |
| * only in 2016-17 | 20 | 811 | 20 | 889 |
| * only in 2017-18 | 20 | 915 | 20 | 948 |

Source: Ban-related variables are authors’ calculations; all other variables are from the California Department of Education STAR research files: <https://star.cde.ca.gov/starresearchfiles.asp>.

The key independent variables concern the presence of a suspension ban. All districts adopted suspension bans between 2013 and 2015, a timeframe we can call the ban uptake period. We want to compare growth for the four adopting districts before and after that uptake period. The uptake period coincides with California’s transition from one assessment system that ended in 2012-13, to an unreported field test in 2013-14, to a new assessment in 2014-15. In order to compare growth before and after the update period, we only need to exclude one year from the analysis: 2014-15. Data from that year attempts to estimate growth from the former assessment system to the new system two years later. There are many factors besides suspension bans that could influence that change, so including that year in the analysis would open our results to the critique that we are in fact measuring some other factor related to the transition to the new assessments. It is fortuitous that we have methodological reasons to exclude that year from the analysis.

Therefore, the most straightforward measure for the ban is a binary variable that is one in years 2015-16 to 2017-18 for the four implementing districts and zero for all other observations. Another approach is to divide that one binary variable into three variables, one for each year: 2015-16, 2016-17, and 2017-18. Alternatively, we can divide that one binary variable into four variables, one for each implementing district. Since the analyses are fixed effect regressions, the independent variables compare academic performance before (2006-07 to 2012-13) and after (2015-16 to 2017-18) the ban uptake period for each grade in each district.

We also create a district-level dataset specifically designed to estimate the impact of bans on suspension rates. California provides annual total and unduplicated (i.e. each student counted only once) suspension rates starting in 2011-12. The section below explains how we specify that model as well as the models that estimate academic impact.

*The Models*

The analyses aim to estimate the effect of implementing suspension ban in four districts after an exogenous uptake period. Given the chronological structure of the data, we use the *xtreg* time series regression program in Stata. We cluster the standard errors by district because we anticipate that districts may have similar patterns across grades. Using random effects – which would be a difference-in-differences analysis – would force us to assume that the districts that implemented bans are similar to districts that did not implement bans.[[11]](#footnote-11) Since we do not believe that assumption is true, we utilize fixed effects. Using fixed effects allows us to use each grade level in each district as its own treatment (after the ban uptake period) and control (before the ban uptake period). This approach does not require us to assume that the four uptake districts are randomly selected or representative of other California districts.

The first set of models estimate the impact bans had on suspension rates. Models 1 and 2 use total suspension rate as the dependent variable; Models 3 and 4 use unduplicated suspension rate as the dependent variable. Models 1 and 3 use one binary variable to measure the impact of a ban. In contrast, Models 2 and 4 include separate variables for each of the four districts. All models remove the data from the ban uptake period – 2013-14 and 2014-15 – and therefore compare suspension rates from before and after that period. Additionally, all models include time and time squared in order to control for statewide trends in suspension rates. This is important because suspension rates have fallen every year since the state began reporting the data publicly in 2011-12. Additionally, all analyses exclude very tiny districts with less than 30 students.

The second set of models look at the impact of bans on academics. The first two models look at English and the second two look at Math. Odd-numbered models utilize the district dataset, while even-numbered models utilize the school dataset. All models include the same variables and exclude observations with fewer than 30 current and prior scores. The dependent variable is the average normalized score (Denoted “Y” in Equation 1 below). The key independent variable is a single binary (B) that indicates whether the district had implemented a ban by that time. Additional analyses divide that into multiple variables in order to analyze the results by district, year, or compliance. At the grade level (G), independent variables include the prior score for that cohort as well as the percent of students with valid scores who are low-income, English Learner, and students with disabilities. At the unit – district or school – level (U), we include the variables mentioned above and the proportion of students in five racial/ethnic categories. Lastly, all models time and time squared (T) in order to account for statewide academic trends. Of course, models include a constant (C) as well as standard errors (E) are clustered by district or school. This is formalized in Equation 1 below, where we estimate academic performance for each grade (g) in each unit (district or school, u) in each year (t):

Equation 1: Ygut = But + Ggut + Uut + Tt + C + Egdt

By controlling for prior score at the grade level (i.e. last year’s 4th grade scores for this year’s 5th graders), our model approximates value added growth models.[[12]](#footnote-12) The inclusion of independent variable B allows us to estimate the impact bans had on academic growth across all grades in those four districts.

Subsequent models break the one independent variable reflecting whether or not districts had implemented a ban (B) into distinct years, districts, or compliance levels. Four models look at the impact by district (Los Angeles, San Francisco, Oakland, and Pasadena) and another four analyze impact by year (2015-16, 2016-17, and 2017-18). In each case, the first pair of models analyzes English and the second pair analyzes Math. Odd-numbered models utilize district data and even-numbered models utilize school data. Scrutinizing the key independent variable in this way tests the extent to which there is a clear pattern across districts and over time.

The last two models only utilize school data, and they divide schools by their degree of compliance with the ban. Schools are completely compliant if they gave no defiance suspensions during the first two years after the ban uptake period; this category constituted 83% of schools in our four districts. Schools are not compliant if they gave defiance suspensions both of those years; this was 6% of schools. Lastly, schools are partially compliant if they gave defiance suspensions in one of those years; the remaining 11% of schools. Analyzing the impact on each group helps us estimate whether academic impact is related to the extent to which staff officially ban suspensions. Models 17 and 18 look at English while Models 19 and 20 look at Math. In Models 17 and 19 we apply two exclusion rules – at least 95% test participation and no more than 5% change in cohort size from the prior year – to approximate consistent cohorts in current and prior years. In Models 18 and 20, we do not apply those exclusion rules.

*Results*

Table 2 shows rough estimates of the impact of the ban on suspensions. Models 1 and 2 show a negative but non-significant decrease in total and unduplicated suspension rates in all four districts. Models 3 and 4 reveal that the impact was negative and significant in Oakland and Pasadena, but positive and significant in Los Angeles and San Francisco. This does not fit with the much more nuanced analysis of Los Angeles by Hashim et al. (2018) that is based on a longer time series and more granular data. Since they found that the ban did cause Los Angeles to decrease its suspension rate, we assume we would have found similar results with more detailed data. This also leads us to believe a more thorough dataset may have revealed a decline in San Francisco’s suspension rate, as well.

**Table 2**

Fixed Effect Regression of Suspension Rates on Ban Adoption

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1:  Unduplicated Suspension Rate | Model 2:  Total Suspension Rate | Model 3:  Unduplicated Suspension Rate | Model 4:  Total Suspension Rate |
| Post-ban period | -0.008 | -0.010 |  |  |
| (0.006) | (0.020) |  |  |
| Los Angeles  x Post-ban period |  |  | -0.001 | 0.021\*\* |
|  |  | (0.001) | (0.005) |
| San Francisco  x Post-ban period |  |  | 0.005\*\* | 0.027\*\* |
|  |  | (0.001) | (0.005) |
| Oakland  x Post-ban period |  |  | -0.016\*\* | -0.023\*\* |
|  |  | (0.001) | (0.005) |
| Pasadena  x Post-ban period |  |  | -0.022\*\* | -0.067\*\* |
|  |  | (0.001) | (0.005) |
| Time | -0.005\*\* | -0.015\*\* | -0.005\*\* | -0.015\*\* |
| (0.001) | (0.003) | (0.001) | (0.003) |
| Time-squared | 0.000\* | 0.001\* | 0.000\*\* | 0.001\*\* |
| (0.000) | (0.000) | (0.000) | (0.000) |
| R-squared | 0.016 | 0.012 | 0.015 | 0.012 |
| Observations | 4,888 | 4,888 | 4,888 | 4,888 |

Source: Authors’ calculations and the California Department of Education suspension files: <https://www.cde.ca.gov/ds/sd/sd/filessd.asp>.

Note: Standard errors are clustered by 1,004 districts.

\*\* p < 0.01, \* p < 0.05, + p < 0.10

There are two reasons we are not concerned about obtaining precise estimates for the impact bans have on suspension rates. One is that schools are able to influence those rates by not reporting suspensions. Even when the official data suggests the bans led to lower suspension rates, we have no way of knowing the degree to which that is accurate. The other is that we are not arguing for a link between reduced suspension rates and reduced math performance. In fact, the evidence argues against such a link. In Philadelphia, schools that did not comply with the ban – and did not have a lower suspension rate – had lower academic performance (Steinberg & Lacoe, 2018). In Pittsburgh, the lower math performance was driven by middle school grades, which did not experience a decline a suspension rates (Augustine et al, 2018). While our data is unable to prove what mechanisms link bans to lower math achievement, we do not argue that the link is reduced suspensions. Instead, we hypothesize that it concerns the altered relationships between students, school staff, and district leadership.

Table 3 shifts to the impact of suspension bans on academics. Models 5 and 6 show a negative impact on English, but it is not significant when using district data and it is small and significant when using school data. Since we cannot determine which model provides a more accurate estimate, we assume that there is no significant impact on English. Models 7 and 8 both show a negative and significant impact on Math. The district data has the benefit of including all students who tested, while the school data has the benefit of including observations where we are likely to have current and prior scores for nearly identical groups of students. The fact that both models provide similar estimates gives us confidence that the ban had a true impact. While we are not focused on the results for other independent variables, it is reassuring that their estimates are the direction we would expect and are quite similar across all models.[[13]](#footnote-13) The few exceptions involve unit-level subgroups.[[14]](#footnote-14)

**Table 3**

Fixed Effect Regression of Academic Growth on Ban Adoption

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | English | | Math | |
| Model 5:  District Data | Model 6:  School Data | Model 7:  District Data | Model 8:  School Data |
| Post-ban period | -0.062 | -0.028\* | -0.088\* | -0.150\*\* |
| (0.042) | (0.012) | (0.037) | (0.018) |
| Prior Score | 0.544\*\* | 0.560\*\* | 0.467\*\* | 0.473\*\* |
| (0.008) | (0.005) | (0.009) | (0.006) |
| Time | -0.026\*\* | -0.032\*\* | -0.037\*\* | -0.037\*\* |
| (0.003) | (0.002) | (0.004) | (0.003) |
| Time squared | 0.002\*\* | 0.002\*\* | 0.003\*\* | 0.003\*\* |
| (0.000) | (0.000) | (0.000) | (0.000) |
| Grade-Level  % Low Income | -0.496\*\* | -0.510\*\* | -0.708\*\* | -0.565\*\* |
| (0.056) | (0.037) | (0.074) | (0.050) |
| % with Disabilities | -0.821\*\* | -0.807\*\* | -1.004\*\* | -0.829\*\* |
| (0.057) | (0.048) | (0.097) | (0.062) |
| % English Learner | -0.466\*\* | -0.461\*\* | -0.519\*\* | -0.463\*\* |
| (0.045) | (0.035) | (0.068) | (0.045) |
| Unit-Level  % African American | -1.303\*\* | -0.872\*\* | -1.602\*\* | -0.740\*\* |
| (0.258) | (0.099) | (0.323) | (0.158) |
| % Latino | -0.423\*\* | -0.406\*\* | -0.522\*\* | -0.369\*\* |
| (0.122) | (0.056) | (0.160) | (0.084) |
| % Asian | 0.205 | 0.214\*\* | 0.706\* | 0.494\*\* |
| (0.181) | (0.071) | (0.276) | (0.109) |
| % Other | -0.292\*\* | -0.253\*\* | -0.141 | 0.072 |
| (0.083) | (0.054) | (0.101) | (0.078) |
| % Low Income | 0.371\*\* | 0.387\*\* | 0.685\*\* | 0.405\*\* |
| (0.066) | (0.046) | (0.094) | (0.062) |
| % with Disabilities | 0.335\*\* | 0.302\*\* | 0.244 | 0.127 |
| (0.110) | (0.078) | (0.154) | (0.106) |
| % English Learner | 0.028 | -0.064 | 0.431\*\* | 0.176\*\* |
| (0.080) | (0.047) | (0.106) | (0.062) |
| R-squared | 0.877 | 0.888 | 0.767 | 0.793 |
| Observations | 45,456 | 66,051 | 32,234 | 59,422 |

Source: Authors’ calculations and the California Department of Education STAR research files: <https://star.cde.ca.gov/starresearchfiles.asp>.

Note: For district data, standard errors are clustered by 884 districts for English analyses and 779 districts for Math analyses. For school data, standard errors are clustered by 8,031 schools for English analyses and 7,018 schools for Math analyses.

\*\* p < 0.01, \* p < 0.05, + p < 0.10

Table 4 examines the impact of bans by district. Oakland, and Pasadena have negative and significant impacts on English and Math, when we use district or school data. San Francisco has negative coefficients in all four models, but the results are only significant when using district data. The clear outlier is Los Angeles; the impact of the ban on English is positive and significant in Model 9 and nonsignificant in Model 10. In Math, the result is positive and nonsignificant in Model 11 but negative and significant in Model 12.

However, this nonsignificant result for Math occurred after an enormous decline in performance during the ban uptake period. Appendix B shows that Math achievement for grades three through seven fell from an average of 0.3 standard deviations below the state average in 2012-13 to 0.5 standard deviations below the state average in 2014-15. We exclude 2014-15 from all analyses because it is part of the ban uptake period; however, that drop allowed Los Angeles to start the post-uptake period with very low prior scores. Since our model controls for prior scores, the non-significant coefficient for math means that Los Angeles students only grew as much as we would expect in math – given their extremely low starting scores. If we include 2014-15 in the analyses, the results for Los Angeles (available upon request) using district data are nonsignificant for English and negative but not quite significant for Math (-0.02 SD, p-value is 0.102). Using school data, the results are negative and significant for both English (-0.02 SD, p-value is 0.085) and Math (-0.16 SD, p-value is 0.000). Given the consistently negative and significant Math results for the other three districts, we believe that the enormous decline from 2012-13 to 2014-15 was caused at least partly by the ban. At worst, the ban caused all of that large decline; at best, the ban did not help students catch up from a huge unrelated decline in Math.

**Table 4**

Fixed Effect Regression of Academic Growth on Ban Adoption, by District

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | English | | Math | |
| Model 9:  District Data | Model 10:  School Data | Model 11:  District Data | Model 12:  School Data |
| Los Angeles  x Post-ban period | 0.045\*\* | 0.001 | 0.020 | -0.135\*\* |
| (0.008) | (0.013) | (0.013) | (0.019) |
| San Francisco  x Post-ban period | -0.053\*\* | -0.073 | -0.103\*\* | -0.079 |
| (0.013) | (0.046) | (0.016) | (0.077) |
| Oakland  x Post-ban period | -0.117\*\* | -0.193\*\* | -0.170\*\* | -0.294\*\* |
| (0.018) | (0.048) | (0.021) | (0.075) |
| Pasadena  x Post-ban period | -0.168\*\* | -0.138\*\* | -0.072\*\* | -0.297\* |
| (0.017) | (0.049) | (0.018) | (0.117) |
| R-squared | 0.860 | 0.886 | 0.767 | 0.790 |
| Observations | 45,456 | 66,051 | 32,234 | 59,422 |

Source: Authors’ calculations and the California Department of Education STAR research files: <https://star.cde.ca.gov/starresearchfiles.asp>.

Note: For district data, standard errors are clustered by 884 districts for English analyses and 779 districts for Math analyses. For school data, standard errors are clustered by 8,031 schools for English analyses and 7,018 schools for Math analyses. Models also include all independent variables in Table 3 except for post-ban period.

\*\* p < 0.01, \* p < 0.05, + p < 0.10

Table 5 shows the impact for each year during the post-ban period. The results for both subjects are always negative, but only the Math results are consistently significant. Using both district and school data, the coefficients for Math are smaller in 2017-18 than in 2015-16. This is consistent with the expectation that the impact of a ban – negative or positive – is likely to diminish over time as students and staff adjust to the new policy. The year 2017-18 is four and five years after San Francisco and Los Angeles adopted their bans, respectively. This suggests that the negative Math impact of a ban probably lasts longer than three years.

**Table 5**

Fixed Effect Regression of Academic Growth on Ban Adoption, by Year

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | English | | Math | |
| Model 13:  District Data | Model 14:  School Data | Model 15:  District Data | Model 16:  School Data |
| 2015-16  x Post-ban period | -0.115\* | -0.017 | -0.106\* | -0.146\*\* |
| (0.045) | (0.016) | (0.041) | (0.021) |
| 2016-17  x Post-ban period | -0.051 | -0.011 | -0.072\* | -0.156\*\* |
| (0.048) | (0.015) | (0.031) | (0.021) |
| 2017-18  x Post-ban period | -0.053 | -0.024+ | -0.065+ | -0.129\*\* |
| (0.045) | (0.015) | (0.039) | (0.020) |
| R-squared | 0.860 | 0.886 | 0.767 | 0.791 |
| Observations | 45,456 | 66,051 | 32,234 | 59,422 |

Source: Authors’ calculations and the California Department of Education STAR research files: <https://star.cde.ca.gov/starresearchfiles.asp>.

Note: For district data, standard errors are clustered by 884 districts for English analyses and 779 districts for Math analyses. For school data, standard errors are clustered by 8,031 schools for English analyses and 7,018 schools for Math analyses. Models also include all independent variables in Table 3 except for post-ban period.

*\*\* p < 0.01, \* p < 0.05, + p < 0.10*

Finally, Table 6 evaluates the impact at the school level by compliance. In Philadelphia, the negative impact of a ban was driven entirely by schools that did not comply with the new policy (Steinberg & Lacoe, 2018). While that is the case for our English results, it is not the case for our Math results.[[15]](#footnote-15) In California, even fully compliant schools experienced a significant decline in Math growth after the ban. The negative impact was significantly greater in schools that were not compliant with the ban. This implies that the academic decline is not caused by a decline in suspension rates. In this sense, our findings agree with evaluations of discipline reform in Philadelphia and Pittsburgh (Steinberg & Lacoe, 2018; Augustine et al. 2018). However, our findings also imply that negative consequences can occur even when schools fully implement a ban. The negative impact for fully compliant schools cannot be explained by school staff being unwilling or unable to adhere to the new discipline policy.[[16]](#footnote-16)

**Table 6**

Fixed Effect Regression of Academic Growth on Ban Adoption, by Compliance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | English | | Math | |
| Model 17:  Exclusions | Model 18:  No Exclusions | Model 19: Exclusions | Model 20:  No Exclusions |
| Fully Compliant  x Post-ban period | -0.020 | -0.020+ | -0.134\*\* | -0.120\*\* |
| (0.014) | (0.010) | (0.019) | (0.013) |
| Partially Compliant  x Post-ban period | -0.075\*\* | -0.071\*\* | -0.230\*\* | -0.143\*\* |
| (0.028) | (0.024) | (0.063) | (0.038) |
| Not Compliant  x Post-ban period | -0.066+ | -0.150\*\* | -0.324\*\* | -0.325\*\* |
| (0.035) | (0.036) | (0.093) | (0.060) |
| R-squared | 0.888 | 0.846 | 0.793 | 0.717 |
| Observations | 45,456 | 213,557 | 32,234 | 180,907 |

Source: Authors’ calculations as well as the California Department of Education STAR research files (<https://star.cde.ca.gov/starresearchfiles.asp>) and suspension files (<https://www.cde.ca.gov/ds/sd/sd/filessd.asp>.).

Note: When exclusions are applied, standard errors are clustered by 8,031 schools for English analyses and 7,018 schools for Math analyses. When exclusion are not applied, standard errors are clustered by 9,096 schools for English analyses and 7,416 schools for Math analyses. Models also include all independent variables in Table 3 except for post-ban period.

\*\* p < 0.01, \* p < 0.05, + p < 0.10

**Conclusion**

Our goal is to determine the academic impact of suspension bans in four California school districts. We utilize fixed effects to control for time-invariant factors that could influence academic growth. We exploit the exogenous introduction of bans to the California policy landscape – and include an array of demographics as independent variables – to control for time-varying factors that could influence academic growth. While the results for English are often not significant, the results for Math are consistently negative and significant. We conclude that bans reduced Math growth by approximately 0.09 standard deviations annually for at least three years after their adoption. The cumulative decline of 0.27 standard deviations equates to districts dropping from the 50th percentile to approximately the 39th percentile in statewide performance.

*Limitations*

There are several important limitations to our finding. First, it is based on a non-representative group of four large, urban, and diverse districts in California. The results may not be generalizable to districts that are smaller, more rural, or serve a smaller percentage of historically disadvantaged students. Those four districts may have other special characteristics that influence the impact of a suspension ban. Indeed, leaders in those districts adopted bans because they had a desire to lead in the area of discipline reform. However, we would expect suspension bans to have a relatively positive impact in districts where leaders have interest and experience in suspension reduction efforts. Districts that are forced to adopt a ban and have less expertise may see even more negative effects

The second limitation of this paper is its reliance on grade-level data. In order to control for prior scores, we have to assume that relatively consistent cohorts of students move from grade to grade within districts. For example, the average score for grade three students in the prior year is an independent variable while the average score for grade four students in the current year at the same school is the dependent variable. Ideally, both current and prior scores are based on the same group of students. Bias could occur when those two groups of students are significantly different. As mentioned above, we use two inclusion rules to minimize this concern:

* Current and prior scores each reflect a participation rate of at least 95%
* There is a less than 5% change in cohort size from the prior year to the current year

Applying these inclusion rules to our school dataset greatly increase the probability that we are obtaining current and prior test scores from nearly identical sets of students. Additionally, the fact that we control for grade level demographics helps to account for any cohort changes that do occur. The fact that we obtain similar results with our school data with or without those inclusion rules gives us confidence that the district-level data is not notably biased by cohort mobility.

Perhaps most importantly, the pattern of our results makes an alternative explanation regarding mobility quite implausible. For mobility to explain our results, each of the districts would have to experience an entrance of relatively low-achieving students and/or an exit of relatively high-achieving students – but only in the years after the ban. (Changes in the percent of students in various demographic groups would have to be unable to explain the changes in expected achievement.) In addition, such mobility would need to occur in each of the three years we analyze: 2015-16, 2016-17, and 2017-18. Most implausibly, the most mobility would have to occur at schools that did not comply with the ban at all. While it is theoretically possible for such student mobility to occur, it takes a series of brave assumptions to believe this alternative explanation.

A third limitation is that we are unable to prove what mechanisms underlie our results. This issue is common to all recent research in this area (Steinberg & Lacoe, 2018; Augustine et al., 2018). The fact that schools that do not comply with bans experience the worst results suggest what the mechanism is not: it is not caused by school staff refusing to suspend defiant students who then disrupt classroom learning. We posit that the mechanisms concern bans’ effort to alter relationships between students, school staff, and district leaders. Future researchers will need to investigate whether this hypothesis is correct or if other mechanisms explain the math decline in the four California districts we analyze.

*Implications and Conclusion*

The literature demonstrates that suspensions harm academic growth, and we should make efforts to reduce them. Unfortunately, the evidence in this paper reveals that bans have an unintended negative impact on math growth. Bans sometimes cause a drop in suspension rates, which should result in improved academic growth. However, evidence from California’s K-12 bans finds no improvement for reading and a substantial and sustained negative impact on math. It appears that unintended consequences from suspension bans more than negate the expected academic improvement. Policymakers should know and consider this trade-off.

There are other ways to reduce suspensions and their disproportionalities. In elementary grades, the “SaferSanerSchools” restorative justice program reduced suspension rates while avoiding academic harm (Augustine et al, 2018). However, in middle school grades that same program reduced English and Math performance while not decreasing suspension rates (Augustine et al., 2018). Policymakers wishing to avoid academic harm would be wise to pilot and test other discipline reforms for middle and high schools. Partnerships between researchers and district leaders could produce valuable insights for the entire education community.

We hypothesize that bans harm academics by altering the relationships between students, school staff, and district leaders. We also note that bans do not address the root causes of seemingly unnecessary suspensions. Bans do not specify what school staff should do to replace their practice of suspending defiant students. Schools need procedures and practices in place to maintain order and enable learning (Cornell and Mayer, 2010). After a 2015 ban for grades K-3, a survey found that 86% of California teachers felt they needed more training and support to deal with discipline effectively (Meredith Adams, 2017). It is one thing to show that school staff should not need to suspend students for defiant behavior; it is quite another to help school staff obtain the training and resources needed to deal with defiant behavior. Policies that successfully address root problems are likely to reduce unnecessary suspensions while avoiding unintended negative consequences.

The findings in this paper highlight the need for improved policy evaluation at the school district level. School board members and district leaders do not regularly conduct rigorous evaluations of innovative policies they implement. An array of factors contribute to this problem, ranging from the financial cost of conducting evaluations to the political cost of having to admit that past decisions led to bad outcomes. We are far from making evaluation a routine component of district policy decisions. This paper is part of a broader effort to nudge us in that direction. Future researchers can find ways to reduce suspensions that also increase academic growth. Even more important would be if future researchers regularly help to inform district leaders as they make important decisions concerning the education of our children.

**Appendix A**

Only four districts adopted K-12 bans in California, all between 2013 and 2015. This appendix looks at the English and Math achievement trends in each district from 2007 to 2018 (except for 2014, when the state did not report scores). The vertical line indicates the year in which that district implemented its ban. The purpose of these graphs is to visually inspect whether there appears to be a change in trajectory after the ban. Indeed, each district experienced a rise in Math achievement immediately before the ban and a decline in Math achievement immediately after the ban. This reassures us that our analyses are not accidentally measuring some long-term Math decline that began before the bans were adopted.

*Figure A1: Trends in Achievement for Los Angeles*



*Figure A2: Trends in Achievement for San Francisco*



*Figure A3: Trends in Achievement for Oakland*



*Figure A4: Trends in Achievement for Pasadena*



**References**

Arcia, Emily. (2006). Achievement and Enrollment Status of Suspended Students: Outcomes in a Large, Multicultural School District. Education and Urban Society 38(3):359–69.

Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., & Joseph, A. A. (2018). Can Restorative Practices Improve School Climate and Curb Suspensions?.

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

Cornell, D. G., & Mayer, M. J. (2010). Why do school order and safety matter?. *Educational Researcher*, *39*(1), 7-15.

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program*. Guilford Press.

Curran, F. C. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. *Educational Evaluation and Policy Analysis*, *38*(4), 647-668.

Curran, F. C. (2017). Racial disproportionalities in discipline: The role of zero tolerance policies. In *Discrimination and Diversity: Concepts, Methodologies, Tools, and Applications* (pp. 1251-1266). IGI Global.

D’Orio Wayne. (2018). Is School Discipline Reform Moving Too Fast? *The Atlantic*, Jan 11.

Eden, M. (2017). School discipline reform and disorder: Evidence from New York City Public Schools, 2012-16. *The Education Digest*, 83(1), 22.

Freiberg, H. J., & Lapointe, J. M. (2006). Research-based programs for preventing and solving discipline problems. In C. M. Evertson & C. S. Weinstein (Eds.), Handbook of classroom management (pp. 735– 786). Mahwah, NJ: Lawrence Erlbaum.

Frey, Susan. (2015). Oakland ends suspensions for willful defiance, funds restorative justice. *EdSource*, May 14.

Frey, Susan. (2014). San Francisco Unified eliminates ‘wllful defiance’ as a reason to expel or suspend students. *EdSource*, Feb 26.

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). Do early educators’ implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions? New Haven, CT: Yale Child Study Center.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin?. *Educational Researcher*, *39*(1), 59-68.

Hashim, A. K., Strunk, K. O., & Dhaliwal, T. K. (2018). Justice for All? Suspension Bans and Restorative Justice Programs in the Los Angeles Unified School District. *Peabody Journal of Education*, (just-accepted).

Howell, J. C. (2003). *Preventing and reducing juvenile delinquency: A comprehensive framework*. Sage.

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

Hwang, N. (2018). Suspensions and Achievement: Varying Links by Type, Frequency, and Subgroup. *Educational Researcher*, 0013189X18779579.

Lacoe, J., & Steinberg, M. P. (2018). Do Suspensions Affect Student Outcomes?. *Educational Evaluation and Policy Analysis*.

Liaupsin, C. J., Umbreit, J., Ferro, J. B., Urso, A., & Upreti, G. (2006). Improving academic engagement through systematic, function-based intervention. *Education and Treatment of Children*, 573-591.

Losen, Daniel. (2011). Discipline Policies, Successful Schools, and Racial Justice. UCLA: The Civil Rights Project / Proyecto Derechos Civiles.

Losen, Daniel J.; Hodson, Cheri L.; Keith II, Michael A.; Morrison, Katrina; & Belway, Shakti. (2015). Are We Closing the School Discipline Gap?. K-12 Racial Disparities in School Discipline. UCLA: The Civil Rights Project / Proyecto Derechos Civiles.

Luiselli, J. K., Putnam, R. F., Handler, M. W., & Feinberg, A. B. (2005). Whole‐school positive behaviour support: effects on student discipline problems and academic performance. *Educational Psychology*, *25*(2-3), 183-198.

Martinez, T., Chandler, A., & Latham, N. (2013). Case Study: School Discipline Reform in California. *CA: The California Endowment*.

McFarland, D. A. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance. *American Journal of Sociology*, *107*(3), 612-678.

Meredith Adams, J. (2017). “Most teachers in California say they need more training in alternatives to suspensions, survey finds.” *EdSource*, May 7.

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological science*, *26*(5), 617-624.

Orpinas, P., & Horne, A. M. (2006). *Bullying prevention: Creating a positive school climate and developing social competence*. American Psychological Association.

Payne, A. A., & Welch, K. (2010). Modeling the effects of racial threat on punitive and restorative school discipline practices. *Criminology: An Interdisciplinary Journal*.

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

Putnam, R. P., Horner, R. H., & Algozzine, R. (2006). Academic achievement and the implementation of school-wide behavior support. *Positive Behavioral Interventions and Supports Newsletter*, *3*(1), 1-6.

Raffaele Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New directions for youth development*, *2003*(99), 17-33.

Rausch, M. K., & Skiba, R. (2004). Disproportionality in School Discipline among Minority Students in Indiana: Description and Analysis. Children Left Behind Policy Briefs. Supplementary Analysis 2-A. *Center for Evaluation and Education Policy, Indiana University*.

Rumberger, R. W., & Losen, D. J. (2017). The Hidden Costs of California's Harsh School Discipline: And the Localized Economic Benefits from Suspending Fewer High School Students. *Civil Rights Project-Proyecto Derechos Civiles*.

Skiba, R., & Sprague, J. (2008). Without Suspensions. *Educational leadership*.

Steinberg, M. P., & Lacoe, J. (2018). Reforming school discipline: school-level policy implementation and the consequences for suspended students and their peers. *American Journal of Education*, *125*(1), 29-77.

Swearer, S. M., Espelage, D. L., & Napolitano, S. A. (2009). *Bullying prevention and intervention: Realistic strategies for schools*. Guilford press.

Wallace Jr, J. M., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among US high school students: 1991-2005. *The Negro educational review*, *59*(1-2), 47.

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

Welch, K., & Payne, A. A. (2012). Exclusionary school punishment: The effect of racial threat on expulsion and suspension. *Youth Violence and Juvenile Justice*, *10*(2), 155-171.

1. However, Hashim et al. (2018) find that disproportionalities persisted for African American students and students with disabilities after the 2013 suspension ban and subsequent restorative justice training at Los Angeles Unified School District. [↑](#footnote-ref-1)
2. California only provides statewide suspension data files that combine in- and out-of-school suspensions. For simplicity, “suspensions” with no qualifier refers to both in- and out-of-school suspensions. [↑](#footnote-ref-2)
3. In 2010, The California Endowment launched a multi-million dollar effort called Building Healthy Communities to improve the health of 14 areas in California. While collecting input from stakeholders, staff of The California Endowment were surprised to hear that the abundant use of school suspensions was harming students’ social and emotional health (Martinez et al., 2013). [↑](#footnote-ref-3)
4. Additionally, Azusa Unified School District banned willful defiance over three years, starting in 2014-15. We exclude it from the list because of the staggered nature of its implementation. [↑](#footnote-ref-4)
5. Such an analysis produces significant negative results for Math (available upon request) of an even greater magnitude than our reported analyses. [↑](#footnote-ref-5)
6. For information, see <https://achieve.lausd.net/Page/11927>. [↑](#footnote-ref-6)
7. It is beyond the scope of this paper to determine the exact mechanisms by which bans harm academics. However, anecdotal evidence points to two possibilities. First, surveys indicate that the school climate worsened after New York City’s discipline reforms in 2014-15; a majority of schools had increasing percentages of teachers reporting more disorder and students reporting more violence, drug use, and gang activity (Eden, 2017). A worsening school climate could result in lower academic performance. Second, bans may also increase teacher turnover, which could harm academic growth (D’Orio, 2018). [↑](#footnote-ref-7)
8. Data on assessment scores and tested student demographics is from here: <https://star.cde.ca.gov/starresearchfiles.asp>. Suspension data is from here: <https://www.cde.ca.gov/ds/sd/sd/filessd.asp>. [↑](#footnote-ref-8)
9. Files prior to 2007 do not report the number of students in each grade with valid scores, which can be lower than the number of students tested. California reported no scores for 2013-14 because it conducted a field test of a new statewide assessment in that year. [↑](#footnote-ref-9)
10. Until 2013, starting in grade eight students would take end of course assessments specific to the math course they took. In order for the analysis to have similar groups of students in the current and prior testing years, we need students to take the same assessments in each year. [↑](#footnote-ref-10)
11. A difference-in-differences analysis leads to negative results for Math of an even greater magnitude. [↑](#footnote-ref-11)
12. Another approach would be to control for the lagged score in that same grade: last year’s 5th grade scores for this year’s 5th graders. This approach leads to negative results for Math of an even greater magnitude. [↑](#footnote-ref-12)
13. We see positive coefficients for unit-level percentages for students who are low income, English Learner, or have disabilities because those demographics are included at the both the unit and grade levels. This leads to multicollinearity. However, the results for our independent variable of interest – available upon request – are the same or stronger if we only include those three variables at the grade or unit level. [↑](#footnote-ref-13)
14. The percent of students in a district who are Asian is not significantly related to English growth, but it is significant when using school data. Additionally, the percent of students in a district or school who are an “Other” race or have a disability are significantly related to English growth but not Math growth. The percent of students in a district or school who are English Learners are significantly related to Math growth but not English growth. These unit-level factors are inconsequential to the independent variable of interest. [↑](#footnote-ref-14)
15. Reassuringly, comparing Models 17 and 18 or Models 19 and 20 reveals that our results are similar whether or not we apply exclusion rules. [↑](#footnote-ref-15)
16. Additionally, the degree of compliance with the ban is not randomly distributed across schools. To the extent that policy compliance is related to higher academic growth (e.g. because it indicates stronger school leaders or teachers), we would expect a more positive coefficient for compliant schools. [↑](#footnote-ref-16)