

The Impact of Suspension Reforms on Discipline Outcomes: Evidence From California High Schools

Rui Wang 

Shanghai University of Finance and Economics

Minority students are suspended at a disproportionately higher rate compared with others. To reduce racial suspension gaps, four California school districts banned schools from suspending students for willful defiance, a category consisting of relatively minor disruptive offenses. I evaluate the impact of these policies on high school student discipline outcomes using a difference-in-differences strategy that exploits the temporal variation in the enactment of these policies across school districts. The results suggest that while these policies decreased willful defiance out-of-school suspension rates by around 69%, they did not reduce overall out-of-school suspension rates. In fact, the policies significantly increased out-of-school suspension rates among Black students, particularly in schools with a small share of Black teachers. Taken together, the results suggest that the willful defiance suspension bans failed to address implicit and explicit biases in California schools.

Keywords: *suspension reform, Black students, discipline outcomes*

OUT-OF-SCHOOL suspension (OSS) is one of the most commonly used discipline actions in U.S. schools: over the past 20 years, about 5% of students received at least one OSS each year (de Brey et al., 2019). While there may be justifiable grounds for excluding disruptive students from the classroom, such as protecting other students' safety and learning, a large proportion of OSSs are issued to students committing minor infractions that pose little to no direct threat to their classmates. This misapplication of OSS deprives students of educational opportunities and more generally harms the school learning environment by creating shared stress (Pena-Shaff et al., 2019).

Suspension is associated with poor academic performance, school dropout, crime, and delinquency (Anderson et al., 2019; Bacher-Hicks et al., 2019; Cuellar & Markowitz, 2015; Wald & Losen, 2003). In addition, the overuse of suspension may exacerbate disparities in learning outcomes among students from different demographic and socioeconomic groups (Gopalan & Nelson, 2019; Gregory et al., 2010; Pearman et al., 2019). Black students are at least twice as likely to be suspended as White students (Losen & Gillespie, 2012) and are suspended for longer periods, even when involved in the same type of incidents (Barrett et al., 2019). This suggests that levels of discretion and potential biases in the use of suspension for minor disorder infractions could disproportionately harm Black students, thus widening existing racial achievement gaps.

Several theories examine the impact of suspension on student outcomes both inside and outside of the classroom.

Positive intervention and restorative justice (RJ) theories argue that suspension could lead to unintended long-term consequences, such as deterioration of the school atmosphere, entrenched antisocial behavior, and increases in instances of misconduct (Gonzalez, 2012; Mowen et al., 2020; Pesta, 2018; Sugai & Horner, 2002; Way, 2011). Empirical studies on the incapacitation effect of school suggest that excluding misbehaved students from schools increases their risk of engaging in criminal activities (Cuellar & Markowitz, 2015; Jacob & Lefgren, 2003; Luallen, 2006). Although the deterrence theory posits that stricter discipline can induce students to comply with rules and authorities (Kinsler, 2013; Nagin, 2013), considerable evidence has shown that the use of a less punitive discipline approach increases students' respect for teachers, reduces students' infractions, and improves school climate (Bradshaw et al., 2009; Bradshaw et al., 2010; Okonofua & Eberhardt, 2015).

Over the past two decades, concerns about the negative consequences of punishment and existing racial discipline gaps, especially suspension gaps between Black and White students, have led to many discipline reforms (Education Commission of the States, 2020), most of which focused on reducing suspensions for minor infractions. For example, New York City Middle Schools (Craig & Martin, 2019), the School District of Philadelphia (Lacoe & Steinberg, 2018), and Chicago Public Schools (Stevens et al., 2015) have either replaced OSS with in-school interventions for insubordination and disruptive behavior or reduced the length of OSS for all infractions. However, existing research suggests



that these suspension reforms have little impact on student academic and discipline outcomes (e.g., Craig & Martin, 2019; Lacoé & Steinberg, 2018; Steinberg & Lacoé, 2018; Stevens et al., 2015).

This article studies a specific kind of suspension reform implemented by four California school districts, which banned schools from issuing OSS for willful defiance. Specifically, it asks whether schools in these districts complied with these willful defiance suspension bans (WDBs) and estimates their impact on student discipline outcomes.

The California Department of Education defines willful defiance as behavior that “disrupted 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.” In practice, however, teachers interpret willful defiance differently, applying this generic label to suspend students for infractions ranging from eyerolling or backtalking to sleeping during class. Consequently, willful defiance accounted for more than 40% of OSSs before the implementation of WDBs. During the same period, OSS rates for Black students reached 15.4%, compared with 5.6% for White students (California Department of Education, n.d.).

Acknowledging OSS being overused for willful defiance and minority students being suspended for higher rates than students of other races, four school districts in California (San Francisco Unified School District [SFUSD], Pasadena Unified School District [PUSD], Azusa Unified School District [AUSD], and Oakland Unified School District [OUSD]) explicitly eliminated willful defiance as a reason for suspending K–12 students out of schools after the 2014–2015 school year.¹ While their approach mimicked previous reforms in Philadelphia and New York, the potential impact was greater. The California WDBs affected about 40% of total OSSs, compared with 15% to 25% in Philadelphia and New York reforms (Craig & Martin, 2019; Lacoé & Steinberg, 2018; Steinberg & Lacoé, 2018).

Although high schools are most likely to utilize OSS, most previous studies have focused on elementary and/or middle schools (Anderson et al., 2019; Gopalan & Nelson, 2019; Lacoé & Steinberg, 2018; Steinberg & Lacoé, 2018; Stevens et al., 2015). Therefore, this study extends this body of literature by estimating the impacts of school discipline reform in high schools. Using publicly available school-level data from the California Department of Education, I start by examining how WDBs affect student OSS rates and how this effect varies by student race to establish the extent of racial barriers and inequality in the California school system.² Then, I study the impact of WDBs on the prevalence of suspension by distinguishing between “willful defiance OSS” and “nonwillful defiance OSS.” To account for the influence of unobserved confounding factors, I exploit the temporal variation in the implementation of WDBs across school districts, using a difference-in-differences (DD) estimation strategy.

Several studies have considered the impact of suspension reforms on student discipline outcomes. Most studies find that these reforms were not correlated with decreases in OSS rates (Baker-Smith, 2018; Lacoé & Steinberg, 2018; Steinberg & Lacoé, 2018; Stevens et al., 2015), although a study on a suspension reform in Los Angeles showed evidence of reduced OSS rates (Hashim et al., 2018). However, these findings should be interpreted with caution for several reasons. First, most of these studies used research designs that could not account for the potentially nonrandom adoption of suspension reforms. Second, some only rely on data from the reformed school district, raising concerns over external validity (Baker-Smith, 2018; Hashim et al., 2018; Stevens et al., 2015). A related strand of research has focused on the impact of suspension reforms on student performance, with mixed findings (e.g., Craig & Martin, 2019; Steinberg & Lacoé, 2018). Compared with those studies, the current analysis of WDBs draws from a far greater number of reformed and unreformed schools, resulting in a much larger sample size. Furthermore, this study examines three specific explanations that may contribute to the negative effect of WDBs on student discipline outcomes: substitution of suspension, lack of same-race teachers, and negative changes in student behaviors.

My findings suggest that WDBs reduced willful defiance OSS rates by around 69%. However, these WDBs did not reduce overall OSS rates: schools simply changed the reasons given when suspending students. More important, Black students were disproportionately affected by this shift, and hence by WDBs. Supplemental analysis using data from the Youth Risk Behavior Surveillance System (YRBSS) shows that the increases in OSS cannot be attributed to more student infractions. Taken together, the findings suggest that WDBs failed to address biases against Black students in California schools.

The remainder of the article is organized as follows. The Policy Background section provides background on WDBs in California. Then, the Method section introduces the data and empirical strategy. The following sections present the main results and results from the supplemental analysis. The final section concludes with a summary of the findings and policy implications.

Policy Background

In 2011, a voluntary agreement between LAUSD and the U.S. Department of Education banned willful defiance as a reason for suspension district-wide (Aron, 2013; Blume, 2012; Hashim et al., 2018). Three years later, the California State Legislature introduced a statewide WDB for Grades K–3 (A.B. 420) in response to the prevalent and disproportionate use of willful defiance OSS among Black and Hispanic students.³ More recently, California extended the WDB to cover students through Grade 8 by July 2025, but

high school students will remain subject to willful defiance OSS.⁴

In 2014, SFUSD became the first to implement a full WDB, applied to students in all grades. Prior to this move, SFUSD primarily relied on RJ and school-wide positive behavior intervention and support programs to combat high suspension rates. Yet, these programs failed to address the district's concerns over the disproportionate suspensions of Black students. As a result, starting from the 2014–2015 school year, SFUSD banned willful defiance OSS to support previous efforts (SFUSD, 2014).

Later, PUSD and OUSD enacted full WDBs to support their existing suspension reduction programs. In the spring of 2015, seeing that positive behavior intervention and support programs (PBIS) failed to reduce willful defiance OSSs, PUSD extended the state's WDB to all K–12 students. OUSD, in response to a 2012 resolution with the Department of Education to promote equitable discipline practices via specific goals, also implemented WDB in the 2016–2017 school year (U.S. Department of Education, 2012). Instead of revising discipline regulations, AUSD informally adopted a WDB in its 2015 annual accountability plan, with the goal to eliminate willful defiance suspension by SY 2016–2017 (AUSD, n.d.).

Notably, the WDBs in all four school districts were followed by alternative programs such as RJ and PBIS, designed to be implemented simultaneously and complementarily (Riestenberg, 2015). According to the California PBIS Coalition, the number of schools adopting PBIS increased from around 500 to more than 3,000 (about 33% of all traditional K–12 schools) from SY 2011–2012 to SY 2018–2019 (California PBIS, n.d.), indicating that some unreformed schools also implement RJ and PBIS programs. While the expansion of RJ and PBIS programs usually requires additional funding, only SFUSD and OUSD among the reformed districts explicitly agreed to provide it (Frey, 2013, 2015).

WDBs prohibited schools from suspending students out of school for willful defiance, requiring that offenders receive class suspension or in-school suspension instead (Frey, 2014). Nevertheless, WDBs did not eliminate the use of OSS for willful defiance for several reasons. First, teachers could suspend students in this category for other reasons, such as exhibiting violent behaviors or using profane language (Lasnover, 2015). Second, while academic outcomes draw intense scrutiny from the public and other authorities, OSS use attracts less attention. This lack of monitoring, training, and accountability systems might reduce schools' willingness to comply with WDB requirements. Third, three out of the four districts (PUSD, SFUSD, and OUSD) implemented WDBs immediately after passage by their education boards. Schools did not have sufficient time to change their faculties' perception that willful defiance was a necessary and sufficient reason to suspend students. Empirical

evidence documented that around 55% of LAUSD teachers still viewed OSS as a legitimate consequence for willful defiance, even 2 years after implementing the ban (Lasnover, 2015).

This study estimates the impact of WDBs on student discipline outcomes using detailed data on OSS by race and suspension reason across four treated and 265 untreated school districts. I test whether (1) suspension reforms were met with full compliance and (2) whether schools substituted willful defiance OSS with nonwillful defiance OSS after WDBs. This work also examines the role of same-race teachers to explore whether student–teacher race match may affect the implementation of WDBs. Because California is a diverse state with large variations in student characteristics across schools, it must be noted that these findings can only be generalized to other California school districts with characteristics similar to those reformed school districts.

Method

Data

I used publicly available school-level data from the California Department of Education from SY 2011–2012 to SY 2018–2019. The data are reported at the school level. The number of OSSs issued was reported for each student racial group across six mutually exclusive categories: violent incidents leading to injury (violent injury), violent incidents that did not lead to injury (violent no injury), weapon possession, incidents related to illicit drugs (drug-related), defiance-only (willful defiance), and others (miscellanies). Supplemental Table A1 (available in the online version of this article) includes detailed definitions of infractions for each suspension category. After dropping schools with incomplete data, schools without data in the pretreatment period, nontraditional high schools, and schools with grades other than Grades 9 to 12, the remaining sample contained 4,730 school-year observations from 638 high schools in 269 school districts (online Supplemental Table A2 shows that the schools excluded due to incomplete data are similar to those schools with complete data).⁵

Table 1 presents summary statistics on OSS rates, measured by the number of OSSs per 1,000 students by treatment status. On average, around 100 OSSs were issued per 1,000 students, including 40 for willful defiance and 60 for nonwillful defiance reasons. Each year, Black students received 211 OSSs per 1,000 students on average, twice as much as White or Hispanic students. Table 1 also shows slightly lower OSS and willful defiance OSS rates in treated districts before the implementation of WDBs. The nonwillful defiance OSS rates of treated and untreated schools were similar before suspension reforms. Columns 4 and 5 indicate that total suspension and willful defiance OSS rates decreased for all students after WDB implementation. However, the overall OSS rate for Black

TABLE 1

Descriptive Statistics on Suspension Rates by Treatment Status

	Treat				
	All	Control	Treated	Before	After
	(1)	(2)	(3)	(4)	(5)
Suspension rate (Overall)	101.25 (116.45)	101.98 (117.39)	77.61 (76.68)	92.63 (90.33)	65.13 (61.00)
Suspension rate (White)	88.60 (104.69)	89.50 (104.99)	59.43 (90.14)	65.75 (110.90)	54.17 (68.63)
Suspension rate (Hispanic)	107.28 (120.87)	108.66 (122.03)	62.18 (57.94)	75.29 (73.89)	51.28 (37.33)
Suspension rate (Black)	211.31 (241.97)	212.06 (244.42)	186.99 (138.94)	200.40 (147.47)	175.84 (131.37)
WD suspension rate (Overall)	40.69 (94.60)	41.53 (95.75)	13.37 (32.35)	26.37 (44.16)	2.56 (7.11)
WD suspension rate (White)	32.74 (78.55)	33.42 (79.38)	10.53 (37.39)	16.61 (50.96)	5.48 (19.09)
WD suspension rate (Hispanic)	44.92 (103.65)	45.97 (104.91)	10.85 (31.73)	21.70 (44.56)	1.83 (5.22)
WD suspension rate (Black)	81.13 (182.26)	82.75 (184.55)	28.42 (54.83)	55.39 (70.50)	6.00 (17.20)
NWD suspension rate (Overall)	60.56 (40.77)	60.45 (40.02)	64.24 (60.39)	66.26 (63.46)	62.56 (58.07)
NWD suspension rate (White)	55.87 (47.14)	56.08 (46.19)	48.90 (71.22)	49.14 (79.42)	48.69 (64.14)
NWD suspension rate (Hispanic)	62.35 (36.00)	62.69 (35.87)	51.33 (38.58)	53.59 (42.67)	49.46 (35.00)
NWD suspension rate (Black)	130.18 (108.00)	129.31 (107.60)	158.57 (117.36)	145.01 (104.96)	169.84 (126.33)
Observation	4,730	4,589	141	64	77
Number of schools	638	619	19	19	19
Number of districts	269	265	4	4	4

Note. Suspension rates are the number of out-of-school suspensions per 1,000 students. Means are reported. Standard deviations are in parentheses. NWD=Nonwillful defiance; WD=Willful defiance.

students only dropped by around 24 per 1,000 students, despite a 50 per 1,000 student reduction in the willful defiance OSSs rate. Such a moderate decline in the total suspension rate was not unexpected: among Black students, nonwillful defiance OSS increased from 145 to 179 per 1,000 students.

Figure 1 plots the trends of overall, willful defiance, and nonwillful defiance OSS rates between SY 2011–2012 and SY 2018–2019. The vertical lines indicate the time of WDB implementation in each treated school district. This visual reveal a few interesting patterns. First, OSS rates vary dramatically across school districts, with PUSD and OUSD maintaining consistently higher levels than SFUSD across all three OSS rates. Furthermore, the average willful defiance OSS rates in the untreated districts were higher than those in the treated districts. Second, willful defiance OSS rates decreased over time in both the treated and untreated districts, while nonwillful defiance OSS rates remained unchanged, perhaps due to rising public concerns about the overuse of willful defiance OSS and the spillover effect of the statewide WDB. Third, Black students were suspended more than White and Hispanic students across all suspension categories. Figure 1 also confirms that schools did not perfectly comply with WDBs, either because of compliance issues or reporting errors (Frey, 2014; Lasnover, 2015). For example, White and Black students together only make up around 4% of

the AUSD sample, with an average range of 0 to 4 willful defiance OSSs per school year. Therefore, the reported increases in this category may be attributed to noise and measurement errors. In addition, these increases indicate that WDBs may be less effective in schools with low willful defiance suspension incident rates due to the high costs of monitoring.

Table 2 presents descriptive statistics on school characteristics by treatment status. In both treated and untreated school districts, Hispanic students made up around 40% of the total population. However, the proportions of Hispanic and White students were smaller in the treated schools than the untreated schools. In addition, while most teachers in the sampled schools were White, the treated schools employed more Black teachers and fewer White teachers than the untreated schools. There were also some small-size variations in student–teacher ratio, with a mean of 23.1, a standard deviation of 3.3, and an overall decrease in the four treated districts after the implementation of WDBs. Last, Table 2 shows that the treated school districts spent more per student and were more likely to have PBIS programs than the untreated schools.

Empirical Strategy

I used a DD strategy to evaluate the causal impact of WDBs on discipline outcomes by comparing the gaps in

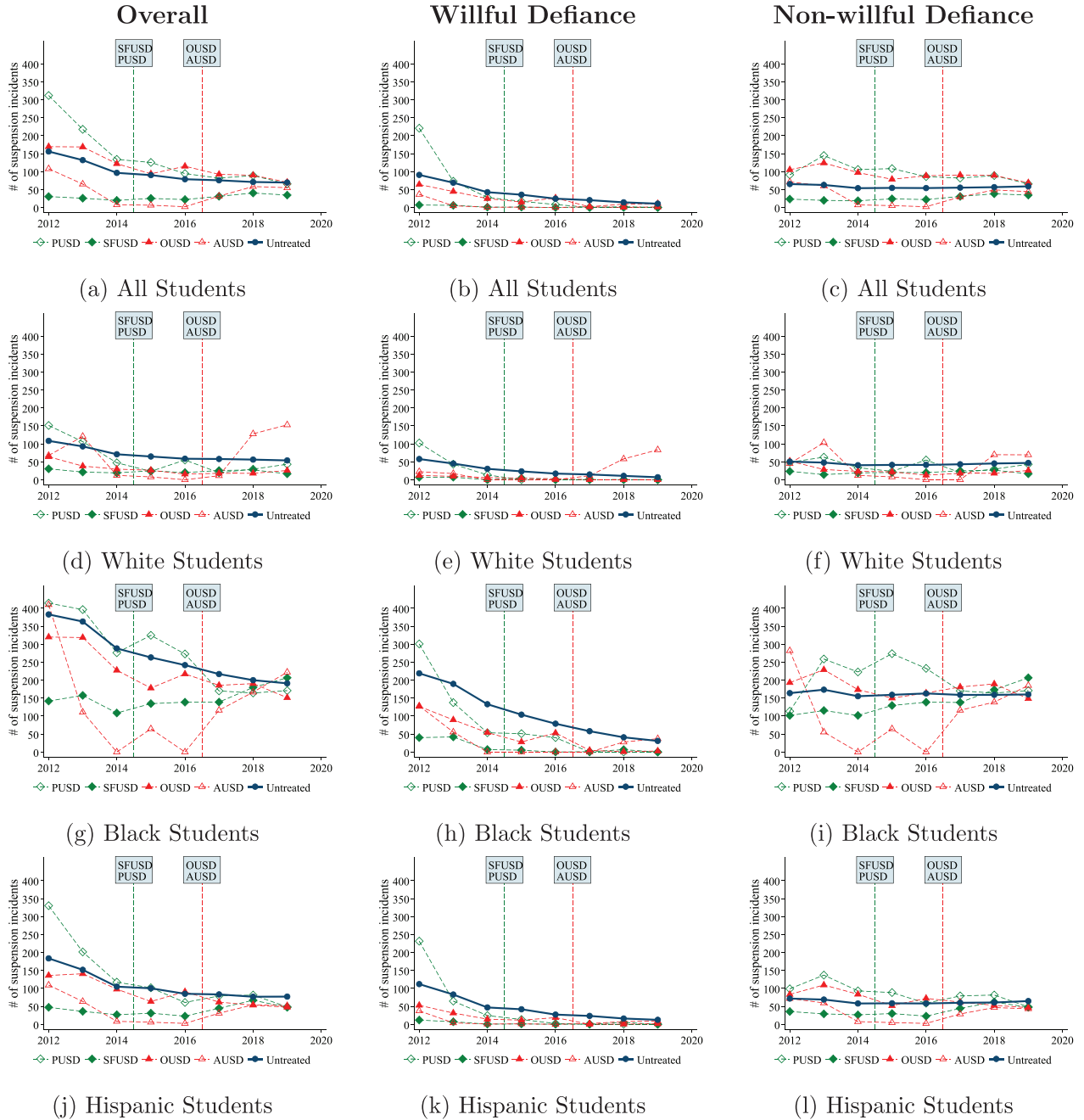


FIGURE 1. Changes in suspension rates over time.

Note. Each dot represents the average suspension rate measured by the number of suspension incidents per 1,000 students. The vertical lines indicate the time of the willful defiance suspension bans. The x-axis indicates the spring of each school year. SFUSD = San Francisco Unified School District; PUSD = Pasadena Unified School District; OUSD = Oakland Unified School District; AUSD = Azusa Unified School District.

OSS rates between the treatment and control districts before and after the implementation of WDBs. The DD strategy relies on the assumption that suspension rates in the treatment and control groups trended in a parallel fashion before the policy change, and that such trends would have

continued if there were no policy change. Therefore, given that my outcome data were reported as count data, I specified a Poisson regression model (Equation 1):

$$E(D_{sdt}) = \exp\left(\alpha_s + \alpha_t + \sum_{k=1}^{k_{\max}} \tau_k \text{Ban}_{d,t-k} + \beta X_{st}\right) \quad (1)$$

TABLE 2

Descriptive Statistics on School Characteristics by Treatment Status

	Treat				
	All	Control	Treated	Before	After
	(1)	(2)	(3)	(4)	(5)
Proportion of White students	0.29 (0.22)	0.30 (0.22)	0.08 (0.06)	0.07 (0.06)	0.08 (0.06)
Proportion of Hispanic students	0.47 (0.25)	0.47 (0.25)	0.39 (0.26)	0.39 (0.27)	0.39 (0.25)
Proportion of Black students	0.06 (0.07)	0.06 (0.06)	0.17 (0.17)	0.19 (0.17)	0.15 (0.16)
Proportion of Other students	0.18 (0.16)	0.17 (0.15)	0.37 (0.27)	0.35 (0.27)	0.38 (0.26)
Proportion of female students	0.51 (0.02)	0.51 (0.02)	0.52 (0.04)	0.52 (0.03)	0.53 (0.04)
Total enrollment	2065.15 (720.52)	2082.94 (715.42)	1486.22 (644.24)	1513.16 (637.07)	1463.83 (653.45)
Proportion of FRPL students	0.45 (0.22)	0.45 (0.22)	0.60 (0.11)	0.62 (0.09)	0.58 (0.12)
Proportion of White teacher	0.69 (0.17)	0.70 (0.17)	0.48 (0.11)	0.49 (0.15)	0.48 (0.08)
Proportion of Hispanic teacher	0.15 (0.10)	0.15 (0.10)	0.15 (0.12)	0.14 (0.12)	0.16 (0.13)
Proportion of Black teacher	0.03 (0.05)	0.03 (0.04)	0.11 (0.12)	0.11 (0.11)	0.11 (0.12)
Proportion of Female teacher	0.54 (0.07)	0.54 (0.07)	0.53 (0.06)	0.53 (0.07)	0.52 (0.06)
Proportion of teachers with advanced degree	0.49 (0.17)	0.50 (0.16)	0.33 (0.17)	0.36 (0.19)	0.30 (0.16)
Student–teacher ratio	23.14 (3.28)	23.26 (3.20)	19.10 (3.32)	19.62 (3.44)	18.67 (3.17)
Per student expenditure	9114.34 (2511.01)	9062.26 (2506.34)	10809.36 (2033.64)	9289.30 (1248.20)	12072.79 (1662.48)
PBIS implementation	0.28 (0.45)	0.27 (0.45)	0.65 (0.48)	0.39 (0.49)	0.86 (0.35)
Observation	4,730	4,589	141	64	77

Note. Means are reported. Standard deviations are in parentheses. FRPL = free or reduced-price lunch; PBIS = positive behavior intervention and support programs.

where D_{sdt} is the discipline outcome, which is the number of OSSs due to a specific reason for school s located in district d in school year t (spring of the indicated school year t). $Ban_{d,t-k}$ are a series of indicator variables equaling one for all schools in district d in the k years after WDB implementation; for schools in districts that never implemented a WDB, all indicator variables equal zero. α_s and α_t are school and year fixed effects that control for systematic differences across schools and secular shocks over time.

In addition, the DD strategy assumes no other contemporaneous policy changes during the treatment period, which may be violated due to the adoption of PBIS and RJ in some school districts. To reduce these concerns, Equation (1) includes a series of time-variant controls (X_{st}) at the school level, such as students' and teachers' racial and gender compositions, student–teacher ratio, percentage of students with free or reduced-price lunch, percentage of teachers with a master's degree or higher, average per-pupil spending, and, most important, whether PBIS was implemented. Average per-pupil spending is included to control for the adoption of any other programs (e.g., RJ), as is a dummy variable measuring the availability of PBIS programs. For the current analyses, RJ program adoption cannot be directly controlled because school-level RJ adoption data are not available.

However, controlling for PBIS program implementation should at least partially address concerns about RJ adoption, as PBIS and RJ programs are usually adopted together (Riestenberg, 2015). Overall, the inclusion of time-variant controls should relax concerns over the influence of contemporaneous policy changes.

For Equation (1), I estimated both an event study (nonparametric) version and a two-way fixed effect (parametric) version of the DD model because the nonparametric model allows treatment effects to vary over time. For example, the slow transition of suspension practices may mean that the first year's treatment effects could be smaller than those in later years.

Equation (1) was also extended into Equation (2) to formally test the validity of the parallel trends assumption (Freyaldenhoven et al., 2019; Lafortune et al., 2018; Lindo & Packham, 2017):

$$E(D_{sdt}) = \exp \left(\alpha_s + \alpha_t + \sum_{k=0}^{k_{max}} \tau_k Ban_{d,t-k} + \sum_{r=r_{min}}^{r=-1} \pi_r Ban_{d,t-r} + \beta X_{st} \right) \quad (2)$$

$Ban_{d,t-r}$ is a series of lead indicators of the treatment and takes a value of one for all schools in district d in r years before WDB implementation. The lead dummies beyond the third year before implementation were combined into one

aggregated dummy indicator. Observations in the first year before WDBs and schools in districts that never adopted WDBs were assigned to the reference group.

Poisson models are preferred when data are measured in counts, and they do not require the transformation of data to accommodate zeros (online Supplemental Table A3 shows that around 5% of schools reported zero total and nonwillful defiance suspension rates, and a relatively larger proportion of schools reported zero willful defiance suspensions rates; Lindo & Packham, 2017; McClellan & Tekin, 2017; Osgood, 2000; Wooldridge, 2010). The Poisson estimates can be interpreted as changes in suspension rates by including school enrollments as controls and restricting their coefficients equal to one.

In addition, recent research suggests that DD analysis with staggered treatment timing can lead to biased estimates when treatment effects vary by cohorts (Goodman-Bacon, 2021; Sun & Abraham, 2020). While Goodman-Bacon (2021) proposed using decomposition to check the source of bias in DD coefficients, decomposition can only be applied to a linear model. Therefore, I calculated the interaction-weighted (IW) estimators, following Sun and Abraham (2020), to check for biases due to staggered timing in policy adoption.

Last, standard errors were clustered at the district-year level to allow within cluster correlation of error terms (results using cluster-robust standard errors at the district level are presented in the online Supplemental Appendix; Abadie et al., 2017). However, because only four school districts were treated, the cluster-robust standard errors might over reject the null hypothesis. Recent research suggests that p values from subcluster wild bootstrap (WR) and randomization inference (RI) procedures perform better when the number of treated groups is substantially smaller than that of untreated groups in DD designs (Conley & Taber, 2011; MacKinnon & Webb, 2018, 2020; Roodman et al., 2019; Young, 2019). Therefore, I also calculated p values using both the RI method and the subcluster restricted WR method at the district-year level. Specifically, I calculated randomization inference p values based on both coefficients (RI-c) and t statistics (RI-t), in line with Young (2019) and Mackinnon and Webb (2020), who found RIs based on t statistics to be more robust. For each p value, I followed the procedure mentioned in Heß (2017) and Pfeifer et al. (2020) with 1,000 permutations. Even though RI p values are superior to the cluster-robust standard errors, Mackinnon and Webb (2020) showed that RI p values under reject the null when the size of the treated groups is larger than that of the untreated groups. In this analysis, the treated districts contain 35.25 school-year observations, compared with 17.31 in untreated districts. Therefore, it needs to be noted that the RI p values in this study may lead to the underrejection of the null hypotheses.

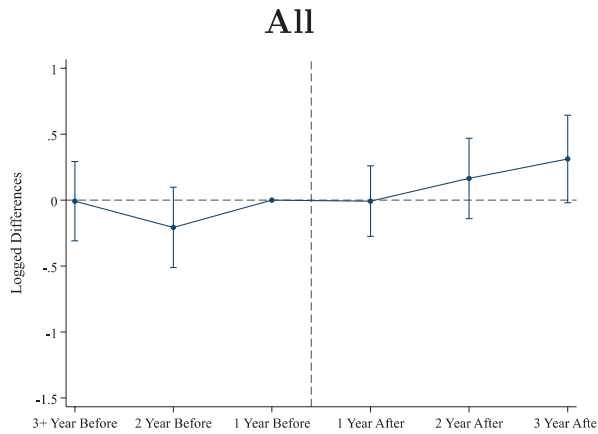
Results

WDBs and Suspension

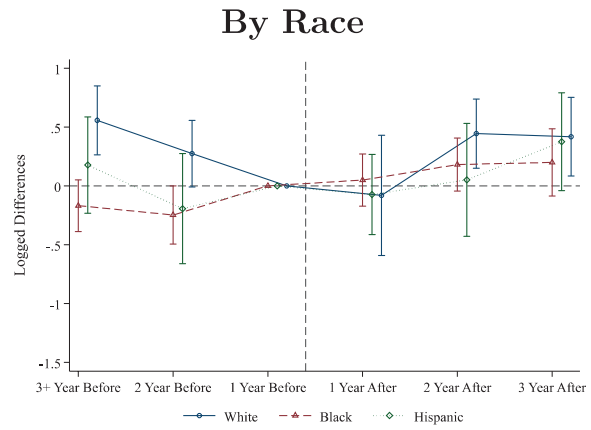
Critically, the DD design assumes that districts with WDBs and those without should have similar trends in suspension rates. Before presenting the main results, Figure 2 plots the event study coefficients and the 95% confidence intervals using Equation (2). Figure 3 presents the event study coefficients based on IW estimators proposed by Sun and Abraham (2020). Figures 2 and 3 show no pretrends for all event study figures except for White students. Notably, the treated school districts implemented WDBs because existing suspension reduction programs failed to reduce suspension rates among Black and Hispanic students. The pretrend in outcomes for White students could be attributed to those early efforts in reducing suspension rates. In estimating Equation (1), I accounted for this differential trend in White students' outcomes across treated and untreated districts by extrapolating a linear trend from the two periods immediately preceding the WDBs following the practice of Dobkin et al. (2018) and Freyaldenhoven et al. (2019; see online Supplemental Table A5 for a reestimation of the main results using this alternative specification, which are consistent with the main results).⁷

Panel A of Table 3 displays the nonparametric DD estimates of Equation (2) as well as WR p values by race and reason. Estimates in Columns 1 to 3 indicate that while WDBs significantly reduced willful defiance OSS rates, they had little effect on overall and nonwillful defiance OSS rates, as suggested by the large p values. Particularly, the nonparametric estimates reflect the dynamic changes in willful defiance OSS rates after the implementation of WDBs; willful defiance OSS rates saw a relatively small decrease in the first year (67% decrease, $e^{-1.123} - 1 \approx -0.67$) and a larger decrease in later years (69% decrease, $e^{-1.184} - 1 \approx -0.69$, $e^{-1.182} - 1 \approx -0.69$, in the second year and after). Panel B of Table 3 presents the parametric DD estimates as well as WR, RI-c, and RI-t p values. The results in Columns 1 to 3 are consistent with the nonparametric estimates; Column 2 of Panel B reports that WDB decreased willful defiance OSS rates by 69% ($e^{-1.158} - 1 \approx -0.69$) on average, which is also at least statistically significant at the 10% level, except for RI-t p value ($p \approx .103$).

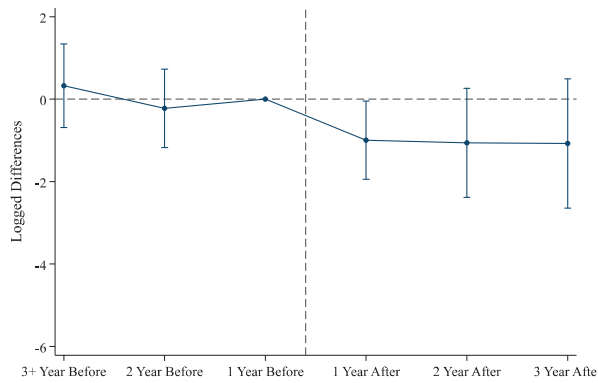
The remaining columns in Panels A and B explore the heterogeneous impact of WDBs across student racial groups. According to the nonparametric estimates in columns 4 and 6, the changes in overall OSS rates for White and Hispanic students are not significant at the conventional level ($p > .05$). However, overall OSS rates for Black students increased by around 22% to 41% ($e^{0.199} - 1 \approx 0.22$, $e^{0.346} - 1 \approx 0.41$) following the implementation of WDBs. These results are similar to the parametric DD estimates (34% increase, $e^{0.29} - 1 \approx 0.34$) in Panel B. Although the estimated coefficients are not statistically significant based on the RI-c and



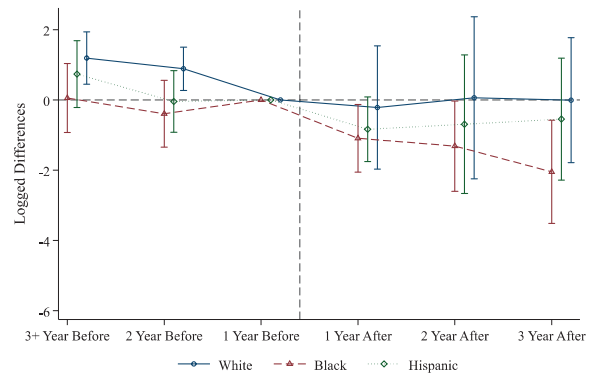
(a) Overall Suspension Rates



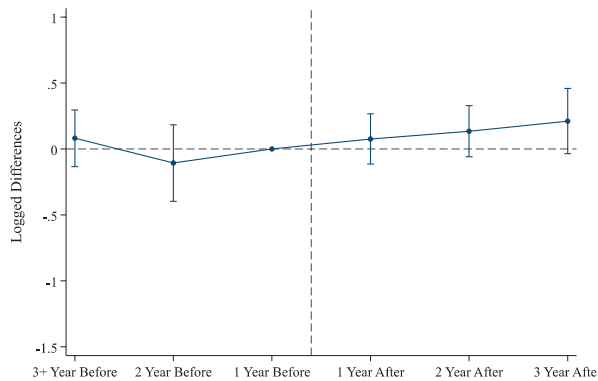
(b) Overall Suspension Rates



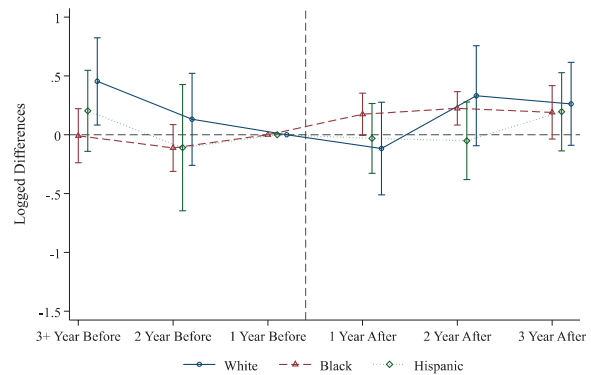
(c) Willful Defiance Suspension Rates



(d) Willful Defiance Suspension Rates



(e) Non-willful Defiance Suspension Rates



(f) Non-willful Defiance Suspension Rates

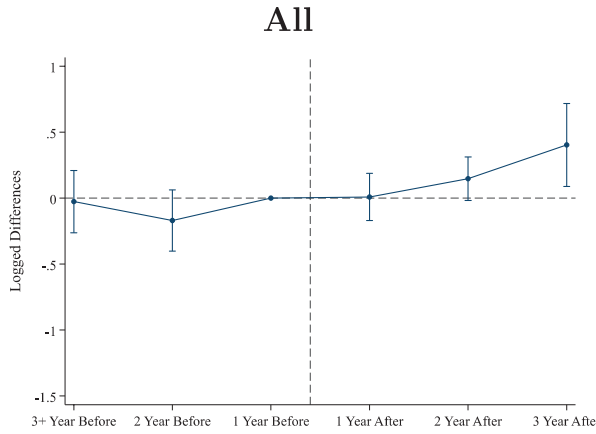
FIGURE 2. *Event study figures.*

Note. Each subfigure presents the coefficients estimated from Poisson regressions and 95% confidence intervals computed using analytical standard errors.

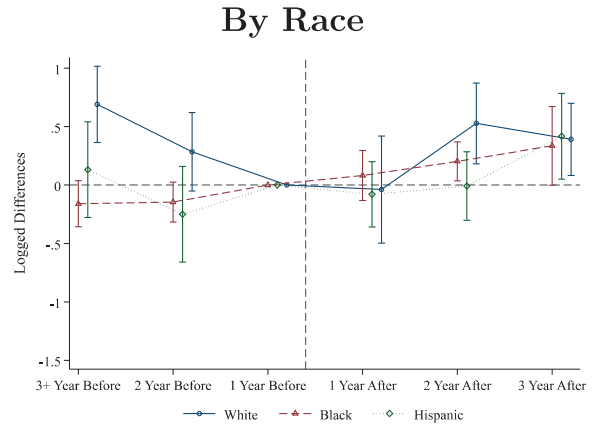
RI- t p values, the WR p values are statistically significant at the 1% level for the parametric DD estimates.

According to results in columns 7 to 9 of both Panels A and B, willful defiance OSS rates fell sharply compared with the contemporary changes in other districts for students of all races. Specifically, according to the parametric DD

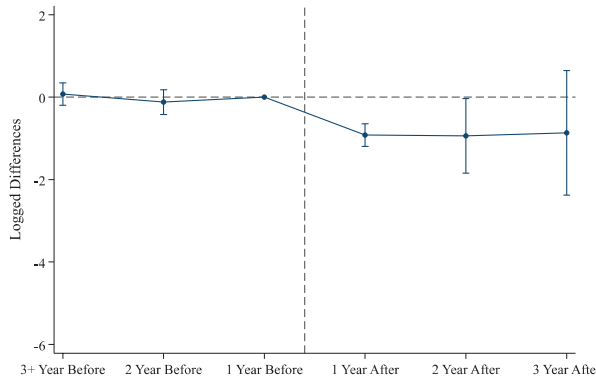
estimates, willful defiance OSS rates decreased by 78% for White students, 73% for Black students, and 67% for Hispanic students ($e^{-1.538} - 1 \approx 0.78$, $e^{-1.296} - 1 \approx 0.73$, $e^{-1.114} - 1 \approx 0.67$). Although most of the RI p values are larger than 0.1, the RI p values on the estimated coefficient for Black students are consistently lower than



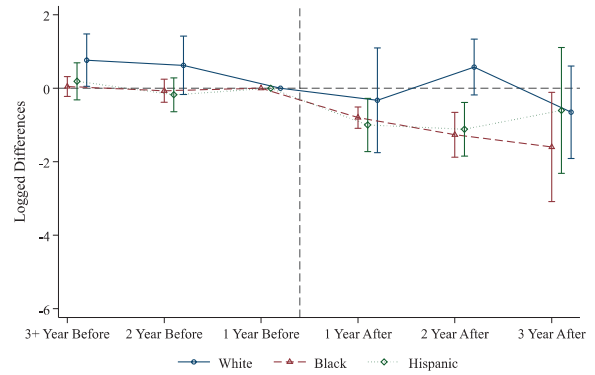
(a) Overall Suspension Rates



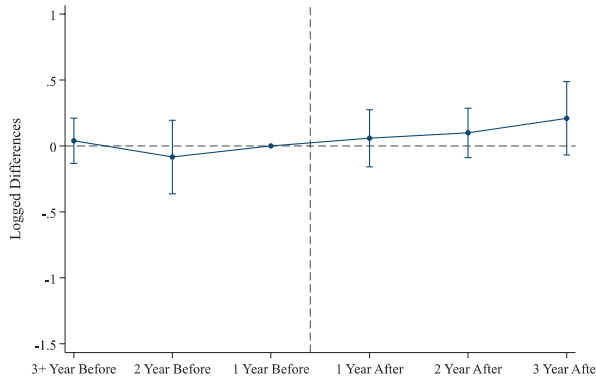
(b) Overall Suspension Rates



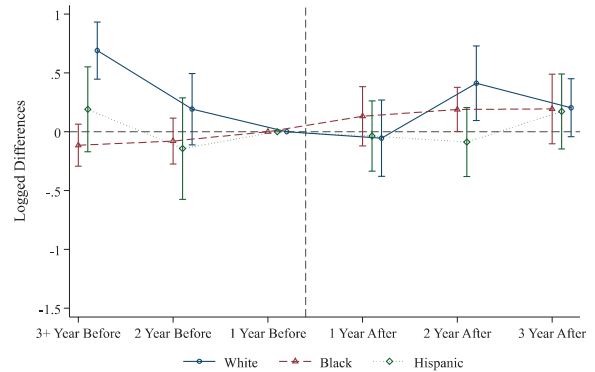
(c) Willful Defiance Suspension Rates



(d) Willful Defiance Suspension Rates



(e) Non-willful Defiance Suspension Rates



(f) Non-willful Defiance Suspension Rates

FIGURE 3. *Event study figures using interaction-weighted (IW) estimates.*

Note. Each subfigure presents the IW coefficients estimated from Poisson regressions and 95% confidence intervals computed using analytical standard errors.

0.1, suggesting that the WDBs were particularly effective in reducing Black student willful defiance OSS rates.

Estimates for nonwillful defiance OSS rates by race are displayed in columns 10 to 12 of Panels A and B. The non-parametric estimates imply that nonwillful defiance OSS rates for White students decreased by around 32%

($e^{-0.396} - 1 \approx -0.32$) immediately after the implementation of WDBs, but this drop vanished in later years. Surprisingly, nonwillful defiance OSS rates for Black students increased by around 23, 30, and 25% ($e^{0.211} - 1 \approx 0.23$, $e^{0.260} - 1 \approx 0.30$, $e^{0.227} - 1 \approx 0.25$) in the first, second, and

TABLE 3
The Impact of Suspension Bans on Suspension Rates

	By race											
	Total			Overall			Willful			Nonwillful		
	Overall	Willful	Nonwillful	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
<i>Panel A: Nonparametric model</i>												
1 Year after	0.052 (0.100)	-1.123*** (0.228)	0.067 (0.082)	-0.501* (0.279)	0.199** (0.099)	-0.122 (0.126)	-1.534 (0.953)	-1.009*** (0.259)	-1.273*** (0.307)	-0.396** (0.168)	0.211** (0.091)	-0.103 (0.115)
WR	[0.633]	[0.118]	[0.472]	[0.121]	[0.211]	[0.400]	[0.225]	—	[0.176]	[0.024]	[0.147]	[0.403]
RI-c	[0.847]	[0.136]	[0.691]	[0.160]	[0.522]	[0.704]	[0.106]	[0.197]	[0.125]	[0.137]	[0.393]	[0.605]
RI-t	[0.780]	[0.064]	[0.619]	[0.290]	[0.305]	[0.615]	[0.410]	[0.119]	[0.092]	[0.132]	[0.183]	[0.617]
2 Years after	0.223* (0.125)	-1.184** (0.523)	0.125 (0.082)	-0.019 (0.252)	0.327*** (0.109)	0.001 (0.208)	-1.555 (1.377)	-1.229** (0.519)	-1.126 (0.920)	0.033 (0.216)	0.260*** (0.075)	-0.124 (0.132)
WR	[0.221]	[0.159]	[0.271]	[0.937]	[0.110]	[0.996]	[0.287]	—	[0.401]	[0.896]	[0.065]	[0.414]
RI-c	[0.465]	[0.111]	[0.523]	[0.963]	[0.386]	[0.998]	[0.163]	[0.145]	[0.159]	[0.895]	[0.37]	[0.589]
RI-t	[0.342]	[0.300]	[0.413]	[0.969]	[0.194]	[0.998]	[0.547]	[0.305]	[0.569]	[0.906]	[0.104]	[0.630]
3+ Years after	0.375*** (0.145)	-1.182* (0.691)	0.205* (0.113)	-0.010 (0.208)	0.346*** (0.132)	0.336* (0.183)	-1.535* (0.885)	-1.950*** (0.624)	-0.958 (0.820)	-0.016 (0.156)	0.227** (0.115)	0.130 (0.131)
WR	[0.022]	[0.150]	[0.070]	[0.960]	[0.024]	[0.090]	[0.180]	[0.027]	[0.287]	[0.920]	[0.034]	[0.314]
RI-c	[0.281]	[0.133]	[0.315]	[0.982]	[0.390]	[0.382]	[0.198]	[0.034]	[0.237]	[0.963]	[0.367]	[0.594]
RI-t	[0.244]	[0.430]	[0.324]	[0.978]	[0.226]	[0.397]	[0.326]	[0.182]	[0.572]	[0.958]	[0.215]	[0.631]
<i>Panel B: Parametric model</i>												
Average effects	0.236** (0.103)	-1.158*** (0.328)	0.144* (0.074)	-0.195 (0.216)	0.290*** (0.095)	0.125 (0.141)	-1.538** (0.744)	-1.296*** (0.308)	-1.114** (0.465)	-0.143 (0.152)	0.231*** (0.078)	0.006 (0.103)
WR	[0.038]	[0.017]	[0.082]	[0.399]	[0.005]	[0.375]	[0.124]	—	[0.087]	[0.343]	[0.008]	[0.964]
RI-c	[0.411]	[0.077]	[0.402]	[0.578]	[0.363]	[0.702]	[0.107]	[0.061]	[0.117]	[0.542]	[0.288]	[0.974]
RI-t	[0.269]	[0.103]	[0.281]	[0.610]	[0.139]	[0.677]	[0.236]	[0.055]	[0.254]	[0.537]	[0.069]	[0.974]
Observations	4,730	4,720	4,730	4,726	4,709	4,727	4,630	4,429	4,680	4,720	4,694	4,727

Note. Panel A reports results from the nonparametric DD model using Equation (1); Panel B reports parametric DD results. Results for White students are estimated using models with a pretrend included. All models include covariates, time fixed effects, and school fixed effects. Covariates include the percentage of White, Black, Hispanic, and female students, percentage of students receiving free or reduced-price lunch, percentage of teachers who are White, Black, Hispanic, female, and hold a master's degree, per-student expenditure, PBIS implementation, and student-teacher ratio. Two dummy variables are included to control for the missing in-school expenditure and the student-teacher ratio. Subcluster restricted wild bootstrap (WR) p values clustered at the district-year level and randomized p values (RI-c and RI-t) are shown in square brackets. Cells with missing WR p value indicate the WR process fails to generate p values. Robust standard errors and p values clustered at the district-year level are shown in parentheses and as asterisks. RI-t = randomization inference p values based on coefficients; RI-c = randomization inference p values based on t statistics; DD = difference-in-differences; PBIS = positive behavior intervention and support programs.
* $p < .1$. ** $p < .05$. *** $p < .01$.

third+ years after adoption. The parametric DD estimates suggest the positive effect of WDB on Black students' non-willful defiance OSS rates is 26% ($e^{0.231} - 1 \approx 0.26$) and are statistically significant (both the analytical and WR p values are less than .05; the RI- t p values is only slightly larger than .05). One may intuit that nonwillful defiance OSSs became substitutes for willful defiance OSSs after the implementation of WDBs (see online Supplemental Table A4 for additional evidence that violent injury and miscellaneous OSSs may replace willful defiance OSSs). Table 4 also presents the IW estimates, following Sun and Abraham (2020). These coefficients are similar to those in Table 3, suggesting that different weightings across treatment cohorts do not affect the results of the current study. Online Supplemental Table A5 presents results using alternative specifications, online Supplemental Table A6 presents results using alternative bins, and online Supplemental Table A7 presents results without controls, all of which report similar coefficients.

Table 5 presents differences in the parametric DD coefficients across each racial group.⁶ For all three OSS rates, I tested differences in the treatment effects between Black and White students, Black and Hispanic students, and Hispanic and White students. The effects of WDBs on nonwillful defiance OSS rates are significantly different between Black and White students and Black and Hispanic students ($p < .05$ for both analytical and WR p values). The treatment effects on overall OSS rates were also different between Black and White students ($p < .05$ for analytical p value, $p < .06$ for WR p value). However, Table 5 shows no evidence that the effects of WDBs on willful defiance OSS rates vary by race. Specifically, after WDBs were implemented, the nonwillful defiance OSS rates for Black students significantly increased, by 64% ($e^{0.497} - 1 \approx 0.64$) relative to changes in suspensions rates for White students, and by 25% ($e^{0.224} - 1 \approx 0.25$) relative to changes in suspensions rates for Hispanic students in the same school districts.

The Presence of Same-Race Teachers

If the substitution of suspension reasons was used to cope with WDBs, the prevalence of the strategy might depend on the characteristics of referring teachers. Research on student-teacher racial match has shown that the presence of same-race teachers could improve students' academic and behavioral outcomes (Dee, 2004, 2005; Gershenson & Papageorge, 2018; Holt & Gershenson, 2019; Papageorge et al., 2020). Redding (2019) summarized three pathways for this phenomenon. First, same-race teachers may hold higher expectations of same-race students than do teachers of other races (Gershenson et al., 2016; Grissom et al., 2015; Ouazad, 2014), adjust instructions to meet the needs of same-race students (Aronson & Laughter, 2016), and build stronger relationships with students and their parents (Saft & Pianta, 2001). Second, students may respond more favorably to same-race teachers and learn from them to overcome

negative racial stereotypes (Dee & Penner, 2019; Steele, 1997). Third, same-race teachers may advocate for changes in school policies and teacher behavior, benefiting students in their racial groups (Grissom et al., 2009).

Therefore, a shared cultural understanding motivates teachers to act or think in a way that could benefit students of their races in making disciplinary decisions. For example, teachers are less likely to escalate a negative response to the behavior of students of their own races (Okonofua & Eberhardt, 2015). However, the extent to which teachers could actively benefit same-race students depends on their ability and desire. Empirical research in other fields has shown such desire and ability are limited by factors such as organizational culture, socialization, and policy environment (Keiser et al., 2002; Wilkins & Williams, 2008). Before the implementation of WDBs, suspending disruptive students was the default and a consensus among school employees (Lasnover, 2015). Under such norms, teachers who preferred to avoid suspending disruptive students might have been pressured by their colleagues and/or risked violating school discipline policies. The implementation of WDBs, thus, served as a nudge and reduced the social and mental costs that such teachers faced before reforms. Although WDBs may also change the use of OSS among teachers of a different race, the lack of cultural understanding and racial stereotypes may encourage coping behavior during implementation by suspending students using other suspension reasons (Lasnover, 2015). Thus, we expect same-race teachers to use willful defiance suspensions and cite other suspension reasons as substitutes less frequently than do teachers of other races. These favorable behaviors among teachers would trigger student behavioral improvements and further reduce overall OSS rates among same-race students. Consequently, I hypothesize that same-race teachers mitigate the negative impacts of WDBs on students.

Table 6 examines how the impact of WDBs on student discipline outcomes varies by the presence of same-race teachers.⁸ Each cell reports a coefficient on an interaction term between the WDB dummy and the percentage of same-race teachers in each school. The results show that increases in the percentage of White and Hispanic teachers did not lead to additional decreases in OSS rates for White and Hispanic students. This may follow from the fact that White and Hispanic students represent majorities and are thus subject to fewer perceptual biases in California (Riddle & Sinclair, 2019; Stewart et al., 2009). Yet, for Black students, 1 percentage point increase in Black teachers was associated with a 0.8% ($e^{0.008} - 1 \approx 0.008$) decrease in nonwillful defiance OSS rates after the WDBs ($p < .1$ for analytical and WR p values). Online Supplemental Table A8 shows that, compared with Black students in the same districts but at a different school, Black students' OSS rates for violent no injury, miscellaneous, and drug-related incidents also decreased as the percentage of Black

TABLE 4
The Impact of Suspension Bans on Suspension Rates Using IW Estimators

	By race															
	Total				Overall				Willful				Nonwillful			
	Overall	Willful	Nonwillful		White	Black	Hispanic		White	Black	Hispanic		White	Black	Hispanic	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	
1 Year after	0.032 (0.092)	-1.180*** (0.194)	0.079 (0.088)	-0.469* (0.267)	0.167* (0.099)	-0.139 (0.117)	-1.290 (0.806)	-1.048*** (0.200)	-1.340*** (0.383)	-0.397** (0.188)	0.227* (0.119)	-0.094 (0.111)				
RI-c	[0.897]	[0.094]	[0.656]	[0.173]	[0.580]	[0.642]	[0.144]	[0.124]	[0.089]	[0.133]	[0.388]	[0.632]				
RI-t	[0.878]	[0.067]	[0.674]	[0.360]	[0.436]	[0.594]	[0.469]	[0.084]	[0.217]	[0.263]	[0.342]	[0.688]				
2 Years after	0.177** (0.087)	-1.154** (0.487)	0.113 (0.069)	0.084 (0.278)	0.289*** (0.084)	-0.061 (0.120)	0.366 (0.416)	-1.475*** (0.348)	-1.407*** (0.386)	0.039 (0.223)	0.275*** (0.084)	-0.157 (0.099)				
RI-c	[0.561]	[0.120]	[0.572]	[0.846]	[0.386]	[0.863]	[0.715]	[0.036]	[0.092]	[0.886]	[0.335]	[0.493]				
RI-t	[0.388]	[0.401]	[0.444]	[0.873]	[0.144]	[0.835]	[0.663]	[0.154]	[0.205]	[0.904]	[0.152]	[0.490]				
3+ Years after	0.367** (0.158)	-1.436* (0.788)	0.259** (0.106)	0.033 (0.214)	0.357** (0.149)	0.298 (0.192)	-1.927** (0.912)	-2.172*** (0.773)	-1.259 (0.890)	0.027 (0.154)	0.323** (0.127)	0.150 (0.132)				
RI-c	[0.305]	[0.091]	[0.252]	[0.944]	[0.362]	[0.443]	[0.111]	[0.006]	[0.161]	[0.934]	[0.250]	[0.537]				
RI-t	[0.374]	[0.481]	[0.252]	[0.940]	[0.283]	[0.536]	[0.271]	[0.230]	[0.589]	[0.932]	[0.153]	[0.628]				

Note. Results are estimated using IW methods; results for White students are estimated using models with a pretrend included. All models include covariates, time fixed effects, and school fixed effects. Covariates include the percentage of White, Black, Hispanic, and female students, percentage of students receiving free or reduced-price lunch, percentage of teachers who are White, Black, Hispanic, female, and hold a master's degree, per-student expenditure, PBIS implementation, and student-teacher ratio. Two dummy variables are included to control for the missing in-school expenditure and the student-teacher ratio. Randomized p values (RI-c and RI-t) are shown in square brackets. Cells with missing WR p value indicate the WR process fails to generate p values. Robust standard errors and p values clustered at the district-year level are shown in parentheses and as asterisks. IW = interaction-weighted. RI-c = interaction-weighted. RI-t = interaction-weighted. RI-c = randomization inference p values based on coefficients; RI-t = randomization inference p values based on t statistics; PBIS = positive behavior intervention and support programs.

* $p < .1$. ** $p < .05$. *** $p < .01$.

TABLE 5
Differential Impacts of Suspension Bans Across Student Subgroups

	Overall	Willful	Nonwillful
	(1)	(2)	(3)
Black vs. White	0.547** (0.229)	-0.275 (0.809)	0.497** (0.195)
WR <i>p</i> value	[.041]	—	[.021]
Hispanic vs. White	0.264 (0.189)	-0.430 (0.687)	0.158 (0.181)
WR <i>p</i> value	—	—	[.427]
Black vs. Hispanic	0.164 (0.125)	-0.192 (0.452)	0.224** (0.094)
WR <i>p</i> value	[.242]	—	[.039]

Note. This table compares the differences in treatment effects between Black and White, Black and Hispanic, and Hispanic and Black students. The coefficient in each cell is estimated from a separate regression. The coefficients are the interaction terms between treatment and race dummies for models that all controls are fully interacted with race dummies on a sample combined from two race groups. All models include covariates, time fixed effects, and school fixed effects. A pretrend is included when White students are included in the comparison. Covariates include the percentage of White, Black, Hispanic, and female students, percentage of students receiving free or reduced-price lunch, percentage of teachers who are White, Black, Hispanic, female, and hold a master’s degree, per-student expenditure, PBIS implementation and student-teacher ratio. Two dummy variables are included to control for the missing in-school expenditure and the student-teacher ratio. Subcluster restricted wild bootstrap (WR) *p* values clustered at the district-year level are shown in brackets. Cells with missing WR *p* value indicate the WR process fails to generate *p* values. Robust standard errors and *p* values clustered at the district-year level are shown in parentheses and as asterisks.

p* < .1. *p* < .05. ****p* < .01.

TABLE 6
The Presence of Same-Race Teacher and the Impact of WDBs

	Total	Willful	Nonwillful
	(1)	(2)	(3)
Treat × % White teacher	-0.014 (0.019)	-0.139 (0.209)	-0.004 (0.014)
WR <i>p</i> value	[.474]	[.620]	[.765]
Treat × % Black teacher	-0.005 (0.006)	0.009 (0.023)	-0.008* (0.004)
WR <i>p</i> value	[.433]	[.780]	[.084]
Treat × % Hispanic teacher	-0.003 (0.008)	0.097* (0.057)	-0.002 (0.007)
WR <i>p</i> value	[.750]	[.238]	[.822]

Note. Coefficients in this table are the interaction terms between the treatment dummies and the percentage of White, Black, and Hispanic teachers, estimated from a separate regression. All models include covariates, time fixed effects, and school fixed effects. A pretrend is included when White students are included in the comparison. Covariates include the percentage of White, Black, Hispanic, and female students, percentage of students receiving free or reduced-price lunch, percentage of teachers who are White, Black, Hispanic, female, and hold a master’s degree, per-student expenditure, PBIS implementation and student-teacher ratio. Two dummy variables are included to control for the missing in-school expenditure and the student-teacher ratio. Subcluster restricted wild bootstrap (WR) *p* values clustered at the district-year level are shown in brackets. Cells with missing WR *p*-value indicate the WR process fails to generate *p* values. Robust standard errors and *p* values clustered at the district-year level are shown in parentheses and as asterisks. WDBs = willful defiance suspension bans; PBIS = positive behavior intervention and support programs.

p* < .1. *p* < .05. ****p* < .01.

teachers increased. Recent research also suggests that the effects of student-teacher race match grow even larger once the proportion of teachers reaches a “critical mass” (Grissom et al., 2017). The overall effects of same-race teacher presence are small and not statistically significant, possibly because the proportions of Black teachers in these schools are below the level of “critical mass.” Therefore, current estimates in the student-teacher race match analysis provide a lower bound of student-teacher race match effects and are limited to schools with a lower proportion of Black teachers.

Robustness Checks

I checked the robustness of the main results by reestimating the parametric models in Panel B of Table 3 using ordinary least squares (OLS) and weighted least square (WLS), as both are considered equivalent to the Poisson model. However, the natural log of suspension rates is undefined for some outcomes, because some observations contain zero OSS rates. I addressed this issue by transforming the OSS rates per 1,000 students using the inverse hyperbolic function. WLS estimates were weighted by student enrollments

in each racial group. I also followed Dufló (2001) and Bhuller et al. (2013) in interacting the districts' average baseline outcomes and covariates in SY 2011–2012 with a linear time trend for each school district. These specifications allow the implementation of WDBs to be related to the underlying time trends depending on the outcomes or the characteristics of the districts before the implementation of WDBs. The specification for the baseline outcomes is

$$E(Y_{sdt}) = \exp(\alpha_s + \alpha_t + \delta WDB_{dt} + \gamma_j t * Y_{k,2012} + \beta X_{st}) \quad (3)$$

and the specification for the baseline covariates is

$$E(Y_{sdt}) = \exp\left(\alpha_s + \alpha_t + \delta WDB_{dt} + t \sum_{j=k} \gamma_j X_{k,2012} + \beta X_{st}\right) \quad (4)$$

Last, I added LAUSD and dated its WDB to 2013, the year of its formal announcement, to test for sensitivity to the inclusion of this earliest adopter.

Online Supplemental Tables A10 to A12 in present results from these robustness checks, along with the main results for overall, willful defiance, and nonwillful defiance OSS rates. While the OLS and WLS estimates in the online Supplemental Tables A10 and A12 are less precise, both report similar coefficients to the main results (OLS and WLS results were omitted for willful defiance OSS rates, since an excessive number of schools reported zero willful defiance suspensions after implementing WDBs). Columns 4 and 5 in online Supplemental Tables A10 to A12 present the estimates based on Equations (3) and (4). Again, these estimates are similar, indicating that the main results are robust to including interacted time trends with baseline outcome and covariates. Last, the results in column 6 show that the main results are not sensitive to adding LAUSD in the analysis.

Student-Level Analysis

Data and Empirical Strategy

The results in previous section suggest a potential explanation for the unintended increase in nonwillful defiance OSS rates for Black students after the implementation of WDBs: school employees (e.g., teachers) continued suspending students for these infractions using different labels. While the increased presence of same-race teachers could mitigate such undesirable consequences, deterrence theory implies that increases in nonwillful defiance suspension rates after WDBs may be attributed to student behavioral changes (Nagin, 2013; Pesta, 2018).

To empirically test whether changes in OSS rates were due to behavioral changes, the main analysis was supplemented with analysis of individual-level student data from the biannual High School Youth Risk Behavior Surveys (YRBS), in LAUSD, SFUSD, and San Diego Unified School District (SDUSD), between 2001 and 2017. These YRBSs

are part of the YRBSS, developed in 1990, to monitor health behaviors among youth and young adults, including those that contribute to unintentional injuries and violence. YRBSs include national surveys administered by the Centers for Disease Control and Prevention, state surveys conducted by state governments, and local surveys administered by local governments (e.g., local school districts). The district-level YRBS data in this study are from three of the 28 local YRBSs across the United States (LAUSD, SFUSD, and SDUSD). All three school districts have YRBS data before the implementation of WDBs, with a survey response rate above 60%. In the three included school districts, YRBSs are implemented among representative district samples of 9th-through 12th-grade students, during the spring semester, every 2 years between 2001 and 2017. This article uses data from the 2001, 2003, 2005, 2007, 2009, 2011, 2013, and 2017 surveys.

I measured the level of disruptiveness by creating a series of dummy variables, including whether students were (1) involved in a fight on school property, (2) brought weapons to school, (3) were offered illicit drugs on school property, and (4) used marijuana based on survey questions (see online Supplemental Appendix B for the original questions in the YRBS survey). Overall, across 44,577 observations, Hispanic students made up about 43% of respondents, and Black and White students accounted for only about 7% and 12%, respectively. The remaining respondents represented other minority groups (see online Supplemental Table A13 for the characteristics of surveyed students).

I used a parametric DD strategy and estimated a linear probability model to measure changes in student behavior after the implementation of WDBs:

$$E(Y_{idt}) = \alpha_d + \alpha_t + \delta Ban_{dt} + \beta X_{it} \quad (5)$$

where i indexes individuals, d indexes school districts, and t indexes the year of YRBS surveys. Y_{idt} are the outcome variables, which are a series of self-reported behavior dummies. Because privacy issues affected the availability of school identifiers for district-level YRBSs, Equation (5) only includes district fixed effects α_d . α_t are year fixed effects. Ban_{dt} equals one for school districts that implemented the WDBs, otherwise Ban_{dt} equals zero. X_{it} is a series of controls on students' age, grade, gender, and race. To interpret δ as the causal impact of WDBs on students' behavior, Equation (5) needs to meet the same parallel-trends and no-contemporaneous-shock assumptions as the models in the previous section. Specifically, event study results in the online Supplemental Appendix C suggest no violation of the parallel trends assumption. WR p values were reported along with analytical p values to avoid potential biases in the standard errors (MacKinnon & Webb, 2018). All standard errors were clustered at the district-year level.

TABLE 7
The Impact of WDBs on Students' Behavior by Race

	All	White	Black	Hispanic
	(1)	(2)	(3)	(4)
Panel A: Physical fight in schools ^a				
Sample mean	12.26	8.47	17.68	13.01
Treat	-0.011 (0.009)	-0.010 (0.009)	-0.013 (0.027)	-0.008 (0.016)
WR <i>p</i> value	[.369]	[.555]	[.769]	[.740]
Panel B: Carry weapons in schools ^b				
Sample mean	4.50	3.50	5.41	4.74
Treat	-0.011*** (0.003)	0.008 (0.010)	-0.034 (0.022)	-0.005 (0.007)
WR <i>p</i> value	[.023]	[.665]	[.280]	[.631]
Panel C: Offered illicit drugs in schools ^a				
Sample mean	33.80	32.63	30.08	36.30
Treat	-0.050*** (0.013)	0.013 (0.021)	0.017 (0.028)	-0.073*** (0.019)
WR <i>p</i> value	[.021]	[.715]	[.697]	[.020]
Panel D: Used marijuana ^b				
Sample mean	19.67	23.03	22.73	20.66
Treat	-0.021** (0.008)	-0.007 (0.016)	-0.051 (0.035)	-0.039** (0.015)
WR <i>p</i> value	[.081]	[.775]	[.356]	[.103]

Note. Means are based on weighted data from the LAUSD, SFUSD, and SDUSD YRBS for available years. Some questions were not asked in certain years. Each coefficient is estimated using a separate linear probability model that controls students' age, grade, sex, race, year fixed effects, and district fixed effects. Standard errors are calculated using clustered robust standard errors at the district-year level shown in parentheses. The subcluster restricted WR *p* value at the district-year level are present in brackets. Asterisks are based on analytical *p* value clustered at district-year level. LAUSD = Los Angeles Unified School District; SFUSD = San Francisco Unified School District; SDUSD = San Diego Unified School District; YRBS = Youth Risk Behavior Surveys; WR = wild bootstrap.

^aIn the past 30 days. ^bWithin the last year or the past 12 months.

p* < .1. *p* < .05. ****p* < .01.

The Results From the YRBS Data

Table 7 presents the sample averages and the treatment effects estimated by Equation (5) on self-reported behavior by race. As shown in Panel A, while 12% of all students were involved in a fight on school property, Black students were around nine percentage points more likely to be involved in a fight than White students. The DD estimates consistently show that the implementation of WDBs led to decreases in the likelihood of being involved in a fight for White, Black, and Hispanic students, although the estimates are not statistically significant. Online Supplemental Table A14 further shows that the DD estimates of Black students are also not statistically different from those of White and Hispanic students. Panel B of Table 7 displays the change in likelihood of carrying weapons in school. Even though Black students were around two percentage points more likely to carry a weapon than White students, the WDBs affected neither of the two. Panels C and D present the likelihood of being offered illicit drugs and using marijuana. Although the sample average indicates that Black students behaved similarly to White and Hispanic students, only Hispanic students experienced a decrease in the likelihood of being offered illicit drugs and using marijuana after WDBs were implemented.

These findings reveal no evidence that students of a certain race became more disruptive after WDB enactments.

Although one could argue that the WDBs generated positive behavior changes among Black students (because disruptors were suspended for nonwillful defiance reasons), this cannot explain why only Black students faced more nonwillful defiance OSSs and White students experienced similar behavior improvements without experiencing increases in suspensions. In addition, student-level analysis suggests that White and Hispanic students experienced some behavior improvements after WDBs were implemented. The purpose of WDBs is to reduce OSS rates when student behavior remains the same. In other words, OSS rates should decrease or remain the same even if student behavior becomes worse. However, the findings show that student behavior improved while OSS rates remained constant, or even increased, indicating that local WDBs failed to achieve their purpose.

To summarize, evidence in Table 7 suggests that, following the implementation of WDBs, student behavior across all racial groups improved instead of deteriorated. Therefore, the increase in nonwillful defiance OSS rates for Black students is at odds with improvements in their behavior.

Conclusion

This study examines the impact of WDBs on the use of OSSs in four California school districts. The results indicate that WDBs effectively reduced willful defiance OSS rates

by around 67%. However, the WDBs did not affect overall OSS rates: an increase in nonwillful defiance OSS rates offset the decreases in willful defiance OSS rates.

Furthermore, the findings suggest that the impact of WDBs varied by student race. Despite receiving OSS at much higher rates before adoption, Black students benefited less from WDBs than White and Hispanic students. Particularly, WDBs increased nonwillful defiance OSS rates for Black students by around 26%, which contributed to increases in overall OSS rates. There were no significant changes in overall OSS rates for White and Hispanic students following the implementation of WDBs. Students' behavior changes could not explain such heterogeneity in treatment effects by race. Rather, it is possible that some broadly defined nonwillful defiance category, particularly "violent no injury" or "miscellaneous," replaced willful defiance as the reason for suspension. In addition, the substitution of suspension reasons among Black students were more salient in schools with fewer Black teachers, who can be assumed to hold fewer perceptual biases against Black students (see results in Tables 6 and A8). These findings are consistent with previous qualitative research by Lasnover (2015), who found that some teachers in LAUSD did suspend students under nonwillful defiance reasons when they no longer had willful defiance as a legitimate reason.

It is important to acknowledge that the analyses in this article suffer from several limitations. First, one should be cautious about applying this article's findings to schools outside California because of the state's unique culture and demographic composition. Second, due to data limitations, my analyses only focus on the use of OSS. If WDBs also led to more in-school suspensions, my estimates could be biased downward, and my results would be the lower bounds of the true effects of WDBs on student discipline outcomes. Future studies with data on student-teacher race match at the individual-level could complement this study. Last, I could not test the impact of WDBs on test scores because the "end of semester tests" after the 2014-2015 school year were not comparable to the previously offered "end of course tests." I tested the impact of WDBs on students' graduation and dropout rates but found no significant changes in either outcome (see online Supplemental Table A15 for these results).

Steinberg and Lacoé (2017) categorized discipline reforms into program-based interventions and policy-based interventions. Program-based interventions aim to improve school climates, encourage positive behavior, and reduce violence among students through nonpunitive approaches such as mentoring, group therapy, and teacher training, while policy-based reforms, such as WDBs, revise discipline policies. My findings, along with previous research on Chicago and Philadelphia suspension reforms, suggest that policy-based reforms are ineffective in changing the use of suspension in schools (Lacoé & Steinberg, 2018; Sartain

et al., 2015; Steinberg & Lacoé, 2017). This study shows that overall OSS rates did not change after WDB reform, and, therefore, suggests that policy-based discipline reforms might make only limited contributions to improvements in academic performance.

This study also sheds light on the dangers of designing policy without accounting for unintended consequences and the critical role of frontline workers in policy implementation. The WDBs produced such unexpected consequences across three specific aspects. First, the policies did not consider that OSSs in the nonwillful defiance category could serve as substitutes for willful defiance OSSs. Second, the abrupt changes in discipline policy did not eliminate biases against Black students. Third, school districts failed to provide enough support for policy implementation. Policy makers looking to extend the current California statewide WDB in the next 5 to 10 years should consider restricting the use of all broadly or vaguely defined suspension reasons and add action items to support implementation, such as increasing the number of minority teachers, providing teacher training on effective discipline, and systematically adopting program-based discipline reforms like PBIS or RJ.

Acknowledgments

The author thanks Erdal Tekin, Seth Gershenson, Anna Amirkhanyan, Robert Shand, and three anonymous reviewers and editors for their detailed and constructive feedback on earlier drafts of this article. The author is grateful for the help from Michael Lombardo and Luke Anderson of the California PBIS Coalition, who generously shared their data on PBIS implementation. The author is also grateful for the codes and insights shared by James MacKinnon, Manudeep Bhuller, Liyang Sun, Katharine Strunk, Ayesha Hashim, and Tasminda Dhaliwal. Youth Risk Behavior Surveillance System data were kindly provided by the San Francisco Unified School District, Los Angeles Unified School District, and Centers for Disease Control and Prevention and were used with permission. Opinions and errors are the sole responsibility of the author.

ORCID iD

Rui Wang  <https://orcid.org/0000-0002-3305-2881>

Open Practices

The data and analysis files for this article can be found at <https://doi.org/10.3886/E154821V2>

Notes

1. I consulted multiple sources to identify school districts that implemented WDBs, including district policy manuals, district board meeting documents, local newspapers, and direct contacts with local school districts. I omitted Los Angeles Unified School District (LAUSD), which banned suspension for willful defiance in SY 2011-2012, before which I have no data. The identified reformed school districts are consistent with the list of reformed school districts provided by California State Senator Nancy

Skinner, who proposed Senate Bill 419, which eliminated willful defiance suspensions for K–8 students statewide and were signed by California Governor Newsom.

2. This study focuses on White, Black, and Hispanic students for two reasons. First, privacy concerns prevent the California Department of Education from releasing school-level data for student groups with fewer than ten enrollments. As a result, expanding the current analysis to Asian Americans would lead to a substantive reduction in sample size. Second, previous research has shown that Asian American students usually experience fewer suspensions than students of other races (Morgan & Wright, 2018). Due to data limitations and the rarity of suspension incidents, I decided to exclude Asian American students from this study.

3. See A.B. 420, 2013–2014 Reg. Sess. (Cal. 2014).

4. See S.B. 419, 2019–2020 Reg. Sess. (Cal. 2019).

5. I restricted the analytical sample to high schools that only offer classes from Grades 9 to 12 for two reasons. First, this ensures that the suspension data across schools represent students from the same grade ranges. Second, it would prevent any spillover effects due to discipline policy changes in lower grades (e.g., A.B. 420 banned willful defiance suspension for all students in kindergarten to Grade 3). In addition, alternative, juvenile justice, virtual teaching, and magnet schools were dropped from the analytical sample because they operate under different goals and mainly serve special groups of students. Due to the requirements of the Family Privacy Act, the California Department of Education only publishes the data if the enrollment of a specific group is larger than 10. Therefore, I excluded schools from the analysis if suspension data for any racial group in a school was missing.

6. I compared the treatment effects between two racial groups by testing the coefficient of the interaction term between the treatment dummy and the race dummy from a model in which all controls, including school and year fixed effects, are fully interacted with the race dummy. Although it is possible to construct a model to estimate the differences in treatment effects across three racial groups, the number of interactions between the school fixed effects and race dummies prevent Poisson regressions from converging when calculating WR p values.

7. Following Dobkin et al. (2018) and Freyaldenhoven et al. (2019), this work includes the linear trend from the two periods immediately preceding the WDBs for White students using the following equation:

$$E(D_{sdt}) = \exp \left(\alpha_s + \alpha_t + \sum_{k=1}^{k_{\max}} \tau_k \text{Ban}_{d,t-k} + \beta X_{st} + \Omega p_{st} 1(-1 \leq p_{st} \leq 2) \right) \text{ where}$$

p_{st} is the time relative to the implementation of WDBs for each district. p_{st} equals zero if school s is in a district that never implemented a WDB or observation is more than 2 years before or after the WDBs. Intuitively, the linear trend allows the treated districts to evolve on a separate secular trend from the untreated school districts 2 years before and after the WDBs. Therefore, the identifying assumption is that, conditional on having a WDB and the included controls, the timing of the WDBs is uncorrelated with the deviation of OSS rates from a linear trend in event time.

8. Data were reported at the school level, and teacher identifiers were redacted for confidentiality considerations, prohibiting the identification of teachers who assigned OSS. I operationalized the presence of same-race teachers by calculating the percentage of same-race full-time teachers in each school.

References

- A.B. 420, 2013–2014 Reg. Sess. (Cal. 2014).
- Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (NBER Working Paper w24003). National Bureau of Economic Research. <https://doi.org/10.3386/w24003>
- Anderson, K. P., Ritter, G. W., & Zamarro, G. (2019). Understanding a vicious cycle: The relationship between student discipline and student academic outcomes. *Educational Researcher*, 48(5), 251–262. <https://doi.org/10/ghqp25>
- Aron, H. (2013, October 14). LA Unified suspension rate accelerating down, to 1.5 percent. *LA School Report*. <http://laschoolreport.com/la-unified-suspension-rate-accelerating-down-to-1-5-percent/>
- Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research*, 86(1), 163–206. <https://doi.org/10/gcv66z>
- Azusa Unified School District. (n.d.). *Local control and accountability plan*. <https://willard.lacoe.edu/lcap/view.pl?5K19012m>
- Bacher-Hicks, A., Billings, S., & Deming, D. (2019). *The school to prison pipeline: Long-run impacts of school suspensions on adult crime* (Working Paper No. w26257). National Bureau of Economic Research. <https://doi.org/10.3386/w26257>
- Baker-Smith, E. C. (2018). Suspensions suspended: Do changes to high school suspension policies change suspension rates. *Peabody Journal of Education*, 93(2), 190–206. <https://doi.org/10/ghqqxb>
- Barrett, N., McEachin, A., Mills, J. N., & Valant, J. (2019). Disparities and discrimination in student discipline by race and family income. *Journal of Human Resources*, 56(3), 711–748. <https://doi.org/10/ghqp3r>
- Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband internet: An information superhighway to sex crime? *Review of Economic Studies*, 80(4), 1237–1266. <https://doi.org/10/f5fgb7>
- Blume, H. (2012, March 6). Suspension figures called “alarming.” *Los Angeles Times*. <https://www.latimes.com/archives/la-xpm-2012-mar-06-la-me-lausd-data-20120306-story.html>
- Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide positive behavioral interventions and supports: Findings from a group-randomized effectiveness trial. *Prevention Science*, 10(2), 100–115. <https://doi.org/10/b5qh87>
- Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, 12(3), 133–148. <https://doi.org/10/dqzv36>
- California Department of Education. (n.d.). *Dataquest*. <http://dq.cde.ca.gov/dataquest/>
- California PBIS. (n.d.). *PBIS growth on California schools*. California PBIS Coalition. <https://pbisca.org/pbis-california-growth>
- Conley, T. G., & Taber, C. R. (2011). Inference with “difference in differences” with a small number of policy changes. *Review of Economics and Statistics*, 93(1), 113–125. <https://doi.org/10/fhws6f>

- Craig, A. C., & Martin, D. C. (2019). *Discipline reform, school culture, and student achievement* (Job Market Paper). <https://scholar.harvard.edu/david-martin/publications/discipline-reform-school-culture-and-student-achievement>
- Cuellar, A. E., & Markowitz, S. (2015). School suspension and the school-to-prison pipeline. *International Review of Law and Economics*, 43(August), 98–106. <https://doi.org/10/ghqp3b>
- de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). *Status and trends in the education of racial and ethnic groups 2018*. NCES 2019-038. National Center for Education Statistics, U.S. Department of Education. <https://nces.ed.gov/pubs2019/2019038.pdf>
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1), 195–210. <https://doi.org/10/d3nd3d>
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2), 158–165. <https://doi.org/10/c7gxn5>
- Dee, T. S., & Penner, E. (2019). *My brother's keeper? The impact of targeted educational supports* (Working Paper No. w26386). National Bureau of Economic Research. <https://doi.org/10.3386/w26386>
- Dobkin, C., Finkelstein, A., Kluender, R., & Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review*, 108(2), 308–352. <https://doi.org/10/gc4v9j>
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4), 795–813. <https://doi.org/10/dv5ssw>
- Education Commission of the States. (2020). *State education policy tracking*. <https://www.ecs.org/state-education-policy-tracking/>
- Frey, S. (2013). San Francisco considers eliminating “willful defiance” as reason for suspensions. *EdSource*. <https://edsources.org/2013/san-francisco-considers-eliminating-willful-defiance-as-reason-for-suspensions/5383>
- Frey, S. (2014). San Francisco Unified eliminates ‘willful defiance’ as a reason to expel or suspend students. *EdSource*. <https://edsources.org/2014/san-francisco-unified-eliminates-willful-defiance-as-a-reason-to-expel-or-suspend-students/58105>
- Frey, S. (2015). Oakland ends suspensions for willful defiance, funds restorative justice. *EdSource*. <https://edsources.org/2015/oakland-ends-suspensions-for-willful-defiance-funds-restorative-justice/79731>
- Freyaldenhoven, S., Hansen, C., & Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, 109(9), 3307–3338. <https://doi.org/10/gf7gjt>
- Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52(June), 209–224. <https://doi.org/10/f8sc6d>
- Gershenson, S., & Papageorge, N. (2018). The power of teacher expectations: How racial bias hinders student attainment. *Education Next*, 18(1), 64–71. <https://www.educationnext.org/power-of-teacher-expectations-racial-bias-hinders-student-attainment/>
- Gonzalez, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the school to prison pipeline. *Journal of Law & Education*, 41(2), 281–335. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2658513
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277. <https://doi.org/10.1016/j.jeconom.2021.03.014>
- Gopalan, M., & Nelson, A. A. (2019). Understanding the racial discipline gap in schools. *AERA Open*, 5(2). <https://doi.org/10.1177/2332858419844613>
- 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. <https://doi.org/10.3102/0013189X09357621>
- Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The “representative bureaucracy” in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*, 44(3), 185–192. <https://doi.org/10.3102/0013189X15580102>
- Grissom, J. A., Nicholson-Crotty, J., & Nicholson-Crotty, S. (2009). Race, region, and representative bureaucracy. *Public Administration Review*, 69(5), 911–919. <https://doi.org/10/b7rdkq>
- Grissom, J. A., Rodriguez, L. A., & Kern, E. C. (2017). Teacher and principal diversity and the representation of students of color in gifted programs: Evidence from national data. *Elementary School Journal*, 117(3), 396–422. <https://doi.org/10.1086/690274>
- 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*, 93(2), 174–189. <https://doi.org/10.1080/0161956X.2018.1435040>
- Heß, S. (2017). Randomization inference with Stata: A guide and software. *Stata Journal: Promoting Communications on Statistics and Stata*, 17(3), 630–651. <https://doi.org/10.1177/1536867X1701700306>
- Holt, S. B., & Gershenson, S. (2019). The impact of demographic representation on absences and suspensions. *Policy Studies Journal*, 47(4), 1069–1099. <https://doi.org/10.1111/psj.12229>
- Jacob, B. A., & Lefgren, L. (2003). Are idle hands the devil’s workshop? Incapacitation, concentration, and juvenile crime. *American Economic Review*, 93(5), 1560–1577. <https://doi.org/10.1257/000282803322655446>
- Keiser, L. R., Wilkins, V. M., Meier, K. J., & Holland, C. A. (2002). Lipstick and logarithms: Gender, institutional context, and representative bureaucracy. *American Political Science Review*, 96(3), 553–564. <https://www.jstor.org/stable/3117929>
- Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, 54(1), 355–383. <https://doi.org/10.1111/j.1468-2354.2012.00736.x>
- Lacoe, J., & Steinberg, M. P. (2018). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes. *Peabody Journal of Education*, 93(2), 207–227. <https://doi.org/10.1080/0161956X.2018.1435047>
- Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, 10(2), 1–26. <https://doi.org/10.1257/app.20160567>
- Lasnover, S. (2015). *The early of effects of the removal of willful defiance from the discipline policy at urban high schools* [Doctoral dissertation, University of California]. <https://escholarship.org/uc/item/385513xg>

- Lindo, J. M., & Packham, A. (2017). How much can expanding access to long-acting reversible contraceptives reduce teen birth rates? *American Economic Journal: Economic Policy*, 9(3), 348–376. <https://doi.org/10.1257/pol.20160039>
- Losen, D. J., & Gillespie, J. (2012). *Opportunities suspended: The disparate impact of disciplinary exclusion from school* (Civil Rights Project/Proyecto Derechos Civiles). <https://files.eric.ed.gov/fulltext/ED534184.pdf>
- Luallen, J. (2006). School's out . . . forever: A study of juvenile crime, at-risk youths and teacher strikes. *Journal of Urban Economics*, 59(1), 75–103. <https://doi.org/10.1016/j.jue.2005.09.002>
- MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *Econometrics Journal*, 21(2), 114–135. <https://doi.org/10.1111/ectj.12107>
- MacKinnon, J. G., & Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2), 435–450. <https://doi.org/10.1016/j.jeconom.2020.04.024>
- McClellan, C., & Tekin, E. (2017). Stand your ground laws, homicides, and injuries. *Journal of Human Resources*, 52(3), 621–653. <https://doi.org/10.3368/jhr.52.3.0613-5723R2>
- Morgan, M. A., & Wright, J. P. (2018). Beyond Black and White: Suspension disparities for Hispanic, Asian, and White youth. *Criminal Justice Review*, 43(4), 377–398. <https://doi.org/10.1086/670398>
- Mowen, T. J., Brent, J. J., & Boman, J. H., IV. (2020). The effect of school discipline on offending across time. *Justice Quarterly*, 37(4), 739–760. <https://doi.org/10.1080/07418825.2019.1625428>
- Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice*, 42(1), 199–263. <https://doi.org/10.1086/670398>
- Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, 26(5), 617–624. <https://doi.org/10.1177/0956797615570365>
- Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology*, 16(1), 21–43. <https://doi.org/10.1023/a:1007521427059>
- Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance & Policy*, 9(3), 334–372. https://doi.org/10.1162/EDFP_a_00136
- Papageorge, N. W., Gershenson, S., & Kang, K. M. (2020). Teacher expectations matter. *Review of Economics and Statistics*, 102(2), 234–251. https://doi.org/10.1162/rest_a_00838
- Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are achievement gaps related to discipline gaps? Evidence from national data. *AERA Open*, 5(4). <https://doi.org/10.1177/2332858419875440>
- Pena-Shaff, J. B., Bessette-Symons, B., Tate, M., & Fingerhut, J. (2019). Racial and ethnic differences in high school students' perceptions of school climate and disciplinary practices. *Race Ethnicity and Education*, 22(2), 269–284. <https://doi.org/10.1080/13613324.2018.1468747>
- Pesta, R. (2018). Labeling and the differential impact of school discipline on negative life outcomes: Assessing ethno-racial variation in the school-to-prison pipeline. *Crime & Delinquency*, 64(11), 1489–1512. <https://doi.org/10.1177/0011128717749223>
- Pfeifer, G., Reutter, M., & Strohmaier, K. (2020). Goodbye smokers' corner: Health effects of school smoking bans. *Journal of Human Resources*, 55(3), 1068–1104. <https://doi.org/10.3368/jhr.55.3.1217-9246R3>
- Redding, C. (2019). A teacher like me: A review of the effect of student-teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, 89(4), 499–535. <https://doi.org/10.3102/0034654319853545>
- Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, 116(17), 8255–8260. <https://doi.org/10.1073/pnas.1808307116>
- Riestenberg, N. (2015). The restorative implementation: Paradigms and practices. *Restorative Practices in Action Journal*. https://www.iirp.edu/images/pdf/Nancy_NY-Riestenberg-final2.pdf
- Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using bootstest. *Stata Journal*, 19(1), 4–60. <https://doi.org/10.1177/1536867X19830877>
- Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, 16(2), 125–141. <https://doi.org/10.1521/scpq.16.2.125.18698>
- San Francisco Unified School District. (2014). *Establishment of a safe and supportive schools policy in the San Francisco Unified School District*. San Francisco Unified School District and County Office of Education. <http://www.fixschool-discipline.org/wp-content/uploads/2014/12/SFUSD-Safe-and-Supportive-Schools-Resolution-2.14.pdf>
- Sartain, L., Allensworth, E. M., Porter, S., Levenstein, R., Johnson, D. W., Huynh, M. H., & Steinberg, M. P. (2015). *Suspending Chicago's students: Differences in discipline practices across schools*. <https://consortium.uchicago.edu/publications/suspending-chicagos-students-differences-discipline-practices-across-schools>
- S.B. 419, 2019–2020 Reg. Sess. (Cal. 2019).
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- Steinberg, M. P., & Lacoë, J. (2017). What do we know about school discipline reform? Assessing the alternatives to suspensions and expulsions. *Education Next*, 17(1), 44–53. https://go.gale.com/ps/i.do?id=GALE%7CA474717787&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=15399664&p=AONE&sw=w&userGroupName=tel_oweb&isGeoAuthType=true
- Steinberg, M. P., & Lacoë, 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. <https://doi.org/10.1086/699811>
- Stevens, W. D., Sartain, L., Allensworth, E. M., Levenstein, R., Guiltinan, S., Mader, N., Huynh, M. H., & Porter, S. (2015). *Discipline practices in Chicago schools: Trends in the use of suspensions and arrests*. <https://consortium.uchicago.edu/publications/discipline-practices-chicago-schools-trends-use-suspensions-and-arrests>
- Stewart, E. A., Baumer, E. P., Brunson, R. K., & Simons, R. L. (2009). Neighborhood racial context and perceptions of police-based racial discrimination among black youth. *Criminology*, 47(3), 847–887. <https://doi.org/10.1111/j.1745-9125.2009.00159.x>

- Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy*, 24(1–2), 23–50. https://doi.org/10.1300/J019v24n01_03
- Sun, L., & Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199. <https://doi.org/10.1016/j.jeconom.2020.09.006>
- U.S. Department of Education. (2012, September 28). *U.S. Department of Education announces voluntary resolution of Oakland Unified School District civil rights investigation*. <https://www.ed.gov/news/press-releases/us-department-education-announces-voluntary-resolution-oakland-unified-school-di>
- Wald, J., & Losen, D. J. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development*, 2003(99), 9–15. <https://doi.org/10.1002/yd.51>
- Way, S. M. (2011). School discipline and disruptive classroom behavior: The moderating effects of student perceptions. *Sociological Quarterly*, 52(3), 346–375. <https://doi.org/10.1111/j.1533-8525.2011.01210.x>
- Wilkins, V. M., & Williams, B. N. (2008). Black or blue: Racial profiling and representative bureaucracy. *Public Administration Review*, 68(4), 654–664. <https://doi.org/10.1111/j.1540-6210.2008.00905.x>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2), 557–598. <https://doi.org/10.1093/qje/qjy029>

Author

RUI WANG is an assistant professor at Shanghai University of Finance and Economics. His research focuses on educational equity and students' noncognitive outcomes.