1-1-2022

Do School Suspension Reforms Work? Evidence From Rhode Island

Terry Ann Craigie
Smith College, tcraigie@smith.edu

Do School Suspension Reforms Work? Evidence from Rhode Island

Terry-Ann Craigie

Smith College

April 22, 2022

Forthcoming, Educational Evaluation and Policy Analysis

In Rhode Island, out-of-school suspensions were excessively and disproportionately used to penalize low-level infractions. To address this problem, the Rhode Island General Assembly passed legislation, effective May 2012, prohibiting out-of-school suspensions for attendance-specific infractions. Four years later, they passed additional legislation to curb out-of-school suspensions for disruption-specific infractions. This study examines the impact of these suspension reforms on out-of-school suspension outcomes for treatment infractions and corresponding racial-ethnic disparities. To execute the analyses, the study uses student-level administrative data (AY 2009–2010 to AY 2017–2018) from the Rhode Island Department of Education, along with quasi-experimental estimation. The study finds that only the first reform lowers out-of-school suspension outcomes for attendance-specific infractions and corresponding racial-ethnic disparities.

Keywords: *attendance infractions, disruptive infractions, exclusionary discipline, minor infractions, out-of-school suspensions, quadruple-difference estimation, quasi-experimental models, racial-ethnic disparities, Rhode Island, suspension reforms, triple-difference estimation.*

Introduction

THE Gun Free Schools Act of 1994 mandated a one-year expulsion exclusively for firearm possession on school grounds. Ultimately, it triggered "zero-tolerance" policies for other student infractions, even minor ones. With all fifty states adopting some variation of this legislation (Cuellar & Markowitz, 2015; Henault, 2001), the use of exclusionary discipline became widespread as a result. The main concern about exclusionary discipline is that it instigates adverse consequences for academic and adult outcomes. Myriad studies show that out-of-school suspensions (hereafter, "OSS") and expulsions are negatively associated with academic achievement (e.g., Perry & Morris, 2014; Skiba et al., 2006) and initiate the school-to-prison pipeline (e.g., Bacher-Hicks et al., 2019; Cuellar & Markowitz, 2015) without markedly improving school safety (e.g., Jones, 2018; Knesting & Skiba, 2002, 2010; Skiba et al., 2006). Moreover, exclusionary discipline produces disparate impact – Black students are expelled and suspended at thrice the rate of White students for similar offenses (U.S. Department of Education, Office of Civil Rights, 2014).

To curtail the excessive use of exclusionary discipline and its consequences, several states to date have enacted legislation to ban the use of OSS for low-level student infractions. Rhode Island is one such state. Effective May 2012, Rhode Island barred OSS for all attendance-specific infractions (such as truancy and absenteeism). With the passage of additional legislation effective June 2016, the state also restricted the use of OSS for disruption-specific infractions (such as insubordination, disorderly conduct, and obscene language) unless a student persistently threatens or harms someone, or has repeatedly obstructed learning, and/or other corrective measures have proven ineffective. This legislation, in contrast to its predecessor, explicitly states

that school superintendents must review discipline data to determine whether there are disproportionalities based on race, ethnicity, or disability; thereafter, they must address any such disparities. By implementing these reforms, Rhode Island represents an intriguing landscape for impact evaluation. While the first reform targets a clear treatment group (i.e., students with attendance-specific infractions), the second appears to be an ambiguous intervention. This is because we cannot observe which disruptive infractions are threatening, harmful, or repeatedly obstructive, and thereby warrant OSS penalties. Moreover, threat, harm, and repeated obstruction in the classroom are subjectively determined and tend to perpetuate disproportionalities in discipline and office referrals (e.g., Francis, 2012; Ritter & Anderson, 2018; Sagar & Schofield, 1980; Skiba, 2002).

Thus far, the evidence has been mixed as to whether school suspension reforms lower OSS or mitigate disparate impact induced by OSS (e.g., Anderson, 2018; Anderson et al., 2019; Baker-Smith, 2018; Lacoe & Steinberg, 2018; Steinberg & Lacoe, 2018; Zarecki, 2019). However, these issues have yet to be evaluated in Rhode Island. As such, this study aims to answer the following questions:

Research Question 1: Do Rhode Island suspension reforms lower the use of OSS for attendance- and disruption-specific infractions?

Research Question 2: Do Rhode Island suspension reforms mitigate the disparate impact of OSS for attendance- and disruption-specific infractions?

This study adds to the burgeoning literature by being the first to evaluate the *first-order* effects of dual suspension reforms in the state of Rhode Island. It evaluates the extent to which

OSS for low-level attendance- and disruption-specific infractions changed in response to the two reforms that took effect in May 2012 and June 2016, respectively. For this impact evaluation, the study exploits student-level administrative data from the Rhode Island Department of Education (AY 2009–2010 to AY 2017–2018), along with a triple-difference (DDD) framework, to measure the impact of each suspension reform on OSS outcomes for treatment infractions in treatment schools (i.e., schools and within-school grades that used OSS to penalize treatment infractions, *ex ante*). The study also employs a quadruple-difference (DDDD) model to measure the extent to which racial-ethnic disparities in OSS changed for the treatment groups under each reform.

Using DDD estimation, the study finds that the first reform substantially lowers OSS for attendance-specific infractions in treatment schools. After accounting for a comprehensive set of controls, the probability of OSS declined by approximately 38 percentage points ($p < 0.01$), or 1.6 times the outcome standard deviation. The first reform also lowers OSS duration by approximately half a day ($p < 0.01$) on average, or 80% of the outcome standard deviation. On the other hand, the second reform does not change OSS penalties for disruption-specific infractions in treatment schools. This lack of compliance points to the conflicting objectives of the second reform (e.g., Baier et al., 1986; Cohen et al., 2007; Firestone, 1989): it prohibits OSS penalties for non-threatening, non-harmful, non-repetitive, and tractable forms of disruption in the classroom setting, while permitting OSS for disruptive behaviors antithetical to these stipulations.

The impacts of the suspension reforms on the disproportionality of OSS penalties are mixed. Using DDDD estimation, the first reform lowers the Black-White disparity in the probability of OSS by nearly 17 percentage points ($p < 0.05$) and OSS duration by 0.37 days ($p <$

0.01) on average. The first reform also lowers the Latinx-White disparity in the probability of OSS by 27 percentage points ($p < 0.01$) and OSS duration by 0.53 days ($p < 0.01$) on average. Similarly, the Other-White disparity declines by 30 percentage points ($p < 0.10$) and 0.62 days ($p < 0.10$) on average for OSS probability and duration, respectively.

In contrast, the second reform is not shown to reduce racial-ethnic disparities in OSS penalties. In fact, the Black-White disparity in OSS duration rises by 0.22 days ($p < 0.10$) on average. These findings suggest that although there has been some progress in reducing the disparate impact of OSS in Rhode Island, there is still much to be accomplished, especially for Black students, who are the most likely group to be disciplined in this fashion (Anderson & Ritter, 2017; Gopalan & Nelson, 2019; Morris & Perry, 2016).

Hereafter, the article is structured as follows: Background Literature, which describes previous findings on the consequences of OSS and the impact of suspension bans in other jurisdictions; Disciplinary Reform in Rhode Island, which discusses the details of Rhode Island's suspension reforms; Data Description, which describes the data and summary statistics; Empirical Framework, which explains the quasi-experimental methods; Results, which interprets the findings; and Conclusion, which summarizes the results and discusses implications for policy and future research.

Background Literature

The Consequences of OSS

Numerous studies exhibit strong correlations between OSS and academic progress. OSS, as a form of exclusionary discipline, is associated with lower student grades and test scores (e.g.,

Arcia, 2006; Beck & Muschkin, 2012; Cobb-Clark et al., 2015; Davis & Jordan, 1994; Lacoe & Steinberg, 2019; Noltemeyer et al., 2015; Raffaele-Mendez, 2003; Raffaele-Mendez et al., 2002; Rausch & Skiba, 2005; Skiba & Rausch, 2004) and a higher likelihood of grade retention, high school dropout, and juvenile justice involvement (e.g., American Academy of Pediatrics, 2013; American Psychological Association, 2008; Balfanz et al., 2015; Cobb-Clark et al., 2015; Fabelo et al., 2011; Gregory & Weinstein, 2008; Krezmein et al., 2006; Marchbanks et al., 2014; Nicholson-Crotty et al., 2009; Noltemeyer et al., 2015; Rafaelle-Mendez, 2003; Rumberger & Losen, 2016; Stearns & Glennie, 2006; Swanson et al., 2017; Wald & Kurlaender, 2003).

There are multiple mechanisms through which OSS might influence these outcomes (Lacoe & Steinberg, 2019; Rausch & Skiba, 2005). With less time in the classroom, out-of-school suspended students may find it more difficult to grasp course materials upon their return. School absences generated by OSS may also work to erode human capital accumulated within the classroom. Through these mechanisms, OSS make it difficult to maintain good grades and proficient test scores, increasing the odds of grade retention and high school dropout.

Out-of-school suspended students may also be prevented from forging strong bonds with non-suspended (or even in-school suspended) peers. This could promote disruptive behaviors in the classroom that obstruct the ability to learn, thus depressing academic achievement. Moreover, some studies suggest that suspending students out of school is linked to negative spillovers for non-suspended peers (Lacoe & Steinberg, 2019; Perry & Morris, 2014). Nevertheless, OSS could raise the academic performance of students not penalized with school exclusion. If students who disrupt the learning environment are suspended out of school, then their absence could help maintain an environment more conducive to learning, thereby improving overall academic performance. One recent study supports this theory, showing that

suspending disruptive students led to improvements in mathematics test scores (Hwang & Domina, 2020); however, the authors caution that this evidence comes from schools in which suspensions are relatively rare.

OSS have also been inextricably linked to juvenile delinquency (e.g., Fabelo et al., 2011; Krezmien et al., 2006; Nicholson-Crotty et al., 2009; Sedlak & McPherson, 2010; Skiba et al., 2014) and adult incarceration later in life (e.g., Bacher-Hicks et al., 2019; Cuellar & Markowitz, 2015; Skiba et al., 2006; Wolf & Kupchik, 2017). Time away from school (and possibly away from adult supervision) could promote idleness and thus create a vacuum in which delinquent behavior can thrive (Farrington, 1980; Gavin, 1996, 1997; McCluskey et al., 2004). Moreover, the lack of social bonds with school peers may boost proclivities toward violence, substance abuse, and other forms of deviance (Conger, 1976; Elliott et al., 1985; Longshore et al., 2004). OSS may also serve as self-fulfilling prophecies for marginalized youth, easing the transition to juvenile delinquency.

Black and Latinx youth are especially vulnerable to the consequences of OSS. Marked by antisocial and criminal stereotypes, Black youth are disproportionately more likely to be subject to exclusionary discipline (Anderson & Ritter, 2017; Beck & Muschkin, 2012; Gopalan & Nelson, 2019; Harper et al., 2019; Krezmien et al., 2006; Morris & Perry, 2016; Wilson, 2020). In fact, Black students are suspended at more than twice the rate of their White and Latinx counterparts. Latinx students also have higher suspension rates relative to Whites, but this disparity has seen a significant decline in recent years (Harper et al., 2019).

The Impact of School Suspension Reforms

By May 2015, twenty-two states and Washington, D.C., had implemented some type of reform to curb exclusionary discipline—mostly suspensions (Anderson, 2018; Steinberg & Lacoe, 2017). For instance, Arkansas Act 1329 banned OSS for truancy for all students, while Act 1059 banned OSS for Kindergarten and elementary students (except for narrow exceptions) (Anderson, 2018). From 2009 – 2014, Chicago Public Schools implemented policy changes to help lower the district's high rate of exclusionary discipline (Stevens et al. 2015). In 2012, Philadelphia implemented reforms to restrict OSS for nonviolent and classroom disorder infractions (Lacoe & Steinberg, 2018; Steinberg & Lacoe, 2018). From 2009 onward, NYC made provisions for suspension alternatives in its disciplinary code, although implementing these alternatives at the school level was discretionary. In California, four school districts banned suspensions for willful defiance between 2013 and 2015 (Zarecki, 2019).

Despite success in implementing suspension reforms at state and school-district levels, compliance with these reforms is largely mixed. Anderson et al. (2019) used a difference-in-difference approach to show that in Arkansas, Act 1329 of 2013—which banned OSS for truancy—produced uneven compliance by schools across the state. In fact, schools with relatively more students of color were less likely to comply with the reform (Anderson 2018). Using fixed effects analysis, Zarecki (2019) found that in California, staff continued to illegally issue suspensions for willful defiance infractions, despite their training in restorative justice and de-escalation practices. By contrast, Lacoe and Steinberg (2018) and Steinberg and Lacoe (2018) used difference-in-difference approaches to show that suspension reforms in Philadelphia lowered OSS for non-violent and classroom disorder infractions, respectively. Similarly, Baker-

Smith (2018) showed that a suspension ban for low-level infractions in NYC high schools is negatively associated with OSS (although the likelihood of OSS is shown to increase after the first offense). Stevens et al. (2015) also illustrated that there were lower OSS rates and shorter suspension durations after exclusionary discipline reforms were implemented in Chicago.

Compliance with school suspension reforms is key to effectively ameliorating the unfavorable consequences of OSS. Early policy implementation studies assert that policy adoption will be ineffectual without compliance (e.g., Berman, 1978; Firestone, 1989; Matland, 1995), and this is no different for school disciplinary reforms. Recent evidence underscores that the lack of compliance with suspension reforms produces less desirable academic outcomes. For instance, Anderson (2019) found trivial improvements in test scores among disadvantaged students in response to OSS bans for truancy—a result attributed to limited compliance with the reform. Similarly, Zarecki (2019) found negative growth in math achievement for fourth through seventh grades in the California school districts that failed to comply with suspension bans for willful defiance.

Adding to this burgeoning literature, this study evaluates the efficacy of school suspension reforms in the Rhode Island context by:

(i) measuring the extent of compliance with the reforms via the change in OSS for attendance- and disruption-specific infractions; and

(ii) measuring the extent to which the reforms change the racial-ethnic disparities in OSS for such infractions.

Disciplinary Reform in Rhode Island

In Rhode Island, exclusionary discipline for student infractions chiefly includes out-of-school suspensions (OSS), in-school suspensions (ISS), and alternate program placements. Expulsions are rarely administered in Rhode Island, with only 41 cases occurring in the entire state from AY 2009–2010 to AY 2017–2018. Each year, the Rhode Island KIDS COUNT Factbook breaks down the total number of disciplinary infractions in Rhode Island public schools by demographic composition, type of infraction, type of suspension, and school district. These annual publications also emphasize the significance of suspensions and prevalence among students from disadvantaged backgrounds.[i] Before the suspension reforms, Rhode Island KIDS COUNT showed that there were well over 40,000 student infractions in Rhode Island, two-thirds of which were due to low-level offenses related to attendance, insubordination/disrespect, and disorderly conduct. OSS comprised approximately 60% of all disciplinary actions, and one-third of OSS were administered for attendance-specific infractions. The Rhode Island American Civil Liberties Union also reported persistent racial-ethnic disproportionalities in the use of OSS for disruption-specific infractions (such as insubordination/disrespect, disorderly conduct, or obscene language) (ACLU 2015), revealing grave implications for disparate impact (U.S. Department of Education, Office of Civil Rights 2014).

In light of the prevalence of exclusionary discipline, the Rhode Island General Assembly passed two laws to mitigate OSS for low-level infractions. Effective May 30, 2012, the first piece of legislation (§ 16-19-1) prohibited schools from using OSS as discipline for attendance-specific infractions (such as truancy and absenteeism). However, schools were still permitted to use ISS and alternate program placements as alternative disciplinary actions.

Effective June 29, 2016, the second piece of legislation—the Right to A Safe School Act (§ 16-2-17)—prohibited OSS penalties for disruptive students except when a student persistently threatens or harms someone, has repeatedly obstructed learning, and/or other corrective measures have proven ineffective. Like the first reform, this legislation is not enforceable for ISS or alternate program placements. Unlike the first reform, however, this legislation explicitly states that school superintendents must review discipline data to determine whether there is a disproportionate impact based on race, ethnicity, or disability, then address any disparities. Put simply, the second legislation aims to curb the use of OSS for disruption-specific infractions that constitute minor student misbehavior, while ensuring that no disparate impact takes place. However, in contrast to the first reform, which outlines a clear target infraction and corresponding penalties, there is much ambiguity in the aim of the language of the second policy that may complicate efforts to implement it. Specifically, it prohibits the general use of OSS penalties for disruption-specific infractions while simultaneously sanctioning their use for said infractions perceived as threatening, harmful, repeatedly obstructive, and intractable.

Therefore, in contrast to the first reform, the impact of the second reform on OSS for disruptive infractions is ambiguous in this context. Although the second reform aims to curb OSS penalties for disruption-specific infractions, it lacks the specificity to do so. OSS penalties for threatening, harmful, obstructive, or intractable behaviors could increase as OSS penalties for non-threatening, non-harmful, non-obstructive, or tractable behaviors concurrently decline. What constitutes threatening in the classroom is also open to scrutiny. Prior research concludes that Black students are often perceived as more assailing in the classroom compared to their White peers and are at a higher risk of exclusionary discipline and office referrals (e.g., Francis, 2012; Ritter & Anderson, 2018; Sagar & Schofield, 1980; Skiba, 2002).

Data Description

This study uses comprehensive student-level administrative data from the Rhode Island Department of Education (AY 2009–2010 to AY 2017–2018) for K-12 public schools. For each semester-year of enrollment, they provide information on each RI student's infraction, type of disposition or penalty, cumulative number of infractions, demographic characteristics (race-ethnicity and gender), free or reduced-price lunch status, disability status (i.e., individualized educational programs [IEP]), current grade, school, and school district.

The study categorizes student infractions as either attendance-specific (i.e., tardy, truancy, or absenteeism) or disruption-specific (i.e., insubordination/disrespect, disorderly conduct, and obscene/abusive language aimed at students and teachers). It constructs two key binary indicators of interest: (i) a binary indicator equal to one if the student has an attendance-specific infraction, and zero otherwise; and (ii) a binary indicator equal to one if the student has a disruption-specific infraction, and zero otherwise.

The study uses the suspension data from the full universe of students over the study period to construct two different outcome measures: (i) OSS probability, measured by a binary indicator equal to one if the student received OSS for a given infraction, and zero otherwise; and (ii) OSS duration, measured by the number of days suspended out of school for a given infraction and zero otherwise. Because the dual reforms might influence changes in student behavior and how they are reported, the analysis sample is comprised of students with and without reported infractions, allowing the analyses to account for changes in that regard (see the *Sensitivity Checks* subsection for further examination).

Table 1 presents summary means and standard deviations of the OSS outcomes, treatment infractions, and student characteristics. These descriptive results are shown for the full sample as well as the sample conditioned on OSS status. Within the universe of nearly 3 million student records, there were close to 283,000 RI students enrolled during that period, with a little over 15% of them suspended out of school.

For the full sample, 3% have an attendance-specific infraction and 4% have a disruption-specific infraction. Further, the average number of cumulative infractions per semester (treatment and otherwise) is about 0.4. Fifty percent of students are free or reduced-price lunch eligible, and close to 20% have a disability. Fifty-nine percent of students are White, 24% are Latinx, 9% are Black, and 7% are from other racial-ethnic groups. Over 50% of students are male and about 45% are in pre-school or elementary school.

Table 1 also presents summary statistics by OSS status. For the conditional sample receiving OSS, the average OSS duration is 2 days. Thirteen percent have attendance-specific infractions while an additional 46% have disruption-specific infractions. The average number of cumulative infractions per semester in the OSS sample (treatment and otherwise) is 3.3. Students penalized with OSS also display higher levels of disadvantage compared to the full sample: 75% are eligible for free or reduced-price lunch, and 32% have a disability. Black and Latinx students also comprise 17% and 34% of the OSS sample, respectively. In addition, 70% are male, and about 15% are in preschool or elementary school.

For the conditional sample receiving no OSS, 2% have attendance-specific infractions while an additional 2% have disruption-specific infractions. Moreover, the average number of cumulative infractions per semester (treatment and otherwise) is below 0.30. A little less than half are eligible for free or reduced-price lunch, and 17% have a disability. There are also

disproportionately fewer students of color, with Black students comprising 9% and Latinx students comprising 24% of the sample. Nearly 45% of the sample is in preschool or elementary school, suggesting that younger students face significantly less exposure to OSS penalties. Although 52% of the sample is male, this is substantially less than the OSS sample, which is 70% male.

These descriptive statistics indicate that students penalized with OSS are disproportionately poor, students of color, male, in higher grades, and characterized by disability. The dependent sample (paired) t-test confirms that these characteristic differences between the OSS and non-OSS samples are statistically different from zero.

Empirical Framework

The first suspension reform (effective May 2012) specifically targets attendance-specific infractions, and the second (effective June 2016) targets disruption-specific infractions. Evaluating the impact of each suspension reform on OSS outcomes in Rhode Island is well suited to quasi-experimental methods. The difference-in-difference (DD) estimation strategy measures the outcome gap between treatment and comparison groups, pre- and post-suspension reforms. The DD model for the Rhode Island context can be specified as follows:

$$Y_{ist} = \beta_0 + \beta_{1j} Tr_{ist}^{j} + \beta_{2j} Post_{t}^{j} + \delta_j \, Tr \cdot Post_{ist}^{j} + X_{ist}\,\beta_3 + \Gamma(t) + \iota_g + \lambda_s + \tau_t + \varepsilon_{ist} \quad (1)$$

where Y represents the OSS outcomes, i denotes the individual student, s denotes the school, t denotes the semester-year, and $j \in \{1,2\}$ denotes each suspension reform (estimated in separate regressions). Tr^1 is a binary indicator equal to one if the student has an attendance-specific infraction (treatment), and zero otherwise (comparison). $Post^1$ is equal to one if the semester-year is Spring 2012 or later, and zero if the semester-year is Fall 2011 or earlier. $Tr \cdot Post^1$, the variable of interest, is a binary indicator equal to one for all students with attendance-specific infractions in Spring 2012 or later, and zero otherwise. δ_1, the DD estimator, measures how the first reform influences the OSS outcomes of students with attendance-specific infractions, *ex post*.

Similarly, Tr^2 is a binary indicator equal to one if the student has a disruption-specific infraction (treatment), and zero otherwise (comparison). $Post^2$ is equal to one if the semester-year is Spring 2016 or later, and zero if the semester-year is Fall 2015 or earlier. $Tr \cdot Post^2$, the variable of interest, is a binary indicator equal to one for all students with disruption-specific infractions in Spring 2016 or later, and zero otherwise. δ_2, the DD estimator, measures how the second reform influences the OSS outcomes of students with disruption-specific infractions, *ex post*.

The DD model also accounts for student characteristics, X (race-ethnicity, gender, free or reduced-price lunch status, IEP status, and cumulative number of infractions per semester), school-specific and general linear time trends ($\Gamma(t)$) (to capture OSS outcome changes over time for individual schools and Rhode Island in general), as well as grade (ι_g), school (λ_s), and semester-year (τ_t) fixed effects.

However, the DD approach in this context is less preferred because it does not account for across and within school differences in the *ex ante* use of OSS for treatment infractions. For

instance, prior to the first reform, some schools never used OSS for attendance-specific infractions or may have excluded younger students from OSS for such infractions altogether. To address this concern, the study incorporates another treatment group into a triple-difference (DDD) framework to account for school- and grade-level variation in the *ex ante* use of OSS penalties for attendance- and disruption-specific infractions. By defining a second treatment group to account for the schools and grades that applied OSS for treatment infractions prior to the reforms, the DDD approach is expected to produce more precise estimates of reform impacts. Therefore, the DDD strategy incorporates variation from this additional treatment group and can be specified as follows:

$$Y_{igst} = \acute{\alpha}_0 + \acute{\alpha}_{1j} Tr_{ist}^{j} + \acute{\alpha}_{2j} Post_{t}^{j} + \acute{\alpha}_3 Sch_{gs}^{j} + \acute{\alpha}_{4j} Tr{\cdot}Sch_{igst}^{j} + \acute{\alpha}_{5j} Post{\cdot}Sch_{gst}^{j} + \acute{\alpha}_{6j} Tr{\cdot}Post_{ist}^{j} + \acute{\omega}_j Tr{\cdot}Sch{\cdot}Post_{igst}^{j} + X_{ist}\acute{\alpha}_7 + \Pi(t) + \acute{\imath}_g + \acute{o}_s + \acute{\eta}_t + \xi_{igst} \quad (2)$$

where *g* denotes grade-level, *Sch* is a binary indicator equal to one if the school used OSS to penalize any student for a treatment infraction in a particular grade prior to the respective reform (hereafter, "treatment schools") and zero for schools that did not. $Tr{\cdot}Sch{\cdot}Post^{j}$, our variable of interest, is a binary indicator equal to one if the student with a current treatment infraction is enrolled in a treatment school, *ex post*, and zero otherwise. $\acute{\omega}_j$, measures the net impact of each suspension reform on the OSS outcomes of students with treatment infractions enrolled in treatment schools, *ex post*. The DDD model also accounts for school-specific and general linear time trends ($\Pi(t)$), as well as grade ($\acute{\imath}_g$), school ($\acute{o}_s$), and semester-year ($\acute{\eta}_t$) fixed effects.

Despite the advantages of the DDD model over the DD model, it can only generate causal impacts if the pre-reform outcome trends of the treatment and comparison groups are parallel. To test for parallel outcome trends, the study employs the following event-study DDD design:

$$Y_{igst} = \alpha_0 + \alpha_{1j} Tr_{ist}^{j} + \alpha_{1j} Tr{\cdot}Sch_{igst}^{j} + \sum_{k=\underline{-a}}^{\overline{b}} \nu_{jk} Sch{\cdot}D_{gst}^{jk} + \sum_{k=\underline{-a}}^{\overline{b}} \psi_{jk} Tr{\cdot}D_{ist}^{jk} \quad (3)$$

$$+ \sum_{k=\underline{-a}}^{\overline{b}} \gamma_{jk} Tr{\cdot}Sch{\cdot}D_{igst}^{jk} + X_{ist}\alpha_2 + \Omega(t) + \psi_g + \theta_s + \pi_t + \zeta_{igst}$$

where D^{jk} denotes pre-reform semester-years ($\underline{-a} \leq k < -1$) and post-reform semester-years ($-1 < k \leq \bar{b}$) for each suspension reform, *j*. This specification essentially captures the relative timing of each reform. $D^{j(-1)}$ represents the semester-year prior to the one in which suspension reform, *j*, took effect and is the excluded category. Since the first suspension reform took effect in Spring 2012, the excluded category for this model is Fall 2011; likewise, the excluded category for the disruptive-infractions model is Fall 2015 since the second reform took effect in Spring 2016. The interpretation of γ_{jk} is slightly different from δ_j and $\acute{\omega}_j$. γ_{jk} measures the impact of each suspension reform on the OSS outcomes of the treatment groups in semester-year *k*, relative to the semester-year immediately preceding the one in which the reform took effect. The model also accounts for school-specific and general linear time trends ($\Omega(t)$), as well as grade (ψ_g), school (θ_s), and semester-year (π_t) fixed effects.

To evaluate how the suspension reforms impact racial-ethnic disparities in OSS, equation (2) can be modified by interacting three binary indicators for race-ethnicity (equal to one if the

student is Black, Latinx, or other racial-ethnicity, and zero if the student is White) with all triple-wise, pairwise, and binary indicators of interest from equation (2). Adding these interactions to the equation transforms the triple-difference (DDD) model to a quadruple-difference (DDDD) model, measuring the relative impact of each reform on the OSS outcomes of Black, Latinx, or other racial-ethnic students with treatment infractions enrolled in treatment schools, *ex post*.

The study therefore implements difference-in-difference (DD), triple-difference (DDD), and quadruple-difference (DDDD) models to evaluate the impacts of school suspension reforms on OSS outcomes and their corresponding racial-ethnic disparities. The study also presents heterogeneous impacts for DDD and DDDD models by student characteristics—free- and reduced-price lunch status, IEP status, gender, grade-level, and cumulative number of infractions per semester. To show these results, DDD and DDDD regressions are estimated by restricting the analysis sample to each subgroup of interest. To mitigate heteroskedasticity and serial correlation from all specifications, standard errors are two-way clustered at the school and semester-year levels.

Although this is not the first study in the education literature to use DDD estimation (e.g., Basant & Gitanjali, 2020; Bosio & Origo, 2020; Datta Gupta et al., 2018; Gjefsen et al., 2020), it is the first study to use the DDD strategy to evaluate school disciplinary reforms. It is also the first study to use DDDD estimation in the education literature in general. Therefore, the study makes an important contribution to the literature by using both DDD and DDDD estimation to support causal inference and measure racial-ethnic disparities in the impacts of suspension reforms.

Results

Overall Effects of Suspension Reforms (DD and DDD Results)

Table 2 presents the DD estimates from equation (1) for the first suspension reform. Column (1) indicates that without individual-level controls and linear time trends, the first suspension reform is associated with a 35-percentage point lower probability of OSS for those with attendance-specific infractions ($p < 0.01$). This estimate represents about 1.5 times the outcome standard deviation. In addition, Column (2) indicates that OSS duration fell by close to 0.5 days ($p < 0.01$) per student (hereafter "on average"), or about 80% of the outcome standard deviation. By adding individual-level controls (Columns 3 and 4), the estimates suggest that the first reform is associated with a 38-percentage point lower probability of OSS and a 0.56-day lower OSS duration on average ($p < 0.01$). Adding school-specific and general linear time trends to the model (Columns 5 and 6) does not change the findings substantially: the first reform is associated with a near 39-percentage point lower probability of OSS and a 0.56-day lower OSS duration on average ($p < 0.01$).

For disruption-specific infractions, Table 3 Column (1) shows that without individual-level controls and linear time trends, the second suspension reform is associated with a 1.1-percentage point higher probability of OSS ($p < 0.05$), or 5% of the outcome standard deviation. Adding individual-level controls and linear time trends (Columns 3 and 5) increases this estimate marginally, showing that the second suspension reform is associated with a 1.5-percentage point higher probability of OSS ($p < 0.05$). Table 3 also shows that the second suspension reform is positively associated with OSS duration, but estimates are not statistically different from zero.

However, these DD results might be biased if the *ex ante* use of OSS for treatment infractions differs across or within schools. To address this concern, Tables 4 and 5 present the DDD estimates from equation (2) measuring the net impact of each suspension reform on OSS outcomes of students with treatment infractions enrolled in treatment schools.[ii] Table 4 indicates that when individual-level controls and linear time trends are included in the model, the first reform lowers OSS probability by about 37.6 percentage points ($p < 0.01$) and OSS duration by a little less than half a day ($p < 0.01$) on average.

Although these DDD results are smaller than the DD estimates in Table 2, they still suggest strong compliance with the first reform as opposed to an endogenous shift in the way schools code their infractions. If school administrators were changing their internal coding systems to continue issuing OSS for attendance-specific infractions (Steinberg & Lacoe, 2018), then OSS outcomes for attendance-specific infractions would be systematically lower in these schools, *ex post*, and would produce larger negative DDD estimates as a result. However, the DDD estimates presented in Table 4 are smaller than the DD estimates in Table 2, allaying concerns that coding modifications may be driving these findings. To further test for the systematic recoding of attendance-specific infractions *ex post*, the study employs a comparative interrupted time series (CITS) model following Dee and Jacob (2011) and Anderson (2019).[iii] The total "effects" of the reform presented in Online Appendix Table D do not support the hypothesis that treatment schools recode attendance-specific infractions as disruption-specific infractions. Rather, these findings suggest that by Fall 2015 (i.e., the semester-year before the second reform took effect) and Spring 2018 (i.e., the final semester-year of the study period), treatment schools report statistically significantly fewer disruption-specific infractions after the passage of the first reform.

Table 5 indicates that the second reform raises OSS duration by 0.08 days on average ($p < 0.10$). However, including individual-level controls and linear time trends in the model eliminates the statistical significance of this finding, suggesting that the second reform does not statistically significantly change either of the OSS outcomes.

The study bolsters the descriptive findings from Anderson et al., (2019), showing that the statewide ban on suspensions in Arkansas lowered the use of OSS for truancy. On the other hand, the study's findings on the second reform contradict Lacoe and Steinberg (2018) and Steinberg and Lacoe (2018), which showed modest declines in OSS for nonviolent and classroom disorder infractions in response to Philadelphia's suspension reforms. The contradictory evidence points to the stark difference in compliance across jurisdictions. Rhode Island schools exhibit strong compliance with the first reform but weak compliance with the second reform, perpetuating conflicting results within the cross-jurisdictional literature.

Still, the validity of the DDD findings hinges on the parallel trend assumption that outcomes of the treatment and comparison groups are parallel prior to the effective date of the reform. Event-study DDD estimates from equation (3) test whether this assumption holds for each reform. For the first reform, plots (A) and (B) of Figure 1 illustrate event-study DDD estimates and corresponding 95% confidence intervals for the probability of OSS and OSS duration, respectively. Pre-reform outcome trends are shown to be parallel for the first suspension reform, with impact estimates that are not statistically different from zero in any of the pre-reform semester years. Moreover, the F-test for joint statistical significance (not shown) confirms that pre-reform impact estimates jointly fail to reject the null hypothesis, thereby reinforcing parallel trends. Post-reform estimates are generally large, negative, and statistically

significantly different from the pre-reform estimates (despite the uptick in Fall 2012). Therefore, these results underscore that, *ex post,* OSS penalties decline in response to the first reform.

For the second reform, plots (C) and (D) of Figure 1 illustrate event-study DDD pre-reform and post-reform estimates for each OSS outcome. Except for the estimates on Fall 2012 and earlier, all pre-reform impact estimates are statistically equivalent to zero, with small and non-robust t-statistics and joint significance F-statistics (not shown). Post-reform impact estimates are generally large and negative (despite the uptick in Fall 2016 and Spring 2017), suggesting that OSS penalties for disruption-specific infractions are generally trending downward, *ex post*. However, because the general DDD impact estimates from Table 5 are not statistically different from zero, these post-reform negative estimates have not yet superseded *ex ante* outcomes. That said, the event-study DDD estimates demonstrate parallel outcome trends up to four semesters prior to each reform, which supports causality within the DDD framework.

Racial-Ethnic Disparities (DDDD Results)

To evaluate the extent to which school suspension reforms change racial-ethnic disparities in OSS outcomes, the study employs a quadruple-difference (DDDD) design. Figure 2 plot (A) illustrates that for the first reform, Black students in treatment schools experience a decline in the probability of OSS by nearly 17 percentage points ($p < 0.01$) relative to White peers. Similarly, OSS duration shows a relative decline of 0.38 days on average ($p < 0.01$). Relative to White peers, Latinx students in treatment schools experience lower OSS odds and duration by 27 percentage points ($p < 0.01$) and 0.53 days ($p < 0.01$) on average, respectively. Further, other racial-ethnic students in treatment schools experience a decline in OSS odds and

duration by 30 percentage points ($p < 0.10$) and 0.62 days ($p < 0.10$) on average relative to White peers. These findings indicate that students of color saw disproportionately bigger reductions under the first suspension reform compared to their White peers, suggesting that the first reform has had a significant ameliorative effect on the racial-ethnic divide in OSS penalties.

The second reform, on the other hand, does not have similar effects. Figure 2 plot (B) indicates that the Black-White, Latinx-White, and Other-White disparities in OSS outcomes for disruption-specific infractions do not decline in response to the second reform. Relative to their White peers, Black students in treatment schools even receive longer OSS durations by 0.22 days ($p < 0.10$) on average. These findings suggest that the second reform has not successfully curbed the racial-ethnic divide in OSS penalties.

Heterogeneous Effects by Student Characteristics (DDD Results)

The study also investigates how the impact of each reform differs by student characteristics. Figure 3 illustrates the DDD impact estimates and their corresponding 95% confidence intervals by free or reduced-price lunch status, IEP status, gender, grade-level, and the cumulative number of infractions per semester. Under the first reform, plot (A) shows that the negative impact estimates are both substantial and statistically significant at the 1% level for each subgroup of interest. Most notably, students eligible for free or reduced-price lunch, students with disabilities, and male students experience the largest declines in OSS probability and duration by approximately 40 percentage points and over 0.50 days on average, respectively. This evidence represents encouraging news for the effectiveness of the first reform, given that students with these characteristics face higher OSS probabilities and longer OSS durations.

The impact estimates for younger students (i.e., preschool and elementary students) appear to be remarkably large in these analyses. Upon further inspection, however, the number of students subject to OSS for attendance-specific infractions at this level is quite small, leading to superficially elevated impact estimates from small sample bias. The small sample limitation, unfortunately, cannot be mitigated using maximum likelihood estimation. Maximum likelihood (ML) alternatives to the LPM framework (e.g., probit and logit) also produce large coefficients (>1), reinforcing that ML estimation does not work well in small samples (Long, 1997).

On the other hand, the second reform is ineffective at reducing OSS for most of the heterogeneous samples. Plot (B) shows that there are no robust impacts of the second reform on OSS outcomes for students eligible and ineligible for free or reduced-price lunch, students with disabilities, male students, and students in middle or high school. Students with multiple infractions per semester have higher OSS odds by about 12 percentage points ($p < 0.01$) and longer OSS durations by 0.28 days ($p < 0.01$) on average. Students without disabilities or with only one infraction experience longer OSS durations by 0.15 and 0.31 days on average, respectively. By contrast, female students are less likely to be penalized with OSS by 9.5 percentage points ($p < 0.10$). Another group experiencing a decline in OSS penalties in response to the second reform is younger students in preschool or elementary school, for whom the probability of OSS declines by 17 percentage points ($p < 0.05$) on average.

Heterogeneous Effects by Intersectional Student Subgroups (DDDD Results)

Evaluating the racial-ethnic differences in heterogeneous effects reveals even more nuanced patterns. Figure 4 shows that for the first reform, Black students generally have larger

declines in OSS penalties compared to their White peers. These impacts are most pronounced for those who are ineligible for free or reduced-price lunch, without disabilities, male, in higher grade levels, or charged with only one infraction. Except for the subgroup with only one infraction, Latinx students from all other subgroups experience lower OSS penalties in response to the first reform compared to their White peers. Students of other racial-ethnic groups who are male, have multiple infractions, and in any grade level experience relatively lower OSS penalties as well.

On the other hand, Figure 5 illustrates that Black students are further disadvantaged by the second reform. Black students who are free or reduced-price lunch eligible, who have no disabilities, who are female, or who have only one infraction, experience longer OSS durations on average relative to their White peers. The second reform also raises OSS penalties for Latinx students who are without disabilities, who are in younger grades, or who have one or more infractions relative to their White peers. Students of other racial-ethnic groups who have no disabilities, who are female, who are in higher-grade levels, or have one or more infractions also experience relatively worse OSS outcomes.

In summary, these findings reinforce that although suspension reforms aim to curb racial-ethnic divides, they can actually exacerbate them. The Rhode Island case shows that generally, the first suspension reform reduces the racial-ethnic divide in OSS penalties for students of color across heterogeneous samples. However, the second reform has been mostly ineffective, stagnating or exacerbating racial-ethnic disparities in OSS penalties for most heterogeneous samples. Further research dedicated to understanding these nuances is highly encouraged.

Sensitivity Checks

In contrast to attendance-specific infractions, what constitutes a minor disruptive infraction may vary significantly across school settings. Therefore, the null effect of the second reform on OSS outcomes may not be uniform across types of disruption-specific infractions (i.e., disorderly conduct, insubordination/disrespect, obscene language directed towards students, and obscene language directed towards teachers). Online Appendix Table A shows DDD impact estimates for each type of disruption-specific infraction. A pointed evaluation of insubordination/disrespect and student-aimed obscene language infractions does not indicate statistically significant changes in OSS outcomes in response to the second reform. However, when disorderly conduct is differentiated from the other disruption-specific infractions, the impact of the second reform on OSS duration is now statistically different from zero, increasing the average OSS duration by 0.17 days ($p < 0.01$) on average. It is also noteworthy that OSS penalties for teacher-aimed obscene language decline in response to the second reform, with the probability of OSS falling by about 7 percentage points ($p < 0.10$) and OSS duration declining by 0.30 days ($p < 0.01$) on average. These findings suggest that although the second reform appears to have no robust effects on disruption-specific infractions in general, infraction-specific impacts are much more nuanced, with disorderly conduct inducing longer OSS durations on average, and teacher-aimed obscenities drawing lower OSS penalties. Therefore, what constitutes threatening, harmful, repeatedly obstructive, and intractable behaviors is likely influenced by diverse interpretations and contexts, meriting further research to understand how OSS penalties are assigned.

Although the DDD model accounts for school- and grade-level variation in the *ex ante* use of OSS for treatment infractions, changes in infraction reports over time might still work to confound the results. Therefore, the DDD model is modified to include infraction-specific linear and quadratic trends as well as infraction-by-year fixed effects. Online Appendix Table B indicates how infraction-specific trends and infraction-by-year fixed effects change the general DDD impact estimates. The revised DDD estimates in Panels A and B are generally larger in magnitude compared to the original DDD estimates, suggesting that changes to infraction reports over time attenuate the impacts of both suspension reforms toward zero.

Controlling for cumulative infractions per semester may be a source of reverse causality or simultaneity bias in the models given that OSS penalties could influence infractions over time. To test the sensitivity of the findings to this control variable, Online Appendix Table C presents DDD results with cumulative infractions per semester excluded from the model (Panel A) as well as DDD results with cumulative infractions per semester included in the model but as a lagged measure (Panel B). These results remain statistically similar to the original findings.

Finally, the impact of the second reform could be confounded by the impact of the first reform, and vice versa. As such, to re-evaluate the impact of the first reform, the study restricts the analysis sample to only semester-years before the second reform took effect. Similarly, to re-evaluate the impact of the second reform, the study restricts the sample to only the semester-years after the first reform took effect. Panel C shows that the results are statistically similar to the results presented in Tables 4 and 5. To further evaluate whether the reform impacts are confounded in the model, the study measures the impacts of both reforms using a single regression. Panel D shows that the DDD effect sizes are similar to those presented in Tables 4 and 5, with estimates for the second reform now statistically significant at the 5% level.

However, by incorporating numerous pairwise and triple-wise interactions in the same regression, the model exhibits symptoms of collinearity (primarily with the time interactions). As such, the original model, albeit empirically conservative, is still preferred in light of this concern.

Conclusion

The study finds that the first reform lowers the probability of OSS by 37.6 percentage points ($p < 0.01$) and OSS duration by about half a day ($p < 0.01$) on average for students with attendance-specific infractions enrolled in treatment schools. In general, the second reform does not change either of the OSS outcomes statistically significantly. These findings suggest that the second reform may be less effective in reducing OSS outcomes relative to its predecessor; however, event-study DDD estimates show that impacts on OSS outcomes are trending downward, suggesting that the second reform may eventually help curb OSS for disruption-specific infractions over time.

The impacts of the dual reforms on racial-ethnic disparities in OSS penalties are also uneven. The first reform lowers the Black-White disparity in OSS probability by about 17 percentage points ($p < 0.01$), the Latinx-White disparity by 27 percentage points ($p < 0.01$), and the Other-White disparity by 30 percentage points ($p < 0.10$). Black-White, Latinx-White, and Other-White disparities in average OSS duration fall by 0.38, 0.53, and 0.62 days, respectively. These findings suggest that the first reform substantially reduces the racial-ethnic divide in OSS penalties in Rhode Island. Conversely, the second reform does not show ameliorative effects on racial-ethnic disparities. In general, Black-White, Latinx-White, and Other-White disparities in

OSS outcomes do not decline in response to the second reform, with the Black-White disparity in OSS duration increasing by 0.22 days ($p < 0.10$) on average.

There are multiple theories that might explain the apparent inefficacy of the second reform. Given that it is implemented at the state-level with multiple school districts, this may account for higher variance in the interpretation of the policy (as opposed to implementation in a single school district such as Philadelphia (Lacoe & Steinberg, 2018; Steinberg & Lacoe 2018)). Yet, it is essential to acknowledge that the second reform may be working as intended, with fewer OSS penalties for some disruption-specific infractions, and concurrently more OSS penalties for other disruption-specific infractions. The study uncovers evidence to bolster this theory. OSS penalties for obscenities directed towards teachers declined in response to the second reform, even though OSS penalties for other types of disruption-specific infractions (i.e., disorderly conduct, insubordination/disrespect, and obscene language directed towards students) increased or remained statistically unchanged. Students with multiple infractions are also more likely to receive OSS penalties presumably because other corrective measures have been ineffective. Nevertheless, such "high-flyer" or high-risk students spend more time away from the classroom, which could lead to delinquency (e.g., Farrington, 1980; Gavin, 1996, 1997; McCluskey et al., 2004) and other antisocial behaviors (Conger, 1976; Elliott et al., 1985; Longshore et al., 2004).

In addition, the policy ambiguity of the second reform cannot be ignored in this context. The first reform targets a distinct infraction (i.e., attendance) and thus may be easier to implement relative to the second reform, where school administrators might interpret the notion of "disruption" with less uniformity. The policy implementation literature has long documented that there can be stark discrepancies between the language of a policy and how it is executed

(e.g., Baier et al., 1986; Cohen et al., 2007; Firestone, 1989; Matland, 1995; Sabatier & Mazmanian, 1979). The second reform also appears to exemplify an adoption-to-implementation conflict: it aims to curb OSS for disruptive behaviors while legally supporting OSS if student disruptions are considered threatening, harmful, repeatedly obstructive, or intractable.[iv] As such, the rise in OSS penalties for disruption-specific infractions for certain subgroups may be on account of this ambiguous policy language or the subjective interpretation of this language (e.g., Baier et al., 1986; Cohen et al., 2007; Firestone, 1989). For instance, disorderly conduct may be more harshly punished relative to teacher-aimed obscenities because its broad classification can be interpreted as severe in a myriad of ways.

It is also unclear whether school superintendents have reviewed disciplinary data on the disproportionate impact of the reform—a mandate of the second legislation. The existing literature suggests that students of color are punished more harshly for minor infractions (e.g., Anderson & Ritter, 2017; Beck & Muschkin, 2012; Gopalan & Nelson, 2019; Harper, 2019; Krezmien et al., 2006; Morris & Perry, 2016). It can thus be argued that disruptive behaviors classified as non-threatening or non-harmful for White students may not be perceived as such for students of color (e.g., Francis, 2012; Ritter & Anderson, 2018; Sagar & Schofield, 1980; Skiba, 2002).

Still, there is much to be done to reduce racial-ethnic disparities in the use of OSS in Rhode Island. The fact that racial-ethnic disparities have not declined in general for disruption-specific infractions bolsters the existing literature that shows students of color continue to be unequally punished relative to their White peers (e.g., Anderson & Ritter, 2017; Beck & Muschkin, 2012; Gopalan & Nelson, 2019; Gregory et al., 2010; Harper, 2019; Kinsler, 2011; Krezmien et al., 2006; Morris & Perry, 2016; Owens & McLanahan, 2019; Skiba et al., 2014;

Skiba et al., 2002; Skiba & Williams, 2014). Therefore, an annual external review of OSS patterns may help guard against disparate impact. Further, qualitative research investigating how OSS penalties are assigned to disruption-specific infractions is also key to understanding disparate impact in this context.

Notwithstanding, policymakers and practitioners must join forces to address the conflicting objectives of the second legislation. The ACLU of Rhode Island hailed § 16-2-17 (i.e., the second reform) as a vital instrument for curbing OSS for minor infractions such as "disorderly conduct" or "disrespect" (ACLU of Rhode Island 2016). However, the study does not uncover evidence to fully support this hypothesis—OSS penalties for most disruption-specific infractions increase or do not change under the second reform. If the intent of the reform is indeed to reduce OSS for all disruption-specific infractions, then the language of the policy must be revisited and revised. Legislative language that clearly states which disruption-specific infractions should be penalized with OSS may help school administrators operationalize these reforms more effectively and equitably. Some theorists also argue that a congruence of "top-down" and "bottom-up" perspectives is requisite for effective policy implementation (e.g., Cohen et al., 2007; Matland, 1995). As such, melding solutions from both policymakers and school administrators could help stem the inefficacies of this policy effort.

Other in-school policies and programs aimed directly at reducing disruption-specific infractions may also help curb racial-ethnic disparities. Recently, two Rhode Island school districts (Central Falls and Westerly) implemented restorative justice practices, although the outcomes of this implementation have yet to be released (see Liberman & Katz, 2017). Still, numerous descriptive studies tout the effectiveness of restorative justice practices in reducing OSS (e.g., Armour, 2013, 2016; Baker 2008; Gonzalez, 2012, 2015; Gregory et al., 2016; Jain et

al., 2014; Lewis, 2009; Liberman & Katz, 2017; McCold, 2008; McMorris et al., 2013; Mirsky, 2007; Simson, 2012; Sumner et al., 2010). This literature suggests that restorative justice practices are not only linked to lower suspensions, but also lower Black-White suspension gaps (e.g., Gonzalez, 2015; Jain et al., 2014; Simson, 2012). The impact evaluation of restorative justice practices, however, reveal more nuanced findings (Augustine et al., 2018). Elementary, African-American, and low-income students experience improvements in suspension outcomes, but not middle school and IEP students. There is also no improvement in achievement in schools exposed to restorative justice practices relative to control schools.

Besides restorative justice, some studies argue that ISS can be a more suitable alternative to OSS if it is effectively structured (Jones, 2019; Gonzalez, 2012). However, other studies show an inverse link between ISS and academic achievement, albeit smaller than the inverse link for OSS (Anderson et al., 2019; Noltemeyer et al., 2015). Still, there are alternatives to restorative justice and ISS that have been successful in reducing OSS penalties as well as associated racial-ethnic gaps (see Welsh & Little, 2018). For instance, culturally relevant interventions (e.g., Manhood Development Program (MDP)) and social-emotional learning (e.g., Promoting Alternative THinking Strategies (PATHS) program) have helped lower OSS, along with other forms of disciplinary incidents for students of color (e.g., Osher et al., 2015; Watson, 2014; Welsh & Little, 2018).

Nevertheless, the study provides causal evidence in support of the first suspension reform (§ 16-19-1) to reduce OSS penalties and their corresponding racial-ethnic disparities, thereby enriching the growing literature in this area (e.g., Anderson, 2018; Anderson et al., 2019; Zarecki, 2019). The conflicting evidence from the study strongly suggests that impact evaluations in other jurisdictions can shed more light on the efficacy of suspension reforms.

Moreover, suspension reforms are likely to have significant impacts on other student outcomes such as grade retention and high-school dropout. Future research would strongly benefit from exploring the impact of suspension reforms on these outcomes as well.

Acknowledgements

The author would like to thank the editor and two anonymous reviewers for their excellent suggestions for improving this manuscript. She would also like to thank Smith College and Connecticut College for their support in completing this project; and the ACLU of Rhode Island for answering pertinent questions related to the dual suspension reforms. The author is also grateful to the Population Studies and Training Center at Brown University, which receives funding from the NIH (P2C HD041020), for general support. Data for this project were obtained with the assistance of DataSpark, University of Rhode Island.

References

American Academy of Pediatrics. (2013). Out-of-school suspension and expulsion. *Pediatrics*, *131*, e1000–e1007.

American Civil Liberties Union of Rhode Island. (2015). Blacklisted: 2013-2014 racial bias in school suspensions in Rhode Island in the 2013-2014 school year.

American Civil Liberties Union of Rhode Island. (2016). ACLU applauds new law restricting school-suspension use. https://www.aclu.org/press-releases/aclu-applauds-new-law-restricting-school-suspension-use

American Psychological Association. (2008). Are zero tolerance policies effective in schools? An evidentiary review and recommendations. *American Psychologist*, *63*, 852–862.

Anderson, K. P. (2018). Inequitable compliance: Implementation failure of a statewide student discipline reform. *Peabody Journal of Education*, *93*(2), 244-263.

Anderson, K. P. (2020). Academic, attendance, and behavioral outcomes of a suspension reduction policy: Lessons for school leaders and policy makers. *Educational Administration Quarterly*, 56(3), 435-471.

Anderson, K. P., Egalite, A. J., & Mills, J. N. (2019). Discipline reform: The impact of a statewide ban on suspensions for truancy. *Journal of Education for Students Placed at Risk (JESPAR)*, *24*(1), 68-91.

Anderson, K. P., & Ritter, G. W. (2017). Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a U.S. state. *Education Policy Analysis Archives*, 25(49).

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.

Arcia, E. (2006). Achievement and enrollment status of suspended students: Outcomes in a large, multicultural school district." *Education and Urban Society* 38, 359–69.

Armour, M. (2013). Ed. white middle school restorative discipline evaluation: implementation and impact, 2012/2013 Sixth and Seventh Grade, Austin, Tex.: University of Texas, The Institute for Restorative Justice and Restorative Dialogue.

Armour, M. (2016). Ed. white middle school restorative discipline evaluation: implementation and impact, 2014/2015, Sixth, Seventh, and Eighth Grades, Austin, Tex.: University of Texas, The Institute for Restorative Justice and Restorative Dialogue, 2016.

Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., & Joseph, A. A. (2018). *Can restorative practices improve school climate and curb suspensions. An evaluation of the impact of restorative practices in a mid-sized urban school district.* Santa Monica, CA: Rand.

Bacher-Hicks, A., Billings, S. B., & Deming, D. J. (2019). *The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime* (No. w26257). National Bureau of Economic Research.

Baier, V. E., March, J. G., & Saetren, H. (1986). Implementation and ambiguity. *Scandinavian Journal of Management Studies*, 2(3-4), 197-212.

Baker, Myriam L. (2008). DPS restorative justice project executive summary 2007–2008, Denver, Colo.: Outcomes, Inc.

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.

Balfanz, R., Byrnes, V., & Fox, J. H. (2015). Sent home and put off track: The antecedents, disproportionalities, and consequences of being suspended in the 9th grade. In D. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 17–30). New York City: Teachers College Press.

Basant, R., & Sen, G. 2020. Quota-based affirmative action in higher education: Impact on other backward classes in India. *Journal of Development Studies, 56(2),* 336–60.

Beck, A. N., & Muschkin,C.G. (2012). The enduring impact of race: Understanding the disparities in student disciplinary infractions and achievement. *Sociological Perspectives*, *55*, 637–662.

Berman, P. (1978). The study of macro- and micro- implementation. *Public Policy 26,* 157–84.

Bosio, Giulio, & Origo, F. 2020. Who gains from active learning in higher education? *Education Economics*, *28(3),* 311–31.

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*, 133–148.

Cobb-Clark, D. A., Kassenboehmer, S. C., Le, T., McVicar, D., & Zhang, R. (2015). Is there an educational penalty for being suspended from school? *Education Economics*, *23*, 376–395.

Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education*, *113(4),* 515–548.

Conger, R. D. (1976). Social control and social learning models of delinquent behavior: A synthesis. *Criminology*, *14*, 17–39.

Cuellar, A. E., & Markowitz, S. (2015). School suspension and the school-to-prison pipeline. *International Review of Law and Economics*, *43*, 98–106.

Datta Gupta, N., Dubey, A., & Simonsen, M. (2018). Rising school attendance in Rural India: An evaluation of the effects of major educational reforms." *Education Economics*, *26*(2), 109–28.

Davis, J. E., & Jordan, W. J. (1994). The effects of school context, structure, and experiences on African American males in middle and high school. *The Journal of Negro Education, 63,* 570–587.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30, 418-446.

DiSalvo, R. (2018). The effects of suspension policy on learning: evidence from gender gaps in exceptional districts. *Available at SSRN 3249835*.

Elliott, D., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use*. Beverly Hills, CA: Sage.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. New York, NY: Council of State Governments Justice Center.

Farrington, D. (1980). Truancy, delinquency, the home, and the school. In L. Hersov & I. Berg (Eds.), *Out of school: Modern perspectives in truancy and school refusal* (pp. 49-63). New York: John Wiley.

Firestone, W. A. (1989). Using reform: Conceptualizing district initiative. *Educational Evaluation and Policy Analysis*, *11(2),* 151–164.

Francis, D. V. (2012). Sugar and spice and everything nice? Teacher perceptions of Black girls in the classroom. *The Review of Black Political Economy*, *39*(3), 311–320.
Garry, E. M. (1996). Truancy: First step to a lifetime of problems. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.
Gavin, T. (1997). Truancy: Not just kids' stuff anymore. FBI Law Enforcement Bulletin, *66*, 8 – 14.
Gjefsen, H.M., & Gunnes, T. 2020. How school accountability affects teacher mobility and teacher sorting. *Education Economics*, *28*(5), 455–73.
Gopalan, M., & Nelson, A. A. (2019). Understanding the racial discipline gap in schools. *AERA Open*, *5*(2), 2332858419844613.
González, T. (2012). Keeping kids in schools: restorative justice, punitive discipline, and the school to prison pipeline. *Journal of Law & Education*, *41*(2), 281–335.
González, T. (2015). Socializing schools: addressing racial disparities in discipline through restorative justice. In Daniel J. Losen, ed., *Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion*, New York: Teachers College Press.
Gregory, A., Skiba, R.J., & Pedro A. Noguera. (2010). The achievement gap and the discipline gap: two sides of the same coin? *Educational Researcher*, *39*(1), 59–68.
Gregory, A., Clawson, K., Davis, A., and Gerewitz, J. (2016). Restorative practices to transform teacher-student relationships and achieve equity in school discipline. *Journal of Educational and Psychological Consultation*, *26*(4), 325–353.
Harper, K., Ryberg, R., & Temkin, D. (2019). Black students and students with disabilities remain more likely to receive out-of-school suspensions, despite overall declines. *Child Trends.*
Henault, C. (2001). Zero tolerance in schools. *J. Law Educ*., *30*, 547–553.
Hwang, N., & Domina, T. (2020). Peer disruption and learning: Links between suspensions and the educational achievement of non-suspended students. *Education Finance and Policy*, 1–44.
Jain, S., Bassey, H., Brown, M.A., & Kalra, P. (2014). *Restorative Justice in Oakland Schools: Implementation and Impacts, An Effective Strategy to Reduce Racially Disproportionate Discipline, Suspensions, and Improve Academic Outcomes*, Oakland, Calif.: Data in Action.
Jones, E.P. (2018). The link between suspensions, expulsions, and dropout rates. https://www.americaspromise.org/opinion/link-between-suspensions-expulsions-and-dropout-rates
Kinsler, J. (2011). Understanding the Black–White school discipline gap. *Economics of Education Review* 30(6), 1370-83.
Knesting, K., & Skiba, R. J. (2002). *Zero tolerance, zero evidence: An analysis of school disciplinary practices. New Directions for Youth Development.* Bloomington, IN: Indiana University Education Policy Center.
Krezmien, M.P., Leone, P.E., & Achilles, G.M. (2006). Suspension, race, and disability: Analysis of state-wide practices and reporting. *Journal of Emotional and Behavioral Disorders*, *14*, 217–226.
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.

Lacoe, J., & Steinberg, M. P. (2019). Do suspensions affect student outcomes?. *Educational Evaluation and Policy Analysis*, *41*(1), 34-62.
Lewis, S. (2009). Improving school climate: findings from schools implementing restorative practices, Bethlehem, Pa.: International Institute for Restorative Practices.
Liberman, A., & Katz, M. (2017). Implementing restorative justice in Rhode Island schools. Urban Institute.
Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences*. Thousand Oaks, CA: Sage.
Longshore, D., Chang, E., Hsieh, S. C., & Messina, N. (2004). Self-control and social bonds: A combined control perspective on deviance. *Crime & Delinquency*, *50*(4), 542–564.
Marchbanks, M. P., III, Blake, J. J., Booth, E. A., Carmichael, D., Seibert, A. L., & Fabelo, T. (2014). The economic effects of exclusionary discipline on grade retention and high school dropout. In D. Losen (Ed). *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 59–74). New York City: Teachers College Press.
Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, 5(2), 145–174.
Mirsky, L. (2007). Safer saner schools: Transforming school culture with restorative practices. *Reclaiming Children and Youth: The Journal of Strength-Based Interventions*, *16*(2), 5–12.
Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, *63*(1), 68–86.
McCluskey, C. P., Bynum, T. S., & Patchin, J. W. (2004). Reducing chronic absenteeism: An assessment of an early truancy initiative. *Crime & Delinquency*, *50*(2), 214–234.
McCold, Paul. (2008). Evaluation of a restorative milieu: Restorative practices in context. In Holly Ventura Miller, ed., *Restorative Justice: From Theory to Practice (Sociology of Crime, Law and Deviance (Vol. 11)*, Bingley, UK: Emerald Group Publishing Limited, pp. 99–137.
McMorris, B. J., Beckman, K.J., Shea, G., Baumgartner, J., & Eggert, R.C. (2013). *Applying Restorative Justice Practices to Minneapolis Public Schools Students Recommended for Possible Expulsion: A Pilot Program Evaluation of the Family and Youth Restorative Conference Program*, Final Report, Minneapolis, Minn.: University of Minnesota.
Nicholson-Crotty, S., Birchmeier, Z., & Valentine, D. (2009). Exploring the impact of school discipline on racial disproportion in the juvenile justice system. *Social Science Quarterly*, *90*(4), 1003–1018.
Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*, 224–240.
Osher, D., Poirier, J., Jarjoura, R., Brown, R., & Kendziora, K. (2015). Avoid simple solutions and quick fixes: Lessons learned from a comprehensive district-wide approach to improve conditions for learning. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 192–206). New York, NY: Teachers College Press.
Owens, J., & McLanahan, S. S. (2019). Unpacking the drivers of racial disparities in school suspension and expulsion. *Social Forces*, *98*(4), 1548-1577.

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

Raffaele-Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New Directions for Youth Development, 2003*(99), 17–33.

Raffaele-Mendez, L. M., Knoff, H. M., & Ferron, J.M. (2002). School demographic variables and out-of-school suspension rates: A quantitative and qualitative analysis of a large, ethnically diverse school district. *Psychology in the Schools, 39,* 259–277.

Rausch, K. M., & Skiba, R. J. (2005). The academic cost of discipline: The contribution of school discipline to achievement. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Rhode Island Kids Count Factbook. (2011). Providence, RI: Rhode Island KIDS COUNT. http://www.rikidscount.org/Portals/0/Uploads/Documents/2011%20Rhode%20Island%20Kids%20Count%20Factbook.pdf

Ritter, G. W., & Anderson, K. P. (2018). Examining disparities in student discipline: Mapping inequities from infractions to consequences. *Peabody Journal of Education, 93,* 161 – 173.

Rumberger, R. W., & Losen, D. J. (2016). The high cost of harsh discipline and its disparate impact. The Center for Civil Rights Remedies at The Civil Rights Project / Proyecto Derechos Civile.

Sabatier, P., & Mazmanian, D. (1979). The conditions of effective implementation: A guide to accomplishing policy objectives. *Policy analysis,* 481–504.

Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology, 39*(4), 590.

Sedlak, A. J., & McPherson, K. (2010). Survey of youth in residential placement: Youth's needs and services. SYRP Report. Rockville, MD: Westat.

Simson, D. (2012). Restorative justice and its effects on (racially disparate) punitive school discipline. In *7th Annual Conference on Empirical Legal Studies Paper*.

Skiba, R. (2010). Zero tolerance and alternative discipline strategies. *National Association of School Psychologists Communique, 39*(1), 28–30.

Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education, 47*(4), 546–564.

Skiba, R. J., Chung, C. G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal, 51*(4), 640–670.

Skiba, R., Reynolds, C. R., Graham, S., Sheras, P., Conoley, J. C., Garcia-Vazquez, E., & Edmiston, A. (2006). Are zero tolerance policies effective in the schools? An evidentiary review and recommendations. An official report of the APA.

Skiba, R. J., & Rausch, M. K. (2004). *The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores*. Bloomington, IN: Center for Evaluation and Education Policy, Indiana University.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, *34*(4), 317–342.

Skiba, R. J., and Williams, N.T. (2014). Are Black kids worse? Myths and facts about racial differences in behavior. In *The Equity Project at Indiana University*. Bloomington, IL: Indiana University.

Stearns, E., & Glennie, E. J. (2006). When and why dropouts leave school. *Youth & Society*, *38*, 29–57.

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

Stevens, W. D., Sartain, L., Allensworth, E. M., & Levenstein, R., with Guiltinan, S., Mader, N., Huynh, M. H., & Porter, S. (2015). *Discipline practices in Chicago schools: Trends in the use of suspensions and arrests.* Chicago, IL: University of Chicago Consortium of Chicago School Research.

Sumner, M. D., Silverman, C. J., & Frampton, M. L. (2010). School-based restorative justice as an alternative to zero-tolerance policies: Lessons from West Oakland. University of California, Berkeley, School of Law.

Swanson, E., Erickson, H. H., & Ritter, G. W. (2017). *Examining the impacts of middle school disciplinary policies on 9th grade retention.* (EDRE Working Paper 2017-11).

US Department of Education and US Department of Justice. (2014). Dear colleague letter on the nondiscriminatory administration of school discipline.

Wald, J., & Kurlaender, M. (2003). Connected in Seattle? An exploratory study of student perceptions of discipline and attachments to teachers. *New Directions for Youth Development*, *2003*(99), 35–54.

Watson, V. (2014). *The Black sonrise: Oakland Unified School District's commitment to address and eliminate institutionalized racism (an evaluation report)*. Oakland, CA: Office of African American Male Achievement, Oakland Unified School District.

Welsh, R. O., & Little, S. (2018). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research*, *88*(5), 752-794.

Wilson, A.C. (2020). Carceral schools and college expectations: Evidence from the National Crime Victimization School Crime Supplement. Job Market Paper.

Wolf, K. C., & Kupchik, A. (2017). School suspensions and adverse experiences in adulthood. *Justice Quarterly*, *34*(3), 407–430.

Zarecki, D. (2019). Banning math progress: The academic impact of California's suspension bans. *Available at SSRN 3251581*.

FIGURE 1

Event-Study DDD Impact Estimates

Note. The figure presents event-study DDD impact estimates from equation (3) showing the impact of each suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))). The impact estimates are shown along with corresponding 95% confidence intervals. Each event-study regression controls for gender, free or reduced-lunch status, IEP status, cumulative number of infractions per semester, school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects. The coefficient on Fall 2011, the excluded category, is normalized to zero for the first reform (top panel). The coefficient on Fall 2015, the excluded category, is normalized to zero for the second reform (bottom panel).

TABLE 1

Summary Statistics

	All		OSS Received		No OSS Received		$Pr(\|T\| > \|t\|) = 0$
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
OSS	.05	.23	1	0	0	0	
OSS Duration	.11	.61	2.01	1.73	0	0	.00
Attendance Infraction	.03	.16	.13	.34	.02	.14	.00
Disruptive Infractions	.04	.21	.46	.5	.02	.14	.00
Free/Reduced-Price Lunch	.5	.5	.75	.43	.49	.5	.00
IEP	.18	.38	.32	.47	.17	.37	.00
White	.59	.49	.42	.49	.6	.49	.00
Black	.09	.29	.17	.38	.09	.28	.00
Latinx	.24	.43	.34	.47	.24	.43	.00
Other Race/Ethnicity	.07	.26	.07	.26	.07	.26	.01
Male	.53	.5	.7	.46	.52	.5	.00
Pre-School	.09	.28	.01	.11	.09	.29	.00
Grade 1	.07	.26	.02	.13	.07	.26	.00
Grade 2	.07	.26	.02	.14	.07	.26	.00
Grade 3	.07	.26	.03	.16	.07	.26	.00
Grade 4	.07	.26	.03	.17	.07	.26	.00
Grade 5	.07	.26	.04	.19	.07	.26	.00
Grade 6	.07	.26	.09	.28	.07	.26	.00
Grade 7	.08	.27	.12	.33	.07	.26	.00
Grade 8	.08	.27	.14	.34	.07	.26	.00
Grade 9	.09	.29	.2	.4	.09	.28	.00
Grade 10	.08	.28	.14	.34	.08	.27	.00
Grade 11	.07	.26	.1	.3	.07	.26	.00
Grade 12	.07	.26	.08	.27	.07	.26	.00
Number of Infractions	.42	1.8	3.3	3.1	.26	1.55	.00
Number of Students	282,686		44,226		282,207		
N (Student x Sem.-Years)	2,938,375		159,184		2,779,191		

TABLE 2

DD Estimates of the First Suspension Reform

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)
*Tr*Post*[1]	-0.348***	-0.483***	-0.384***	-0.556***	-0.385***	-0.558***
	(0.005)	(0.008)	(0.008)	(0.015)	(0.008)	(0.015)
Fixed Effects	Y	Y	Y	Y	Y	Y
Individual Controls	N	N	Y	Y	Y	Y
Linear Trends	N	N	N	N	Y	Y
R-Squared	.099	.045	.214	.112	.214	.112
N	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375

Note. The table presents general DD estimates from equation (1) showing the relationship between the first suspension reform and OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))). Each DD regression controls for grade, school, and semester-year fixed effects. Individual controls include race, gender, free or reduced-lunch status, IEP status, and cumulative number of infractions per semester. Linear trends include school-specific and general linear time trends. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE 3

DD Estimates of the Second Suspension Reform

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)
*Tr*Post*^{2}	0.011**	0.002	0.015**	0.010	0.015**	0.010
	(0.005)	(0.011)	(0.006)	(0.013)	(0.006)	(0.013)
Fixed Effects	Y	Y	Y	Y	Y	Y
Individual Controls	N	N	Y	Y	Y	Y
Linear Trends	N	N	N	N	Y	Y
R-Squared	.263	.103	.298	.129	.298	.129
N	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375

Note. The table presents general DD estimates from equation (1) showing the relationship between the second suspension reform and OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS Days))). Each DD regression controls for grade, school, and semester-year fixed effects. Individual controls include race, gender, free or reduced-lunch status, IEP status, and cumulative number of infractions per semester. Linear trends include school-specific and general linear time trends. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE 4

DDD Impact Estimates of the First Suspension Reform

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)
*Tr*Sch*Post*[1]	-0.373***	-0.493***	-0.376***	-0.494***	-0.376***	-0.495***
	(0.009)	(0.017)	(0.025)	(0.053)	(0.025)	(0.053)
Fixed Effects	Y	Y	Y	Y	Y	Y
Individual Controls	N	N	Y	Y	Y	Y
Linear Trends	N	N	N	N	Y	Y
R-Squared	.099	.045	.214	.112	.214	.112
N	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375

Note. The table presents general DDD impact estimates from equation (2) showing the impact of the first suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))). Each DDD regression controls for grade, school, and semester-year fixed effects. Individual controls include race, gender, free or reduced-lunch status, IEP status, and cumulative number of infractions per semester. Linear trends include school-specific and general linear time trends. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE 5

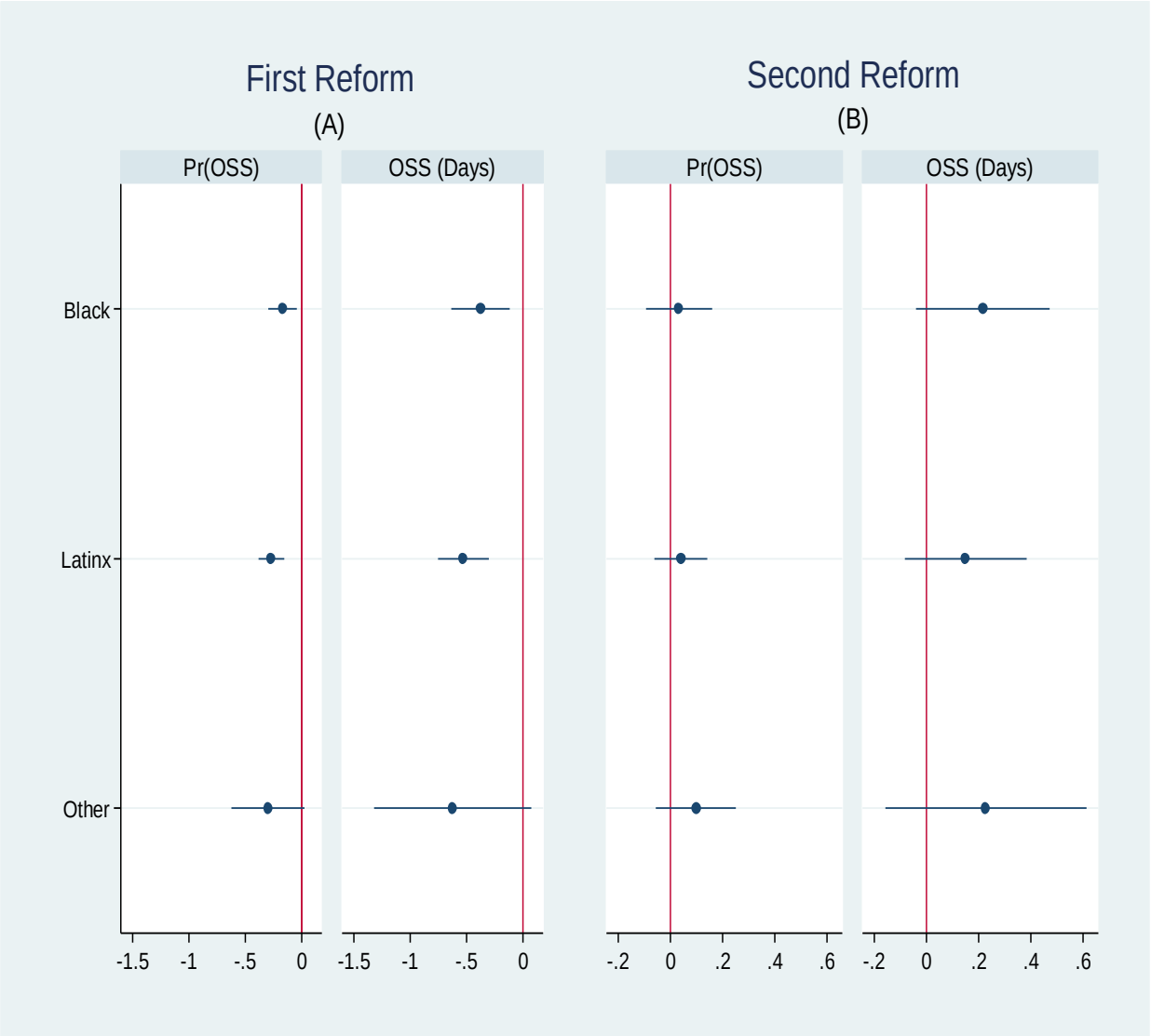
DDD Impact Estimates of the Second Suspension Reform

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)
*Tr*Sch*Post*^{2}	0.010	0.077*	0.011	0.078	0.011	0.077
	(0.021)	(0.046)	(0.022)	(0.049)	(0.022)	(0.049)
Fixed Effects	Y	Y	Y	Y	Y	Y
Individual Controls	N	N	Y	Y	Y	Y
Linear Trends	N	N	N	N	Y	Y
R-Squared	.263	.103	.298	.129	.298	.129
N	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375

Note. The table presents general DDD impact estimates from equation (2) showing the impact of the second suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))). Each DDD regression controls for grade, school, and semester-year fixed effects. Individual controls include race, gender, free or reduced-lunch status, IEP status, and cumulative number of infractions per semester. Linear trends include school-specific and general linear time trends. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

FIGURE 2

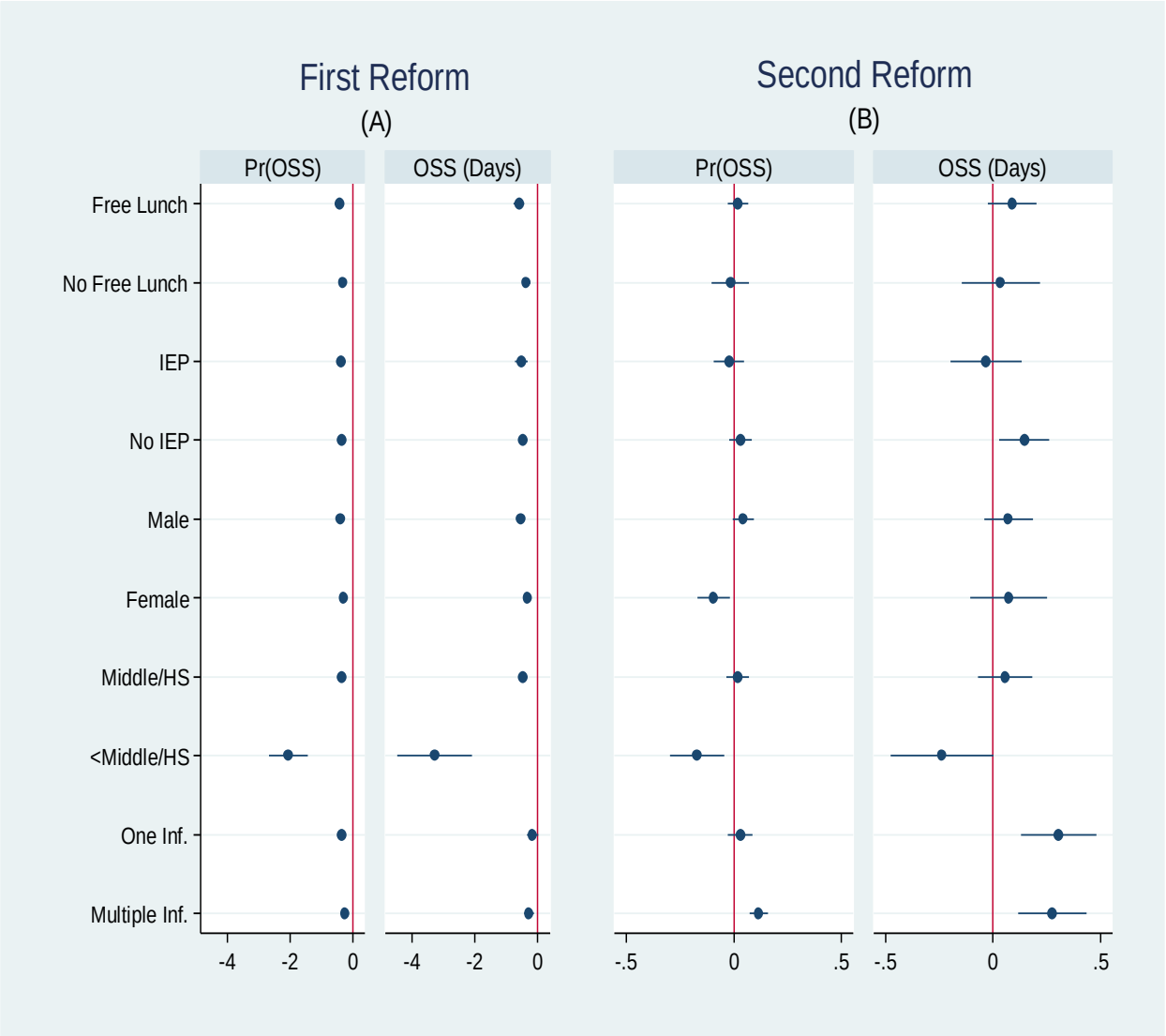
DDDD Impact Estimates (Racial-Ethnic Disparities)



Note. The figure presents general DDDD impact estimates of the impact of each suspension reform on the racial-ethnic disparities in OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))). The impact estimates are shown along with corresponding 95% confidence intervals. Each DDDD regression controls for gender, free or reduced-lunch status, IEP status, cumulative number of infractions per semester, school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects.

FIGURE 3

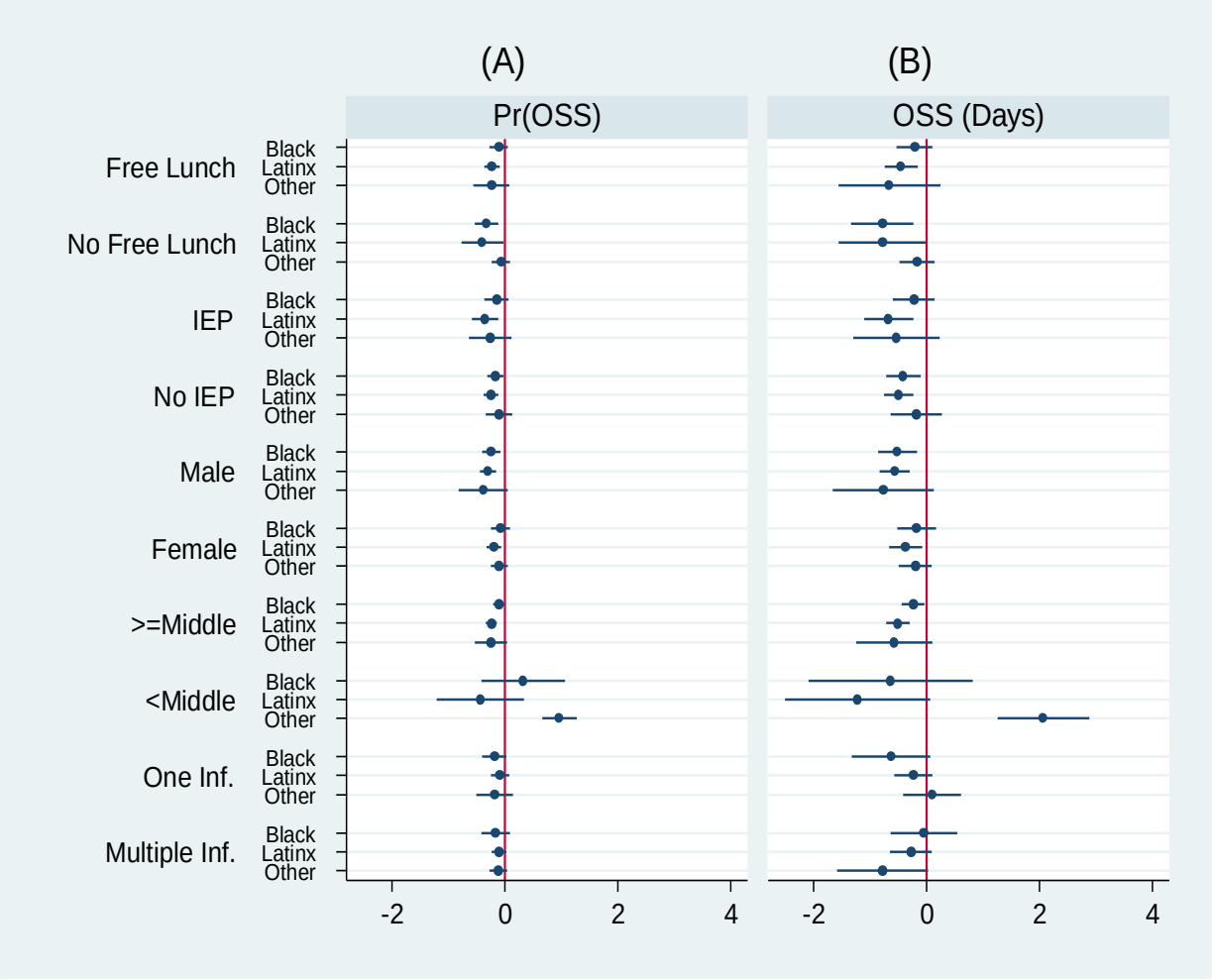
DDD Impact Estimates (*Heterogeneous Effects)*



Note. The figure presents DDD impact estimates from equation (2) showing the impact of each suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))) for heterogeneous samples restricted by free or reduced-lunch status, IEP status, gender, grade-level, and cumulative number of infractions per semester. The impact estimates are shown along with corresponding 95% confidence intervals. Due to small standard errors, some confidence intervals may be difficult to observe. Each DDD regression controls for school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects.

FIGURE 4

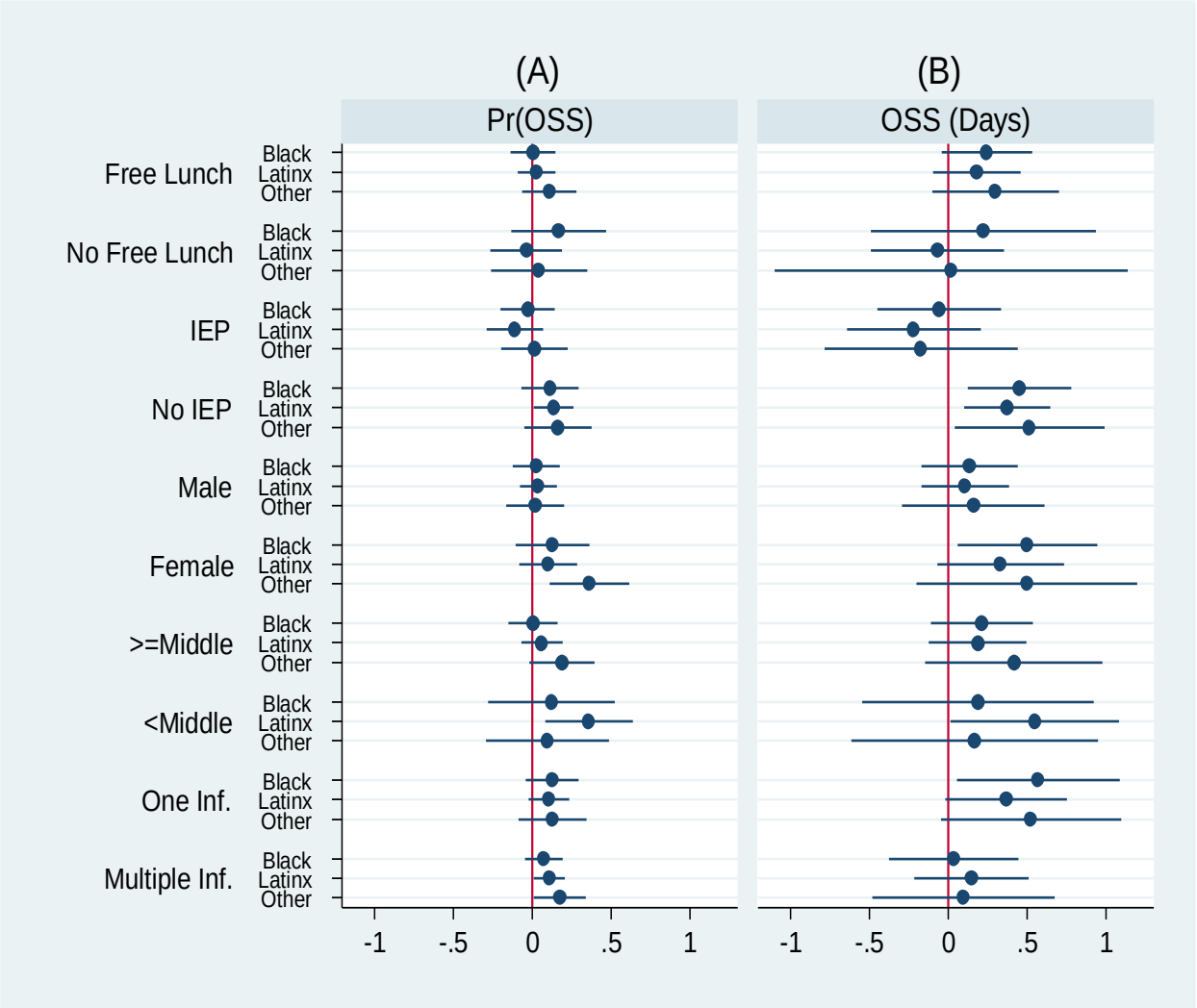
DDDD Impact Estimates of the First Suspension Reform (Heterogeneous Effects)



Note. The figure presents DDDD impact estimates showing the impact of the first suspension reform on Black-White, Latinx-White, and Other-White disparities in OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))) for heterogeneous samples restricted by free or reduced-lunch status, IEP status, gender, grade-level, and cumulative number of infractions per semester. The impact estimates are shown along with corresponding 95% confidence intervals. Each DDDD regression controls for school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects.

FIGURE 5

DDDD Impact Estimates of the Second Suspension Reform (Heterogeneous Impacts)



Note. The figure presents general DDDD impact estimates showing the impact of the second suspension reform on Black-White, Latinx-White, and Other-White disparities in OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS(Days))) for heterogeneous samples restricted by free or reduced-lunch status, IEP status, gender, grade-level, and cumulative number of infractions per semester. The impact estimates are shown along with corresponding 95% confidence intervals. Each DDDD regression controls school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects.

[ONLINE APPENDIX BEGINS HERE]

ONLINE APPENDIX TABLE A

Infraction-Specific DDD Analyses for the Second Suspension Reform

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dis. Conduct		Insub./Disrespect		Obs. Lang – Student		Obs. Lang – Teacher	
	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)
*Tr*Sch*Post*^{2}	-0.010	0.174***	-0.033	0.007	-0.020	-0.045	-0.069*	-0.302***
	(0.026)	(0.060)	(0.028)	(0.058)	(0.039)	(0.088)	(0.042)	(0.113)
R-Squared	.230	.116	.231	.109	.202	.105	.210	.109
N	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375	2,938,375

Note. The table presents general DDD impact estimates from equation (2) showing the impact of the second suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS Days))) for each type of disruption-specific infraction {disorderly conduct, insubordination/disrespect, obscene language used towards students, and obscene language used towards teachers}. Each DDD regression controls for the race, gender, free or reduced-lunch status, IEP status, cumulative number of infractions per semester, school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

ONLINE APPENDIX TABLE B

Accounting for Infraction-Specific Trends and Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Pr(OSS)	Pr(OSS)	Pr(OSS)	OSS(Days)	OSS(Days)	OSS(Days)
Panel A.						
*Tr*Sch*Post*[1]	-0.379***	-0.383***	-0.442***	-0.501***	-0.510***	-0.627***
	(0.022)	(0.018)	(0.016)	(0.046)	(0.037)	(0.025)
Panel B.						
*Tr*Sch*Post*[2]	0.019	0.029	0.116***	0.095**	0.118**	0.249***
	(0.021)	(0.021)	(0.020)	(0.048)	(0.047)	(0.046)
Infraction Linear Trend	Y	Y	Y	Y	Y	Y
Infraction Quadratic Trend	N	Y	Y	N	Y	Y
Infraction x Year FE	N	N	Y	N	N	Y

Note. The table presents general DDD impact estimates from equation (2) showing the impact of each suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS Days))) while accounting for infraction-specific trends and/or infraction-by-year fixed effects. Each DDD regression controls for the race, gender, free or reduced-lunch status, IEP status, cumulative number of infractions per semester, school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

ONLINE APPENDIX TABLE C

Additional Sensitivity Checks

	(1)	(2)	(3)	(4)
	Pr(OSS)	OSS(Days)	Pr(OSS)	OSS(Days)
Panel A.				
*Tr*Sch*Post*^{1}	-0.382***	-0.508***		
	(0.010)	(0.020)		
*Tr*Sch*Post*^{2}			0.012	0.080*
			(0.021)	(0.046)
Panel B.				
*Tr*Sch*Post*^{1}	-0.359***	-0.464***		
	(0.021)	(0.045)		
*Tr*Sch*Post*^{2}			-0.001	0.060
			(0.022)	(0.049)
Panel C.				
*Tr*Sch*Post*^{1}	-0.372***	-0.494***		
	(0.0289)	(0.0595)		
*Tr*Sch*Post*^{2}			0.007	0.018
			(0.030)	(0.067)
Panel D.				
*Tr*Sch*Post*^{1}	-0.386***	-0.508***		
	(0.014)	(0.038)		
*Tr*Sch*Post*^{2}	0.053**	0.111**		
	(0.021)	(0.049)		

Note. The table presents general DDD impact estimates from equation (2) showing the impact of each suspension reform on OSS outcomes (i.e., the probability of OSS (Pr(OSS)) and OSS duration (OSS Days))). Each DDD regression controls for the race, gender, free or reduced-lunch status, IEP status, school-specific and general linear time trends, as well as grade, school, and semester-year fixed effects. Panel A shows DDD impact estimates without cumulative infractions per semester as a control variable. Panel B shows DDD impact estimates with cumulative infractions per semester lagged as a control variable. Panel C shows DDD impact estimates when the analysis sample is restricted to semester-years before the second reform took effect (*Tr*Sch*Post*^{1}) and after the first reform took effect (*Tr*Sch*Post*^{2}), respectively. Panel D shows DDD impact estimates for both reforms using a single regression. Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

ONLINE APPENDIX TABLE D

Comparative Interrupted Time Series (CITS) Model

	(1)
	Number of Dis. Infractions
T_s*$REFORM_t$	11.7197***
	(2.6821)
T_s*$YEAR\_SINCE\_REFORM_t$	-2.1552***
	(0.2877)
Total "Effect" by Fall 2015	-5.5219***
	(2.0055)
Total "Effect" by Spring 2018	-14.1427***
	(2.5928)
Outcome Mean	21.49
R-squared	0.65
Number of Schools	388
Total Number of Observations	6072

Note. The table presents key estimates from a school-level comparative interrupted time series (CITS) analysis, estimating the difference in the number of disruption-related infractions between treatment and comparison schools, before vs. after the first reform (expressed as a binary indicator equal to 1 for all semester-years after the first reform (*REFORM*) as well as all semester-years after the first reform (*YEAR _ SINCE _ REFORM*)). (See Dee & Jacob (2011) equation (1) and Anderson (2019) equation (2) for the CITS design used here). Standard errors are two-way clustered at the school and semester-year levels and are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

[i] See http://www.rikidscount.org/Issue-Areas/Education for further details.
[ii] "Treatment schools" in the DDD framework is defined as schools that used OSS for *grade-specific* treatment infractions prior to each suspension reform. For the first reform, treatment schools used OSS for *grade-specific* treatment infractions any time prior to Spring 2012. For the second reform, however, treatment schools could not be defined as OSS for *grade-specific* treatment infractions any time prior to Spring 2016. This is because close to 100% of schools used OSS for disruptive infractions, producing symptoms of multicollinearity in the model. To remedy this problem, the second model defines "treatment schools" as schools that used OSS for *grade-specific* treatment infractions at any time from 2013 to 2015.
[iii] See Anderson (2019) equation (2) and Dee & Jacob (2011) equation (1) for the CITS design used here.
[iv] This legislation was introduced and amended using bills: H 7056, H 7057, S 2168.