Discipline Reform and Academic Achievement

Terry-Ann Craigie*

October 26, 2025

Abstract

To curb the excessive and disproportionate use of out-of-school suspensions, the Rhode Island General Assembly passed two key legislative reforms. The first reform, effective May 2012, banned the use of out-of-school suspensions for attendance-related infractions. Four years later, in June 2016, the Assembly passed additional legislation to lower the use of out-of-school suspensions for disruption-related infractions. The study investigates the impact of these reforms on individual and overall academic performance of Rhode Island's elementary and middle-school students, as well as corresponding racial-ethnic and disability achievement gaps. Using quasi-experimental estimation, the paper finds that although the first reform does not change individual and overall academic outcomes, it substantially reduces minority-white achievement gaps. The second reform, on the other hand, improves academic outcomes for treated students, but sacrifices the academic performance of the overall student population. The implications of these findings are discussed in detail.

JEL codes: I24, I28, J15, J16, J24, K42.

Keywords: School Suspensions, Academic Achievement, Proficiency Rates, Race, Ethnicity, Disability, Rhode Island.

---

*Smith College. Contact Email: tcraigie@smith.edu. Mailing Address: 10 Elm Street, Northampton, MA 01063, USA. The author would like to thank Data Spark at the University of Rhode Island for their excellent data assistance; 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. The usual caveat holds.

1 Introduction

Following the Gun-Free Schools Act of 1994, schools across the United States increasingly used zero-tolerance policies to thwart school violence, gang activity, and drug abuse (McNeal and Dunbar Jr, 2010). This approach triggered a sharp rise in the use of exclusionary discipline, even for minor infractions. Numerous studies show that exclusionary discipline – specifically out-of-school suspensions ("OSS") – is inversely linked to student academic achievement (e.g., Perry and Morris (2014), Skiba et al. (2006)) without any distinct improvements to school safety (e.g., Jones (2018), Skiba and Knesting (2002), Skiba et al. (2006)).

Exclusionary discipline not only lowers academic achievement (e.g., Perry and Morris (2014), Skiba et al. (2006)), but also reinforces the school-to-prison pipeline by increasing the likelihood of juvenile arrests and adult incarceration for minority males (e.g., Bacher-Hicks et al. (2024), Cuellar and Markowitz (2015)). Moreover, there is disparate impact in exclusionary discipline — Black students are significantly more likely to be expelled or suspended than White students for similar offenses (*e.g.,* Anderson and Ritter (2017), Gregory et al. (2010), Lhamon and Samuels (2014), Morris and Perry (2016), Skiba et al. (2014), Kinsler (2013)).

Several states have sought to curtail the inordinate use of exclusionary discipline by implementing suspension policies or reforms to various degrees (Anderson, 2018; Bacher-Hicks et al., 2024; Baker-Smith, 2018; Cleveland, 2022; Craig and Martin, 2023; Hashim et al., 2018; Lacoe and Steinberg, 2018; Pope and Zuo, 2023; Stevens et al., 2015; Zarecki, 2019). One such state is Rhode Island, which bans OSS for attendance-specific and disruption-specific infractions. Effective May 2012, Rhode Island barred OSS for attendance-specific infractions (such as truancy, absenteeism, and tardiness). Effective June 2016, the Rhode Island General Assembly passed additional legislation to restrict the use of OSS for disruption-specific infractions (such as insubordination and disorderly conduct).

In contrast to the first reform, the second reform was enacted with conspicuous loopholes.

Under this legislation, the use of OSS may be used for disruption-specific infractions if the student continually threatens or harms another, repeatedly obstructs learning, or alternative disciplinary measures have been ineffective. The second reform also states that discipline data must be reviewed by school superintendents to assess and address any disproportionalities by race, ethnicity, or disability resulting from suspensions.

With the implementation of dual reforms, Rhode Island is an intriguing case study for empirical evaluation. Although the first reform identifies a distinct treatment group (students penalized with attendance-specific infractions), the second is an ambiguous intervention – we cannot observe which disruption-specific infractions are potentially threatening, harmful, or repeatedly obstructive, and would thereby justify the use of OSS. In this way, Rhode Island's dual reforms may ultimately be "dueling" reforms—working against rather than with each other— in terms of their overall impact on academic achievement.

This study evaluates the causal impact of Rhode Island's dual suspension reforms on the academic achievement of students enrolled in public schools from academic year (AY) 2009–10 to AY 2017–18. Over the study period, OSS for attendance-specific infractions declined by 100%, while OSS for disruption-specific infractions declined only marginally. However, the extent to which the dual reforms changed academic achievement, both individually and collectively, has yet to be addressed in the literature. The study exploits administrative student-level data from the Rhode Island Department of Education (RIDE) (AY 2009–10 to AY 2017–18), along with quasi-experimental designs to address three key research problems.

First, the study uses triple-difference (DDD) estimation to evaluate the impact of suspension reforms on the English Language Arts (ELA) and math proficiency rates of individual students charged with treatment infractions but penalized with alternatives to OSS (i.e., in-school suspensions and alternate program placements (hereafter, collectively referred to as "ISS")). Second, the study uses a quadruple-difference (DDDD) framework to measure the net impact of the reforms on both racial-ethnic and disability achievement gaps. Third, the study adopts an instrumented difference-in-difference (DDIV) approach with continuous treatment to evaluate whether there are associated spillover effects of these suspension reforms on ELA and math proficiency rates of students collectively. Therefore, the study uniquely identifies not only the direct effects of the dual reforms on individual student academic proficiency rates, but also indirect spillover effects on peer outcomes.

The causal impact of suspension reform on peer outcomes is challenging to identify because school suspension rates are endogenously linked to latent characteristics (Pope and Zuo, 2023). The DDIV approach allows us to measure the causal impact of the dual suspension reforms while mitigating the endogeneity intrinsic to school suspension rates. Analogous to Pope and Zuo (2023), the study constructs a classic shift-share instrument for school suspension rates using year-to-year shocks in district-specific suspension rates multiplied by school suspension rates from 2009, the initial sample period. This instrument relies on key assumptions to produce consistent estimates. Year-to-year shocks in district-level suspension rates are assumed to be exogenous to latent factors that drive academic achievement. Further details on the construction and validity of the instrument are outlined in Section 4 with falsification tests detailed in Section 5.

The study finds that Rhode Island's dual suspension reforms are in essence, dueling reforms in terms of their impact on ELA and math proficiency outcomes. While the first reform shows no direct effects on individual-student proficiency measures, the second reform increases individual-student math proficiency rates by 3 percentage points (p <0.05) or 7% of the outcome mean. Although the first reform reduces Black-White, Hispanic-White, and Other-White math proficiency gaps for the treated population by at least 20 percentage points, the second reform has no statistically significant effect on race-ethnic or disability achievement gaps. Notwithstanding, the analysis of the spillover effects on academic outcomes reveals the prominent "dueling" nature of the reforms. The study uncovers no spillover effects of the first reform on the likelihood of ELA or math proficiency; however, the second reform engenders a substantial negative spillover effect on math proficiency rates of the student body as a whole. The spillover effect is not only negative, but also outweighs the small gains it induces for the math proficiency rates of individual students in the treatment group.

In light of these findings, policymakers, administrators, and practitioners might consider collaborating on the revision of the legislative language of Rhode Island's second suspension reform. Although this reform aims to curb OSS for disruption-specific infractions, the criteria under which the policy is implemented (i.e., non-threatening, non-harmful, and non-obstructive behaviors) are ambiguous, thereby creating loopholes that allow OSS to persist. If the objective of the second reform is to reduce OSS, then stipulating alternative disciplinary actions in the legislation would likely reduce both OSS rates and the negative spillover effect on math proficiency outcomes.

The existing literature evaluating suspension reforms offers mixed results (*e.g.,* Anderson (2018), Baker-Smith (2018), Cleveland (2022), Craig and Martin (2023), Craigie (2022), Hashim et al. (2018), Lacoe and Steinberg (2018), Liberman and Katz (2017), Stevens et al. (2015), Zarecki (2019)), with relatively few identifying causal impacts on academic outcomes. Recently, Pope and Zuo (2023) used data from the Los Angeles Unified School District (LAUSD) to demonstrate that policies implemented to reduce the use of suspensions improved the academic outcomes of suspended students but led to adverse academic outcomes among peers in general. Craig and Martin (2023) evaluated the 2012 suspension reform implemented in New York City public middle schools to eliminate suspensions for non-violent, disorderly infractions. They found that the reform improved math and reading scores, particularly for students in schools with high suspension rates. Lacoe and Steinberg (2018) showed that Philadelphia's 2012 suspension reform lowered suspensions for nonviolent infractions but also raised suspensions for more serious infractions; as a result, math and ELA achievement outcomes fell statistically significantly. Cleveland (2022) also finds that Massachusetts' statewide suspension reform improves ELA test scores for some groups.

My study brings a novel dual case study to this burgeoning literature. The first contribution of this study is to identify the direct impact of both suspension reforms for individual students. Second, I introduce a DDDD approach to measure how racial-ethnic and disability disparities in academic achievement shift in response to both reforms. The Rhode Island context represents an important example of how academic achievement gaps persist. As of AY 2021–22, average ELA and math proficiency rates for third-grade White students were more than twice as high as for their Black and Hispanic counterparts (Matos, 2023). This study is the first to quantify the extent to which these reforms have alleviated these disparities. Finally, the study adopts a novel DDIV (with continuous treatment) design to identify the causal spillover effects of each reform while mitigating biases from latent school-, student-, and principal-specific characteristics. Therefore, the study presents a comprehensive analysis that aims to inform educators, administrators, and other stakeholders on the efficacy of the dual suspension reforms passed in Rhode Island.

2 School Suspension Reforms in Rhode Island

Before suspension reforms were enacted in Rhode Island, there were more than 40,000 student infractions (Bryant et al., 2011). Of the disciplinary actions administered, 60% of them were OSS — a third were for attendance-specific infractions and another third for disruption-specific infractions.[1] These high OSS rates disproportionately affect racial-ethnic minorities and provide corroborative evidence of disparate impact in Rhode Island. According to the 2011 Rhode Island Kids Count Factbook, racial-ethnic minorities accounted for approximately 50% of all disciplinary actions despite comprising less than a third of the student population.

In light of this evidence, the Rhode Island General Assembly passed two laws with the aim of curbing OSS for low-level offenses. On May 30, 2012, the first law (§ 16-19-1) banned OSS for attendance-related infractions such as truancy, absenteeism, and tardiness. Alternative disciplinary actions (i.e., ISS) were not affected by this legislation.

[1] Attendance-specific infractions are classified as offenses related to leaving school grounds without permission, tardiness, truancy, and skipping class or detention or in-school suspension. Disruption-specific infractions are classified as offenses related to insubordination/disrespect, disorderly conduct, and obscene/abusive language targeting students and teachers.

"The 2016 Right to a Safe School Act" (§ 16-2-17) was the second piece of legislation enacted. This reform banned OSS for disruption-specific infractions unless the student repeatedly threatened or harmed another student, repeatedly obstructed learning, or other disciplinary measures were ineffective. Akin to the first reform, this piece of legislation does not affect other types of disciplinary actions (i.e., ISS). In contrast to the first reform, however, this piece of legislation states that disciplinary data must be reviewed by school superintendents to ascertain racial-, ethnic-, and disability-specific disproportionalities and subsequently address them. Put succinctly, the second reform seeks to restrict OSS for disruption-specific offenses while preventing disparate impact.

The second reform — unlike the first — is therefore an ambiguous policy. The first reform clearly outlines the treatment infractions and the corresponding penalties permitted. However, the language of the second legislation might confound implementation efforts. The second reform bans OSS for disruption-specific infractions, but it also sanctions them for the same infractions if students are perceived as persistently threatening, harmful, obstructive, or intransigent. Moreover, the classification of these behaviors is subjective and may be influenced by race, ethnicity, and disability — the very disparities the reform was enacted to mitigate. Studies find that Black students are perceived as more aggressive than their White counterparts as evidenced by their increased risk of office referrals and exclusionary discipline (*e.g.,* Francis (2012), Liu et al. (2022), Lhamon and Samuels (2014), Ritter and Anderson (2018), Sagar and Schofield (1980), Skiba and Knesting (2002)).

Craigie (2022) evaluated the effectiveness of these dual reforms in transforming how OSS penalties are administered in Rhode Island. The author used quasi-experimental estimation to show that the first reform lowered OSS and corresponding racial-ethnic disparities. The second reform had no statistically significant effect on OSS but exacerbated the Black-White disparity in average OSS durations.

The causal literature predominantly shows that suspension reforms improve the academic outcomes of individual students (*e.g.,* Craig and Martin (2023), Pope and Zuo (2023), Cleveland (2022)); however, we have limited evidence on whether suspension reforms improve or worsen academic outcomes of the student body as a whole (*e.g.,* Pope and Zuo (2023)). To expand this literature, this study focuses on the direct-individual and indirect-collective impacts of the dual reforms on the academic achievement of elementary and middle-school students in Rhode Island.

3 Data

The study uses student-specific administrative data from the Rhode Island Department of Education (RIDE) from AY 2009–10 to AY 2017–18 for K–12 public schools. Files with enrollment, infraction, and state-test data are provided separately and merged using unique student identifiers. Enrollment and state-test data are obtained from student-by-year files that provide information on demographic characteristics (age, gender, and race-ethnicity), free/reduced-price lunch status, disability status (i.e., enrolled in an individualized educational programs (IEP)), current grade, school, and school-district, as well as ELA and math state-test outcomes. Infraction data are obtained from student-by-semester-by-year files that include information on each student's infraction type, type of disposition or penalty, as well as the cumulative number of infractions and suspensions per academic year.

To seamlessly merge all three files, I convert the complete student-by-semester-by-year infractions file into a student-by-year file, by constructing binary indicators for whether a student had a treatment infraction — categorized as either attendance-specific (i.e., truancy, absenteeism, or tardiness) or disruption-specific (i.e., insubordination/disrespect, disorderly conduct) — in each academic year. For students with a treatment infraction, I also construct binary indicators for whether the student was issued an ISS as the disciplinary action corresponding to the infraction.[2] The study retained the academic-year measures for the cu-

[2]There is full compliance with the first reform, *ex post*, and thus a binary indicator for OSS would be perfectly collinear in the model.

mulative number of infractions and suspensions for each student. This converted infraction student-by-year file was then merged with the enrollment and state-test files using unique student identifiers. The merged sample is subsequently restricted to elementary and middle-school students (i.e., students in grades 3 through 8) for which we can analyze ELA and math state-test scores.

Within the analysis sample, 60% of students are White, 8% are Black, 25% are Hispanic, and 7% are from other racial-ethnic groups. More than half of the students are male, and more than half are eligible for free/reduced-price lunch. Additionally, 15% of the students are classified as having a disability. About 15% of students had attendance-specific infractions while approximately 50% had disruption-specific infractions.

Over the study period, Rhode Island instituted three different state tests: the New England Common Assessment Program (NECAP) from AY 2009–10 to 2013–14, the Partnership for Assessment of Readiness for College and Careers (PARCC) from AY 2014–15 to AY 2016–17, and the Rhode Island Comprehensive Assessment System (RICAS) in AY 2017–18. These are "criterion-referenced" assessments with varying scoring mechanisms, making normalized scores non-comparable across tests. To harmonize all three assessments for this analysis, for each assessment, the study identifies the proficiency performance levels required to transition to the next grade in ELA and mathematics.

Table 1: Proficiency Levels for Academic Tests in Rhode Island

NECAP	PARCC	RICAS
–	Did not Meet Expectations	–
Substantially below Proficient	Partially Met Expectations	Not Meeting Expectations
Partially Proficient	Approached Expectations	Partially Meeting Expectations
Proficient	Met Expectations	Meeting Expectations
Proficient with Distinction	Exceeded Expectations	Exceeding Expectations

Notes: This table illustrates the proficiency levels of each academic test administered in Rhode Island from AY 2009–10 to AY 2017–18 for students grades 3 through 8. Proficiency levels outlined in red indicate ELA and math proficiency scores sufficient to transition to the next grade.

As shown in Table 1, the NECAP assessment has four performance levels: (1) substantially below proficient, (2) partially proficient, (3) proficient, and (4) proficient with distinction. Students who are proficient or proficient with distinction are deemed prepared to move on to the next grade. The PARCC assessment has five performance levels: Level 1 – did not meet expectations, Level 2 – partially met expectations, Level 3 – approached expectations, Level 4 – met expectations, and Level 5 – exceeded expectations. Students achieving Level 4 or 5 scores are viewed as proficient and able to transition to the next grade.[3] Finally, the RICAS assessment defines four simple performance levels: (1) not meeting expectations, (2) partially meeting expectations, (3) meeting expectations, and (4) exceeding expectations. Students meeting or exceeding expectations are able to transition to the next grade. On average, 57% of students are proficient in ELA, while 48% are proficient in mathematics.

Figure 1 illustrates the minority and disability achievement disparities in ELA and math proficiency rates within the analysis sample. White students have the highest average ELA proficiency rate at a little over 65%. Meanwhile, Black, Hispanic, and Other-ethnic students experience much lower average proficiency rates at 43%, 40%, and 55%, respectively. Similarly, approximately 60% of White students are proficient in math on average compared to about 30% of Black, 30% of Hispanic, and nearly 50% of Other-ethnic students, respectively.

[3] A PARCC score of 3 is subject to additional review by administrators to determine whether a student is able to transition to the next grade. To resolve this ambiguity, students with a score of 3 are not classified as proficient in this study.

Larger achievement gaps can be observed by student disability status defined as enrollment in an individualized educational program (IEP). Non-IEP students have average proficiency rates that loom over their IEP counterparts. Nearly 65% of non-IEP students are proficient in ELA on average compared to only about 20% of IEP students. Similarly, close to 55% of non-IEP students are proficient in math on average relative to about 15% of IEP students. Given that minorities and students with disabilities are disproportionately penalized with OSS (*e.g.,* Anderson and Ritter (2017), Gregory et al. (2010), Lhamon and Samuels (2014), Morris and Perry (2016), Morgan et al. (2019), Owens and McLanahan (2020), Skiba et al. (2014), Kinsler (2013)), these yawning disparities in proficiency outcomes reinforce the need to explore whether penalizing students with OSS —in lieu of more lenient options—is potentially the reason.

Figure 2 illustrates how OSS and ISS changed from AY 2009–10 to AY 2017–18 for attendance-specific and disruption-specific infractions, respectively. For attendance-specific infractions, OSS remained between 40% and 60% of all cases from 2009 to 2011. However, after the first reform became effective in AY 2012-13, the percentage of OSS assigned to attendance-specific infractions fell to zero, while the proportion of ISS rose to 100%, indicating full compliance of all Rhode Island public schools with the first reform. For disruption-specific infractions, the percentage of OSS penalties remained high until AY 2016-17. However, the implementation of the second reform produced volatile changes in OSS and ISS for treatment infractions over the study period.[4] The language used in the second legislation may account for why a complete switch from OSS to ISS observed after the first reform was not observed after the second. Notwithstanding, the figures illustrate that both reforms reduce the issuance of OSS while increasing the issuance of ISS over the study period. The study uses quasi-experimental methods to investigate whether the switch from presumably more severe OSS to more lenient ISS changed student proficiency rates in ELA and mathematics

[4]Beginning in AY 2013-14, there was a steep decline in the proportion of OSS and steep rise in the proportion of ISS that remains unexplained. To test the sensitivity of the results to this change, I restrict the analysis sample for the second reform analyses to 2013 and later. The findings do not change statistically significantly.

directly (at the individual-student level) and indirectly (for the student-body collectively).

4 Empirical Strategy

The study seeks to accomplish three main empirical objectives: (1) identify the causal impacts of Rhode Island's dual reforms on individual-student academic achievement (i.e., direct effects); (2) examine the extent to which the dual reforms change student achievement disparities based on racial-ethnicity and disability status (i.e., achievement gaps); and (3) measure achievement externalities induced by these reforms for the student body as a whole (i.e., indirect effects). The study adopts a different quasi-experimental method to achieve each objective.

Direct Effects

For the first objective, the study uses the following triple-difference (DDD) equation:

$$\begin{aligned} Y_{igst} =& \alpha + \delta_j \, Tr \cdot ISS \cdot Post^j_{it} + \beta_{j1} \, Tr^j_i + \beta_2 \, ISS_i + \beta_{j2} \, Post^j_t + \beta_{j3} \, Tr \cdot ISS^j_i + \\ & \beta_{j5} \, Tr \cdot Post^j_{it} + \beta_{j6} \, ISS \cdot Post^j_{it} + \beta_7 \, X_{igt} + \Gamma_{sg} + \Pi_{ot} + \tau + \gamma_t + \lambda_g + \phi_s + \epsilon_{igst} \end{aligned} \quad (1)$$

where Y denotes binary indicators for ELA and mathematics proficiency, i denotes individual student, g denotes grade-level, s denotes school, t denotes academic-year. $j \in \{1,2\}$ represents each suspension reform estimated as separate regressions, where 1 represents the first reform and 2 represents the second reform.

For the first reform, Tr^1 is a binary indicator equal to 1 if a student has an attendance-related infraction (treatment) and zero otherwise (comparison). ISS is a binary indicator equal to 1 if the school issued an ISS to a student for a treatment infraction. $Post^1$ is equal to 1 if the academic year is AY 2012-13 or later. $Tr \cdot ISS \cdot Post^1$, the variable of interest, is a binary indicator equal to 1 if the student was penalized with ISS for an attendance-specific infraction, *ex post*. Therefore, δ_1, the DDD estimator, measures the impact of the first suspension reform on the proficiency outcomes of students penalized with ISS for an attendance-specific infraction, *ex post*.

Analogously, for the second reform, Tr^2 is a binary indicator equal to 1 if a student has a disruption-related infraction (treatment) and zero for any other non-attendance-specific infraction (comparison).[5] $Post^2$ is equal to 1 if the academic-year is AY 2016-17 or later, and 0 otherwise. $Tr \cdot ISS \cdot Post^2$, the variable of interest, is equal to 1 if the student was penalized with ISS for a disruption-specific infraction, *ex post*. δ_2, the DDD estimator, measures the impact of the second suspension reform on the proficiency outcomes of students penalized with ISS for a disruption-specific infraction, *ex post*.[6]

Both models account for student-level characteristics such as age, race-ethnicity, gender, free or reduced-price lunch status, IEP status, type of infraction, and cumulative number of infractions and suspensions (X), school-grade fixed effects (Γ_{sg}) to capture within-school grade differences across time, as well as infraction-by-year fixed effects (Π_{ot}) to capture year-specific changes in infractions; grade-specific, school-specific, and general linear time trends to capture within-grade, school, and overall changes over time (τ); and grade (λ_g), school (ϕ_s), and academic-year (γ_t) fixed effects. To address heteroskedasticity and serial correlation concerns, standard errors are clustered at the school level.

Effects on Achievement Gaps

Next, to determine the impact of the dual reforms on achievement gaps in Rhode Island, the study adopts the following quadruple-difference (DDDD) equation:

[5] Attendance-specific infractions are excluded because they are already "treated" would conflate the findings.

[6] To ensure that the estimate of δ_2 is not confounded with the effects of the first reform, the analysis sample excludes students who had both attendance- and disruption-related infractions in an academic year.

$$
\begin{aligned}
Y_{igst} = &\acute{\alpha} + \acute{\delta}_j\, Tr \cdot ISS \cdot Gr \cdot Post^j_{it} + \acute{\beta}_{j1}\, Tr^j_i + \acute{\beta}_2\, ISS_i + \acute{\beta}_3\, Gr_i + \acute{\beta}_{j4}\, Post^j_t \\
&+ \acute{\beta}_{j5}\, Tr \cdot Gr \cdot Post^j_{it} + \acute{\beta}_{j6}\, Tr \cdot ISS \cdot Post^j_{it} + \acute{\beta}_{j7}\, Tr \cdot Gr \cdot ISS^j_i \\
&+ \acute{\beta}_{j8}\, ISS \cdot Gr \cdot Post^j_{it} + \acute{\beta}_{j9}\, ISS \cdot Post^j_{it} + \acute{\beta}_{10}\, ISS \cdot Gr_i \\
&+ \acute{\beta}_{j11}\, Post \cdot Gr^j_t + \acute{\beta}_{j12}\, Tr \cdot ISS^j_i + \acute{\beta}_{j13}\, Tr \cdot Gr^j_i + \acute{\beta}_{j14}\, Tr \cdot Post^j_{it} \\
&+ \acute{\beta}_{15}\, X_{igt} + \acute{\Gamma}_{sg} + \acute{\Pi}_{ot} + \acute{\tau} + \acute{\gamma}_t + \acute{\lambda}_g + \acute{\phi}_s + \acute{\epsilon}_{igst}
\end{aligned} \tag{2}
$$

where Gr is a binary indicator equal to 1 if the student is a racial-ethnic minority or has a disability (i.e., enrolled in an individualized education program (IEP)), and 0 if the student is White or does not have a disability (i.e., not enrolled in an IEP), respectively. $Tr \cdot ISS \cdot Gr \cdot Post$, the variable of interest, is a binary indicator equal to 1 if the minority or IEP student was penalized with ISS for a treatment infraction, *ex post*. Therefore, $\acute{\delta}$, the DDDD estimator, measures the net impact of each suspension reform on the proficiency rates of minority or IEP students penalized with ISS for treatment infractions relative to their majority counterparts (i.e., White or non-IEP, respectively), *ex post*. As in the previous model, standard errors are clustered at the school level.

Indirect Effects

To measure the indirect or spillover effects of each reform, the study uses a modified version of the Pope and Zuo (2023) approach that measures the impact of suspension rates on academic achievement. I convert their ordinary least squares (OLS) model to a continuous difference-in-difference (DD) model to measure the effect of each suspension reform. This modified equation can be written as:

$$
\begin{aligned}
Y_{igst} = &\rho + \omega_j\, ISSRate \cdot Post^j_{gst} + \theta_{j1}\, ISSRate^j_{gst} + \theta_{j2}\, Post^j_t \\
&+ \theta_3\, \chi_{igt-1} + \theta_4\, S_{sgt-1} + \Lambda_{sg} + \Upsilon_{ot} + \eta_t + \mu_g + \psi_s + \zeta_{igst}
\end{aligned} \tag{3}
$$

where $ISSRate$ represents the fraction of students suspended in-school for a treatment infraction, j, calculated by grade, school, and year. $ISSRate \cdot Post$, the variable of interest, quantifies the fraction of students suspended via ISS for a treatment infraction, *ex post*. ω_j, the DD estimator, measures the impact of infraction-specific ISS rates on the proficiency outcomes of all students, *ex post*. The model accounts for χ_{isgt-1}, a vector comprising lagged student-level indicators for ELA and math proficiency, cumulative number of infractions in an academic year, and cumulative number of suspensions in an academic year. It also accounts for S_{sgt-1}, a vector comprising lagged average ELA and math proficiency outcomes within grade, g, and school s. Similar to the earlier equations, standard errors are clustered at the school level.

There is warranted concern that ω may be biased given that ISS rates are systematically different for each school and may depend on latent characteristics that concurrently influence academic achievement (Pope and Zuo, 2023). To mitigate this concern, the study adopts an instrumented difference-in-difference (DDIV) model using the following instrumental variable:

$$\widetilde{ISSRate}^{j}_{gst} = ISSRate^{j}_{sg2009} * G^{j,-s}_{gt} \tag{4}$$

where $ISSRate^{j}_{sg2009}$ is the infraction-specific ISS rate by school and grade in 2009 (the first academic year of the analysis sample). $G^{j,-s}_{gt}$ is defined as $ISSRate^{j,-s}_{gt}/ISSRate^{j,-s}_{gt-1}$, capturing the year-to-year change in overall ISS rates per grade between the current year, t, and the previous year, $t-1$, but excluding school, s. Excluding school, s, from this measure of the growth rate, attenuates finite sample bias that may emerge from using idiosyncratic school information in the first stage analyses (Pope and Zuo, 2023; Goldsmith-Pinkham et al., 2020). In this way, the instrument is similar in principle to Bartik's shift-share instrument described in the existing literature (*e.g.,* Aizer and Doyle Jr (2015); Currie and Gruber

(1996); Gruber and Saez (2002)).

The first-stage equations using this instrument can be written as:

$$\begin{aligned} \widehat{ISSRate}^{j}_{gst} =& \dot{\rho} + \dot{\theta_{j1}}\, \widetilde{ISSRate}^{j}_{gst} + \dot{\theta_{j2}}\, Post^{j}_{t} + \dot{\theta}_{3}\, \chi_{igt-1} + \dot{\theta}_{4}\, S_{sgt-1} \\ & + \dot{\Lambda_{sg}} + \dot{\Upsilon_{ot}} + \dot{\eta_{t}} + \dot{mu_{g}} + \dot{\psi}_{s} + \dot{\zeta_{igst}} \end{aligned} \tag{5}$$

$$\begin{aligned} \widehat{ISSRate \cdot Post}^{j}_{gst} =& \check{\rho} + \check{\theta_{j1}}\, \widetilde{ISSRate \cdot Post}^{j}_{gst} + \check{\theta_{j2}}\, Post^{j}_{t} + \check{\theta}_{3}\, \chi_{igt-1} \\ & + \check{\theta}_{4}\, S_{sgt-1} + \check{\Lambda_{sg}} + \check{\Upsilon_{ot}} + \check{\eta_{t}} + \check{mu_{g}} + \check{\psi}_{s} + \check{\zeta_{igst}} \end{aligned} \tag{6}$$

Substituting the first-stage equations into the second stage, we obtain:

$$\begin{aligned} Y_{igst} =& \acute{\rho} + \acute{\omega}_{j}\, \widehat{ISSRate \cdot Post}^{j}_{gst} + \acute{\theta_{j1}}\, \widehat{ISSRate}^{j}_{gst} + \acute{\theta_{j2}}\, Post^{j}_{t} \\ & + \acute{\theta_{3}}\, \chi_{igt-1} + \acute{\theta_{4}}\, S_{sgt-1} + \acute{\Lambda_{sg}} + \acute{\Upsilon_{ot}} + \acute{\eta_{t}} + \acute{mu_{g}} + \acute{\psi}_{s} + \acute{\zeta_{igst}} \end{aligned} \tag{7}$$

Similar to any instrumental variable (IV) model, this instrument must satisfy relevance and exclusion restriction conditions. The relevance condition can be empirically tested by showing the correlation of the instrument with the ISS rates defined in Equation (1). Appendix Figure A1 illustrates binned scatterplots of the strong positive relationship between the instrument and endogenous ISS rates. Moreover, the first-stage coefficients are statistically meaningful and the first-stage F statistics are quite large (see Appendix Table A1), confirming a strong relationship between the instrument and instrumented variable. Appendix Figure A2 also depicts a strong inverse relationship between the instrumental variable and proficiency measures for each reform, thereby satisfying the monotonicity assumption (see figure notes for further details).

The exclusion condition is much more difficult to corroborate as it mandates that the instrument be orthogonal to the error term. However, by using school- and grade-specific pre-period ISS rates interacted with a growth quotient that excludes own-school information, the instrumental variable is likely to satisfy the exclusion condition. Still, caution is required

– pre-period ISS rates may partially explain future ISS rates and academic outcomes, thereby violating the exclusion restriction. Therefore, the study uses falsification tests to support this measure as a valid instrument for causal DDIV estimation (see Section 5).

5 Results

Direct Effects

Tables 2 and 3 report the DDD estimates of the direct impact of the first and second suspension reforms on individual-student ELA and math proficiency rates, respectively. Without any additional control variables, columns (1) and (2) of Table 2 indicate that the first suspension reform has no impact on ELA proficiency rates for treated students but lowers their math proficiency rates by 5.5 percentage points (p <0.10) or about 10% of the outcome mean. Adding fixed effects, student characteristics, and trend variables (columns 3 – 8), eliminates the marginally significant effect on math proficiency showing that on average, the first reform has no robust effect on the proficiency outcomes of students penalized with ISS for attendance-specific infractions.

For the second reform, accounting for fixed effects, student characteristics, and trend variables, shows that the reform changes treated students' academic achievement. When all controls are included in the model (Table 3, columns 7 and 8), DDD estimates indicate that the second reform has no statistically significant impact on treated students' ELA proficiency but improves their math proficiency by 3.3 percentage points (p <0.05) or about 7% of the outcome mean.

Sensitivity Checks

For these estimates to be causal, key conditions must be satisfied. First, the parallel trend assumption states that outcome trends of treated and comparison groups should be parallel prior to the effective dates of each reform. The study presents graphs from the corresponding event-study specification to Equation (1)[7]; Figures 3a and 3b support the hypothesis that the parallel trend assumption holds: individual pre-reform estimates and the overall significance of these estimates (measured by F-statistics) are statistically equivalent to zero.

Recent research suggests that there may be threats to causality in the DDD framework even when the parallel trends assumption is satisfied. Roth and Rambachan (2022) propose the honest difference-in-differences (DiD) method, which provides treatment effect bounds for possible parallel trend assumption violations. This method produces "breakdown values" ($\overline{M}$) that indicate where estimated treatment effects overlap with zero by a factor of potential pre-trends. To perform this analysis, the study uses the full analysis sample to assess threats to validity under the dual suspension model.

For the first reform, Figure A3a reveals that $\overline{M}$ does not exceed zero for ELA proficiency, suggesting that the confidence intervals for treatment effects would not be statistically different from zero (McConnell et al., 2024). Figure A3b reveals the "breakdown value" ($\overline{M}$) exceeds 1 for math proficiency, suggesting that if post-reform violations were at least as large as pre-period violations, the confidence intervals for treatment effects would still exceed zero. For the second reform, Figure A3d reveals that $\overline{M}$ exceeds 0.5 for both outcomes suggesting that if post-reform violations were half as large as pre-period violations, the confidence intervals for treatment effects would still exceed zero.

Beyond the parallel trend assumption, other checks are important to support causal inference. First, it must be acknowledged that not all Rhode Island schools issue OSS for the treatment infractions; this suggests that our analysis sample that incorporates all schools,

[7] The corresponding event-study specification is:

$$\begin{aligned} Y_{igst} =& \alpha + \sum_{l \in k} \delta_j^l \, Tr \cdot ISS \cdot Post_{it}^{jl} + \beta_{j1} \, Tr_i^j + \beta_2 \, ISS_i + \beta_{j2} \, Post_t^j + \beta_{j3} \, Tr \cdot ISS_i^j + \\ & \beta_{j5} \, Tr \cdot Post_{it}^j + \beta_{j6} \, ISS \cdot Post_{it}^j + \beta_7 \, X_{igt} + \Gamma_{sg} + \Pi_{ot} + \tau + \gamma_t + \lambda_g + \phi_s + \epsilon_{igst} \end{aligned}$$

where $k = \{-2, -1, 0, 1, 2, 3, 4, 5, 6\}$ with -2 denoting two years before and 6 denoting six years after the first reform and $k = \{-6, -5, -4, -3, -2, -1, 0, 1, 2\}$ with -6 denoting six years before and 2 denoting two years after the second reform.

grades 3 through 8 could attenuate the estimates of interest. To address this concern, the study restricts the analysis sample to only schools that issue OSS for treatment infractions. Corresponding DDD estimates presented in Appendix Table A2 are statistically similar to those in Tables 2 and 3.

Achievement Gaps

Tables 4 to 5 report DDDD estimates of the impacts of the dual reforms on achievement gaps in ELA and math proficiency outcomes. Column (2) of Table 4 shows that in response to the first reform, Black students penalized with ISS for attendance-specific infractions have higher math proficiency rates by 20 percentage points (p <0.05) relative to their White counterparts, *ex post*. To evaluate how the Hispanic-White achievement gap responded to the first reform, corresponding DDDD estimates are presented in Columns (3) and (4) of Table 4. In response to the first reform, Hispanic students penalized with ISS for attendance-specific infractions have higher math proficiency rates by 22 percentage points (p <0.01) relative to their White counterparts, *ex post*.

Analogous to the findings on Black-White and Hispanic-White achievement gaps, column (6) of Table 4 indicates that in response to the first reform, students of Other-ethnic backgrounds penalized with ISS for attendance-specific infractions have higher math proficiency rates by 26 percentage points (p <0.05) relative to their White counterparts, *ex post*. There are no statistically significant changes in Black-White, Hispanic-White, and Other-White achievement gaps in response to the second reform (see Table 5). These findings suggest that while the first reform did not change academic outcomes on average, there are statistically meaningful benefits to racial-ethnic minorities relative to Whites, especially as it relates to student proficiency in mathematics.

DDDD estimates corresponding to disability achievement gaps are also presented in Tables 4 to 5. In contrast to the racial-ethnic achievement gaps, neither reform lowers the disability achievement gap in ELA and math proficiency rates.

Sensitivity Checks

To ensure causal inferences can be drawn from these findings, the study uses the event-study specification of Equation(2)[8] and the honest DiD approach to test for potential pre-trend violations. Figures 3 through 5 show that all the statistically significant achievement gap findings are supported by parallel pre-reform trends: individual pre-reform estimates and overall significance F-statistics are not statistically different from zero for Black-White, Hispanic-White, and Other-White achievement gap models.

Findings from the honest quadruple-difference analyses are presented in Appendix Figures A4 to A7. Appendix Figure A4 illustrates that $\overline{M}$ exceeds 2 for the Black-White achievement gap in math proficiency rates in response to the first reform, suggesting that if post-reform violations were twice as large as pre-period violations, confidence intervals for treatment effects would exceed zero. Figures A5 and A6 also reveal that $\overline{M}$ corresponding to Hispanic-White and Other-White achievement gaps in math proficiency rates are at least 0.5, suggesting that if post-reform violations were at least half as large as pre-period violations, confidence intervals for treatment effects would still exceed zero.

In summary, the study indicates that Black-White, Hispanic-White, and Other-White achievement gaps in math proficiency rates decline in response to the first reform, whereas the second reform does not appear to produce causal effects on racial-ethnic or disability achievement gaps.

[8] The corresponding event-study specification is:

$$
\begin{aligned}
Y_{igst} = \acute{\alpha} &+ \sum_{l \in k} \acute{\delta}_j^l \, Tr \cdot ISS \cdot Gr \cdot Post_{it}^{jl} + \acute{\beta}_{j1} \, Tr_i^j + \acute{\beta}_2 \, ISS_i + \acute{\beta}_3 \, Gr_i + \acute{\beta}_{j4} \, Post_t^j \\
&+ \acute{\beta}_{j5} \, Tr \cdot Gr \cdot Post_{it}^j + \acute{\beta}_{j6} \, Tr \cdot ISS \cdot Post_{it}^j + \acute{\beta}_{j7} \, Tr \cdot Gr \cdot ISS_i^j \\
&+ \acute{\beta}_{j8} \, ISS \cdot Gr \cdot Post_{it}^j + \acute{\beta}_{j9} \, ISS \cdot Post_{it}^j + \acute{\beta}_{10} \, ISS \cdot Gr_i \\
&+ \acute{\beta}_{j11} \, Post \cdot Gr_t^j + \acute{\beta}_{j12} \, Tr \cdot ISS_i^j + \acute{\beta}_{j13} \, Tr \cdot Gr_i^j + \acute{\beta}_{j14} \, Tr \cdot Post_{it}^j \\
&+ \acute{\beta}_{15} \, X_{igt} + \acute{\Gamma}_{sg} + \acute{\Pi}_{ot} + \acute{\tau} + \acute{\gamma}_t + \acute{\lambda}_g + \acute{\phi}_s + \acute{\epsilon}_{igst}
\end{aligned}
$$

where $k = \{-2, -1, 0, 1, 2, 3, 4, 5, 6\}$ with -2 denoting two years before and 6 denoting six years after the first reform and $k = \{-6, -5, -4, -3, -2, -1, 0, 1, 2\}$ with -6 denoting six years before and 2 denoting two years after the second reform.

Indirect (Spillover) Effects

To identify the indirect effects of Rhode Island's dual reforms, the study quantifies how school- and grade-level ISS practices for treatment infractions change average ELA and math proficiency rates for the overall student population in each academic year. Table 6 presents first-stage DD estimates of Equation (3) with the full set of control variables. Columns (1) and (2) indicate that the fraction of students penalized with ISS for attendance-specific infractions does not change peer ELA or math proficiency rates statistically significantly for the first reform. However, columns (3) and (4) show that the second reform does change student academic performance statistically significantly. A 10 percent increase in ISS rates for disruption-specific infractions at the school-grade level lowers the probability of ELA and math proficiency by 1.1 and 1 percentage points respectively, *ex post* (p <0.01).

Nevertheless, we should assume that these DD estimates are biased because the ISS rate measure is endogenously linked to latent student-, principal-, and school-specific characteristics. The preferred model is therefore the DDIV model, which instruments for this endogenous variable using a Bartik-style suspension rate measure. Table 7 presents the DDIV findings, showing that for the second reform, a 10 percent increase in ISS rates for disruption-specific infractions at the school-grade level lowers the likelihood of math proficiency by about 2.1 percentage points (or 47% of the outcome mean) (p <0.05), *ex post*.

In summary, Rhode Island's first suspension reform that uses ISS to penalize attendance-related infractions induces neither negative nor positive spillover effects on student academic proficiencies. However, the second suspension reform, which aims to increase the use of ISS for disruption-specific infractions lowers math proficiency rates of the student-body, on average. Therefore, the small gains in math proficiency rates for treated students are negated by the adverse impact of the reform on the overall student population. This finding aligns with Kinsler (2013) indicating that more punitive disciplinary policies produce positive externalities on academic achievement. However, the ambiguity in the legislative language associated with the second reform should not be overlooked. The Rhode Island General Assembly might consider revising the language of the second legislative reform to improve transparency in its implementation as well as address its disparate impact.

Sensitivity Checks

The study performs additional testing to support causal inference in the DDIV framework. The event-study specification corresponding to Equation (7)[9] confirms that the pre-reform outcome trends are all parallel (see Figure 6). The parameters from the first-stage equations (Equations 5 and 6) are presented in Appendix Table A1 and show a strong relationship between the instrument and the endogenous variable. Appendix Figure A2 also illustrates binned scatter plots visualizing the reduced-form relationship between academic proficiency outcomes and the instrumental variable defined in Equation (4).

Since the measure of ISS rates is continuous, using a continuous treatment variable in the DD framework simply amalgamates treatment effects over all doses of the treatment, making the interpretation of the DDIV estimator less intuitive (Callaway et al., 2024). To show how the impact of ISS rates changes over the treatment distribution, I use a dose-response function, allowing for nonlinearities in the treatment effect. Figure 7 shows that for schools with high ISS rates (above 80%), ELA and math proficiency rates are approximately three times lower than in schools with low ISS rates (less than 40%), *ex post*.

To check the validity of these findings, I also perform falsification tests to assess the extent to which the residualized future instruments, $\widehat{ISSRate}^{j}_{gst+k}$, for $k \in \{\ 1,\ 2\ \}$, predict Y_{igst} conditional on $\widehat{ISSRate}^{j}_{gst}$ (see Pope and Zuo (2023)). We would not expect future

[9]The corresponding event-study equation is:

$$\begin{aligned} Y_{igst} =& \acute{\rho} + \sum_{l \in k} \acute{\omega}^{l}_{j}\, \widehat{ISSRate \cdot Post}^{jl}_{gst} + \acute{\theta}_{j1}\, \widehat{ISSRate}^{j}_{gst} + \acute{\theta}_{j2}\, Post^{j}_{t} \\ &+ \acute{\theta}_{3}\, \chi_{igt-1} + \acute{\theta}_{4}\, S_{sgt-1} + \acute{\Lambda}_{sg} + \acute{\Upsilon}_{ot} + \acute{\eta}_{t} + \acute{mu}_{g} + \acute{\psi}_{s} + \acute{\zeta}_{igst} \end{aligned}$$

where $k = \{-2, -1, 0, 1, 2, 3, 4, 5, 6\}$ with -2 denoting two years before and 6 denoting six years after the first reform and $k = \{-6, -5, -4, -3, -2, -1, 0, 1, 2\}$ with -6 denoting six years before and 2 denoting two years after the second reform.

values of the instrument to predict contemporaneous academic achievement, except through current values of the instrument. Therefore, the test replaces the contemporaneous measure of ISS rates with the residualized $\widehat{ISSRate}^{j}_{gst+k}$, which excludes the information in the contemporaneous instrument, $\widehat{ISSRate}^{j}_{gst}$. The findings in Appendix Tables A5 and A6 show no statistically significant relationship between the future values of the instrument and the contemporaneous proficiency outcomes, thereby reinforcing instrument validity.

6 Conclusion

This paper evaluates dual suspension reforms enacted in the state of Rhode Island and their impact on student academic achievement. These dual reforms have become 'dueling' reforms in terms of their individual and overall impacts on ELA and math proficiency rates.

The study uses a DDD design to show that while the first suspension reform has no statistically significant impacts on individual-student academic outcomes, the second suspension reform raises individual-student math proficiency rates by about 3 percentage points ($p < 0.05$), *ex post*, representing 7% of the outcome mean. Next, the study uses a DDDD design to show that while the first suspension reform lowers the Black-White, Hispanic-White, and Other-White math proficiency gaps by more than 20 percentage points, the second reform does not change racial-ethnic or disability achievement gaps statistically significantly. Finally, the study uses a DDIV design to show that even though the first reform does not change average academic outcomes for the overall student population, the second reform substantially lowers math proficiency rates for the entire student population. These findings are all robust to pre-reform parallel trends, sensitivity checks, falsification tests, and the honest DiD approach (Roth and Rambachan, 2022) modified for DDD, DDDD, and DDIV models.

Why the second suspension reform produces negative indirect or spillover effects remains a topic for future research. Although the legislative language clearly states its intent to curb OSS for disruption-specific infractions, the conditions under which this reform can be implemented (i.e., under non-threatening, non-harmful, and non-obstructive conditions) are less clear, which may account for the negative spillover effects we observe for Rhode Island students on average. Policymakers, administrators, and practitioners interested in addressing these effects of the second reform might reconsider the language of the legislation and perhaps stipulate alternative disciplinary actions.

References

Aizer, A. and J. J. Doyle Jr (2015): "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges," The Quarterly Journal of Economics, 130, 759–803.

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

Anderson, K. P. and G. W. Ritter (2017): "Disparate use of exclusionary discipline: Evidence on inequities in school discipline from a US state," Education Policy Analysis Archives, 25, 49–49.

Bacher-Hicks, A., S. B. Billings, and D. J. Deming (2024): "The school-to-prison pipeline: Long-run impacts of school suspensions on adult crime," American Economic Journal: Economic Policy, 16, 165–193.

Baker-Smith, E. C. (2018): "Suspensions suspended: Do changes to high school suspension policies change suspension rates?" Peabody Journal of Education, 93, 190–206.

Bryant, E. B., C. Walsh, L. Barrett, J. Beckwith, and others. (2011): "Rhode Island KIDS COUNT Factbook, 2011," .

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2024): "Difference-in-differences with a continuous treatment," Tech. rep., National Bureau of Economic Research.

Cleveland, C. (2022): "Rethinking discipline: The effects of school disciplinary reform laws on adult and student behavior," Tech. rep., Working Paper.

Craig, A. C. and D. Martin (2023): "Discipline reform, school culture, and student achievement," Tech. rep., IZA Discussion Papers.

Craigie, T.-A. (2022): "Do school suspension reforms work? Evidence from Rhode Island," Educational Evaluation and Policy Analysis, 44, 667–688.

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

Currie, J. and J. Gruber (1996): "Health insurance eligibility, utilization of medical care, and child health," The Quarterly Journal of Economics, 111, 431–466.

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, 311–320.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik instruments: What, when, why, and how," American Economic Review, 110, 2586–2624.

Gregory, A., R. J. Skiba, and P. A. Noguera (2010): "The achievement gap and the discipline gap: Two sides of the same coin?" Educational researcher, 39, 59–68.

Gruber, J. and E. Saez (2002): "The elasticity of taxable income: evidence and implications," Journal of public Economics, 84, 1–32.

Hashim, A. K., K. O. Strunk, and T. K. Dhaliwal (2018): "Justice for all? Suspension bans and restorative justice programs in the Los Angeles Unified School District," Peabody Journal of Education, 93, 174–189.

Jones, E. (2018): "The link between suspensions, expulsions, and dropout rates," Center for Promise, 1–7.

Kinsler, J. (2013): "School discipline: A source or salve for the racial achievement gap?" International Economic Review, 54, 355–383.

Lacoe, J. and M. P. Steinberg (2018): "Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes," Peabody Journal of Education, 93, 207–227.

Lhamon, C. and J. Samuels (2014): "Dear colleague letter on the nondiscriminatory administration of school discipline," Washington, DC: US Department of Education Office of Civil Rights & US Department of Justice Civil Rights Division.

Liberman, A. and M. Katz (2017): "Implementing restorative justice in Rhode Island schools," Urban Institute.

Liu, J., M. S. Hayes, and S. Gershenson (2022): "JUE insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline," Journal of Urban Economics, 103453.

Matos, L. G. S. (2023): "Racial and Ethnic Disparities in K-16 Education in Rhode Island," .

McConnell, B., K. T. K. Tan, and M. Zapryanova (2024): "How do parole boards respond to large, societal shocks? Evidence from the 9/11 terrorist attacks," Journal of Public Economics, 238, 105206.

McNeal, L. and C. Dunbar Jr (2010): "In the eyes of the beholder: Urban student perceptions of zero tolerance policy," Urban Education, 45, 293–311.

Morgan, P. L., G. Farkas, M. M. Hillemeier, Y. Wang, Z. Mandel, C. De-Jarnett, and S. Maczuga (2019): "Are students with disabilities suspended more frequently than otherwise similar students without disabilities?" Journal of school psychology, 72, 1–13.

Morris, E. W. and B. L. Perry (2016): "The punishment gap: School suspension and racial disparities in achievement," Social Problems, 63, 68–86.

Owens, J. and S. S. McLanahan (2020): "Unpacking the drivers of racial disparities in school suspension and expulsion," Social Forces, 98, 1548–1577.

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

Pope, N. G. and G. W. Zuo (2023): "Suspending suspensions: The education production consequences of school suspension policies," The Economic Journal, 133, 2025–2054.

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

Roth, J. and A. Rambachan (2022): "A more credible approach to parallel trends," Rev. Econ. Stud., 1–37.

Sagar, H. A. and J. W. Schofield (1980): "Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts." Journal of personality and social psychology, 39, 590.

Skiba, R., C. R. Reynolds, S. Graham, P. Sheras, J. C. Conoley, E. Garcia-Vazquez, Z. T. T. Force, A. Staff, R. Subotnik, H. Sickler, et al. (2006): "Are Zero Tolerance Policies Effective in the Schools? An Evidentiary Review and Recommendations An official report of the APA," .

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

Skiba, R. J. and K. Knesting (2002): Zero tolerance, zero evidence: an analysis of school disciplinary practice., Jossey-Bass/Wiley.

Stevens, W. D., L. Sartain, E. M. Allensworth, and R. Levenstein (2015): "Discipline practices in Chicago schools," The University of Chicago Consortium on Chicago School Research, 52.

Zarecki, D. (2019): "Banning math progress: The academic impact of California's suspension bans," Available at SSRN 3797279.

Tables

Table 2: DDD Estimates for the Impact of the First Reform

	(1) English	(2) Math	(3) English	(4) Math	(5) English	(6) Math	(7) English	(8) Math
Attend*ISS*Post	-0.0550*	-0.0344	-0.0252	-0.0045	-0.0074	-0.0249	-0.0075	-0.0250
	(0.0303)	(0.0242)	(0.0287)	(0.0259)	(0.0332)	(0.0297)	(0.0332)	(0.0298)
$\overline{Y}$	0.57	0.49	0.57	0.49	0.57	0.49	0.57	0.49
R^2	0.18	0.18	0.28	0.25	0.28	0.26	0.28	0.26
N	573418	570230	573418	570230	573418	570230	573418	570230
Various FE	Y	Y	Y	Y	Y	Y	Y	Y
Control Variables	N	N	Y	Y	Y	Y	Y	Y
School-Grade FE	N	N	N	N	Y	Y	Y	Y
Trend Variables	N	N	N	N	N	N	Y	Y

Note: The table presents DDD estimates from Equation (1). The coefficient on $Attend * ISS * Post$ indicates the impact of being penalized with an ISS for an attendance-related infraction on the probability of ELA and math proficiency, *ex post* (expressed in percentage points). The models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: DDD Estimates for the Impact of the Second Reform

	(1) English	(2) Math	(3) English	(4) Math	(5) English	(6) Math	(7) English	(8) Math
Disrupt*ISS*Post	-0.0237	-0.0109	-0.0379**	-0.0218	-0.0085	0.0331**	-0.0085	0.0330**
	(0.0172)	(0.0188)	(0.0155)	(0.0160)	(0.0171)	(0.0165)	(0.0172)	(0.0165)
$\overline{Y}$	0.57	0.49	0.57	0.49	0.57	0.49	0.57	0.49
R^2	0.18	0.18	0.28	0.25	0.28	0.26	0.28	0.26
N	573418	570230	573418	570230	573418	570230	573418	570230
Various FE	Y	Y	Y	Y	Y	Y	Y	Y
Control Variables	N	N	Y	Y	Y	Y	Y	Y
School-Grade FE	N	N	N	N	Y	Y	Y	Y
Trend Variables	N	N	N	N	N	N	Y	Y

Note: The table presents DDD estimates from Equation (1). The coefficient on $Disrupt * ISS * Post$ indicates the impact of being penalized with an ISS for a disruption-related infraction on the probability of ELA and math proficiency, *ex post* (expressed in percentage points). The models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: DDDD Estimates of the Impact of the First Reform on Achievement Gaps

	(1) English	(2) Math	(3) English	(4) Math	(5) English	(6) Math	(7) English	(8) Math
Attend*ISS*Black*Post	0.0649	0.2024**						
	(0.0928)	(0.0975)						
Attend*ISS*Hisp*Post			0.0484	0.2202***				
			(0.0686)	(0.0596)				
Attend*ISS*Other*Post					0.0492	0.2638**		
					(0.1074)	(0.1032)		
Attend*ISS*IEP*Post							0.0321	-0.0045
							(0.0523)	(0.0540)
$\overline{Y}$	0.63	0.55	0.59	0.50	0.65	0.57	0.57	0.49
R^2	0.28	0.25	0.29	0.26	0.27	0.24	0.29	0.26
N	399035	393692	484684	481226	391800	386116	573418	570230
Various FE	Y	Y	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y	Y	Y	Y	Y

Note: The table presents DDDD estimates from Equation (2). The coefficient on $Attend * ISS * Gr * Post$ indicates the net impact of a minority or IEP student being penalized with an ISS for an attendance-related infraction on the probability of ELA and math proficiency relative to white counterparts, *ex post* (expressed in percentage points). The models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: DDDD Estimates of the Impact of the Second Reform on Achievement Gap

	(1) English	(2) Math	(3) English	(4) Math	(5) English	(6) Math	(7) English	(8) Math
Disrupt*ISS*Black*Post	0.0481	-0.0105						
	(0.0488)	(0.0362)						
Disrupt*ISS*Hisp*Post			0.0529	0.0247				
			(0.0400)	(0.0300)				
Disrupt*ISS*Other*Post					-0.0081	0.0290		
					(0.0545)	(0.0511)		
Disrupt*ISS*IEP*Post							0.0534	-0.0049
							(0.0340)	(0.0242)
$\overline{Y}$	0.63	0.55	0.59	0.50	0.65	0.57	0.57	0.49
R^2	0.28	0.25	0.29	0.26	0.27	0.24	0.29	0.26
N	399035	393692	484684	481226	391800	386116	573418	570230
Various FE	Y	Y	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y	Y	Y	Y	Y

Note: The table presents DDDD estimates from Equation (2). The coefficient on $Disrupt * ISS * Gr * Post$ indicates the net impact of a minority or IEP student being penalized with an ISS for a disruption-related infraction on the probability of ELA and math proficiency relative to white counterparts, *ex post* (expressed in percentage points). The models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: DD Estimates of the Impact of Suspension Rates on Overall Academic Achievement

	(1) English	(2) Math	(3) English	(4) Math
ISS Rate*Post(Attend)	0.0017	0.0178		
	(0.0459)	(0.0351)		
ISS Rate*Post(Disrupt)			-0.1094***	-0.0993***
			(0.0241)	(0.0259)
$\overline{Y}$	0.57	0.49	0.57	0.49
R^2	0.21	0.20	0.21	0.20
N	571998	568846	571998	568846
Various FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y

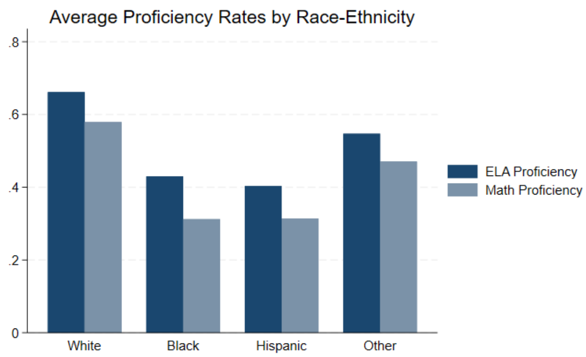
Note: The table presents DD estimates from Equation (3). The DD coefficient on $ISSRate*Post(Attend)$ and $ISSRate*Post(Disrupt)$ indicate the impact of a 1 percentage point increase in infraction-specific ISS rates on the probability of ELA and math proficiency, *ex post*. All models accounts for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: DDIV Estimates of the Impact of Suspension Rates on Overall Academic Achievement

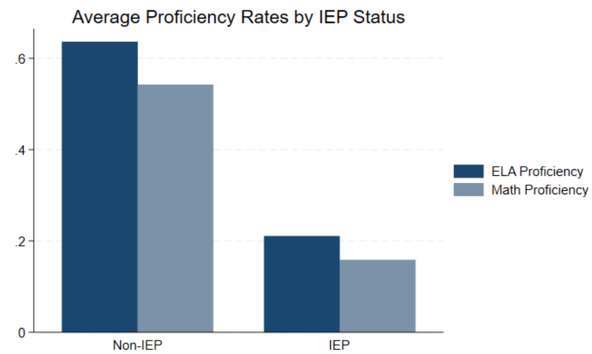
	(1) English	(2) Math	(3) English	(4) Math
ISS Rate*Post(Attend)	-0.0203	-0.0074		
	(0.0330)	(0.0390)		
ISS Rate*Post(Disrupt)			-0.0828	-0.2123**
			(0.1044)	(0.0955)
$\overline{Y}$	0.55	0.45	0.55	0.45
R^2	0.02	0.02	0.02	0.02
N	372864	369576	372864	369576
Various FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y

Note: The table presents DDIV estimates from Equation (7). The DDIV coefficient on $ISSRate * Post(Attend)$ and $ISSRate * Post(Disrupt)$ indicate the impact of a 1 percentage point increase in infraction-specific ISS rates on the probability of ELA and math proficiency, *ex post*. The DDIV estimates use the instrumented ISS rates found in equations (5) and (6). All models accounts for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures



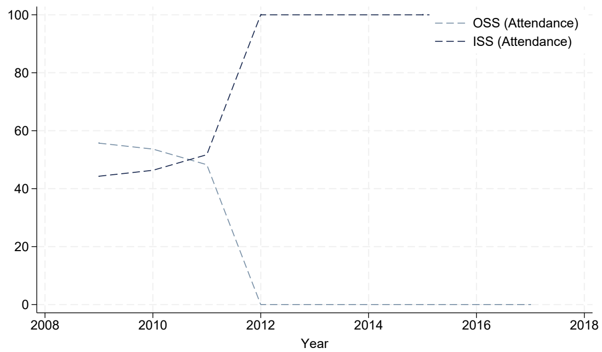
(a) Minority Achievement Gaps



(b) Disability Achievement Gaps

Figure 1: Achievement Gaps.

Note: Achievement gaps in ELA and math proficiency.

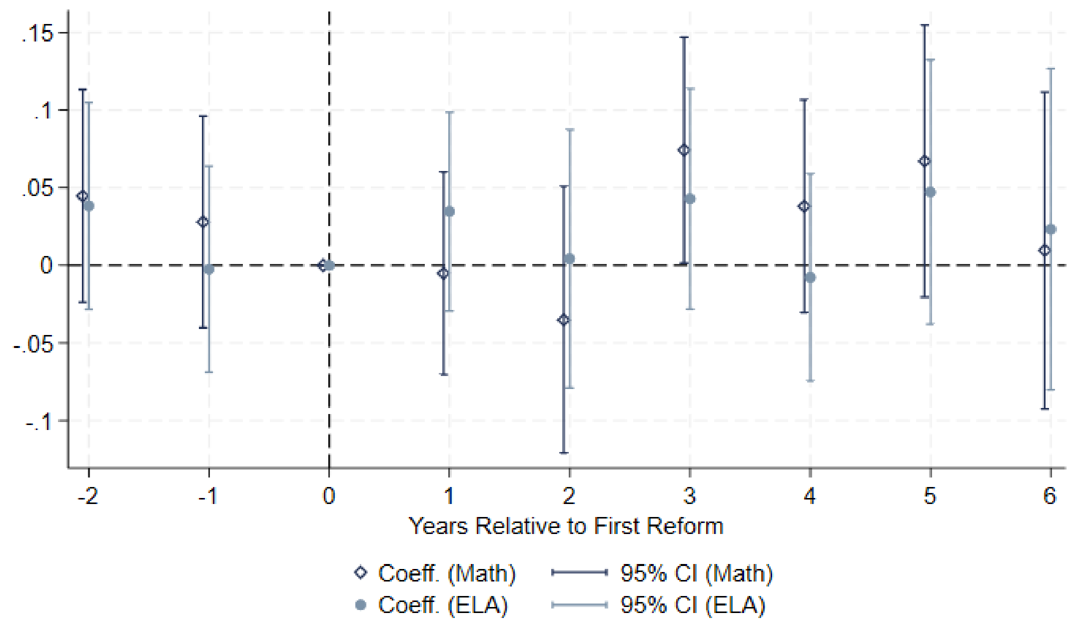


(a) Attendance-Related Infractions

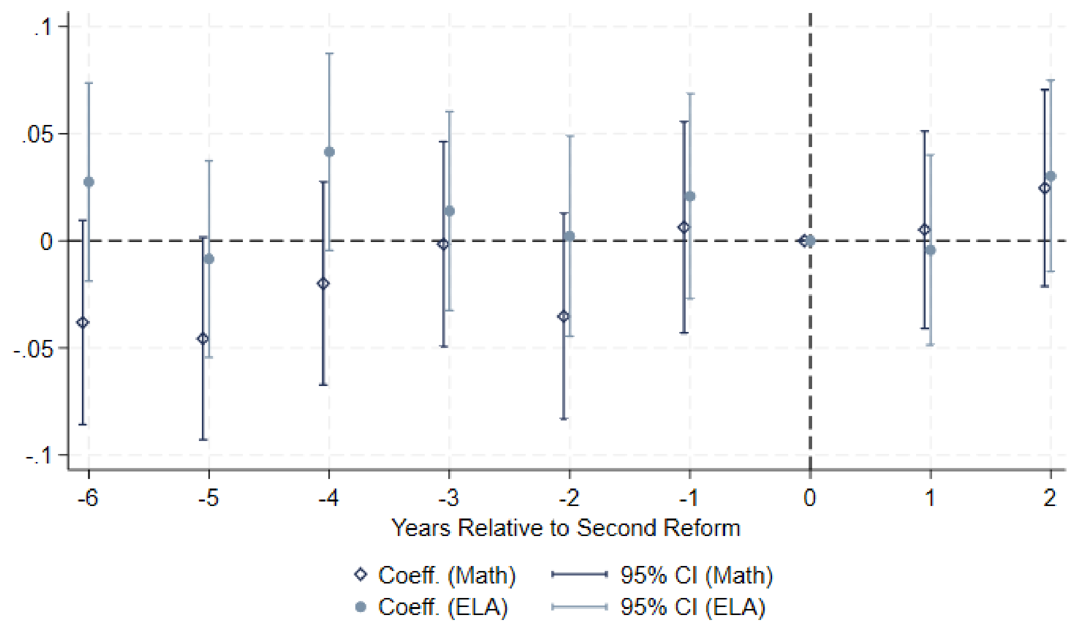
(b) Disruption-Related Infractions

Figure 2: OSS and ISS Trends for Treatment Infractions.

Note: RIDE (2009-2018). This graph illustrates OSS and ISS trends in attendance-related infractions (such as leaving school grounds without permission, tardiness, truancy, and skipping class or detention or in-school suspension) and disruption-related infractions (such as insubordination/disrespect, disorderly conduct, and obscene/abusive language targeting students and teachers) from AY 2009-10 to AY 2017-18.



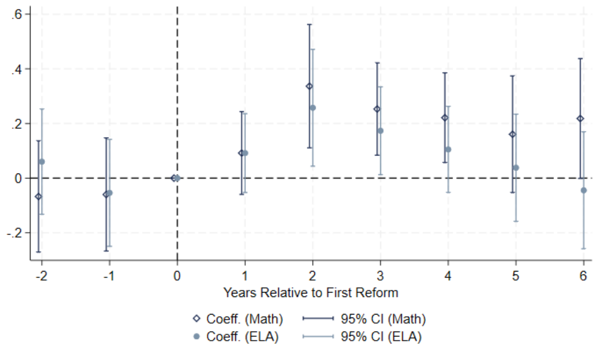
(a) First Reform



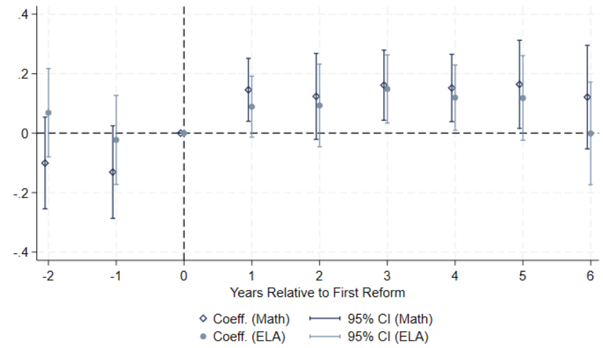
(b) Second Reform

Figure 3: Event-Study Graphs: General Model

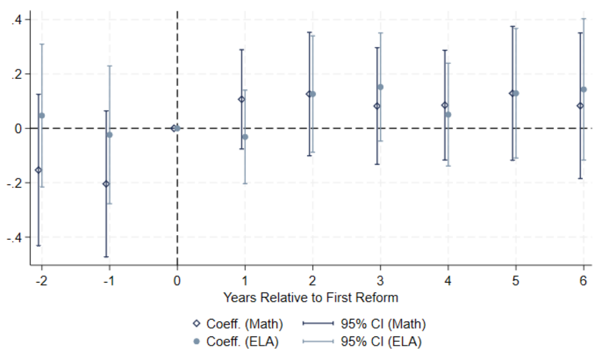
Notes: Figures illustrate the pre- and post-reform DDD estimates on ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The coefficients indicate the impact of being penalized with an ISS for a treatment infraction on the probability of ELA and math proficiency in time t relative to the excluded time period, *ex post* (expressed in percentage points). All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



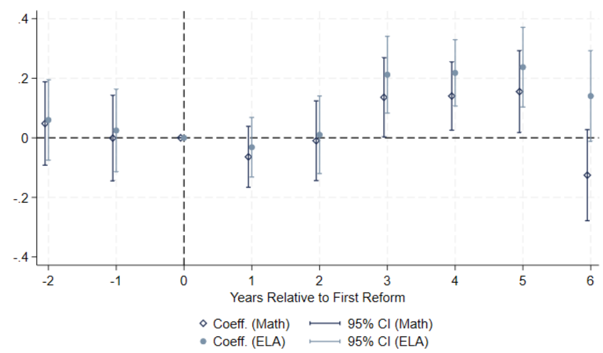
(a) Black-White



(b) Hispanic-White



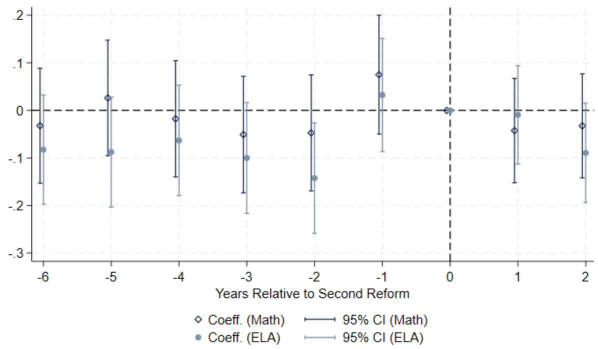
(c) Other-White



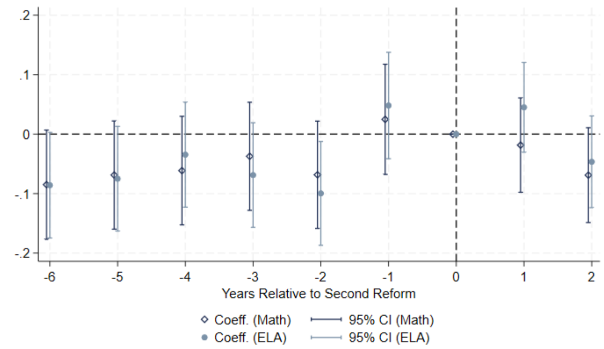
(d) Disability

Figure 4: Event-Study Graphs: Impact of the First Reform on Achievement Gaps

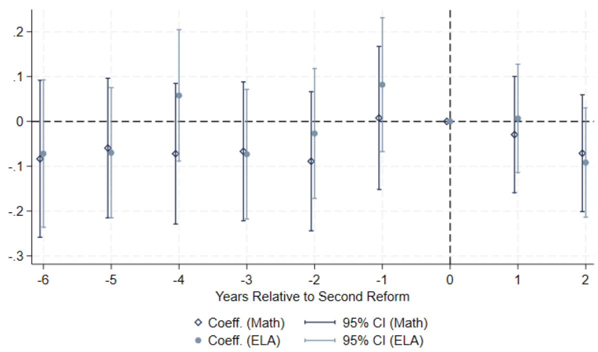
Note: Figures illustrate the pre- and post-reform DDDD estimates of the impact of the first reform on achievement gaps in ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The coefficients indicate the relative impact of a minority or IEP student being penalized with an ISS for a treatment infraction on the probability of ELA and math proficiency in time t relative to the excluded time period, *ex post* (expressed in percentage points). All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



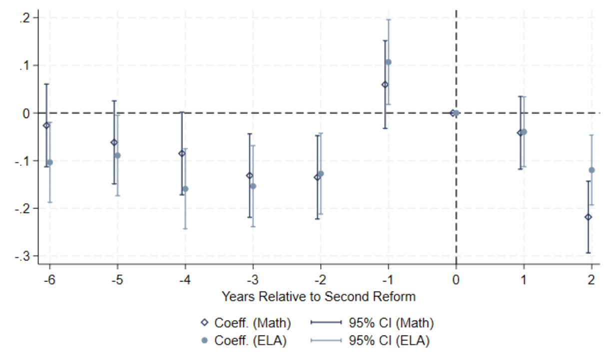
(a) Black-White



(b) Hispanic-White



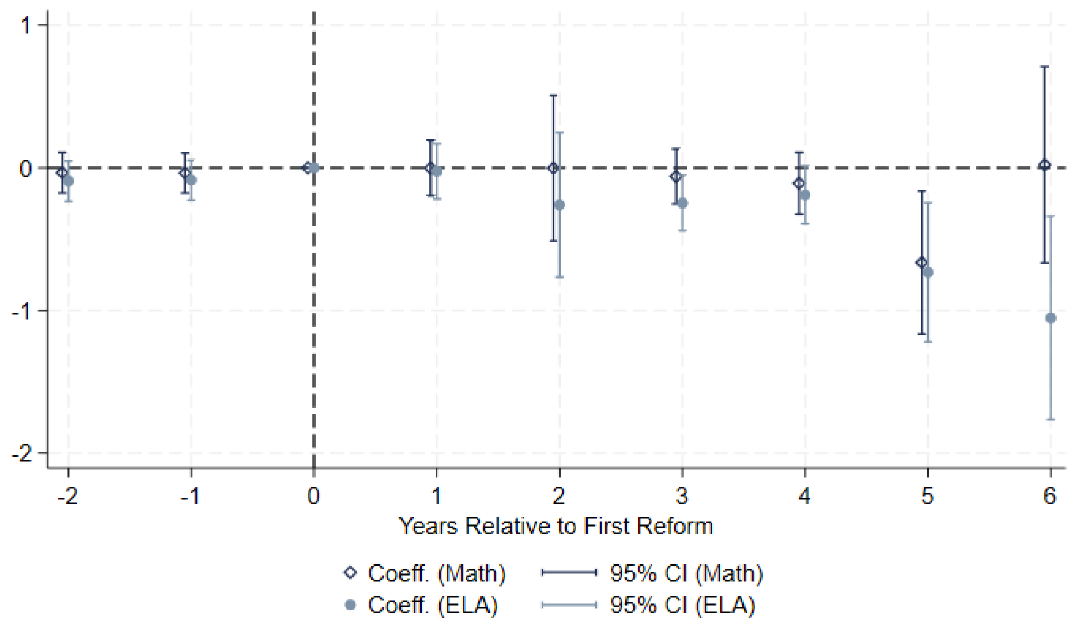
(c) Other-White



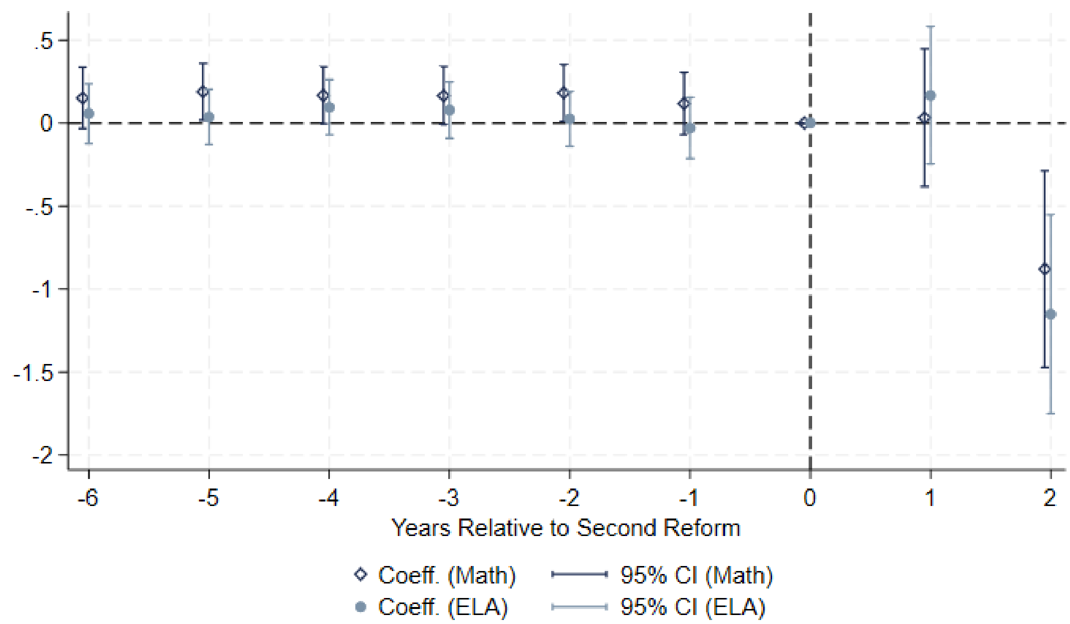
(d) Disability

Figure 5: Event-Study Graphs: Impact of the Second Reform on Achievement Gaps

Note: Figures illustrate the pre- and post-reform DDDD estimates of the impact of the second reform on achievement gaps in ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The coefficients indicate the relative impact of a minority or IEP student being penalized with an ISS for a treatment infraction on the probability of ELA and math proficiency in time t relative to the excluded time period, *ex post* (expressed in percentage points). All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



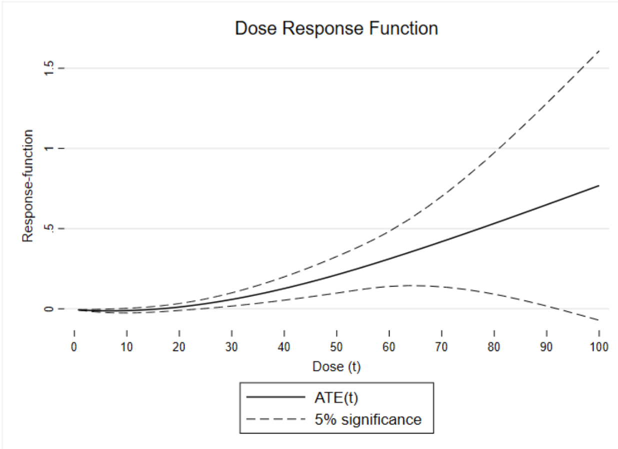
(a) First Reform



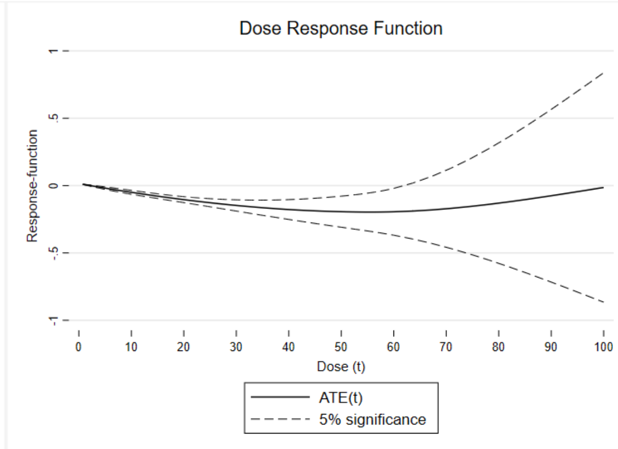
(b) Second Reform

Figure 6: Event-Study Graphs: DDIV Model

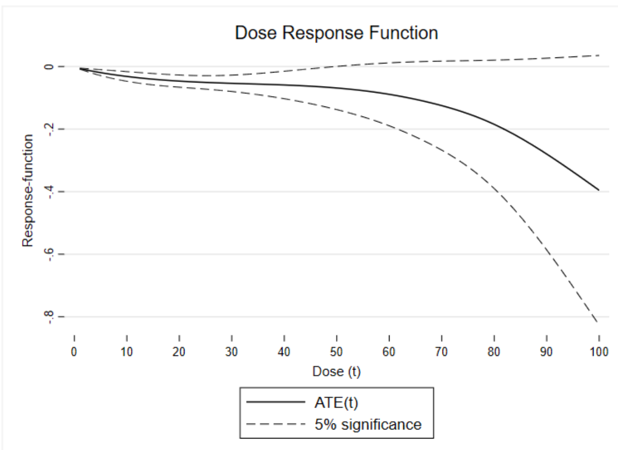
Note: Figures illustrate the pre- and post-reform DDIV estimates on the impact of suspension reforms on ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The coefficients indicate the relative impact of a student being penalized with an ISS for a treatment infraction on the probability of ELA and math proficiency in time t relative to the excluded time period, *ex post* (expressed in percentage points). All models account for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects.



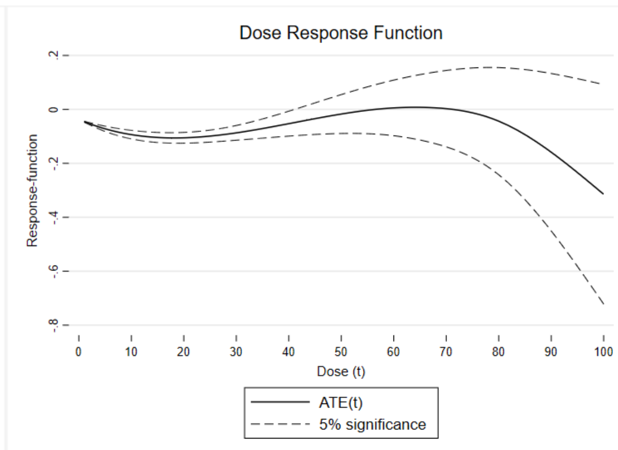
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



(c) Second Reform – ELA Proficient



(d) Second Reform – Math Proficient

Figure 7: Dose Response Functions: DDIV

Note: Figures illustrate the dose response functions for the DDIV regressions. The confidence interval indicates the 95% confidence interval of the DDIV estimates of the average treatment effect (ATE) on ELA and math proficiency rates of the overall student population. All models account for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects.

Appendix – Tables

Table A1: First-Stage IV Estimates – Main Model

	First-Stage: Equation 1			
	(1)	(2)	(3)	(4)
	English	Math	English	Math
ISS Rate*Post (Attend.)	-0.0026***	-0.0024***		
	(0.0001)	(0.0001)		
ISS Rate (Attend.)	0.0902***	0.0905***		
	(0.0001)	(0.0001)		
ISS Rate*Post (Disrupt.)			-0.0336***	-0.0347***
			(0.0003)	(0.0003)
ISS Rate (Disrupt.)			0.0777***	0.0777***
			(0.0002)	(0.0002)
Observations	372864	369576	372864	369576
R^2	0.892	0.893	0.769	0.771
First-Stage F-Statistic	328010	326474	60518	60511

	First-Stage: Equation 2			
	(1)	(2)	(3)	(4)
	English	Math	English	Math
ISS Rate*Post (Attend.)	0.0907***	0.0911***		
	(0.0001)	(0.0001)		
ISS Rate (Attend.)	0.0024***	0.0027***		
	(0.0001)	(0.0001)		
ISS Rate*Post (Disrupt.)			0.0552***	0.0544***
			(0.0002)	(0.0002)
ISS Rate (Disrupt.)			0.0015***	0.0013***
			(0.0001)	(0.0001)
Observations	372864	369576	372864	369576
R^2	0.861	0.863	0.510	0.513
First-Stage F-Statistic	436766	439939	40059	39812
Various FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y

Note: The table presents first-stage IV estimates from Equations (5) and (6). The coefficient on $ISSRate * Post(Attend)$ and $ISSRate * Post(Disrupt)$ indicate the impact of a 1 percentage point increase in infraction-specific ISS rates on the probability of ELA and math proficiency, *ex post*. The model accounts for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: DDD Estimates – Treatment Schools Only

	(1) English	(2) Math	(3) English	(4) Math
Attend*ISS*Post	-0.0094	-0.0167		
	(0.0341)	(0.0308)		
Disrupt*ISS*Post			-0.0182	0.0250
			(0.0183)	(0.0175)
$\overline{Y}$	0.55	0.45	0.57	0.47
R^2	0.31	0.28	0.29	0.27
N	285883	281215	439313	435326
Various FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y

Note: The table presents DDD estimates using the restricted sample of treatment schools defined as schools in Rhode Island that use OSS to penalize treatment infractions either for the student body as a whole or for specific grades. The coefficient on $Attend*ISS*Post$ indicates the impact of being penalized with an ISS for an attendance-related infraction on the probability of ELA and math proficiency, *ex post* (expressed in percentage points). The coefficient on $Disrupt*ISS*Post$ indicates the impact of being penalized with an ISS for an disruption-related infraction on the probability of ELA and math proficiency, *ex post* (expressed in percentage points). All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: DDDD Estimates for the First Reform – Treatment Schools Only

	(1) English	(2) Math	(3) English	(4) Math	(5) English	(6) Math	(7) English	(8) Math
Attend*ISS*Black*Post	0.0551	0.2021**						
	(0.0931)	(0.1004)						
Attend*ISS*Hisp*Post			0.0497	0.2128***				
			(0.0696)	(0.0611)				
Attend*ISS*Other*Post					0.0062	0.2738**		
					(0.1097)	(0.1063)		
Attend*ISS*IEP*Post							0.0333	0.0447
							(0.0530)	(0.0572)
$\overline{Y}$	0.62	0.51	0.57	0.47	0.64	0.54	0.55	0.45
R^2	0.30	0.27	0.31	0.28	0.29	0.26	0.31	0.28
N	195590	190374	240102	235753	189812	184291	285883	281215
Various FE	Y	Y	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y	Y	Y	Y	Y

Note: The table presents DDDD estimates using the restricted sample of treatment schools defined as schools in Rhode Island that use OSS to penalize treatment infractions either for the student body as a whole or for specific grades. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: DDDD Estimates for the Second Reform – Treatment Schools Only

	(1) English	(2) Math	(3) English	(4) Math	(5) English	(6) Math	(7) English	(8) Math
Disrupt*ISS*Black*Post	0.0264	-0.0112						
	(0.0512)	(0.0377)						
Disrupt*ISS*Hisp*Post			0.0482	0.0325				
			(0.0420)	(0.0306)				
Disrupt*ISS*Other*Post					-0.0267	0.0446		
					(0.0553)	(0.0523)		
Disrupt*ISS*IEP*Post							0.0399	-0.0033
							(0.0346)	(0.0256)
$\overline{Y}$	0.63	0.54	0.59	0.49	0.65	0.56	0.57	0.47
R^2	0.29	0.26	0.30	0.27	0.28	0.25	0.30	0.27
N	301006	295588	368242	364199	293903	288142	439313	435326
Various FE	Y	Y	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y	Y	Y	Y	Y

Note: The table presents DDDD estimates using the restricted sample of treatment schools defined as schools in Rhode Island that use OSS to penalize treatment infractions either for the student body as a whole or for specific grades. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: DDIV Estimates – Falsification Test I

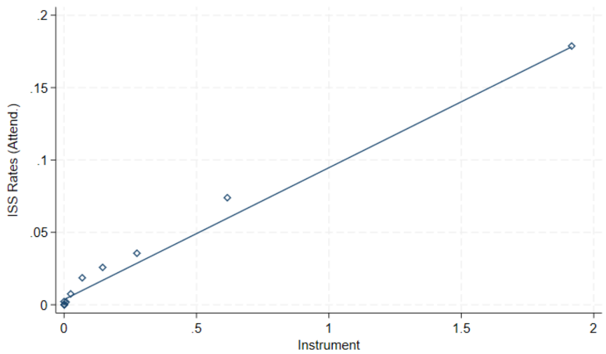
	(1) English	(2) Math	(3) English	(4) Math
ISS Rate*Post(Attend)	0.0062	-0.0220		
	(0.0309)	(0.2060)		
ISS Rate*Post(Disrupt)			0.0235	0.2841
			(0.6421)	(1.4504)
$\overline{Y}$	0.55	0.45	0.55	0.45
R^2	0.02	0.02	0.02	0.02
N	372864	369576	372864	369576
Various FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y

Note: The table presents the results of the falsification test that uses future values of the suspension rates in lieu of lagged values of suspension rates to construct the instrumental variable. The table presents DD estimates from Equation (7). The DDIV coefficient on $ISSRate * Post(Attend)$ and $ISSRate * Post(Disrupt)$ indicate the impact of a 1 percentage point increase in infraction-specific ISS rates on the probability of ELA and math proficiency, *ex post*. The DDIV estimates use the instrumented ISS rates found in equations (5) and (6). All models account for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

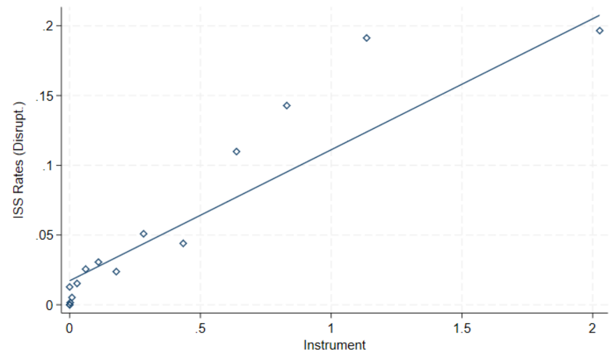
Table A6: DDIV Estimates – Falsification Test II

	(1) English	(2) Math	(3) English	(4) Math
ISS Rate*Post(Attend)	0.0049	-0.0081		
	(0.0358)	(0.1374)		
ISS Rate*Post(Disrupt)			-0.2599	-1.1569
			(1.5202)	(7.7962)
$\overline{Y}$	0.55	0.45	0.55	0.45
R^2	0.02	0.02	0.02	0.01
N	372864	369576	372864	369576
Various FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
School-Grade FE	Y	Y	Y	Y
Trend Variables	Y	Y	Y	Y

Note: The table presents the results of the falsification test that uses twice-lead values of the suspension rates in lieu of lagged values of suspension rates to construct the instrumental variable. The table presents DD estimates from Equation (7). The DDIV coefficient on $ISSRate * Post(Attend)$ and $ISSRate * Post(Disrupt)$ indicate the impact of a 10 percentage point increase in infraction-specific ISS rates on the probability of ELA and math proficiency, *ex post*. The DDIV estimates use the instrumented ISS rates found in equations (5) and (6). All models accounts for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects. Standard errors clustered at the school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



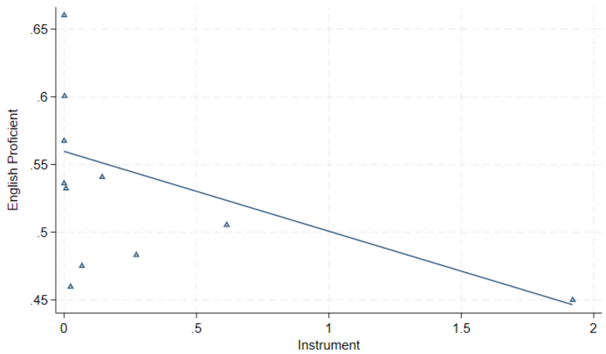
(a) First Reform



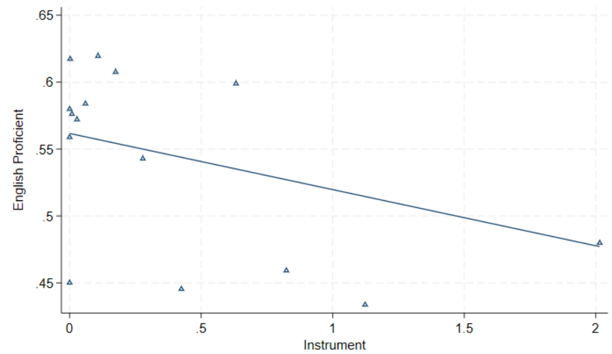
(b) Second Reform

Figure A1: Binned Scatter Plots: First-Stage Equations.

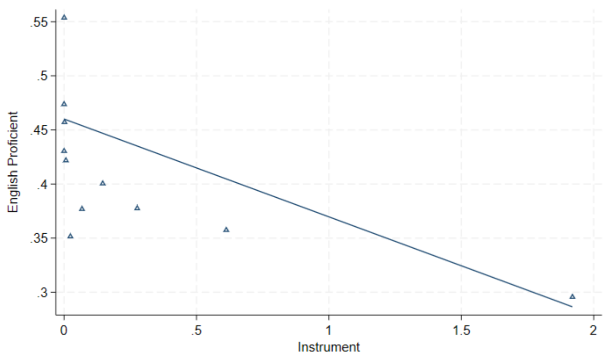
Note: Figures illustrate the binned scatter plots from the first-stage DDIV equations. This figure plots average offense-specific ISS rates against binned values of the instrumental variable. The instrument is calculated based on Equation (4). Both variables are residualized with respect to lagged variables and school-grade fixed effects. To generate the the binned scatterplot, I use the Stata command 'binscatter', which groups the values of the instrument into 30 equal-sized groups, then plots the average values of treatment-specific ISS rates and instrument by bin.



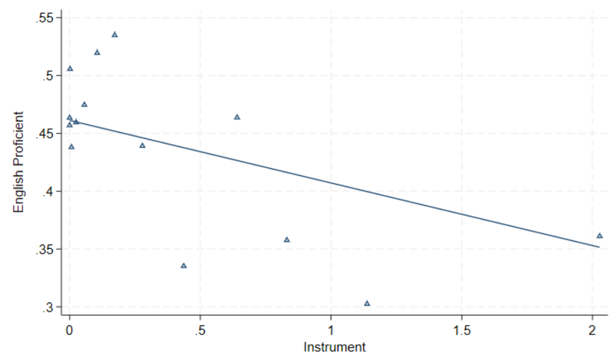
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



(c) Second Reform – ELA Proficient

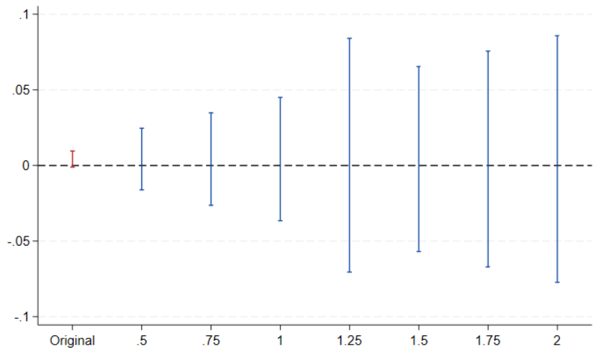


(d) Second Reform – Math Proficient

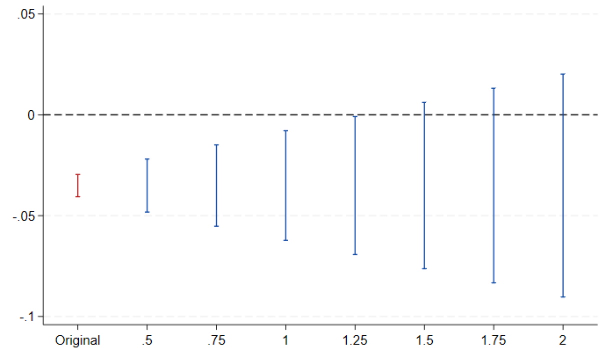
Figure A2: Binned Scatter Plots: Reduced-Form Equations

Note: Figures illustrate the binned scatter plots of the reduced-form equations. This figure plots average ELA and math proficiency rates against binned values of the instrumental variable. The instrument is calculated based on Equation (4). Both variables are residualized with respect to lagged variables and school-grade fixed effects. To generate the the binned scatterplot, I use the Stata command 'binscatter', which groups the values of the instrument into 30 equal-sized groups, then plots the average values of outcome and instrument by bin.

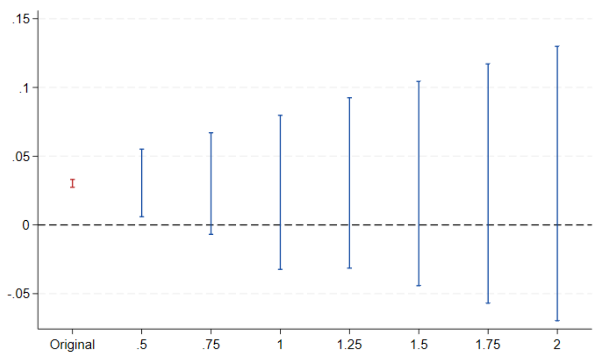
Appendix – Figures



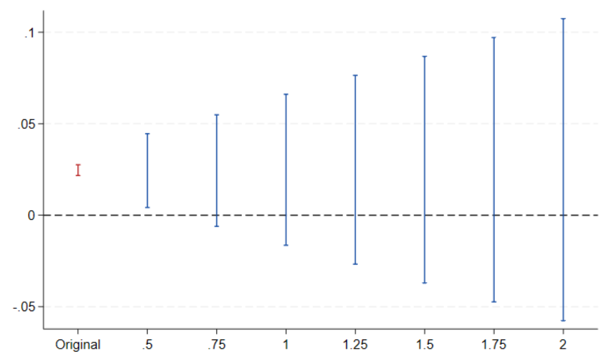
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



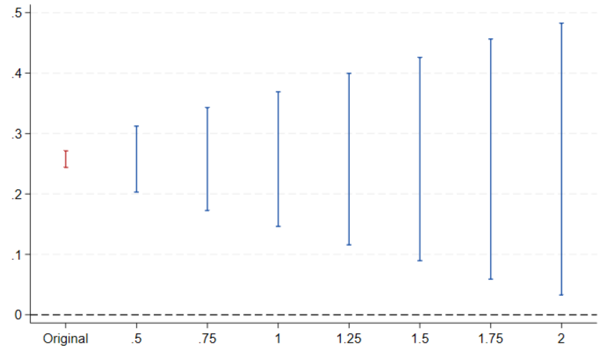
(c) Second Reform – ELA Proficient



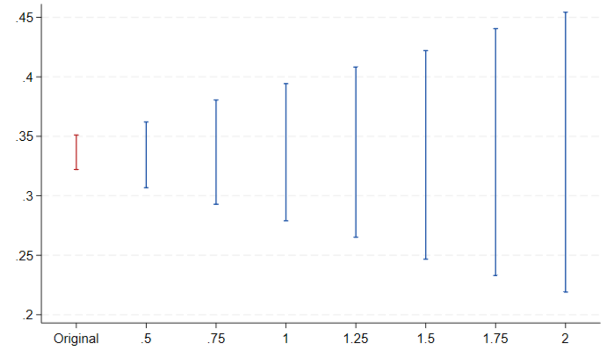
(d) Second Reform – Math Proficient

Figure A3: Honest DDD Graphs: General Model

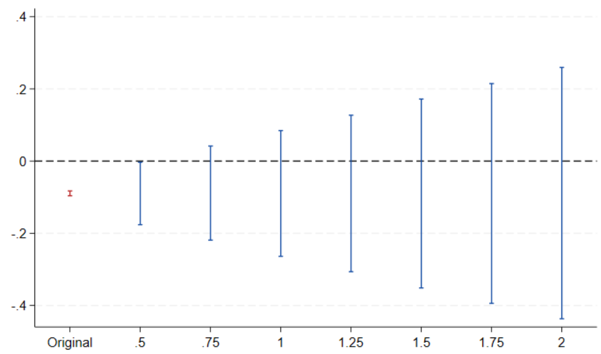
Note: Figures illustrate the Honest DDD estimates on ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The original (i.e, "red") confidence interval indicates the 95% confidence interval of the DDD estimates. The subsequent (i.e., "blue") confidence intervals vary with $\overline{M}$, which represent the factors of the maximum parallel trend violation occurring pre-reform. A confidence interval completely above or below 0 when $\overline{M} = 1$ suggests that if parallel trend violations are as large as each factor, 95% confidence intervals of bounded treatment effects will be statistically different from zero. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



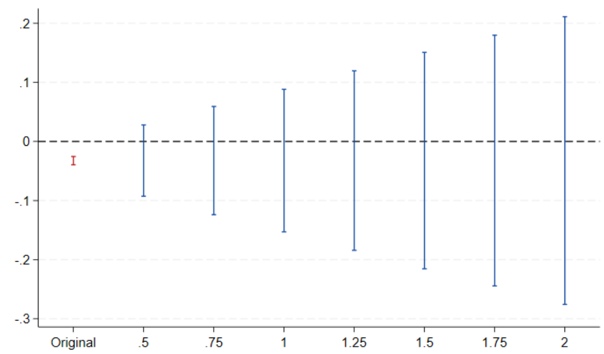
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



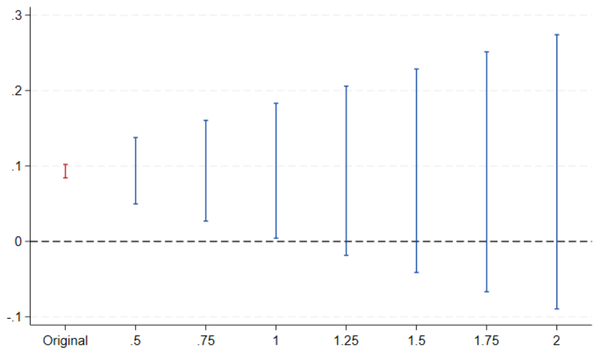
(c) Second Reform – ELA Proficient



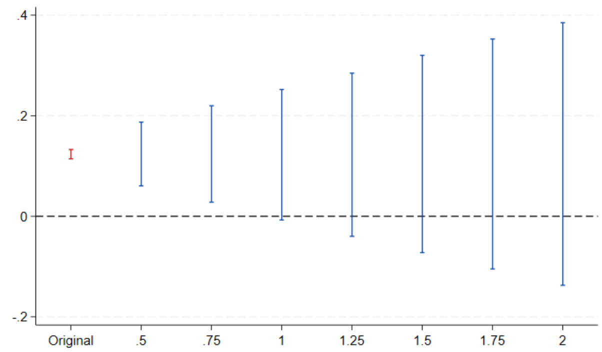
(d) Second Reform – Math Proficient

Figure A4: Honest DDD Graphs: Black-White Achievement Gaps

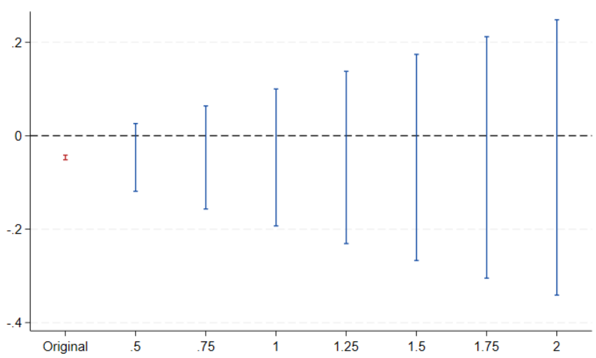
Note: Figures illustrate the Honest DDD estimates on black-white achievement gaps in ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The original (i.e, "red") confidence interval indicates the 95% confidence interval of the DDD estimates. The subsequent (i.e., "blue") confidence intervals vary with $\overline{M}$, which represent the factors of the maximum parallel trend violation occurring pre-reform. A confidence interval completely above or below 0 when $\overline{M} = 1$ suggests that if parallel trend violations are as large as each factor, 95% confidence intervals of bounded treatment effects will be statistically different from zero. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



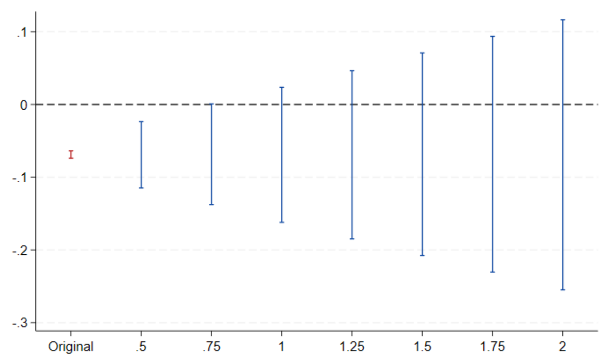
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



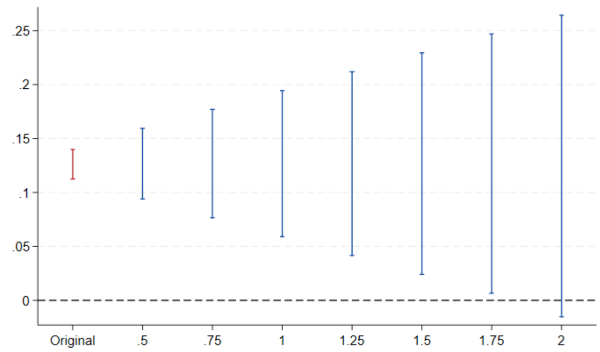
(c) Second Reform – ELA Proficient



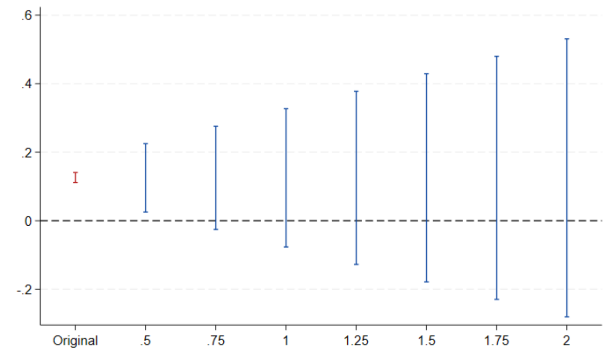
(d) Second Reform – Math Proficient

Figure A5: Honest DDD Graphs: Hispanic-White Achievement Gaps

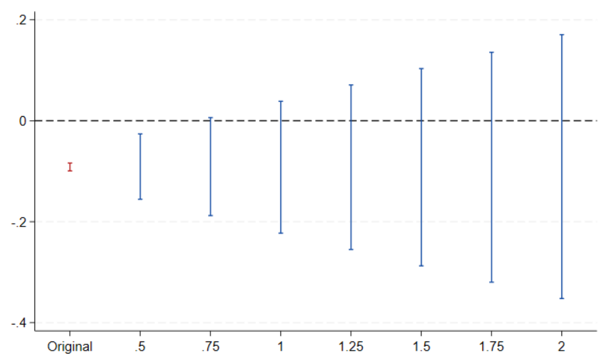
Note: Figures illustrate the Honest DDD estimates on Hispanic-white achievement gaps in ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The original (i.e, "red") confidence interval indicates the 95% confidence interval of the DDD estimates. The subsequent (i.e., "blue") confidence intervals vary with $\overline{M}$, which represent the factors of the maximum parallel trend violation occurring pre-reform. A confidence interval completely above or below 0 when $\overline{M} = 1$ suggests that if parallel trend violations are as large as each factor, 95% confidence intervals of bounded treatment effects will be statistically different from zero. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



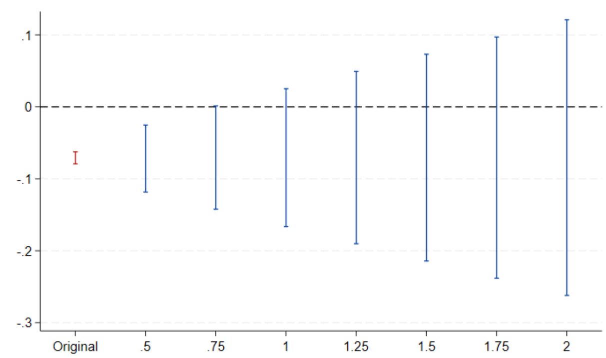
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



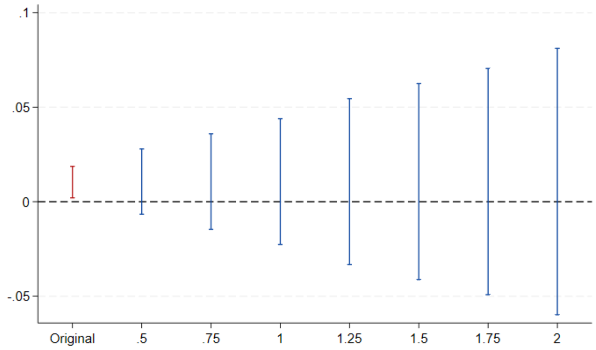
(c) Second Reform – ELA Proficient



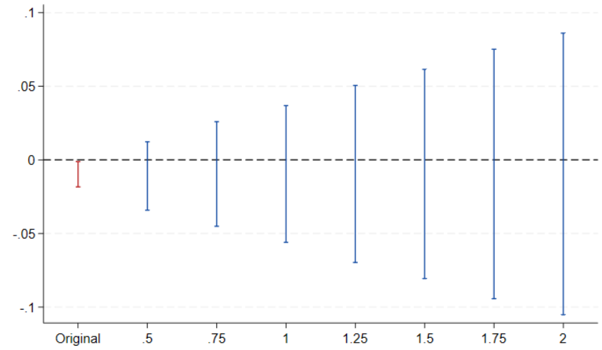
(d) Second Reform – Math Proficient

Figure A6: Honest DDD Graphs: Other-White Achievement Gaps

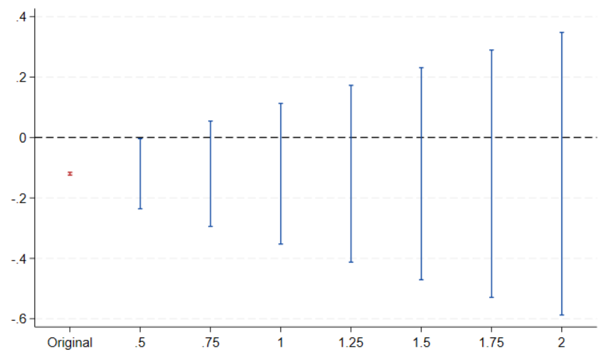
Note: Figures illustrate the Honest DDD estimates on Other-white achievement gaps in ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The original (i.e, "red") confidence interval indicates the 95% confidence interval of the DDD estimates. The subsequent (i.e., "blue") confidence intervals vary with $\overline{M}$, which represent the factors of the maximum parallel trend violation occurring pre-reform. A confidence interval completely above or below 0 when $\overline{M} = 1$ suggests that if parallel trend violations are as large as each factor, 95% confidence intervals of bounded treatment effects will be statistically different from zero. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



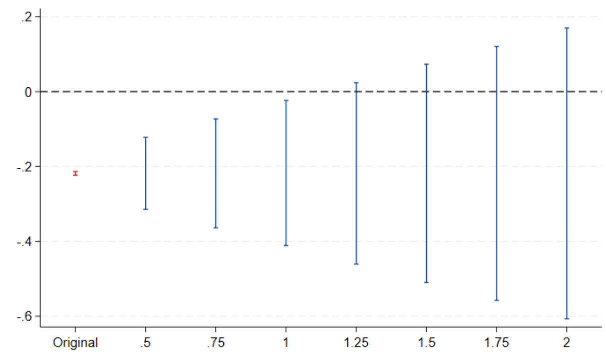
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



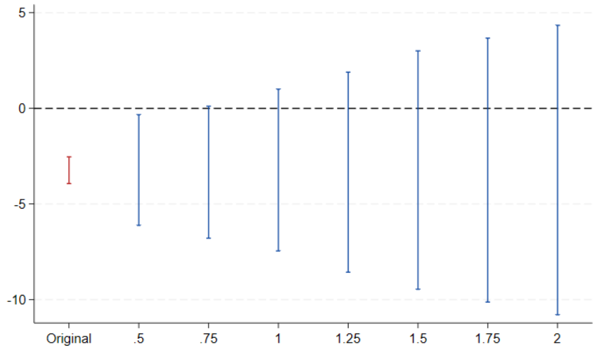
(c) Second Reform – ELA Proficient



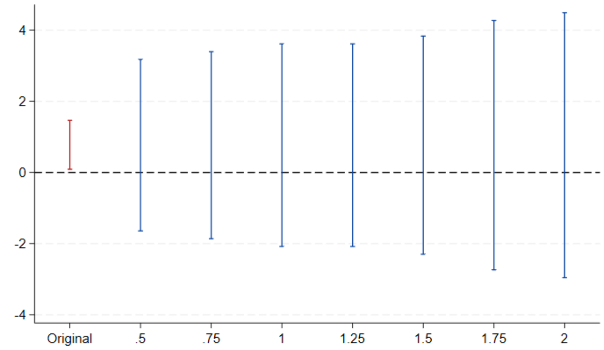
(d) Second Reform – Math Proficient

Figure A7: Honest DDD Graphs: Disability Achievement Gaps

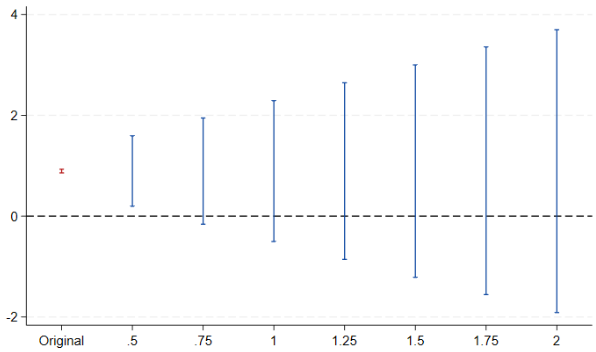
Note: Figures illustrate the Honest DDD estimates on disability achievement gaps in ELA and math proficiency rates for Rhode Island's public school students, grades 3 through 8. The original (i.e, "red") confidence interval indicates the 95% confidence interval of the DDD estimates. The subsequent (i.e., "blue") confidence intervals vary with $\overline{M}$, which represent the factors of the maximum parallel trend violation occurring pre-reform. A confidence interval completely above or below 0 when $\overline{M} = 1$ suggests that if parallel trend violations are as large as each factor, 95% confidence intervals of bounded treatment effects will be statistically different from zero. All models account for school, grade, and year fixed effects, student characteristics, school-grade fixed effects, infraction-year fixed effects, and school-specific, grade-specific, and linear time trends.



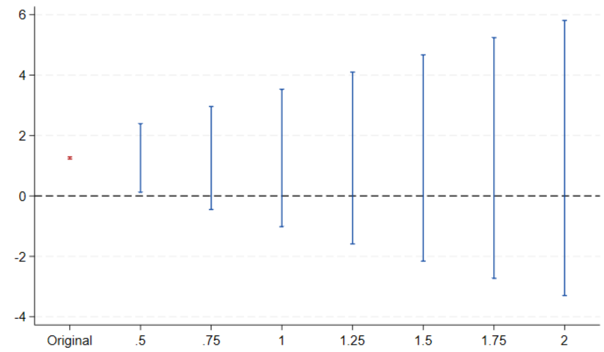
(a) First Reform – ELA Proficient



(b) First Reform – Math Proficient



(c) Second Reform – ELA Proficient



(d) Second Reform – Math Proficient

Figure A8: Honest DiD Graphs: DDIV Models

Note: Figures illustrate the Honest DiD estimates for DDIV regressions. The original (i.e, "red") confidence interval indicates the 95% confidence interval of the DDD estimates. The subsequent (i.e., "blue") confidence intervals vary with $\overline{M}$, which represent the factors of the maximum parallel trend violation occurring pre-reform. A confidence interval completely above or below 0 when $\overline{M} = 1$ suggests that if parallel trend violations are as large as each factor, 95% confidence intervals of bounded treatment effects will be statistically different from zero. All models account for school, grade, and year fixed effects, lagged variables, school-grade fixed effects, and infraction-year fixed effects.