ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-950

Ending Early Grade Suspensions

Ezra Karger
Federal Reserve Bank of Chicago

Sarah Komisarow
Duke University

We investigate the beginning of the school discipline pipeline using a reform in Charlotte-Mecklenburg Schools that limited the use of out-of- school suspension for students in grades K–2. We find that the reform reduced the likelihood of out-of-school suspension by 1.4 percentage points (56%) and had precise null effects on test scores and disciplinary infractions. This leads us to reject a key argument in favor of early-grade suspensions: namely, that early-grade suspensions improve classroom- level outcomes. For high-risk students, we find short-run increases in test scores that persist into third grade. The reform reduced the Black- white out-of-school suspension gap by 79%.

VERSION: May 2024

Suggested citation: Karger, Ezra, and Sarah Komisarow. (2024). Ending Early Grade Suspensions. (EdWorkingPaper: 24-950). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/qprq-vh05

Ending Early Grade Suspensions*

Ezra Karger and Sarah Komisarow†

April 26, 2024

Abstract

We investigate the beginning of the school discipline pipeline using a reform in Charlotte-Mecklenburg Schools that limited the use of out-of-school suspension for students in grades K–2. We find that the reform reduced the likelihood of out-of-school suspension by 1.4 percentage points (56%) and had precise null effects on test scores and disciplinary infractions. This leads us to reject a key argument in favor of early-grade suspensions: namely, that early-grade suspensions improve classroom-level outcomes. For high-risk students, we find short-run increases in test scores that persist into third grade. The reform reduced the Black-white out-of-school suspension gap by 79%.

Keywords: K-12 Education; School Discipline.
JEL: I21; I24; J15.

*We are grateful for helpful suggestions from Briana Ballis, Todd Jones, Ofer Malamud, Kelli Marquardt, and Diane Schanzenbach. Seminar participants at Harvard University, the Duke Children's Law Clinic, Carolina Region Empirical Economics Day (CREED), ASSA 2023, Northwestern University, Penn State University, Ohio State University, Triangle Economists in Applied Microeconomics (TEAM), AEFP 2022, and the University of North Carolina at Chapel Hill Policy Research Group also provided helpful comments. We thank Taegan Mullane for research assistance. All errors and opinions are our own. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System. Komisarow gratefully acknowledges funding from the Duke University Office of the Provost.

†Karger: Federal Reserve Bank of Chicago, Economic Research, 230 S. LaSalle Street, Chicago, IL, 60604. Email: ezra.karger@chi.frb.org. Komisarow: Sanford School of Public Policy, Duke University, 201 Science Drive, Durham, NC 27708. Email: sarah.komisarow@duke.edu.

1 Introduction

There is ongoing debate in education about the use of suspensions as a part of school discipline policy. On the one hand, proponents of suspensions argue that they offer teachers and principals an avenue both to address classroom disruption and violence and to deter their occurrences in the first place (Wall, 2023). On the other hand, opponents of suspensions argue that they are punitive, fail to address root causes of disruptive and/or violent behavior, and are doled out unequally to students of color, male students, and students with disabilities (U.S. Government Accountability Office, 2018; Losen and Martinez, 2020; Leung-Gagné et al., 2022).[1]

Much of the ongoing debate about suspensions as part of school discipline policy lumps together all students in grades K–12. In this paper, however, we focus on grades K–2 (henceforth, "early grades"), where underlying reasons for school suspension look very different than underlying reasons for suspension in older grades. In Figure 1 we plot suspension levels and the distribution of disciplinary infraction types resulting in suspension, separately by grade level, for all K–12 public school students in North Carolina during the 2018/19 school year.[2] Panel (a) highlights that although suspension levels are lower for early-grade students than for students enrolled in all other grades (a point we return to later), the underlying reasons for suspension are very different. Panel (b)

[1]For recent work on within-school disparities in suspension rates by race/ethnicity, gender, and economic background, see, for example, Gopalan and Nelson (2019), Barrett et al. (2021), Liu et al. (2022), and Shi and Zhu (2022).

[2]North Carolina is a state with high quality administrative K–12 education data that provides information on the underlying reason for suspension. We describe these data in more detail in Section 3.

demonstrates that most early-grade suspensions resulted from nonviolent disciplinary infractions. For example, the most common disciplinary infractions resulting in suspension for students enrolled in grades K–2 were aggressive behavior (between 30-40%) and disruptive behavior (between 18-20%).

While early-grade (K–2) suspensions are less common than suspensions for older students, they are not rare. Although there is no national data on K–2 suspensions, we estimate that around 165,000 students in grades K–2 in the U.S. were suspended during the 2017/18 school year.[3] Scaling this estimate by the average length of out-of-school suspension for students in early grades (3.1 days),[4] we estimate that this practice resulted in more than 500,000 lost instructional days for students enrolled in these grades.

Despite a small but growing causal literature in economics on this topic, our paper is the first to investigate the causal effects of suspension policy on early-grade students. Quasi-experimental work by Lacoe and Steinberg (2018), Lacoe and Steinberg (2019), Wang (2022), Craig and Martin (2023), and Pope and Zuo (2023) estimates the impacts of school discipline policy changes on student outcomes for students as young as third grade, but none of these papers considers the impacts on younger students (e.g., students enrolled in kindergarten through second grades).[5] Suspension from school may be particularly

[3]Although the U.S. Civil Rights Data Collection (CRDC) publishes biennial estimates of national- and state-level school suspension rates, these rates are not disaggregated by grade level. To address this shortcoming, we used state-level data from 17 states that published grade-specific counts (or estimates) of suspensions, K–12 enrollment data, and K–12 suspension data to estimate the number of students enrolled in kindergarten through second grades who were suspended in the U.S. during the 2017/18 school year. For more details regarding data collection and estimation methods, please see Appendix A.

[4]Based on authors' calculations. See Appendix Figure B1.

[5]We note that Lacoe and Steinberg (2018) use data aggregated at the district-level (K–12) from the School District of Philadelphia, which includes students enrolled in kindergarten

harmful to students enrolled in the early grades for four reasons. First, lost instructional time may be more detrimental to student learning and socialization in early grades, since students are working on foundational academic skills, forming new relationships with teachers and peers, adapting to the classroom environment and routines, and developing core social-emotional skills such as empathy, cooperation, conflict management, and emotional regulation.[6] This is consistent with a theoretical literature on dynamic skill formation Cunha and Heckman (2007). Second, schooling disruptions induced by suspension may be particularly costly to families and caregivers of young children, who—in the absence of schooling—may have to miss work to care for young children or arrange for alternative childcare. Third, early grade suspensions may catalyze "scarring" effects that persist through the K-12 schooling experience, as past school discipline history may influence the harshness of future consequences and punishments. Finally, harsh exclusionary discipline —including suspensions—has also been causally linked to outcomes in early adulthood, including negative impacts on high school graduation and postsecondary enrollment and increases in criminal justice involvement (Bacher-Hicks et al., 2019).

To address the gap in the existing literature and answer the question of how school discipline policy affect students in the early grades, we take advantage of an abrupt and unexpected reform to school discipline policy in Charlotte-Mecklenburg Schools (CMS) – a large school district serving nearly 140,000

through second grades.

[6]For empirical estimates of student learning gains in early grades, see, for example, Hill et al. (2008) and Fitzpatrick et al. (2011).

students in North Carolina. Specifically, this reform severely limited the use of out-of-school suspension for students enrolled in kindergarten through second grades by adding the requirement that the school district superintendent approve all out-of-school suspensions for students enrolled in these grades. Prior to the reform, there was no requirement for the superintendent to oversee or otherwise approve out-of-school suspension decisions. This reform was passed by the Charlotte-Mecklenburg Board of Education (CMBOE) in August 2017 and went into effect immediately at the beginning of the 2017/18 school year (Doss Helms, 2017a).

Using triple-differences and event study approaches, we demonstrate that the school discipline reform had large, immediate, and permanent impacts on the use of out-of-school suspension in CMS in the early grades. The reform reduced the likelihood that a student enrolled in kindergarten through second grade received an out-of-school suspension by 1.4 percentage points, which translates into a 56 percent decline relative to pre-reform levels. We find little to no evidence of substitution to in-school suspension (unaffected by the reform) or increases in special education placement (which carried the possibility of increased time spent outside of the regular classroom). Finally, we estimate null effects on test scores and reported disciplinary infractions (e.g., fights, disruptive behavior) in the aggregate, implying minimal spillovers of the reform to unaffected students and little support for the claim that limiting suspensions resulted in worsened student behavior or major classroom disruption.

To investigate the possibility of treatment effect heterogeneity underlying

our main results, we partition our sample into subgroups based on predicted student- and classroom-level risk of out-of-school suspension. Our subgroup analyses by student-level risk reveal, unsurprisingly, that the reform reduced out-of-school suspension use the most for high-risk students. We also find suggestive evidence that this group of students experienced reading test score gains due to the reform (0.037 SDs). Our examination of treatment heterogeneity by classroom-level risk revealed larger declines in out-of-school suspension among high-risk classrooms, although we did not detect any statistically significant effects on test scores across classroom groups. These results lead us to conclude that the reform may have benefited high-risk students in terms of reduced suspensions and increased test scores, and at the same time did not harm student achievement for unaffected (low-risk) students.

In addition to short-run results measured in the early grades (K–2), we also measure the longer-run effects of this reform on outcomes in third grade (a grade level unaffected by the reform). Using a modified triple-differences design, we compare students who were affected by the reform in early grades versus those who were not, controlling for time variation using trends for non-CMS students and for within-school variation using older (never treated) students. We find no evidence that the reform affected longer-run (third grade) outcomes related to out-of-school suspension, in-school suspension, special education placement, nor reported disciplinary infractions. Intriguingly, however, we find evidence of increases in reading test scores for high-risk students that persist into third grade. These results imply that the reform affecting early grades may have improved academic achievement over the longer-run; or more

conservatively, we can—in contrast to much of the previous work—confidently reject even small negative effects of the reform on student achievement.

Aside from direct impacts on students, we also investigate how the reform affected teachers and suspension rate gaps. We do not find any evidence of statistically significant effects on teacher turnover. We do find, however, large declines in suspension rate gaps that have been well-documented elsewhere. Namely, we find that the Black-White out-of-school suspension gap declined by 3.7 percentage points (79 percent) and the male-female out-of-school suspension gap declined by 2.0 percentage points (65 percent). We also find that the special education suspension gap declined by 4.6 percentage points (92 percent) and the economically disadvantaged suspension gap declined by 2.2 percentage points (73 percent). These declines in gaps are especially notable in light of the fact that we do not find evidence of negative spillovers onto unaffected peers.

Our work contributes to the small but growing literature on the causal effects of school discipline policy on student outcomes in two ways. First, our paper is the first to use a natural experiment to estimate the causal effects of school discipline policy on students in the early grades. Our findings—which contrast with much of the previous work on this topic—suggest the possibility of positive impacts (reduced out-of-school suspension and increased reading test scores) for high-risk students while providing no evidence of negative spillovers to unaffected students. We emphasize that in our context we are unable to identify the existence of the tradeoff between the outcomes of high- versus low-risk students that has been documented in other work (Pope and

Zuo, 2023). Second, our estimates are generated from a context that dramatically altered school discipline policy. In contrast to much of the previous work that focuses on suspensions bans for specific categories of offenses, we produce policy-relevant estimates emanating from a large-scale change to school discipline policy. This context makes our null findings on in-school suspension, special education placement, and student behavior especially important.

2 Background

2.1 School Discipline Policy in the United States

Starting in the late 1980s and continuing through the 1990s, "zero-tolerance" became the dominant paradigm in school discipline policy in the U.S. Following the introduction of the framework at the federal level, state legislatures and local school districts enacted new laws and policies that significantly expanded the prevalence of the framework (Skiba and Rausch, 2006; Curran, 2016; Ritter and Anderson, 2018; Curran, 2019). As the name implies, zero-tolerance policies mandate punishment—often exclusion from school, such as out-of-school suspension—for student offenses, without consideration of context, mitigating factors, or extenuating circumstances. Some common areas of student behavior that involve zero-tolerance policies include physical violence, weapons possession, drug possession, and bullying. Curran (2016) finds evidence to suggest that expansions of zero-tolerance policies contributed to higher rates of exclusionary discipline and to larger disproportionality in exclusionary discipline for Black students.

Following several decades of zero-tolerance policies, state legislatures and local school districts reversed course and began enacting laws and policies to lessen the influence of this framework. Between 2013 to 2018, the use of school discipline declined nationally: in the 2013/2014 school year, 5.28% of students in U.S. public K–12 schools were given an out-of-school suspension; in the 2017/2018 school year, that number had fallen to 4.96%.[7] In 2014 the U.S. Department of Education (DoE) under the Obama Administration issued federal guidance around school discipline, urging districts and schools to reduce the use of exclusionary discipline and increase the use of alternative approaches that were less punitive, highlighting that "racial discrimination in school discipline is a real problem."[8] Although this guidance was rescinded (Vara-Orta, 2018), many states and school districts around the U.S. continued to pursue school discipline policy reforms with the goal of reducing the use of "zero-tolerance" as an organizing framework. Citing both high levels of and racial disparities in rates of exclusionary discipline – and particularly school suspensions – each of the ten largest school districts in the U.S. enacted at least one major school discipline policy reform. Some examples of these reforms include banning the use of suspension for low-level infractions (e.g., New York City Schools (Associated Press, 2012), Los Angeles Unified School District (National Public Radio, 2013)), adding oversight to suspension decisions

[7]See Table 233.40 from the NCES Digest of Education Statistics. Accessed here: https://nces.ed.gov/programs/digest/d18/tables/dt18233.40.asp and https://nces.ed.gov/programs/digest/d21/tables/dt21233.40.asp.

[8]The letter cautioned that "The increasing use of disciplinary sanctions such as in-school and out-of-school suspensions, expulsions, or referrals to law enforcement authorities creates the potential for significant, negative educational and long-term outcomes, and can contribute to what has been termed the 'school to prison pipeline."' The letter can be found here: https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html.

and lengths (e.g., Hillsborough County Schools (Sokol, 2015), Orange County Public Schools (Ray, 2018)), revising student conduct codes (e.g., Chicago Public Schools (Hinze-Pifer and Sartain, 2018), Palm Beach County Schools (Ross, 2013)), and promoting alternatives to out-of-school suspension (e.g., Miami-Dade County Schools (Veiga, 2015), Broward County Schools (Mack, 2014)). This reversal of zero-tolerance continued until very recently, as state legislatures and school districts have again returned to policy proposals that increase the strictness of school discipline policy (Wall, 2023).

In addition to these large districts reforming school discipline policies, a small number of states and school districts are also beginning to craft policies specifically focused on students in the early grades. Texas, Connecticut, and New Jersey banned suspensions for age groups including K–2 students, except in cases of students bringing weapons or drugs to school (TX), violent behavior (CT), or gun possession (NJ). And California banned suspensions for minor misbehavior, allowing early-grade suspensions for a broader set of more serious offenses.[9]

2.2 School Discipline Reform in Charlotte-Mecklenburg Schools

Charlotte-Mecklenburg Schools (CMS) is the second-largest school district in North Carolina. Located in Mecklenburg County, which is home to the city of Charlotte and several nearby towns, the district serves close to 150,000

[9]The relevant state-level laws are Texas House Bill 674 (2017), Connecticut Public Act No. 15-96 (2015), California Assembly Bill 420 (2013), and New Jersey Senate Bill 2081/Assembly Bill 3790 (2016).

students across 180 elementary, middle, and high schools. In a statewide annual ranking of counties based on wealth, Mecklenburg typically ranks in the top two (out of 100) counties in the state (North Carolina Department of Public Instruction, 2018).

In August 2017, the Charlotte-Mecklenburg Board of Education (CMBOE) amended the district's school discipline policies with the goal of reducing the use of out-of-school suspensions for students in kindergarten through second grades (Doss Helms, 2017b). Following concerns raised about the district's suspension policies for young children at CMBOE meetings in 2015 and 2016 (Charlotte-Mecklenburg Board of Education, 2015, 2016), CMBOE formally took up the issue in early 2017. Following the failure of a proposed reform that would have entirely banned the use of out-of-school suspension for students in the early grades, CMBOE reached a compromise that imposed a "near-ban." This near-ban required the school district superintendent to approve all out-of-school suspensions for students enrolled in these grades (Gwaltney, 2017). In the remainder of the paper, we refer to this near-ban as the "school discipline policy reform" (or simply the "reform"). This measure was passed in August 2017 and took effect weeks later, at the beginning of the 2017/18 school year. In prior school years, out-of-school suspension decisions would have been made by school principals without the requirement of superintendent oversight or approval. For clarity we emphasize that the reform only affected students in kindergarten through second grades and that it only applied to out-of-school suspension; elementary students enrolled in older grades (third through fifth) and in-school suspension (across all grades) were unaffected.

The school discipline reform in CMS resulted in large, immediate, and permanent reductions in out-of-school suspension among students enrolled in kindergarten through second grades. Figure 2 plots rates of out-of-school suspension, separately for students in CMS (Panel (a)) and students enrolled in other districts in North Carolina (Panel (b)). The impact of the reform is visually striking. Although trends within CMS across grades are similar in the school years preceding the reform (depicted to the left of the vertical red line), there is an immediate and large decline in out-of-school suspension for students enrolled in kindergarten through second grades in the school years following the reform (depicted to the right of the vertical line). We do not observe any declines for students enrolled in third through fifth grades. Panel (b) shows the same information for students enrolled in kindergarten through fifth grades in other public school districts in North Carolina. We do not observe declines in out-of-school suspension in kindergarten through second grades outside of CMS; instead, trends in out-of-school suspension are similar (and rising) over time for each grade.

In practice the school discipline reform in CMS resulted in reductions in out-of-school suspension that were large and meaningful. In the school years prior to the reform, the number of students enrolled in kindergarten through second grade who received at least one out-of-school suspension hovered around 925. In the school years following the reform, this number dropped by more than ten-fold: to 83 (2017/18) and 86 (2018/19) students, respectively. These suspensions resulted from the following disciplinary infractions: 40% Aggressive Behavior, 10% Disruptive Behavior, and 8% Insubordination

(2017/18), and 40% were Aggressive Behavior, 11% were Disruptive Behavior, 7% were Fights, and 7% were Assault on School Personnel Not Resulting in Serious Injury (2018/19).[10] Even though these raw declines do not account for CMS-specific trends in out-of-school suspension use nor for relative trends in K-2 versus 3-5 grade out-of-school suspensions elsewhere in North Carolina, we note the large and visually apparent increases here and defer these statistical adjustments to later in the paper.

3 Data

We obtained student-level data on the universe of elementary school (kindergarten through fifth grade) students enrolled in traditional public schools in North Carolina during the 2013/14-2018/19 school years.[11] These administrative data contain information on student characteristics (race/ethnicity, sex, special education status, and economic disadvantage), enrollment (school, district, and grade level), reported disciplinary infractions and their consequences (e.g., in- and out-of-school suspensions), and reading test scores. Reading test scores for kindergarten through second grade students came from the Dynamic

[10]The remainder of suspensions resulted from disciplinary infractions in the following categories: Fight; Assault on School Personnel Not Resulting in Serious Injury; Inappropriate Language / Disrespect; Inappropriate Behavior; Inappropriate Items on School Property; Communicating Threats; Leaving Class Without Permission; Threat of Physical Attack with a Weapon; Indecent Exposure; Assault on Student w/o Weapon & Not Resulting in Serious Injury; Threat of Physical Attack without a Weapon; False Fire Alarm; Theft; Harassment - Sexual; Assault on Student; Bullying; Unacceptable Behavior (Other); Violent Assault Not Resulting in Serious Injury.

[11]These data were obtained from the North Carolina Education Research Data Center (NCERDC). For more information about these data, see https://childandfamilypolicy.duke.edu/north-carolina-education-research-data/.

Indicators of Basic Early Literacy Skills (DIBELS),[12] while reading test scores for third through fifth grade students came from the North Carolina End-of-Grade (EOG) tests.[13] All reading test scores were normalized separately by grade, subject, and school year. We do not examine math test scores in this paper because math tests were not administered to kindergarten through second grade students during our sample period.

Directly relevant to this paper are the student-level data containing the universe of reported disciplinary infractions and their consequences. For each reported disciplinary infraction, we observe infraction type (selected by school administrators from a pre-populated menu of options) and associated consequences (if any), including whether the reported infraction resulted in an in-school or out-of-school suspension. In this paper, we consider several outcomes derived from these data. First, we consider the outcome of whether a student had any reported disciplinary infractions during the school year (0/1). Second, we separately consider whether a student had any reported disciplinary infractions of the six infraction types most common among kindergarten through second grade students: aggressive behavior, disruptive behavior, bus misbehavior, insubordination, inappropriate language, and fights.[14] Finally, we consider whether the student received any in-school (0/1) or out-of-school (0/1)

[12]For more information regarding DIBELS, see: https://dibels.uoregon.edu/.

[13]In Appendix Figure B2 we present percentile-percentile plots relating students' third grade DIBELS reading scores (not used in our analysis) to End-of-Grade (EOG) test scores in math and reading in third through eighth grades. The relationship between the two exams is remarkably strong and stable, particularly in reading.

[14]We identified these six infraction types based on pre-reform (2013/14-2016/17) data. The six most common types among kindergarten through second grade students were: Aggressive Behavior (31%), Disruptive Behavior (22%), Bus Misbehavior (10%), Insubordination (5%), Inappropriate Language/Disrespect (4%), and Fighting (3%).

suspensions during the school year. In robustness checks, we also consider these outcomes along the intensive margin (counts).

Our analytic sample contains all elementary school students enrolled in traditional public schools in North Carolina during the 2013/14-2018/19 school years. In our main analysis, however, we remove third grade students from our analytic sample to avoid crossover between treatment (kindergarten through second grades) and comparison (third through fifth grades) grades. For example, treated students enrolled in second grade during the 2017/18 school year (post-reform) would crossover into comparison (untreated) grades in 2018/19 (post-reform). Failing to remove these students could bias our estimates toward zero by dampening the treatment/comparison contrast; for completeness we explore this in robustness checks by re-estimating our models on the full sample of students enrolled in kindergarten through fifth grades.

Table 1 reports summary statistics for kindergarten through fifth grade students in 2013/14-2016/17 (the pre-reform school years), separately for treated grades (kindergarten through second grades) and comparison grades (third through fifth grades) and by school district (CMS versus all other school districts in North Carolina). Within CMS (Columns (1) and (2)), students in kindergarten through second grades and third through fifth grades have similar demographic characteristics, and the same is true in other NC school districts (Columns (3) and (4)). CMS does differ from other NC school districts in important ways, however, with a significantly larger fraction of non-white students relative to other school districts in NC. 37.3 percent of students in kindergarten through second grades in CMS are Black, relative to 24.2 percent

of students in other districts; and 24.3 percent are Hispanic (relative to 17.9 percent outside of CMS). Kindergarten through second grade students in CMS are also less economically disadvantaged than those students in other NC districts, with 59.5 percent classified as economically disadvantaged, relative to 66.7 percent in other NC school districts.

4 Empirical Strategies

Our main empirical strategies in this paper rely on the following three contrasts. First, we compare students enrolled in treated grades (kindergarten through second) with students enrolled in older grades (third through fifth). Second, we compare Charlotte-Mecklenburg Schools (CMS) to all other school districts (114 in total) in North Carolina. And finally, we consider the school years preceding (pre-reform) and following (post-reform) the school discipline reform. We formally combine these three contrasts in a triple-differences framework using event study and difference-in-difference-in-differences (i.e., triple-differences) approaches. We estimate the following two equations:

$$\begin{aligned} Y_{irdt} = {} & \alpha_0 + \alpha_1 earlygrade_{ir} + \alpha_2 CMS_{id} + \alpha_3 Post_{it} \\ & + \gamma_1(earlygrade_{ir} \times CMS_{id}) + \gamma_2(CMS_{id} \times Post_{it}) + \gamma_3(earlygrade_{ir} \times Post_{it}) \\ & + \beta_1(earlygrade_{ir} \times CMS_{id} \times Post_{it}) + \varepsilon_{irdt} \end{aligned} \tag{1a}$$

$$
\begin{aligned}
Y_{irdt} = {} & \alpha_0 + \alpha_1 earlygrade_{ir} + \alpha_2 CMS_{id} + \alpha_3 Post_{it} \\
& + \gamma_1(earlygrade_{ir} \times CMS_{id}) + \gamma_2(CMS_{id} \times Post_{it}) + \gamma_3(earlygrade_{ir} \times Post_{it}) \\
& + \sum_{\substack{k=-4 \\ k \neq -1}}^{2} \left(\pi_k(earlygrade_{ir} \times CMS_{id} \times \mathbf{1}\left[t - T_d^* = k\right]) \right) + \varepsilon_{irdt}
\end{aligned}
\tag{1b}
$$

Y_{irdt} is an outcome for student i in grade group r in school district d in school year t. $earlygrade_{ir}$ is an indicator equal to one if student i is enrolled in kindergarten through second grades (i.e., early grades) and zero if student i is enrolled in third through fifth grades. CMS_{id} is an indicator equal to one if student i is enrolled in an elementary school in CMS and zero if student i is enrolled in an elementary school in one of the other 114 school districts in North Carolina. $Post_{it}$ is an indicator equal to one in the school years following the enactment of the school discipline reform (2017/18 and 2018/19) and zero in the school years prior (in 2013/2014 to 2016/2017). In the above equations, α_0 is a constant, $\alpha_{\{1,2,3\}}$ are coefficients on the one-way fixed effects, and $\gamma_{\{1,2,3\}}$ are coefficients on the two-way fixed effects. In Equation (1a) β_1 is the coefficient on the three-way fixed effect (i.e., three-way interaction) and thus the triple-differences estimate of the effect of the school discipline reform on student outcomes. In Equation (1b), we replace this coefficient and three-way interaction with a sequence of π_k coefficients for $k = -4, ..., 2$ ($k = -1$

omitted) interacted with the three-way interaction.[15] We define T_d^* as the year in which district d was treated.

Because our setting does not have any staggered policy adoption (i.e., the school discipline reform affected all elementary schools in CMS at the same time), the event-time indicators correspond to specific school years (i.e., $k = -4$ is the 2013/14 school year, $k = -3$ is the 2014/15 school year, and etc.). In the event study plots that follow, $k = 0$ is the 2017/18 school year, which is the first "treated" school year in which the school discipline reform took effect. Coefficients on the sequence of negative event-time indicators provide evidence in support of the parallel trends assumption, while coefficients on the non-negative event-time indicators provide insight into treatment dynamics following the school discipline policy change. ε_{irdt} is the residual.

We begin with the basic triple-differences estimation outlined above in Equation (1a) and then sequentially enrich the specification by replacing binary indicators (and their interactions) with more refined and flexible sets of fixed effects. First, we replace $Post_{it}$ with year fixed-effects and add student-level covariates (gender, race/ethnicity, economic disadvantage). Second, we replace $early_{ir}$ with grade fixed-effects and CMS_{id} with school fixed-effects. Finally, we replace school and grade fixed effects with school-by-grade fixed effects. In our main specifications, we report heteroskedasticity-robust standard errors clustered at the school level.

[15] We only consider two post-reform school years because available administrative data from the 2019/20 school year (and beyond) is not comparable to earlier years due to school closures and the COVID-19 pandemic.

5 Main Results

5.1 Effects of the Reform on Out-of-School Suspension

Panel (a) of Table 2 presents triple-difference estimates for the binary outcome of at least one out-of-school suspension. Column (1) presents an estimate of β_1 from Equation (1a). Columns (2)-(4) present estimates from specifications subsequently augmented with more flexible year, school, grade, and school-by-grade fixed effects, respectively. Estimates across Columns (1)-(4) are remarkably similar to one another. Our preferred estimate in Column (4) indicates that the likelihood a kindergarten through second grade student in CMS received at least one out-of-school suspension decreased by 1.4 percentage points following the reform. This effect translates into a 56 percent reduction relative to the baseline mean of 2.5 percent. In Appendix Tables we report similar results for the full sample of kindergarten through fifth grade students (Appendix Table B1), standard errors clustered at the district level (Appendix Table B2), and two-stage triple-differences estimates (Appendix Table B3).[16]. In Panel (A) of Appendix Table B4 we report estimates for the intensive margin outcome of the number of out-of-school suspensions (count), where we find

[16]Even though our setting is not characterized by staggered adoption, as a robustness check on our main results we implement an approach most closely related to Gardner (2022) and produce triple-differences estimates using a two-stage procedure. This two-stage approach addresses concerns regarding the possibility of bias introduced by the estimation of fixed effects using data on eventually-treated units. In the first stage, we regress outcomes of interest on all relevant fixed effects and covariates specified in Equation 1a, except for the three-way interaction of $earlygrade_{ir} \times CMS_{id} \times Post_{it}$, using a sample of only untreated observations. We then obtain residualized outcomes (net of fixed effects and covariates estimated in the first stage) for all units and regress these residualized outcomes on the three-way interaction $earlygrade_{ir} \times CMS_{id} \times Post_{it}$ of interest. The point estimates from this approach are nearly identical to our main results. We report ninety and ninety-five percent confidence intervals generated using a bootstrap procedure.

that the reform resulted in 0.035 fewer out-of-school suspensions per student per year (70 percent decline relative to baseline of 0.050 suspensions per year).

Panel (a) of Figure 3 plots event-time coefficients from triple-differences event study Equation (1b). We note that in our setting with no staggered policy adoption, the event-time coefficients map onto specific school years (e.g., $k = 0$ is the 2017/18 school year). Coefficient estimates corresponding to school years before the reform are all very close to zero and statistically insignificant, while coefficient estimates for school years following the reform are negative and statistically significant.

Our findings contrast with previous literature on the topic of school discipline policy reform. Lacoe and Steinberg (2018) concluded that a policy designed to reduce suspension for nonviolent disciplinary issues in the School District of Philadelphia had little effect on suspension usage in the district. Wang (2022) investigated the impacts of a suspension ban for "willful defiance" across four California school districts and found that although suspension for this specific offense decreased, this effect was canceled out by increased suspensions for other types of offenses. We emphasize two key points about these previous studies: first, these studies consider school discipline policy changes that are far less stringent than the one we study. Namely, these policy changes only affect specific types of suspensions. Second, these estimates were generated using data primarily focused on older students.[17] In contrast, our results indicate that the reform students resulted in a meaningful, large, and perma-

[17]We note that Lacoe and Steinberg (2018) use district-level data encompassing of students enrolled in kindergarten through twelfth grades, though these data are likely to be dominated by disciplinary outcomes for older students.

nent decrease in out-of-school suspension for early-grade students.

5.2 Effects of the Reform on In-School Suspension and Special Education

To explore the possibility that out-of-school suspension was replaced by other forms of student-time spent outside the classroom, we investigated the effects of the reform on in-school suspension (unaffected by school discipline reform) and special education.[18][19] Our triple-differences estimates in Panels (B) and (C) of Table 2 are small in magnitude and statistically insignificant. In Column (4) of Panel (B) we estimate a precise null effect on in-school suspension. The upper bound of the ninety-five percent confidence interval (0.005) entirely rules out the possibility of one-for-one replacement of out-of-school suspension with in-school suspension and—most unfavorably—suggests replacement on the order of less than one-third.[20] Our results for special education are similarly small and statistically insignificant. Panels (b) and (c) of Figure 3 present event study plots for the same outcomes. Although both panels show very slight evidence of positive pre-trends—which would bias our estimated effects of the reform away from zero—we do not see any visual evidence suggesting

[18] We code special education (0/1) based on whether we observe an Individualized Education Plan (IEP) for student i in year t. Although we do not observe time spent outside of the classroom in our data, recent estimates from National Center for Education Statistics (2018) indicate that only 64 percent of students with disabilities spend more than 80 percent of their time in the classroom. 18 percent spend between 40-79 percent of their time in the classroom, and 13.1 percent spend less than 40 percent of their time in the classroom (the balance of students with disabilities spend time in environments outside of schools).

[19] We present raw plots for these outcomes in Appendix Figures B3 and B4.

[20] We present similar results for the count of in-school suspensions in Panel (B) of Appendix Table B4.

that the reform led to increases in either of these outcomes. Our findings again contrast with those of Lacoe and Steinberg (2018) and Wang (2022), where alternative forms of punishment (i.e., the replacement of banned suspension types with other forms of suspension) increased to offset declines induced by the policy changes.

5.3 Effects of the Reform on Academic Performance

In Panel (D) of Table 2 we present triple-differences estimates regarding the effect of the reform on reading achievement.[21] Our preferred estimate in Column (4) is 0.014 SDs. Although the effect is statistically insignificant, our ninety-five percent confidence interval (−0.015 to 0.043 SDs) in narrow enough to rule out the majority of negative effects and the single positive effect suggested by previous work. The estimates from Pope and Zuo (2023) suggest that suspension reductions of the magnitude we detect would lead to 0.016 SD decrease in English test scores. Lacoe and Steinberg (2019) estimate direct effects of out-of-school suspension of −0.014 SDs in English, although they find no relationship between out-of-school suspensions and test scores for low-level offenses (which is most comparable to our context). We note that both of these papers focus exclusively on students enrolled in third grade and above. Craig and Martin (2023) find that a suspension reform in New York City that eliminated the use of suspension for low-level offenses led to an *increase* in English test scores of 0.03 SDs. Once again this estimate comes from a middle school context, where baseline rates of suspension for low-level offenses are

[21] We present a raw plot for this outcome in Appendix Figure B5.

twice as high as the rate from our context. Panel (d) of Figure 3 depicts associated event study coefficients, which suggest that trends in reading achievement were mostly flat. We note the exception of a small negative shock to reading achievement the school year before the reform, which we believe would bias our estimate away from zero.

5.4 Effects on Disciplinary Infractions

In Panel (E) of Table 2 and Panel (e) of Figure 3 we present triple-differences and event-study estimates of the effect of the reform on the binary outcome of at least one disciplinary infraction.[22] We note that disciplinary infractions are reported independent of consequences (such as in-school or out-of-school suspension). Across Columns (1)-(4) in our table, the point estimates are small and statistically insignificant. The upper bound of our ninety-five percent confidence interval is 0.008 and rules out the large positive effects on disciplinary infractions and other measures of student behavior suggested by previous work. Lacoe and Steinberg (2018) found that a district-level policy designed to reduce suspension for nonviolent offenses in the School District of Philadelphia actually led to more incidents of serious misconduct and increased truancy. Pope and Zuo (2023) found that increased suspension rates (i.e., harsher school discipline policy) in Los Angeles Unified School District led to fewer absences and higher grade point averages (GPAs).

To further investigate this issue, in Table 3 we focused on six types of disciplinary infractions (i.e., the most common types among kindergarten through

[22] We present a raw plot related to this outcome in Appendix Figure B6.

second grade students): aggressive behavior, disruptive behavior, bus misbehavior, insubordination, inappropriate language, and fights. One concern about results related to disciplinary infractions is that they may be contaminated by changes in incentives for teachers, school staff, and administrators to report student misbehavior. If teachers and principals believe that the possibility of an out-of-school suspension is unlikely following the reform, they may choose not to formally report student misbehavior. We believe that this concern, although valid, is least likely to affect fights (Panel (F)) because such reports are required by North Carolina law (State Board of Education, 2014). Our conclusion from these results is that the reform did not lead to any meaningful increases in disciplinary infractions. In fact, most of our point estimates are actually negative, and in the cases of aggressive behavior and insubordination our results are negative and statistically significant. For completeness we report the same results for counts of reported disciplinary infractions overall and by type in Appendix Table B5. Qualitatively these results are similar to our main findings. Taken together our results for specific disciplinary infractions suggest that the deterrent effect associated with the threat of out-of-school suspension is likely to be small to nonexistent in this context, and further, that we can reject the hypothesis that limiting early-grade out-of-school suspensions worsened student-level disciplinary outcomes.

5.5 Heterogeneity by Predicted Student- and Classroom-Level Risk

To investigate the possibility of treatment heterogeneity across students and classrooms, we split our sample based on predicted student- and classroom-level risk that we constructed as follows. First, we predicted student-level risk for out-of-school suspension (0/1) using data from school years preceding the school discipline reform (i.e., 2013/14-2016/17). We estimated our student-level prediction equation as follows:

$$\Pr(Y_i = 1|\mathbf{X_i}) = \alpha + \boldsymbol{\beta}\mathbf{X_i} + \varepsilon_i \tag{2}$$

Y_i is a binary (0/1) variable indicating that student i received at least one out-of-school suspension during the school year. $\mathbf{X_i}$ is a vector of interacted indicators for race/ethnicity (5 categories), sex (2 categories), economic disadvantage (2 categories), grade level (6 categories), month of birth (12 categories), and first-observed DIBELS test score percentile (divided into 25 bins so that each roughly corresponds to 4 percentile points).[23] We then used the estimated parameters from this equation to predict $\hat{P}_i$ for all students in all years (including students who are "out-of-sample" in school years following the reform). We defined a student as "high-risk" if the student's predicted out-of-school suspension risk exceeded the grade-specific median, and "low-risk" otherwise.

[23]For the majority of our sample, the first-observed DIBELS test score corresponds to the DIBELS exam administered during the first week of kindergarten. If we observed the student entering North Carolina public schools later, we used the score corresponding to the student's first-ever instance of the exam.

Second, we predicted classroom-level risk for out-of-school suspension (0/1) with the same data from pre-reform school years and the following equation:

$$\Pr(Y_i = 1|\mathbf{G_i}) = \alpha + \boldsymbol{\omega}\mathbf{G_i} + \varepsilon_i \tag{3}$$

Y_i is a binary (0/1) variable indicating that student i received at least one out-of-school suspension during the school year. $\mathbf{G_i}$ was a vector of bins for pairwise combinations of school-by-grade. Using the estimated parameters from this equation, we predicted $\hat{C}_i$ for all students. We defined a classroom (i.e., school-by-grade cell) as "high-risk" if the classroom's predicted out-of-school suspension risk exceeded the grade-specific median, and "low-risk" otherwise.

In Table 4 we report results investigating treatment heterogeneity for the following four groups: high-risk students versus low-risk students, and high-risk classrooms versus low-risk classrooms. Column (1) reproduces our triple-differences estimate for the full sample, and Columns (2)-(5) present estimates for our four risk groups. In Panel (A) we report triple-differences estimates of the effect of reform on the likelihood of receiving at least one out-of-school suspension. Our estimates in Column (2) and (3) indicate that the likelihood of out-of-school suspension decreased by 3.9 percentage points for high-risk students and by 0.3 percentage points for low-risk students. Our estimates in Column (4) and (5) indicate that the likelihood of out-of-school suspension decreased by 2.5 percentage points for students in high-risk classrooms and by 1.3 percentage points for students in low-risk classrooms. Our findings for substitution to in-school suspension (Panel (B)) and special education placement

(Panel (C)) mirror our main results; we do not detect any statistically significant effects along these margins. Our results for reading achievement in Panel (D) suggest that high-risk students benefited from the reform—namely, we find that reading test scores increased by 0.037 SDs—although we cannot identify the exact mechanism underlying this test score increase. High-risk students could have benefited from more time in the classroom in the absence of out-of-school suspension or changes in teachers' behavior management strategies induced by the reform. Panel (E) reports results for the outcome of disciplinary infractions, although we do not detect any statistically significant effects on this outcome in any subgroup.

We report the same results for specific disciplinary infraction types in Appendix Table B6, for student subgroups partitioned by the interaction of student- and classroom-level risk in Appendix Tables B7 and B8, and by race/ethnicity and sex subgroups in Appendix Tables B9 and B10.

5.6 Longer-Run Effects on Academic Performance

Although our sample period ends in the 2018/19 school year due to the onset of the COVID-19 pandemic, we are able to assess the effects of the reform on student outcomes in the longer-run (i.e., following students' exit from treated grades) using third grade outcomes in 2018/19. To do this, we used a sample composed of third through fifth grade students enrolled in traditional public elementary schools in North Carolina between 2013/14-2018/19. In this sample, only students enrolled in third grade during the 2018/19 school year (who were second grade students in 2017/18) were exposed to the reform in

CMS. To assess the effects of exposure to the reform on student outcomes in third grade, we modified our triple-differences and event study specifications as follows:

$$
\begin{aligned}
Y_{irdt} = {} & \kappa_0 + \kappa_1 thirdgrade_{ir} + \kappa_2 CMS_{id} + \kappa_3 Exposed_{it-1} \\
& + \omega_1(thirdgrade_{ir} \times CMS_{id}) + \omega_2(CMS_{id} \times Exposed_{it-1}) + \omega_3(thirdgrade_{ir} \times Exposed_{it-1}) \\
& + \lambda_1(thirdgrade_{ir} \times CMS_{id} \times Exposed_{it-1}) + \eta_{irdt}
\end{aligned}
\tag{4a}
$$

$$
\begin{aligned}
Y_{irdt} = {} & \kappa_0 + \kappa_1 thirdgrade_{ir} + \kappa_2 CMS_{id} + \kappa_3 Exposed_{it-1} \\
& + \omega_1(thirdgrade_{ir} \times CMS_{id}) + \omega_2(CMS_{id} \times Exposed_{it-1}) + \omega_3(thirdgrade_{ir} \times Exposed_{it-1}) \\
& + \sum_{\substack{k=-5 \\ k \neq -1}}^{0} \left(\phi_k(thirdgrade_{ir} \times CMS_{id} \times \mathbf{1}\left[t - T^*_{d+1} = k\right]) \right) + \eta_{irdt}
\end{aligned}
\tag{4b}
$$

Y_{irdt} is an outcome for student i in grade group r in school district d in school year t. $thirdgrade_{ir}$ is an indicator equal to one if student i is enrolled in the exposed grade group (defined here as third grade) and zero if student i is enrolled in fourth or fifth grade. CMS_{id} is an indicator equal to one if student i is enrolled in an elementary school in CMS and zero if student i is enrolled in an elementary school in one of the other 114 school districts in North Carolina.

$Exposed_{it-1}$ is an indicator equal to one if student i was exposed to the reform in the previous school year, $t-1$. Put differently, this indicator is equal to one in the 2018/19 school year and zero in all other school years. In the above equation, κ_0 is a constant, $\kappa_{\{1,2,3\}}$ are coefficients on the one-way fixed effects, and $\omega_{\{1,2,3\}}$ are coefficients on the two-way fixed effects in the triple-differences research design. λ_1 is the coefficient on the three-way fixed effect (i.e., three-way interaction) and thus the triple-differences estimate of the effect of the reform on student outcomes in third grade. η_{irdt} is the residual. For event study analysis, we replace the three-way fixed effect with a sequence of ϕ_k coefficients and define T^*_{d+1} as the school year following exposure to the reform in district d. Once again, we begin with the basic triple-differences specification and then sequentially enrich the specification by replacing binary indicators (and their interactions) with more refined and flexible sets of fixed effects. For all specifications, we report heteroskedasticity-robust standard errors clustered at the school level.

Table 5 presents triple-differences estimates of the effect of the reform on student outcomes in third grade. We do not find any evidence to suggest that the reform led to any changes in the likelihood of out-of-school suspension, in-school suspension, nor special education placement. In contrast, however, we find suggestive – albeit somewhat weak – evidence of increases in reading test scores. Point estimates across Columns (1)-(4) range from 0.035-0.037 standard deviations and are all statistically significant. We present complementary event study plots in Figure 4.

In addition to exploring effects for the full sample of students, we also esti-

mate effects separately by predicted student- and classroom-level risk groups. Table 6 presents triple-differences estimates of the effect of the reform on student outcomes in third grade, separately for the same four groups as previously presented in our heterogeneity analysis: high- versus low-risk students, and high- versus low-risk classrooms. The results in this table make clear that the weak evidence of increases in reading test scores observed for the full sample are entirely driven by high-risk students. These are the same students for whom we observed short-run reading impacts. We find that reading test scores increased by 0.074 SDs for high-risk students by third grade (recall that the short-run effect was 0.037 SDs). As with our earlier results, these reading test score improvements could be explained by both increased instructional time (coming through reduced out-of-school suspension) or by unobserved changes in teachers' classroom management and instructional strategies induced by the school discipline reform.

5.7 Effects on Teacher Turnover

In Table 7 we report triple-differences estimates of the effect of the reform on teacher turnover, which we define based on observing teacher assignment (to grade group within a school) between time t and $t+1$.[24] Our sample is the universe of elementary school teachers in traditional public schools in North Carolina between 2013/14-2018/19.

We define teacher turnover outcomes related to (1) exiting from public elementary school teaching, (2) leaving the school district, (3) leaving the

[24]We consider early grades (kindergarten through second grades) versus older grades (third through fifth grades).

school (but staying within the same district), (4) leaving the grade group (but staying within the same school and district), and (5) staying in current role.[25] In contrast to previous work on this topic (see e.g., Pope and Zuo (2023)), we do not find any statistically significant effects on teacher turnover outcomes.

6 Effects on Out-of-School Suspension Gaps and Implications for Inequality

As a final empirical exercise, we examine the effects of the reform on several measures of suspension inequality: namely, in gaps in out-of-school suspension rates by race/ethnicity, gender, special education status, and economic status. The existence of these gaps are well-documented across a wide variety of contexts. To estimate the effects of the reform on out-of-school suspension gaps, we collapsed our student-level data into cells defined at the district-by-grade level-by school year (e.g., a cell in our collapsed data would be kindergarten in CMS in 2013/14). Following previous literature, we then calculate the difference in the likelihood of out-of-school suspension for the following pairwise groups: Black and White, Hispanic and White, Male and Female, Special Education and No Special Education, Economically Disadvantaged and Not Economically Disadvantaged.

We present our results in Table 8 and Figure 5.[26] Results in Column (1) of Panel (A) indicate that the Black-White out-of-school suspension gap declined

[25] For clarity we emphasize that the outcomes in Panels (A)-(D) indicate some form of exit/change in teaching position/role relative to the outcome "Stay" in Panel (E).

[26] We provide companion raw plots in Appendix Figures B7-B11.

by 3.7 percentage points. In relative terms, this translates into a 79 percent reduction relative to the baseline gap of 4.7 percentage points. We do not find any evidence of statistically significant declines in the Hispanic-White out-of-school suspension gap relative to baseline (0.1 percentage points) (Column (2)). In Column (3) we find that the Male-Female out-of-school suspension gap declined by 2.0 percentage points (65 percent relative to baseline gap of 3.1 percentage points). In Columns (4) and (5) we find that the special education out-of-school suspension gap declined by 4.6 percentage points (92 percent relative to baseline gap of 5.0 percentage points) and that the economic disadvantage out-of-school suspension gap declined by 2.2 percentage points (73 percent relative to baseline gap of 3.0 percentage points).

7 Conclusion

In this paper, we investigate the effects of a reform to school discipline policy in Charlotte-Mecklenburg Schools (CMS) on the outcomes of students enrolled in early grades (K–2). Our triple-differences and event study results demonstrate that the reform reduced out-of-school suspension among students enrolled in kindergarten through second grades by 56 percent. We do not find convincing evidence of substitution to other forms of exclusion from or time-spent outside the classroom, including in-school suspension (unaffected by the reform) and special education. In the aggregate, we report null effects on test scores and student behavior (as measured by reported disciplinary infractions). These findings suggest little to no spillovers to students unaffected by the policy and

that the threat of out-of-school suspension had little to no deterrent effect on student misbehavior.

Our investigation of treatment heterogeneity reveals differential effects underlying our main results. Specifically, we find that high-risk students experienced the largest declines in out-of-school suspension and reading test score gains (0.037 SDs). Our examination of heterogeneity by classroom-level risk revealed large declines in out-of-school suspension in high-risk classrooms, but no differences in test score effects across classroom groups. These results lead us to conclude that the reform may have benefited high-risk students and that these benefits did not come at the expense of unaffected (low-risk) students.

In the long-run, we do not find evidence to suggest that the reform affected out-of-school suspension, in-school suspension, special education placement, or disciplinary infractions. We do find, however, evidence of increases in reading test scores (0.035 SDs) for affected students in third grade. We tentatively conclude that the reform may have improved academic achievement over the longer-term and more definitively reject negative effects on student achievement over the same time horizon.

Our paper contributes to the existing work on the causal effects of school discipline policy on student outcomes and provides new evidence for education policymakers debating whether and how to craft school discipline policy as it relates to students in the early grades. Our work suggests that much of the previous evidence may be irrelevant for this age group.

References

Associated Press (2012, August). Revised Rules Reduce NYC Student Suspensions. *Associated Press State Wire: New York (NY)*.

Bacher-Hicks, A., S. B. Billings, and D. J. Deming (2019). The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime. NBER Working Paper 26257, National Bureau of Economic Research, Cambridge, MA.

Barrett, N., A. McEachin, J. N. Mills, and J. Valant (2021). Disparities and Discrimination in Student Discipline by Race and Family Income. *Journal of Human Resources 56*(3), 711–748. Publisher: University of Wisconsin Press.

Charlotte-Mecklenburg Board of Education (2015, October). Meeting Minutes for the Regular Meeting of the Charlotte-Mecklenburg Board of Education. Meeting Minutes, Charlotte-Mecklenburg Board of Education, Charlotte, NC.

Charlotte-Mecklenburg Board of Education (2016, August). Meeting Minutes for the Regular Meeting of the Charlotte-Mecklenburg Board of Education. Meeting Minutes, Charlotte-Mecklenburg Board of Education, Charlotte, NC.

Craig, A. C. and D. C. Martin (2023, February). Discipline Reform, School Culture, and Student Achievement. Technical report, Annenberg Institute at Brown University. Publication Title: EdWorkingPapers.com.

Cunha, F. and J. Heckman (2007, April). The Technology of Skill Formation. *American Economic Review 97*(2), 31–47.

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

Curran, F. C. (2019, March). The Law, Policy, and Portrayal of Zero Tolerance School Discipline: Examining Prevalence and Characteristics Across Levels of Governance and School Districts. *Educational Policy 33*(2), 319–349. Publisher: SAGE Publications Inc.

Doss Helms, A. (2017a, August). Superintendent's approval needed now for CMS preK-2 suspensions. *The Charlotte Observer*, 6A.

Doss Helms, A. (2017b, August). Suspending a CMS first-grader is about to get harder. Here's why. *The Charlotte Observer*.

Fitzpatrick, M. D., D. Grissmer, and S. Hastedt (2011, April). What A Difference A Day Makes: Estimating Daily Learning Gains During Kindergarten and First Grade Using a Natural Experiment. *Economics of Education Review 30*(2), 269–279.

Gardner, J. (2022, July). Two-stage differences in differences. arXiv:2207.05943 [econ].

Gopalan, M. and A. A. Nelson (2019, April). Understanding the Racial Discipline Gap in Schools. *AERA Open 5*(2), 2332858419844613. Publisher: SAGE Publications Inc.

Gwaltney, J. (2017, July). Charlotte-Mecklenburg Schools Suspension policy still in play. *The Charlotte Observer*, 4A.

Hill, C. J., H. S. Bloom, A. R. Black, and M. W. Lipsey (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives 2*(3), 172–177. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1750-8606.2008.00061.x.

Hinze-Pifer, R. and L. Sartain (2018, March). Rethinking Universal Suspension for Severe Student Behavior. *Peabody Journal of Education 93*(2), 228–243.

Lacoe, J. and M. P. Steinberg (2018, March). Rolling Back Zero Tolerance: The Effect of Discipline Policy Reform on Suspension Usage and Student Outcomes. *Peabody Journal of Education 93*(2), 207–227. Publisher: Routledge _eprint: https://doi.org/10.1080/0161956X.2018.1435047.

Lacoe, J. and M. P. Steinberg (2019, March). Do Suspensions Affect Student Outcomes? *Educational Evaluation and Policy Analysis 41*(1), 34–62. Publisher: American Educational Research Association.

Leung-Gagné, M., J. McCombs, C. Scott, and D. J. Losen (2022). Pushed Out: Trends and Disparities in Out-of-School Suspension. Technical report, Learning Policy Institute.

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

Losen, D. J. and P. Martinez (2020). Lost Opportunities: How Disparate School Discipline Continues to Drive Differences in the Opportunity to Learn. Technical report.

Mack, S. (2014, January). How A Broward County School District Is Rethinking Zero-Tolerance Policies. *WLRN*. Section: Education.

National Center for Education Statistics (2018). Table 204.60. Percentage distribution of students 6 to 21 years old served under Individuals with Disabilities Education Act (IDEA), Part B, by educational environment and type of disability: Selected years, fall 1989 through fall 2018. Digest of Education Statistics.

National Public Radio (2013, May). LA Schools Throw Out Suspensions For 'Willful Defiance'. *All Things Considered [NPR] (USA)*.

North Carolina Department of Public Instruction (2018). Low Wealth Ranking | NC DPI.

Pope, N. G. and G. W. Zuo (2023, July). Suspending Suspensions: The Education Production Consequences of School Suspension Policies*. *The Economic Journal 133*(653), 2025–2054.

Ray, K. (2018, October). Why are fewer Orange County students being disciplined? *WFTV*.

Ritter, G. W. and K. P. Anderson (2018, March). Examining Disparities in Student Discipline: Mapping Inequities from Infractions to Consequences. *Peabody Journal of Education 93*(2), 161–173. Publisher: Routledge _eprint: https://doi.org/10.1080/0161956X.2018.1435038.

Ross, A. (2013, June). Palm Beach County school district's ongoing struggles with suspension policies hurting certain schools, student groups. *The Palm Beach Post*.

Shi, Y. and M. Zhu (2022, June). Equal time for equal crime? Racial bias in school discipline. *Economics of Education Review 88*, 102256.

Skiba, R. J. and M. K. Rausch (2006). Zero Tolerance, Suspension, and Expulsion. In *Handbook of Classroom Management*. Routledge Handbooks Online.

Sokol, M. (2015, July). Student Conduct Policies Revised. *Tampa Bay Times (FL)*, 1B.

State Board of Education (2014, September). School Violence Acts Defined and the Annual Report of these Crimes.

U.S. Government Accountability Office (2018). K-12 Education: Discipline Disparities for Black Students, Boys, and Students with Disabilities. Technical Report GAO-18-258, United States Government Accountability Office, Washington DC.

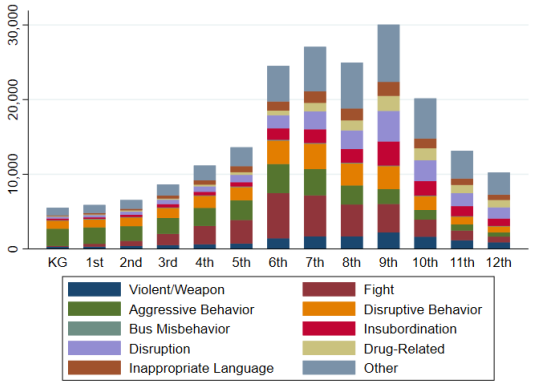
Vara-Orta, F. (2018, December). It's official: DeVos has axed Obama discipline guidelines meant to reduce suspensions of students of color. *Chalkbeat*.

Veiga, C. (2015, August). As school year starts, Miami-Dade rolls out a new approach to student discipline. *Miami Herald, The (FL)*.

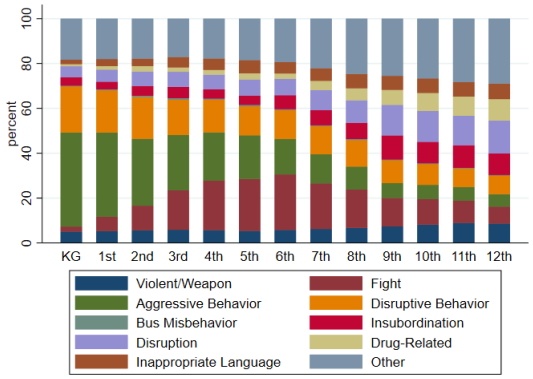
Wall, P. (2023, March). Lawmakers Across U.S. Push for Harsher School Discipline as Safety Fears Rise. *Chalkbeat*. Section: Front Page.

Wang, R. (2022, January). The Impact of Suspension Reforms on Discipline Outcomes: Evidence From California High Schools. *AERA Open 8*, 23328584211068067. Publisher: SAGE Publications Inc.

Figures and Tables



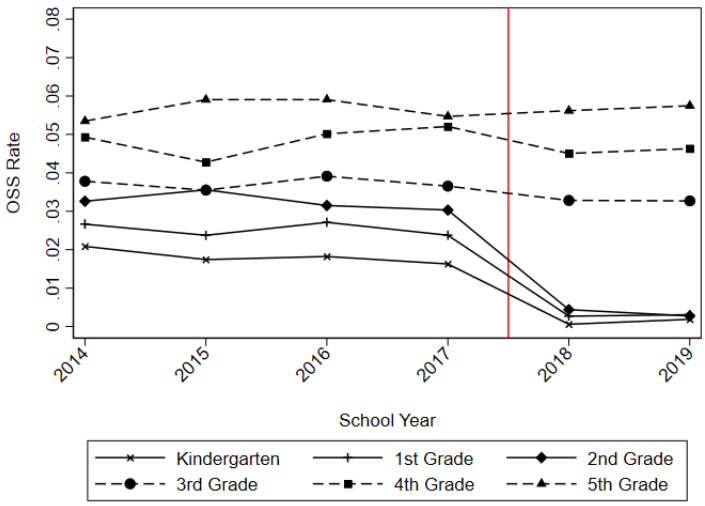
(a) Number of Out-of-School Suspensions, Separately by Grade Level



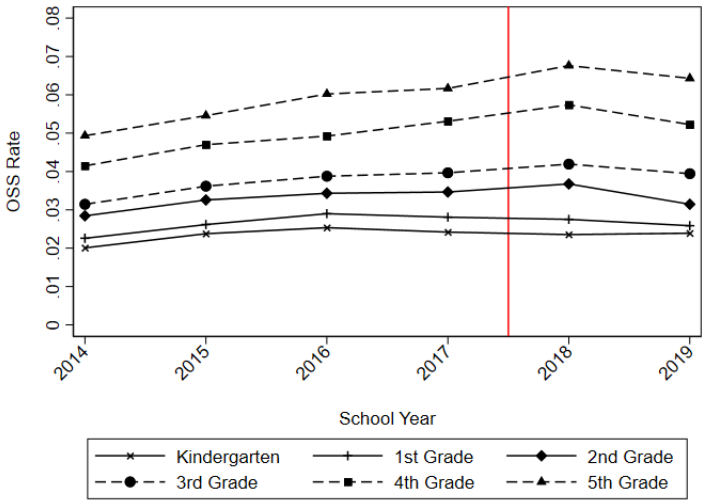
(b) Distribution of Infraction Types Resulting in Out-of-School Suspension, Separately by Grade Level

Figure 1: Out-of-School Suspension in North Carolina

Notes: Panel (a) depicts the number of out-of-school suspensions in North Carolina, separately by grade level. Panel (b) depicts the distribution of infraction types resulting in out-of-school suspension, separately by grade level. Data are from the 2018/19 school year.



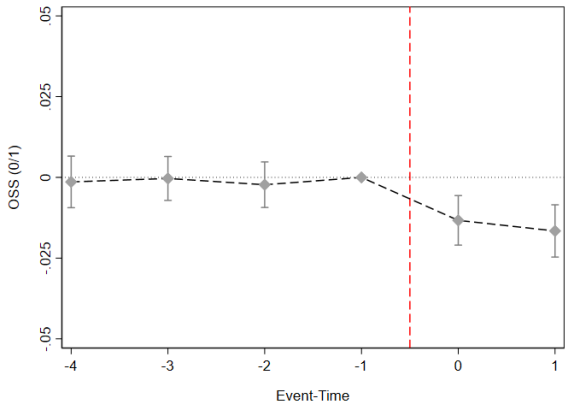
(a) Charlotte-Mecklenburg Schools



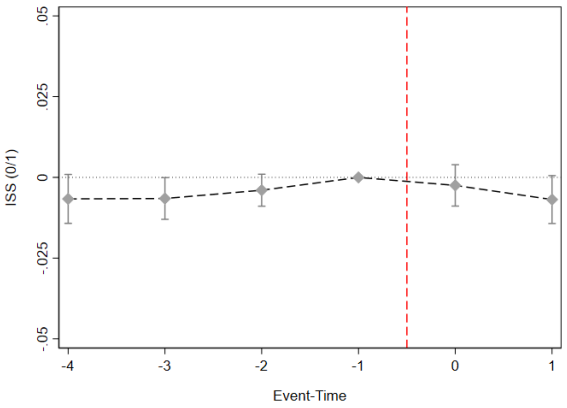
(b) Other North Carolina School Districts

Figure 2: Share of Elementary School Students With At Least One Out-of-School Suspension, Separately by Grade

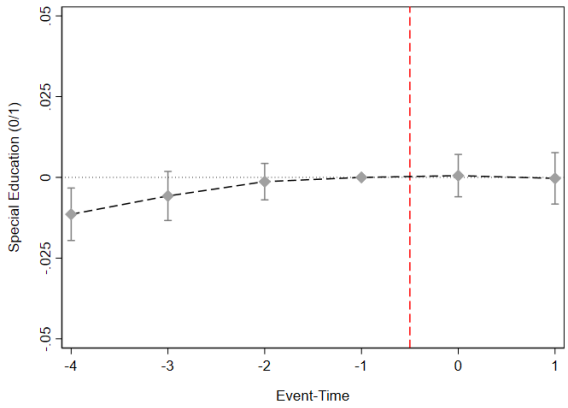
Notes: Panels (a) and (b) depict the share of elementary school students (grades K-5), separately by grade, who received at least one out-of-school suspension (OSS) in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



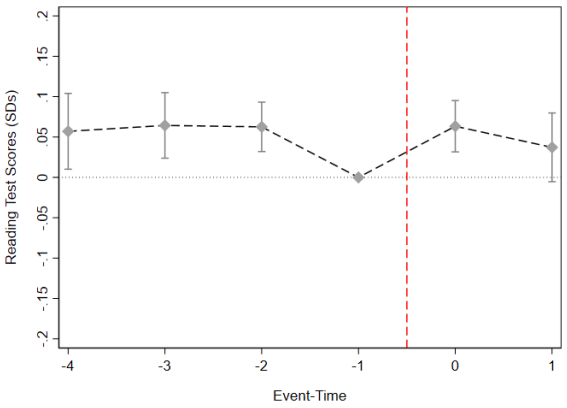
(a) Out-of-School Suspension (0/1)



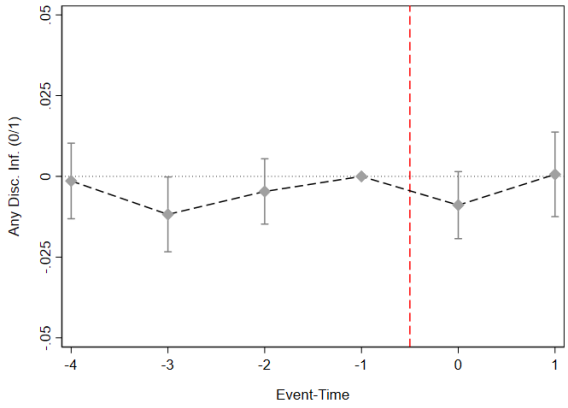
(b) In-School Suspension (0/1)



(c) Special Education (0/1)



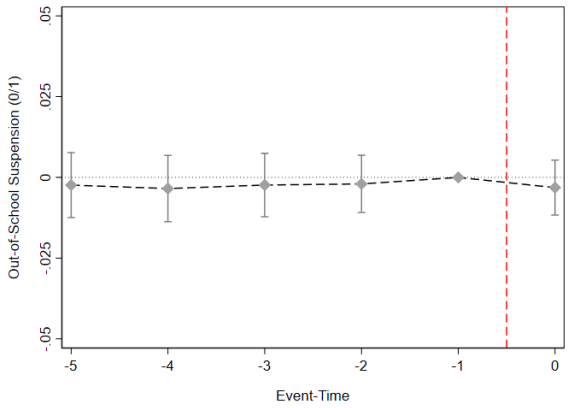
(d) Reading Achievement (SDs)



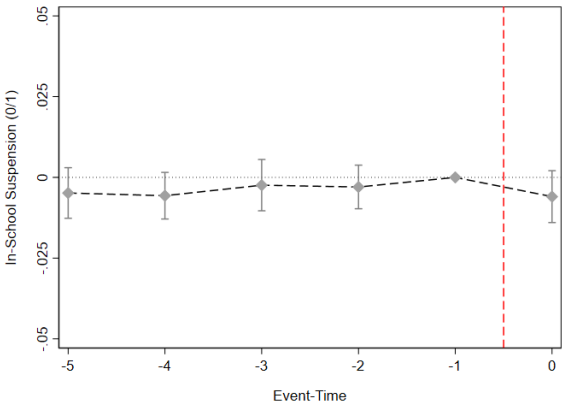
(e) Any Disciplinary Infraction (0/1)

Figure 3: Event Study Plots

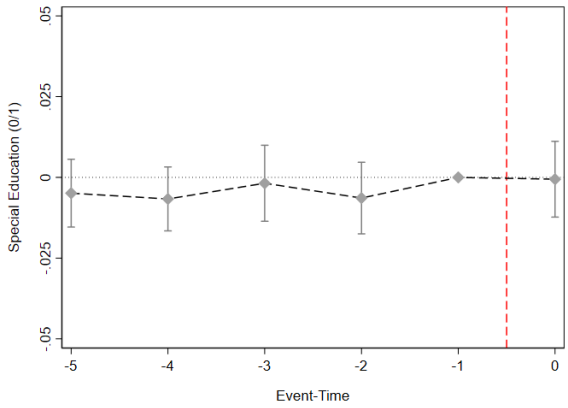
Notes: Panels (a)-(e) depict event study estimates corresponding to Equation (1b). Heteroskedasticity-robust standard errors are clustered at the school level. Event-time zero corresponds to the 2017/18 school year. Error bars represent 95 percent confidence intervals.



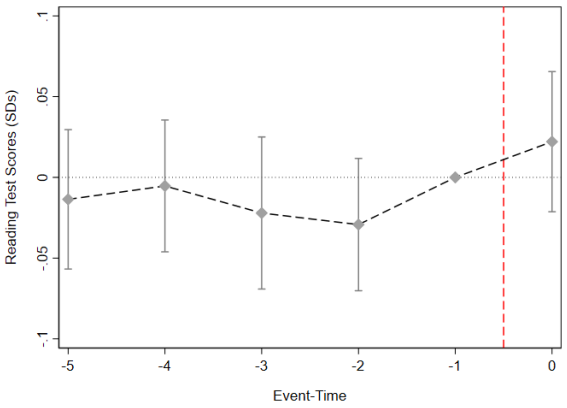
(a) Out-of-School Suspension (0/1)



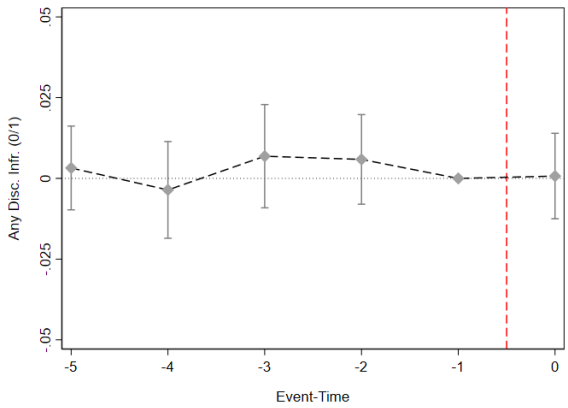
(b) In-School Suspension (0/1)



(c) Special Education (0/1)



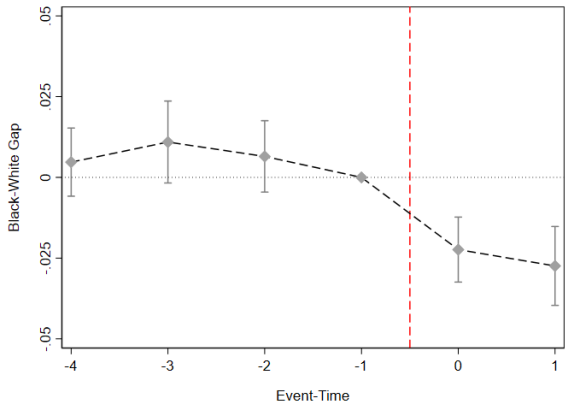
(d) Reading Achievement (SDs)



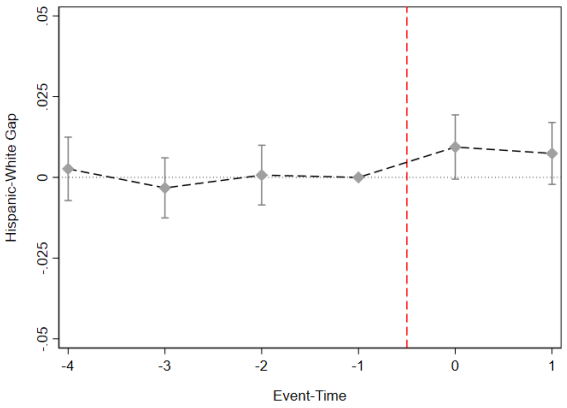
(e) Any Disciplinary Infraction (0/1)

Figure 4: Event Study Plots, Third Grade Outcomes

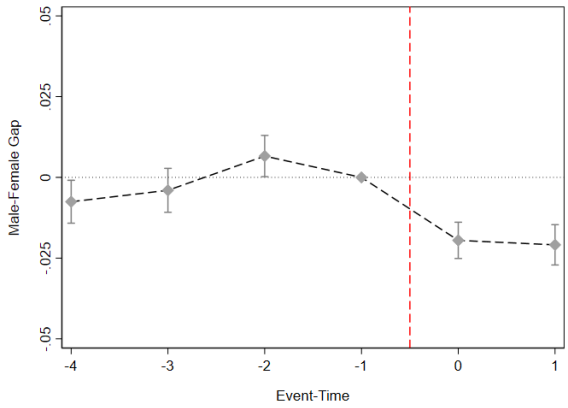
Notes: Panels (a)-(f) depict event study estimates corresponding to Equation (4b). Heteroskedasticity-robust standard errors are clustered at the school level. Event-time zero corresponds to the 2018/19 school year (i.e., the year following the initial policy change). Error bars represent 95 percent confidence intervals.



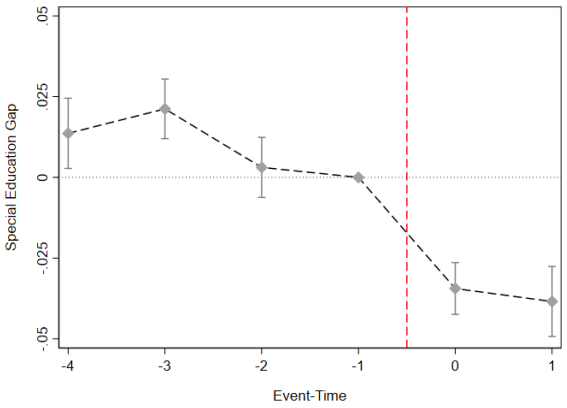
(a) Black-White Gap



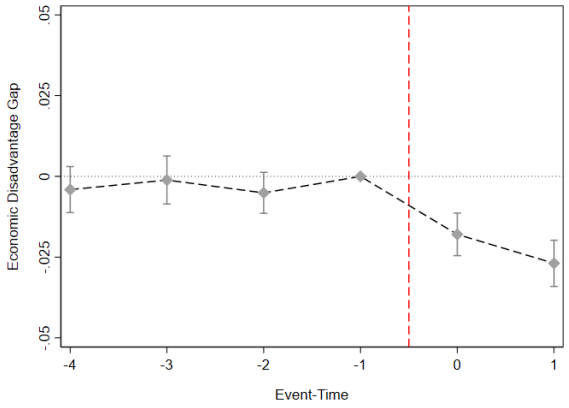
(b) Hispanic-White Gap



(c) Male-Female Gap



(d) Special Education Gap



(e) Economic Disadvantage Gap

Figure 5: Event Study Plots, Out-of-School Suspension Gaps

Notes: Panels (a)-(f) depict event study estimates and ninety-five percent confidence intervals. Heteroskedasticity-robust standard errors are clustered at the district level. Event-time zero corresponds to the 2017/18 school year. The dependent variable from each event study regression is the difference in out-of-school suspension rates at the district by grade level by school year level.

Table 1: Summary Statistics

	Charlotte-Mecklenburg		Other NC Districts	
	(1) Grades K-2	(2) Grades 3-5	(3) Grades K-2	(4) Grades 3-5
Panel A. Student Characteristics				
Female (0/1)	0.489	0.493	0.485	0.489
Asian/Pacific Islander (0/1)	0.071	0.060	0.029	0.028
Black (0/1)	0.373	0.393	0.242	0.233
Hispanic (0/1)	0.243	0.227	0.179	0.171
Other Race/Ethnicity (0/1)	0.026	0.024	0.057	0.057
White (0/1)	0.287	0.296	0.493	0.511
Economically Disadvantaged (0/1)	0.595	0.607	0.667	0.633
Panel B. Student Outcomes				
Out-of-School Suspension (0/1)	0.025	0.047	0.027	0.047
In-School Suspension (0/1)	0.010	0.023	0.016	0.032
Special Education (0/1)	0.069	0.094	0.119	0.147
Reading Test Scores (SDs)	0.115	0.025	-0.000	-0.016
Any Disciplinary Inf. (0/1)	0.074	0.117	0.075	0.112
N	141,921	135,451	1,182,747	1,157,444

Notes: Columns (1) and (2) report mean sample characteristics for students enrolled in traditional public elementary schools in Charlotte-Mecklenburg Schools (CMS) separately for early grades (K-2) and older grades (3-5). Columns (3) and (4) report mean sample characteristics for students enrolled in traditional public elementary schools in all other school districts in North Carolina separately for early grades (K-2) and older grades (3-5). Summary statistics are calculated using pooled student-level data from the 2013/14-2016/17 school years. Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively.

Table 2: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes

	(1)	(2)	(3)	(4)
Panel A. Out-of-School Suspension (0/1)	-0.014***	-0.015***	-0.015***	-0.014***
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.006	0.045	0.071	0.079
Baseline Mean	0.025	0.025	0.025	0.025
Panel B. In-School Suspension (0/1)	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.005	0.024	0.073	0.080
Baseline Mean	0.010	0.010	0.010	0.010
Panel C. Special Education (0/1)	0.005	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.005	0.030	0.041	0.042
Baseline Mean	0.069	0.069	0.069	0.069
Panel D. Reading Test Scores (SDs)	0.004	0.008	0.005	0.014
	(0.018)	(0.016)	(0.015)	(0.015)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.126	0.164	0.179
Baseline Mean	0.115	0.115	0.115	0.115
Panel E. Any Disc. Inf. (0/1)	-0.000	-0.002	-0.001	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.008	0.071	0.115	0.121
Baseline Mean	0.074	0.074	0.074	0.074
Year FE		X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table 3: Triple-Differences Estimates of the Effect of the Suspension Reform on Reported Disciplinary Infractions

	(1)	(2)	(3)	(4)
Panel A. Aggressive Behavior (0/1)	-0.009***	-0.010***	-0.010***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.027	0.050	0.053
Baseline Mean	0.036	0.036	0.036	0.036
Panel B. Disruptive Behavior (0/1)	-0.002	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.027	0.055	0.060
Baseline Mean	0.022	0.022	0.022	0.022
Panel C. Bus Misbehavior (0/1)	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.016	0.054	0.057
Baseline Mean	0.014	0.014	0.014	0.014
Panel D. Insubordination (0/1)	-0.003**	-0.004**	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.012	0.033	0.037
Baseline Mean	0.010	0.010	0.010	0.010
Panel E. Inappropriate Language (0/1)	-0.003*	-0.004*	-0.003*	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.003	0.013	0.025	0.029
Baseline Mean	0.008	0.008	0.008	0.008
Panel F. Fight (0/1)	0.003	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.006	0.025	0.043	0.052
Baseline Mean	0.009	0.009	0.009	0.009
Year FE	X	X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(F) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to disciplinary infractions indicate at least one disciplinary infraction (0/1) of the type indicated. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table 4: Heterogeneous Treatment Effects by Predicted Student- and Classroom-Risk

	(1) Full Sample	(2) High-Risk Student	(3) Low-Risk Student	(4) High-Risk Classroom	(5) Low-Risk Classroom
Panel A. Out-of-School Suspension (0/1)	-0.014***	-0.039***	-0.003***	-0.025***	-0.013***
	(0.003)	(0.005)	(0.001)	(0.005)	(0.003)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.025	0.057	0.002	0.050	0.010
Panel B. In-School Suspension (0/1)	-0.001	-0.003	-0.001	0.003	-0.006
	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.010	0.020	0.002	0.018	0.004
Panel C. Special Education (0/1)	0.004	-0.002	0.001	0.002	0.006
	(0.003)	(0.005)	(0.004)	(0.004)	(0.005)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.069	0.101	0.045	0.076	0.064
Panel D. Reading Test Scores (SDs)	0.014	0.037**	0.003	0.022	0.003
	(0.015)	(0.019)	(0.015)	(0.024)	(0.020)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.115	-0.215	0.358	-0.144	0.280
Panel E. Any Disc. Infr. (0/1)	-0.002	-0.011	-0.002	-0.003	-0.005
	(0.005)	(0.007)	(0.003)	(0.008)	(0.008)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.074	0.137	0.028	0.116	0.048

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Students are divided into groups based on predicted student- and classroom-level risk. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table 5: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes in Third Grade

	(1)	(2)	(3)	(4)
Panel A. Out-of-School Suspension (0/1)	-0.001	-0.001	-0.001	-0.001
	(0.004)	(0.003)	(0.004)	(0.003)
N	1,959,252	1,959,252	1,959,252	1,959,252
R-Squared	0.001	0.054	0.088	0.090
Baseline Mean	0.036	0.036	0.036	0.036
Panel B. In-School Suspension (0/1)	-0.003	-0.003	-0.004	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
N	1,959,252	1,959,252	1,959,252	1,959,252
R-Squared	0.001	0.028	0.084	0.086
Baseline Mean	0.019	0.019	0.019	0.019
Panel C. Special Education (0/1)	0.003	0.003	0.002	0.001
	(0.005)	(0.005)	(0.005)	(0.005)
N	1,959,252	1,959,252	1,959,252	1,959,252
R-Squared	0.003	0.031	0.042	0.041
Baseline Mean	0.092	0.092	0.092	0.092
Panel D. Reading Test Scores (SDs)	0.036**	0.036**	0.037**	0.035**
	(0.017)	(0.015)	(0.015)	(0.015)
N	1,959,252	1,959,252	1,959,252	1,959,252
R-Squared	0.000	0.180	0.224	0.225
Baseline Mean	0.025	0.025	0.025	0.025
Panel E. Any Disc. Infr. (0/1)	-0.002	-0.002	-0.002	-0.001
	(0.006)	(0.005)	(0.005)	(0.005)
N	1,959,252	1,959,252	1,959,252	1,959,252
R-Squared	0.002	0.084	0.133	0.135
Baseline Mean	0.103	0.103	0.103	0.103
Year FE		X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (4a). The sample contains all third through fifth grade students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years. Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table 6: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes in Third Grade, Separately by Predicted Student- and Classroom-Risk

	(1) Full Sample	(2) High-Risk Student	(3) Low-Risk Student	(4) High-Risk Classroom	(5) Low-Risk Classroom
Panel A. Out-of-School Suspension (0/1)	-0.001	-0.005	0.003*	0.002	-0.006
	(0.003)	(0.006)	(0.002)	(0.005)	(0.005)
N	1,959,252	1,096,686	860,755	1,149,798	801,783
Baseline Mean	0.037	0.073	0.003	0.061	0.010
Panel B. In-School Suspension (0/1)	-0.002	-0.003	-0.002	-0.003	-0.003
	(0.003)	(0.005)	(0.002)	(0.005)	(0.005)
N	1,959,252	1,096,686	860,755	1,149,798	801,783
Baseline Mean	0.017	0.031	0.003	0.026	0.007
Panel C. Special Education (0/1)	0.001	-0.007	0.004	-0.005	0.008
	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)
N	1,959,252	1,096,686	860,755	1,149,798	801,783
Baseline Mean	0.089	0.122	0.056	0.094	0.082
Panel D. Reading Test Scores (SDs)	0.035**	0.074***	0.004	0.035*	0.030
	(0.015)	(0.021)	(0.019)	(0.021)	(0.027)
N	1,959,252	1,096,686	860,755	1,149,798	801,783
Baseline Mean	0.029	-0.338	0.387	-0.245	0.340
Panel E. Any Disc. Infr. (0/1)	-0.001	-0.004	-0.001	-0.001	-0.003
	(0.005)	(0.009)	(0.004)	(0.009)	(0.010)
N	1,959,252	1,096,686	860,755	1,149,798	801,783
Baseline Mean	0.102	0.176	0.030	0.146	0.052

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (4a) for four student- and classroom-level risk groups. The sample contains all third through fifth grade students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years. Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table 7: Triple-Differences Estimates of the Effect of the Suspension Reform on Teacher Turnover

	(1)	(2)	(3)	(4)
Panel A. Leave Public NC Elem. Teaching (0/1)	-0.003	-0.002	-0.001	0.001
	(0.015)	(0.015)	(0.015)	(0.015)
N	207,283	207,283	207,283	207,283
R-Squared	0.003	0.005	0.008	0.010
Baseline Mean	0.166	0.166	0.166	0.166
Panel B. Leave District (0/1)	-0.002	-0.001	-0.001	-0.001
	(0.006)	(0.006)	(0.006)	(0.006)
N	207,283	207,283	207,283	207,283
R-Squared	0.001	0.006	0.008	0.007
Baseline Mean	0.026	0.026	0.026	0.026
Panel C. Change School Within District (0/1)	0.010	0.010	0.012	0.013
	(0.016)	(0.016)	(0.016)	(0.016)
N	207,283	207,283	207,283	207,283
R-Squared	0.002	0.002	0.009	0.010
Baseline Mean	0.058	0.058	0.058	0.058
Panel D. Change Grade Group Within School (0/1)	0.015	0.016	0.016	0.015
	(0.011)	(0.011)	(0.011)	(0.010)
N	207,283	207,283	207,283	207,283
R-Squared	0.000	0.004	0.021	0.023
Baseline Mean	0.075	0.075	0.075	0.075
Panel E. Stay (0/1)	-0.021	-0.023	-0.026	-0.028
	(0.019)	(0.019)	(0.019)	(0.019)
N	207,283	207,283	207,283	207,283
R-Squared	0.002	0.009	0.021	0.022
Baseline Mean	0.675	0.675	0.675	0.675
Year FE		X	X	X
Teacher Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all full-time, K-5 teachers from traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years. Teacher turnover (0/1) is defined as leaving (or staying in) the location indicated between time t and $t+1$. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table 8: Triple-Differences Estimates of the Effect of the Suspension Reform on Out-of-School Suspension Gaps

	(1) Black-White Gap	(2) Hispanic-White Gap	(3) Male-Female Gap	(4) Special Education Gap	(5) Economic Disadvantage Gap
Panel A. Full Sample	-0.037***	0.002	-0.020***	-0.046***	-0.022***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
N	3,956	4,111	4,140	4,140	4,137
R-Squared	0.139	0.010	0.128	0.073	0.183
Baseline Mean	0.047	0.001	0.031	0.050	0.030
Panel B. Restricted Sample	-0.038***	0.001	-0.021***	-0.046***	-0.023***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)
N	2,130	2,148	3,648	2,100	2,955
R-Squared	0.198	0.022	0.135	0.113	0.209
Baseline Mean	0.047	0.001	0.031	0.050	0.030

Notes: The sample in Panel (A) contains suspension gaps calculated as the difference in suspension rates between groups in each cell at the district by grade by school year level. The sample in Panel (B) is restricted to district-grade-year cells with at least fifty students in each group. Heteroskedasticity-robust standard errors are clustered at the district level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Appendix A: Estimating the Total Number of Kindergarten Through Second Grade Suspensions in the United States

We estimate that 165,000 K–2nd grade students in U.S. public schools received out-of-school suspensions during the 2017–2018 school year. We relied on the 2017–2018 school year for data because, in our attempts to track down state-level data on general suspension patterns and enrollment information, we found data for all states during that school year. Also, the policy change discussed in this paper occurred during the 2017–2018 school year.

We began by collecting bi-annual data reports from the Office of Civil Rights (OCR) listing state-level counts of grade K–12 students who are given out-of-school suspensions lasting at least one day. We then identified 17 states that report the number of students given out-of-school suspensions in grades K–2. For those 17 states, we calculated the (student) population-weighted fraction of students suspended in grades K–12 who were in grades K–2. We used this fraction to impute the number of grade K–2 suspensions in the other 34 states and D.C. To check the quality of this estimate, we compare the estimated number of students suspended in grades K–2 and the actual number of students suspended for the 17 states that reported disaggregated grade-level data, and we find a correlation coefficient of 0.94 between the estimated and reported values.

We implemented the above procedure with four exceptions; Texas,[27] Connecticut,[28] California,[29] and New Jersey[30] because they passed laws prior to the 2017-2018 school year limiting out-of-school suspensions for early-grade students. Texas, Connecticut, and New Jersey banned suspensions for age groups that included K–2 students, except in cases where students brought weapons or drugs to school, engaged in violent behavior, or were involved in a gun possession incident (respectively). We treated these three states separately from the other 47 states and DC since their laws led to a deviation from naturally occurring suspension rates. TX and NJ reported the number of K–2 students in their states who were suspended, so we used those numbers directly. Due to the similar law change, we used the fraction of K–2 students suspended in TX and NJ to impute estimates for CT.

California only banned suspensions for the most minor incidents. So, while

[27] Texas House Bill 674, 2017

[28] Connecticut Public Act No. 15-96, 2015

[29] California Assembly Bill 420, 2013

[30] New Jersey Senate Bill 2081/Assembly Bill 3790, 2016

we treated TX, CT, and NJ distinctly, we included CA in the pool with the rest of the states. We validated this decision by comparing CA's report of 17,619 (the unduplicated number of K-3 students suspended) to the 16,420 students we estimated had been suspended using the imputation logic described above.

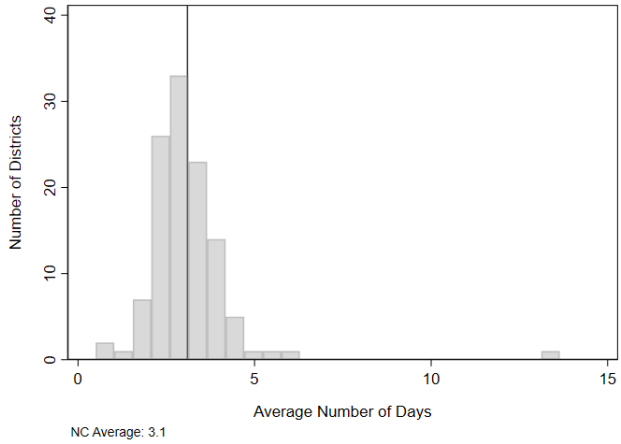
A few other states required special consideration:

1) Alaska:
   Anchorage School District was the only area to report disaggregated suspension counts by grade. They reported the total enrollment, number of students suspended, and suspension rate allowing for the calculation of K–2 enrollment and total number of students suspended. Alaska's statewide enrollment, total K–12 enrollment and K–2 enrollment as well as the total number of students suspended were also reported by the state department of education and OCR data respectively. From those numbers, we calculated the fraction of the statewide count of suspensions consisting of suspensions in Anchorage School District as 37.48%. This matched population patterns, since Anchorage made up just over 36% of the enrolled student population and 37% of the enrolled K–2 population. Assuming that percentage remained constant across the state, we calculated the number of K–2 students suspended as 338 students (by multiplying the number of students suspended in Anchorage by $(1 + 0.3748)$).

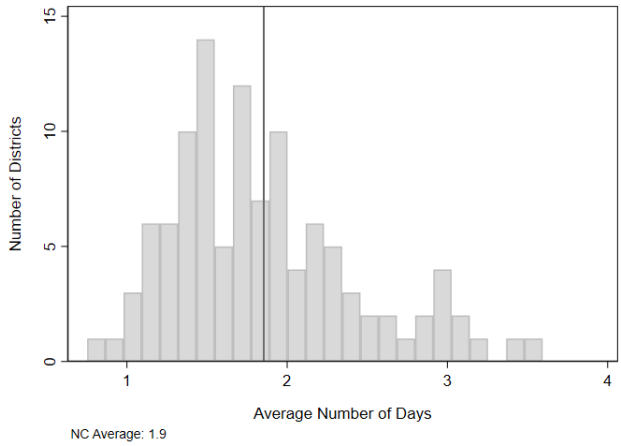
2) New Jersey:
   The Commissioner's Annual Report to the Education Committees of the Senate and General Assembly on Student Safety and Discipline in New Jersey Public Schools' state discipline report listed the suspension rate for Pre-K–2 students. Using grade enrollment data from the state, we calculated the total number of Pre-K–2 students suspended in NJ. We then subtracted the number of Pre-K suspensions reported by OCR from this total to estimate the total number of K–2 students suspended.

3) Vermont:
   Vermont Early Childhood Data & Policy Center's 2022 data brief on exclusionary discipline in the state included a breakdown of suspensions by age. For the purposes of our grade-level analysis, we treated the 5—7-year-olds as our K-2nd graders.

Appendix B: Supplemental Tables and Figures



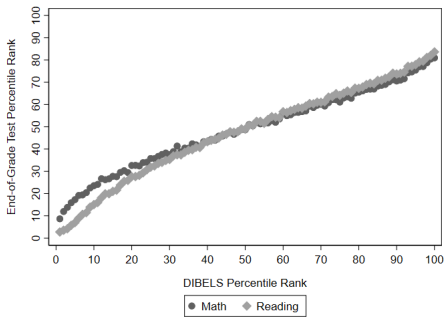
(a) Out-of-School Suspension Length (Days)



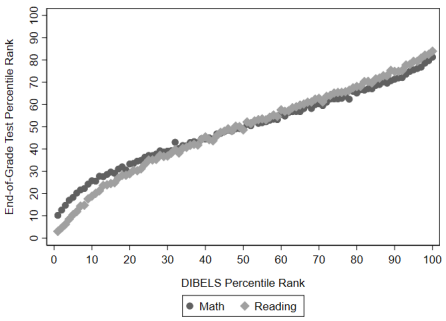
(b) In-School Suspension Length (Days)

Figure B1: Distributions of District-Level Suspension Lengths for Kindergarten through Second Grade Students in North Carolina

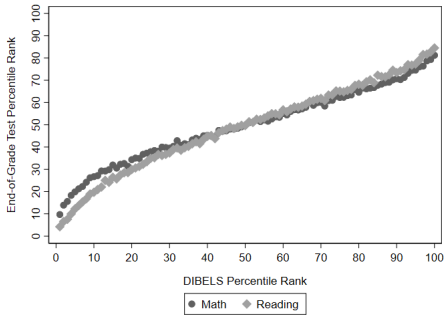
Notes: All suspension lengths are calculated among kindergarten through second grade students. Panels (a) and (b) depict average lengths of out-of-school and in-school suspensions, respectively, among kindergarten through second grade students in North Carolina. Vertical lines in each panel depict the state-level average.



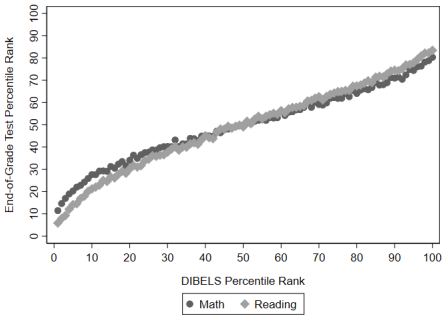
(a) 3rd Grade



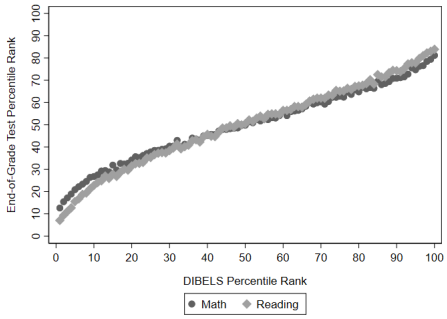
(b) 4th Grade



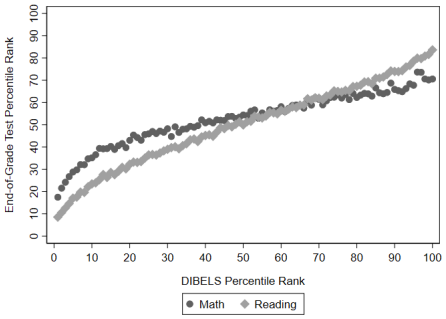
(c) 5th Grade



(d) 6th Grade



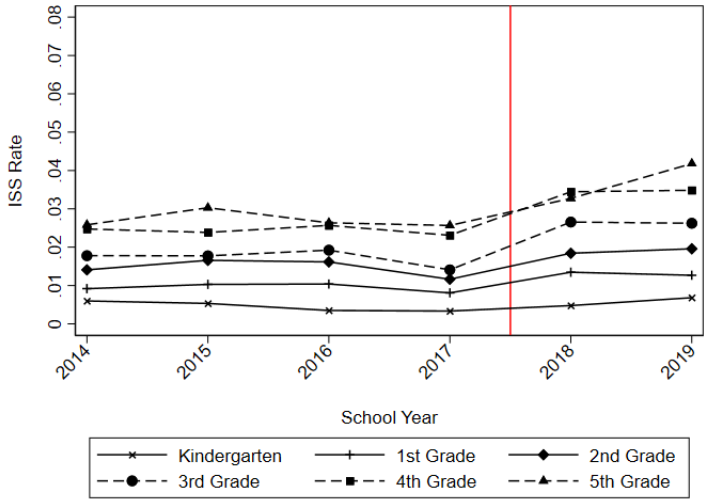
(e) 7th Grade



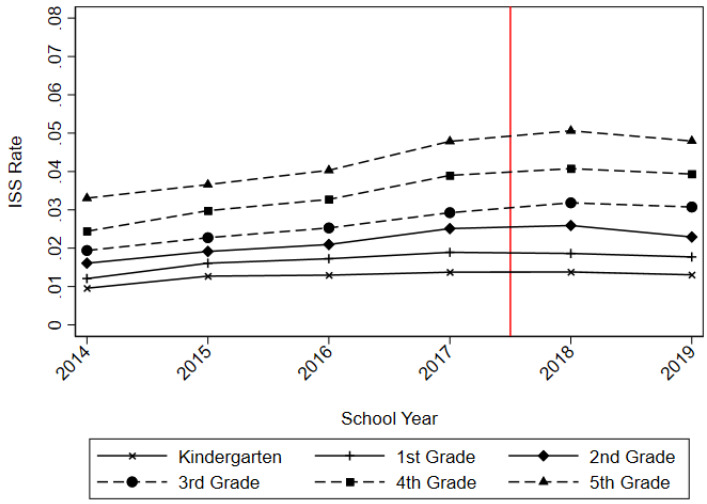
(f) 8th Grade

Figure B2: Correlation Between 3rd Grade DIBELS and EOG Test Scores

Notes: Panels (a)-(f) are percentile-percentile plots for DIBELS test scores in 3rd grade (end of year) and end-of-grade test scores in grades 3-8. Each plot illustrates the average percentile ranking for EOG test scores in math and reading, by DIBELS percentile ranking. The sample is comprised of all North Carolina students enrolled in 3rd grade in 2013/2014 who made on-time grade progress in subsequent school years through 2018/19.



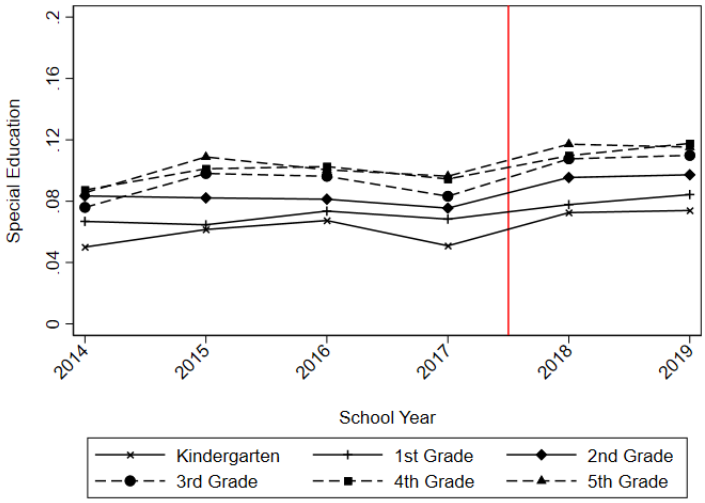
(a) Charlotte-Mecklenburg Schools



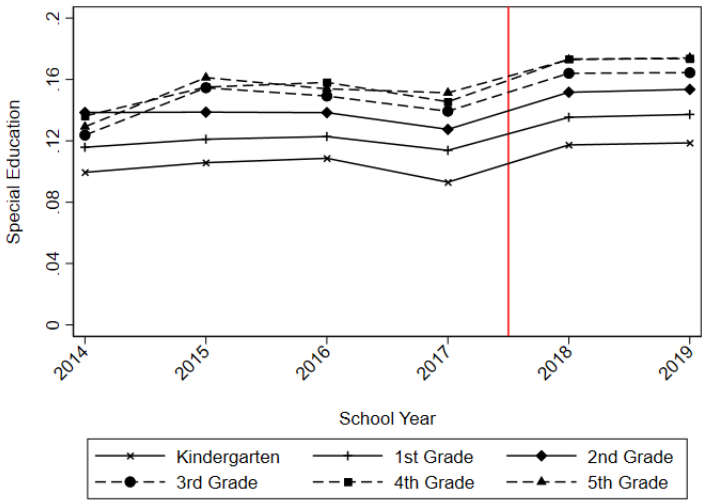
(b) Other North Carolina School Districts

Figure B3: Share of Elementary School Students With At Least One In-School Suspension, Separately by Grade

Notes: Panels (a) and (b) depict the share of elementary school students (grades K-5), separately by grade, who received at least one in-school suspension (ISS)in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



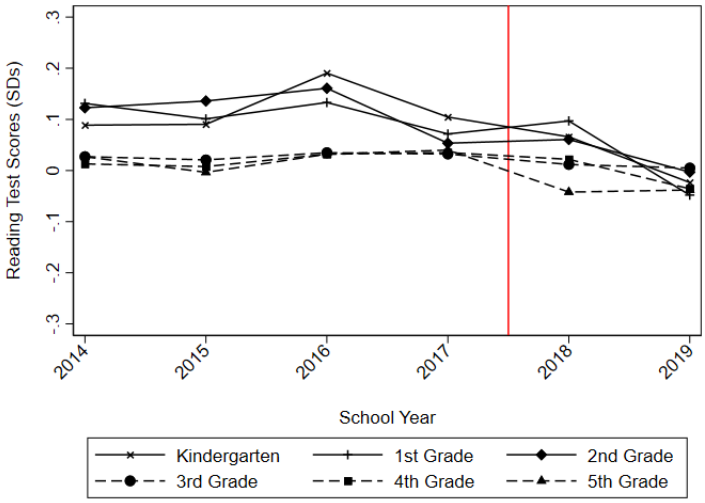
(a) Charlotte-Mecklenburg Schools



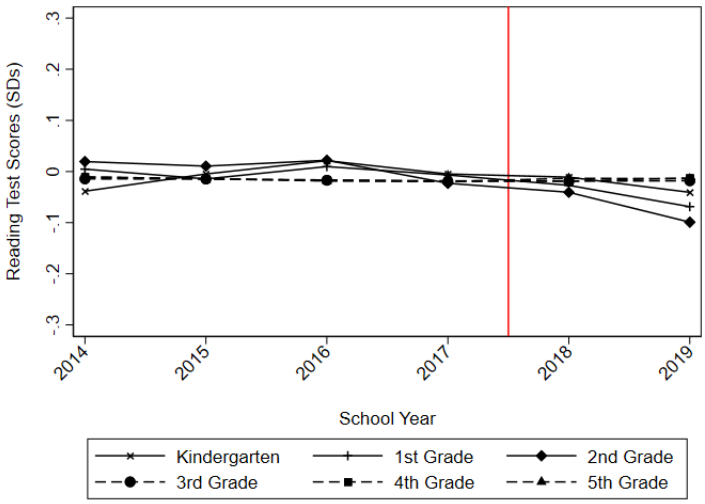
(b) Other North Carolina School Districts

Figure B4: Share of Elementary School Students With Special Education Designation, Separately by Grade

Notes: Panels (a) and (b) depict the share of elementary school students (grades K-5), separately by grade, with a special education designation in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



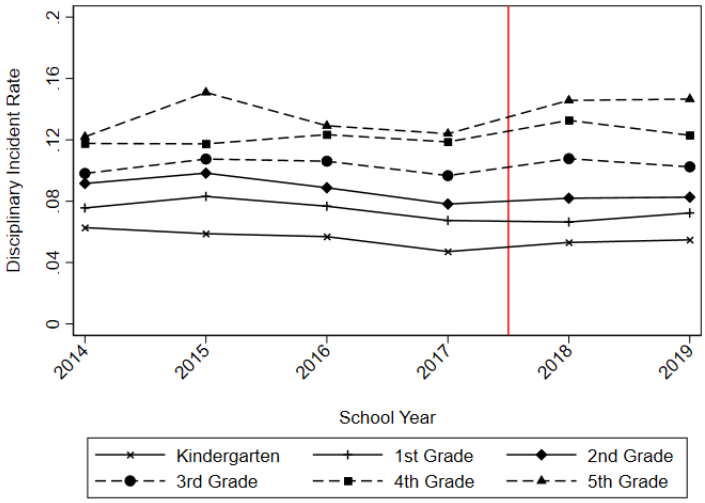
(a) Charlotte-Mecklenburg Schools



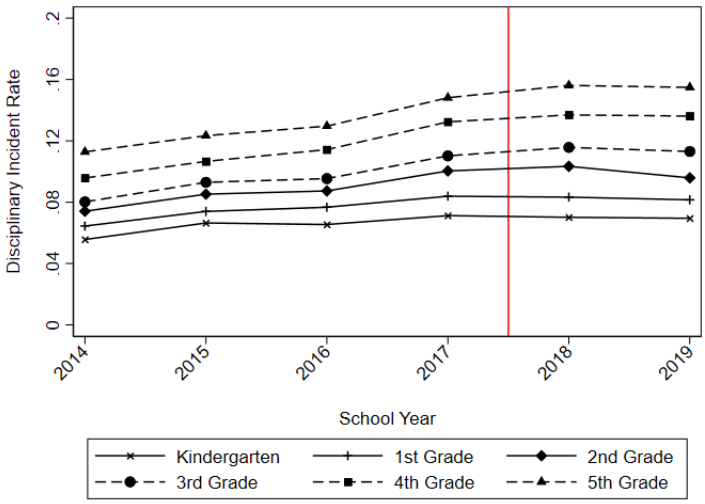
(b) Other North Carolina School Districts

Figure B5: Average Reading Achievement Among Elementary School Students, Separately by Grade

Notes: Panels (a) and (b) depict average reading achievement among elementary school students (grades K-5), separately by grade, in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



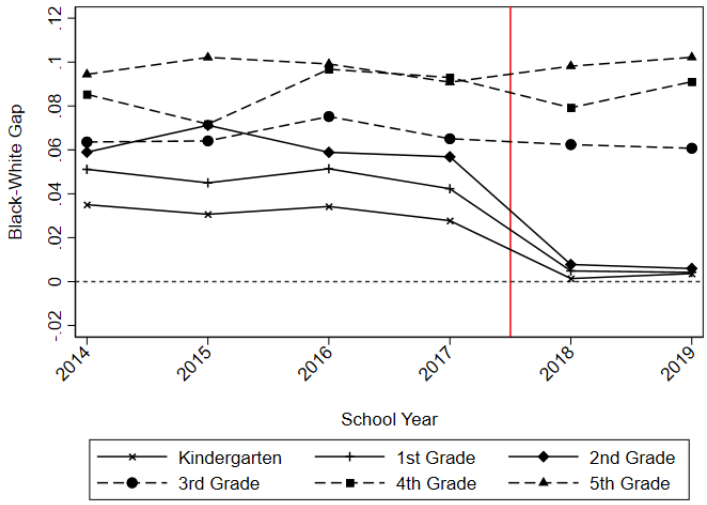
(a) Charlotte-Mecklenburg Schools



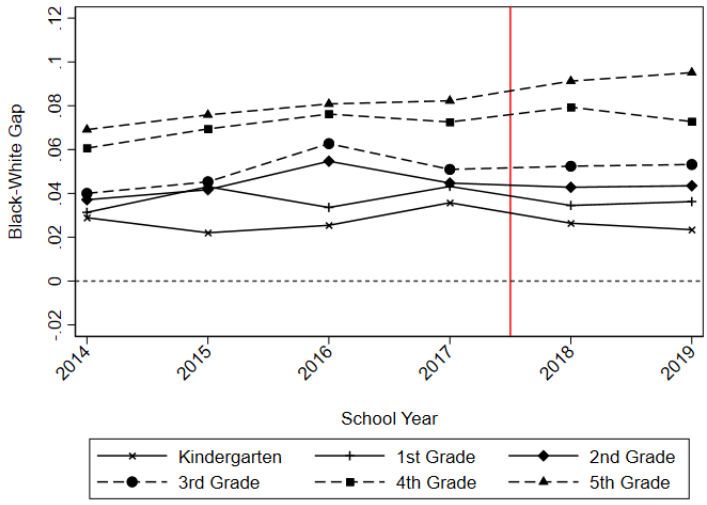
(b) Other North Carolina School Districts

Figure B6: Share of Elementary School Students With Any Disciplinary Infraction, Separately by Grade

Notes: Panels (a) and (b) depict the share of elementary school students (grades K-5), separately by grade, with at least one disciplinary infraction in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



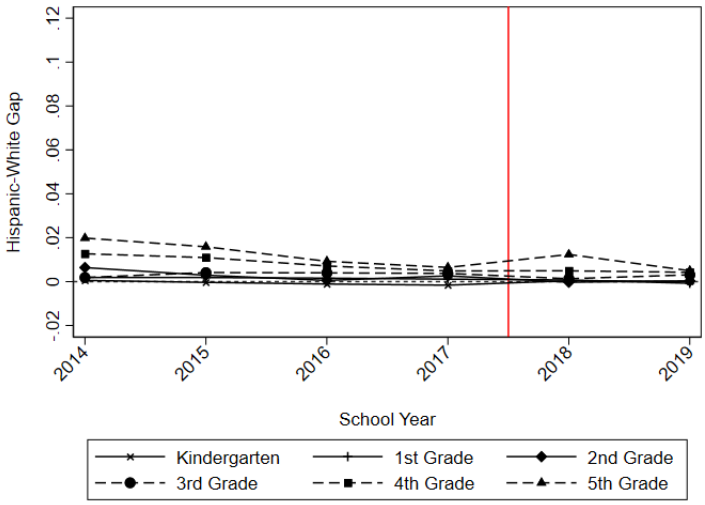
(a) Charlotte-Mecklenburg Schools



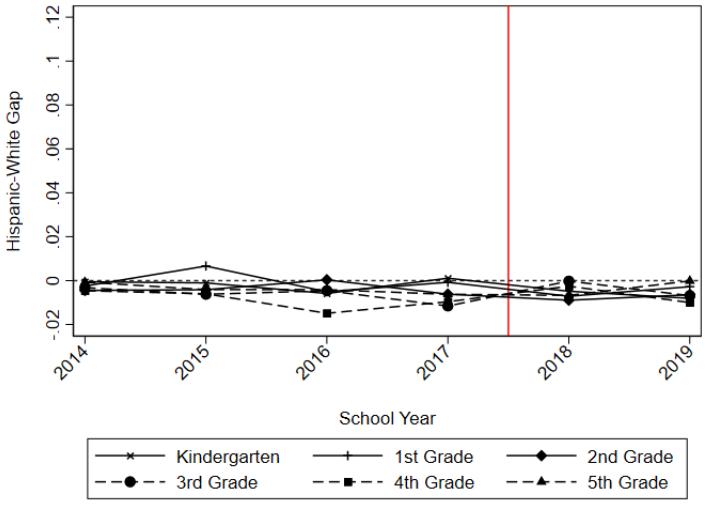
(b) Other North Carolina School Districts

Figure B7: Black-White Out-of-School Suspension Gap, Separately by Grade

Notes: Panels (a) and (b) depict gaps in out-of-school suspension rates in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



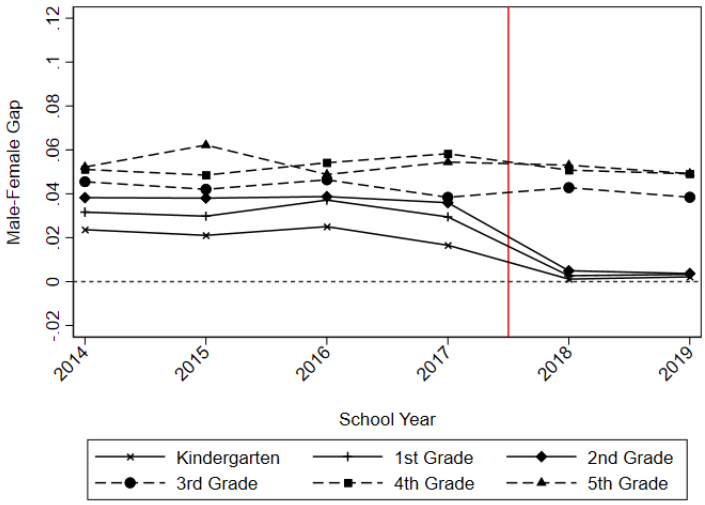
(a) Charlotte-Mecklenburg Schools



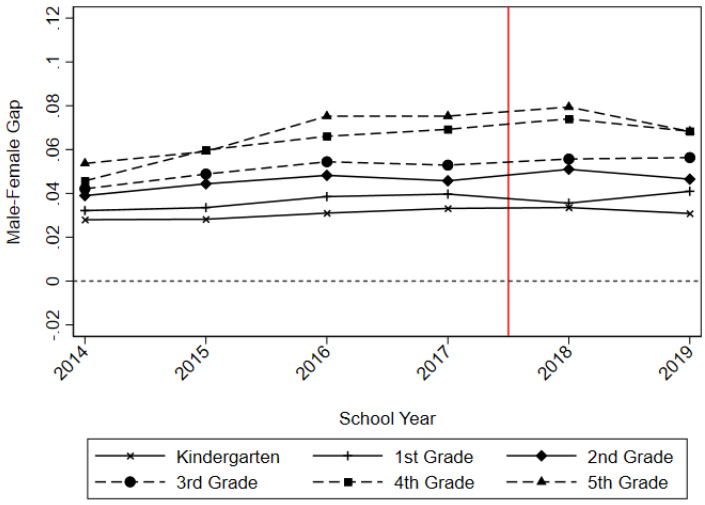
(b) Other North Carolina School Districts

Figure B8: Hispanic-White Out-of-School Suspension Gap, Separately by Grade

Notes: Panels (a) and (b) depict gaps in out-of-school suspension rates in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



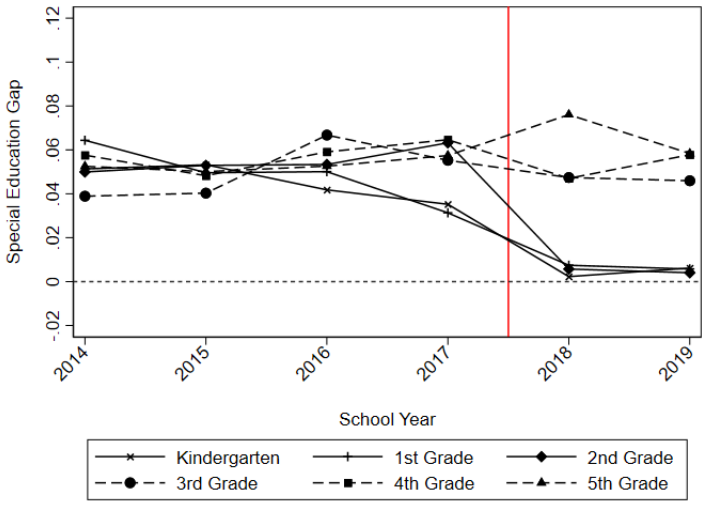
(a) Charlotte-Mecklenburg Schools



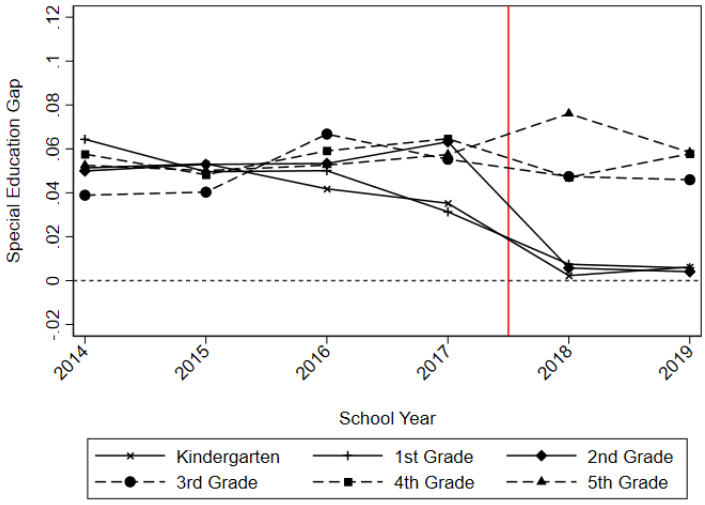
(b) Other North Carolina School Districts

Figure B9: Male-Female Out-of-School Suspension Gap, Separately by Grade

Notes: Panels (a) and (b) depict gaps in out-of-school suspension rates in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



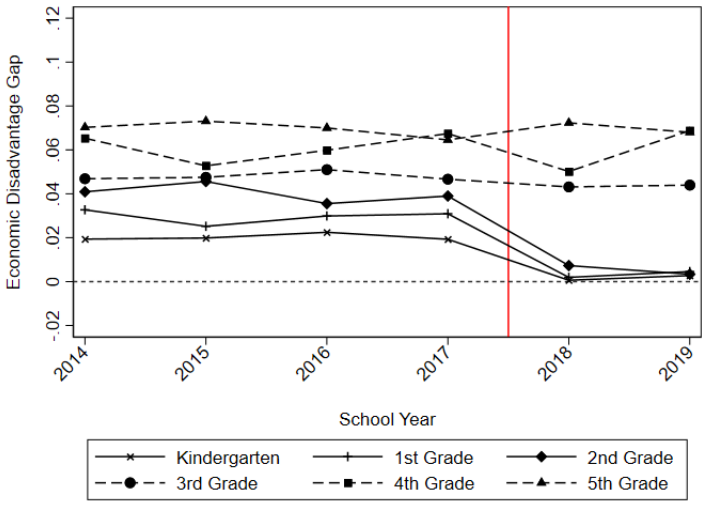
(a) Charlotte-Mecklenburg Schools



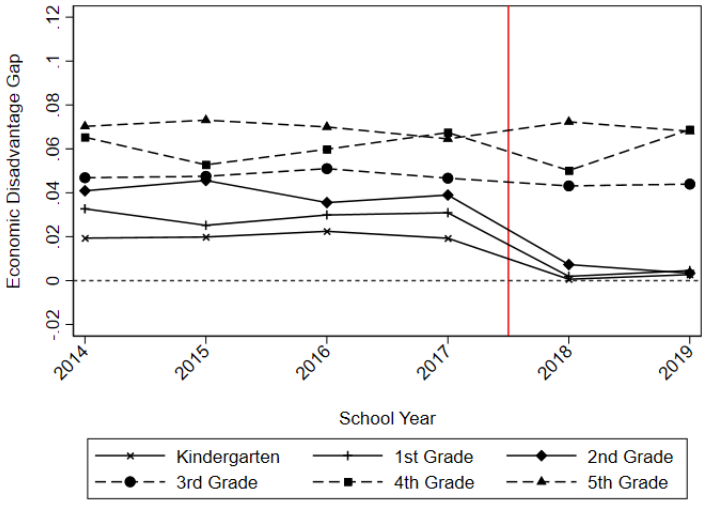
(b) Other North Carolina School Districts

Figure B10: Special Education Out-of-School Suspension Gap, Separately by Grade

Notes: Panels (a) and (b) depict gaps in out-of-school suspension rates in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.



(a) Charlotte-Mecklenburg Schools



(b) Other North Carolina School Districts

Figure B11: Economic Disadvantage Out-of-School Suspension Gap, Separately by Grade

Notes: Panels (a) and (b) depict gaps in out-of-school suspension rates in Charlotte-Mecklenburg Schools (CMS) and other school districts in North Carolina during the 2013/14-2018/19 school years.

Table B1: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes (Augmented Sample)

	(1)	(2)	(3)	(4)
Panel A. Out-of-School Suspension (0/1)	-0.014***	-0.015***	-0.015***	-0.015***
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,888,924	3,888,924	3,888,924	3,888,924
R-Squared	0.004	0.044	0.072	0.078
Baseline Mean	0.025	0.025	0.025	0.025
Panel B. In-School Suspension (0/1)	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,888,924	3,888,924	3,888,924	3,888,924
R-Squared	0.004	0.023	0.073	0.079
Baseline Mean	0.010	0.010	0.010	0.010
Panel C. Special Education (0/1)	0.003	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,888,924	3,888,924	3,888,924	3,888,924
R-Squared	0.005	0.030	0.041	0.042
Baseline Mean	0.069	0.069	0.069	0.069
Panel D. Reading Test Scores (SDs)	-0.005	-0.001	-0.006	0.004
	(0.017)	(0.015)	(0.015)	(0.015)
N	3,888,924	3,888,924	3,888,924	3,888,924
R-Squared	0.001	0.133	0.171	0.185
Baseline Mean	0.115	0.115	0.115	0.115
Panel E. Any Disc. Inf. (0/1)	0.001	-0.000	-0.000	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
N	3,888,924	3,888,924	3,888,924	3,888,924
R-Squared	0.006	0.070	0.115	0.121
Baseline Mean	0.074	0.074	0.074	0.074
Year FE		X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years. Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B2: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes (SEs Clustered at District Level)

	(1)	(2)	(3)	(4)
Panel A. Out-of-School Suspension (0/1)	-0.014***	-0.015***	-0.015***	-0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.006	0.045	0.071	0.079
Baseline Mean	0.025	0.025	0.025	0.025
Panel B. In-School Suspension (0/1)	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.005	0.024	0.073	0.080
Baseline Mean	0.010	0.010	0.010	0.010
Panel C. Special Education (0/1)	0.005	0.004	0.004	0.004
	(0.004)	(0.004)	(0.003)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.005	0.030	0.041	0.042
Baseline Mean	0.069	0.069	0.069	0.069
Panel D. Reading Test Scores (SDs)	0.004	0.008	0.005	0.014
	(0.014)	(0.014)	(0.016)	(0.015)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.126	0.164	0.179
Baseline Mean	0.115	0.115	0.115	0.115
Panel E. Any Disc. Inf. (0/1)	-0.000	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.008	0.071	0.115	0.121
Baseline Mean	0.074	0.074	0.074	0.074
Year FE		X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Heteroskedasticity-robust standard errors are clustered at the district level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B3: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes (Two-Stage Estimation)

	(1)	(2)	(3)	(4)
Panel A. Out-of-School Suspension (0/1)	-0.014	-0.015	-0.015	-0.014
	[-0.021, -0.008]	[-0.021, -0.009]	[-0.021, -0.009]	[-0.020, -0.008]
	{-0.020, -0.009}	{-0.020, -0.010}	{-0.020, -0.010}	{-0.019, -0.009}
N	3,229,777	3,229,777	3,229,750	3,228,340
Baseline Mean	0.025	0.025	0.025	0.025
Panel B. In-School Suspension (0/1)	-0.001	-0.001	-0.001	-0.001
	[-0.007, 0.005]	[-0.007, 0.005]	[-0.008, 0.005]	[-0.007, 0.005]
	{-0.006, 0.005}	{-0.006, 0.004}	{-0.006, 0.004}	{-0.006, 0.004}
N	3,229,777	3,229,777	3,229,750	3,228,340
Baseline Mean	0.010	0.010	0.010	0.010
Panel C. Special Education (0/1)	0.005	0.004	0.004	0.004
	[-0.002, 0.011]	[-0.003, 0.011]	[-0.002, 0.011]	[-0.003, 0.011]
	{-0.001, 0.010}	{-0.001, 0.010}	{-0.002, 0.010}	{-0.002, 0.010}
N	3,229,777	3,229,777	3,229,750	3,228,340
Baseline Mean	0.069	0.069	0.069	0.069
Panel D. Reading Test Scores (SDs)	0.004	0.009	0.005	0.016
	[-0.031, 0.038]	[-0.024, 0.038]	[-0.026, 0.034]	[-0.013, 0.046]
	{-0.026, 0.032}	{-0.018, 0.033}	{-0.022, 0.030}	{-0.008, 0.040}
N	3,229,777	3,229,777	3,229,750	3,228,340
Baseline Mean	0.115	0.115	0.115	0.115
Panel E. Any Disc. Inf. (0/1)	-0.000	-0.002	-0.001	-0.002
	[-0.011, 0.009]	[-0.012, 0.007]	[-0.012, 0.008]	[-0.012, 0.008]
	{-0.008, 0.008}	{-0.010, 0.006}	{-0.009, 0.006}	{-0.010, 0.006}
N	3,229,777	3,229,777	3,229,750	3,228,340
Baseline Mean	0.074	0.074	0.074	0.074
Year FE		X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(E) report two-stage triple-differences estimates. The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Upper and lower bounds of bootstrapped ninety-five percent confidence intervals are reported in brackets. Upper and lower bounds of bootstrapped ninety percent confidence intervals are reported in braces.

Table B4: Triple-Differences Estimates of the Effect of the Suspension Reform on Student Outcomes (Counts)

	(1)	(2)	(3)	(4)
Panel A. Out-of-School Suspensions (Count)	-0.034***	-0.036***	-0.036***	-0.035***
	(0.007)	(0.007)	(0.007)	(0.008)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.003	0.030	0.057	0.063
Baseline Mean	0.050	0.050	0.050	0.050
Panel B. In-School Suspensions (Count)	0.006	0.005	0.005	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.003	0.017	0.065	0.073
Baseline Mean	0.014	0.014	0.014	0.014
Panel C. Disciplinary Infractions (Count)	-0.026	-0.030	-0.030	-0.027
	(0.026)	(0.026)	(0.025)	(0.025)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.003	0.037	0.076	0.082
Baseline Mean	0.217	0.217	0.217	0.217
Year FE		X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(C) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Variables related to suspension and disciplinary infractions are counts. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B5: Triple-Differences Estimates of the Effect of the Suspension Reform on Reported Disciplinary Infractions (Counts)

	(1)	(2)	(3)	(4)
Panel A. Aggressive Behavior (Count)	-0.020***	-0.021***	-0.021***	-0.023***
	(0.008)	(0.008)	(0.008)	(0.008)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.000	0.014	0.042	0.051
Baseline Mean	0.081	0.081	0.081	0.081
Panel B. Disruptive Behavior (Count)	-0.006	-0.007	-0.008	-0.006
	(0.008)	(0.008)	(0.007)	(0.007)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.000	0.015	0.045	0.051
Baseline Mean	0.048	0.048	0.048	0.048
Panel C. Bus Misbehavior (Count)	0.005	0.005	0.005	0.006
	(0.004)	(0.004)	(0.004)	(0.004)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.012	0.045	0.047
Baseline Mean	0.023	0.023	0.023	0.023
Panel D. Insubordination (Count)	-0.012***	-0.013***	-0.012***	-0.013***
	(0.005)	(0.005)	(0.004)	(0.004)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.001	0.007	0.025	0.029
Baseline Mean	0.020	0.020	0.020	0.020
Panel E. Inappropriate Language (Count)	-0.006**	-0.006**	-0.006**	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.003	0.010	0.022	0.026
Baseline Mean	0.010	0.010	0.010	0.010
Panel F. Fight (Count)	0.005	0.005	0.005	0.004
	(0.004)	(0.004)	(0.004)	(0.003)
N	3,229,736	3,229,736	3,229,736	3,229,736
R-Squared	0.005	0.020	0.038	0.047
Baseline Mean	0.012	0.012	0.012	0.012
Year FE	X	X	X	X
Student Covariates		X	X	X
School FE			X	
Grade FE			X	
School X Grade FE				X

Notes: Panels (A)-(F) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to disciplinary infractions are counts. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B6: Heterogeneous Treatment Effects by Predicted Student- and Classroom-Risk

	(1) Full Sample	(2) High-Risk Student	(3) Low-Risk Student	(4) High-Risk Classroom	(5) Low-Risk Classroom
Panel A. Aggressive Behavior (0/1)	-0.010***	-0.020***	-0.002	-0.016***	-0.005
	(0.003)	(0.005)	(0.002)	(0.005)	(0.004)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.036	0.070	0.012	0.059	0.022
Panel B. Disruptive Behavior (0/1)	-0.003	-0.007*	-0.001	-0.003	-0.004
	(0.002)	(0.004)	(0.001)	(0.004)	(0.003)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.022	0.044	0.006	0.039	0.011
Panel C. Bus Misbehavior (0/1)	0.002	0.004	0.001	0.006*	-0.001
	(0.002)	(0.003)	(0.001)	(0.003)	(0.002)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.014	0.026	0.006	0.021	0.011
Panel D. Insubordination (0/1)	-0.004**	-0.007***	-0.001	-0.004	-0.003
	(0.002)	(0.003)	(0.001)	(0.003)	(0.002)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.010	0.020	0.002	0.018	0.004
Panel E. Inappropriate Language (0/1)	-0.003	-0.006*	-0.001	-0.003	-0.003
	(0.002)	(0.003)	(0.001)	(0.003)	(0.002)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.008	0.016	0.002	0.013	0.005
Panel F. Fight (0/1)	0.002	-0.000	-0.000	0.002	-0.003
	(0.002)	(0.004)	(0.001)	(0.003)	(0.002)
N	3,229,736	1,615,182	1,612,759	1,606,714	1,611,550
Baseline Mean	0.009	0.020	0.001	0.019	0.003

Notes: Panels (A)-(F) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to disciplinary infractions indicate at least one disciplinary infraction (0/1) of the type indicated. Students are divided into groups based on predicted student- and classroom-level risk. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B7: Heterogeneous Treatment Effects by Predicted Student- and Classroom-Risk

	(1) Full Sample	(2) High-Risk Stud., High-Risk Class	(3) High-Risk Stud., Low-Risk Class	(4) Low-Risk Stud., High-Risk Class	(5) Low-Risk Stud., Low-Risk Class
Panel A. Out-of-School Suspension (0/1)	-0.014***	-0.052***	-0.035***	-0.006***	-0.003***
	(0.003)	(0.007)	(0.006)	(0.002)	(0.001)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.025	0.083	0.027	0.005	0.002
Panel B. In-School Suspension (0/1)	-0.001	0.000	-0.011	0.003	-0.004*
	(0.003)	(0.006)	(0.007)	(0.002)	(0.002)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.010	0.028	0.011	0.003	0.002
Panel C. Special Education (0/1)	0.004	-0.001	-0.002	-0.004	0.002
	(0.003)	(0.006)	(0.009)	(0.005)	(0.004)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.069	0.099	0.103	0.044	0.046
Panel D. Reading Test Scores (SDs)	0.014	0.037	0.033	0.014	-0.005
	(0.015)	(0.025)	(0.029)	(0.028)	(0.020)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.115	-0.316	-0.100	0.092	0.463
Panel E. Any Disc. Infr. (0/1)	-0.002	-0.012	-0.016	-0.001	-0.005
	(0.005)	(0.010)	(0.014)	(0.006)	(0.006)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.074	0.174	0.095	0.037	0.024

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B8: Heterogeneous Treatment Effects by Predicted Student- and Classroom-Risk

	(1) Full Sample	(2) High-Risk Stud., High-Risk Class	(3) High-Risk Stud., Low-Risk Class	(4) Low-Risk Stud., High-Risk Class	(5) Low-Risk Stud., Low-Risk Class
Panel A. Aggressive Behavior (0/1)	-0.010***	-0.024***	-0.016*	-0.005**	-0.000
	(0.003)	(0.007)	(0.009)	(0.003)	(0.003)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.036	0.091	0.047	0.015	0.010
Panel B. Disruptive Behavior (0/1)	-0.003	-0.008	-0.010*	-0.000	-0.002
	(0.002)	(0.005)	(0.006)	(0.002)	(0.002)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.022	0.062	0.025	0.008	0.005
Panel C. Bus Misbehavior (0/1)	0.002	0.007	0.001	0.004**	-0.002
	(0.002)	(0.005)	(0.004)	(0.002)	(0.002)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.014	0.031	0.022	0.007	0.005
Panel D. Insubordination (0/1)	-0.004**	-0.006	-0.008*	-0.002	-0.001
	(0.002)	(0.004)	(0.004)	(0.001)	(0.001)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.010	0.029	0.010	0.003	0.001
Panel E. Inappropriate Language (0/1)	-0.003	-0.005	-0.006	-0.001	-0.002
	(0.002)	(0.004)	(0.004)	(0.002)	(0.002)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.008	0.020	0.011	0.003	0.002
Panel F. Fight (0/1)	0.002	-0.003	-0.008*	0.000	-0.002*
	(0.002)	(0.005)	(0.004)	(0.001)	(0.001)
N	3,229,736	974,238	634,806	631,402	976,037
Baseline Mean	0.009	0.031	0.007	0.002	0.001

Notes: Panels (A)-(F) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). Binary variables related to disciplinary infractions indicate at least one disciplinary infraction (0/1) of the type indicated. Students are divided into four groups based on predicted student- and classroom-level risk. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B9: Heterogeneous Treatment Effects by Race/Ethnicity and Gender

	(1) Full Sample	(2) Black Males	(3) Black Females	(4) White Males	(5) White Females	(6) Hispanic Males	(7) Hispanic Females
Panel A. Out-of-School Suspension (0/1)	-0.014***	-0.067***	-0.011*	-0.005*	0.001	-0.003	0.001
	(0.003)	(0.010)	(0.006)	(0.003)	(0.001)	(0.004)	(0.002)
N	3,229,736	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.025	0.086	0.022	0.012	0.003	0.015	0.002
Panel B. In-School Suspension (0/1)	-0.001	-0.007	0.002	-0.001	0.000	0.002	-0.000
	(0.003)	(0.009)	(0.005)	(0.004)	(0.001)	(0.004)	(0.002)
N	3,229,736	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.010	0.032	0.009	0.004	0.001	0.006	0.001
Panel C. Special Education (0/1)	0.004	-0.007	0.005	0.018*	0.001	0.001	-0.004
	(0.003)	(0.008)	(0.007)	(0.009)	(0.006)	(0.009)	(0.008)
N	3,229,736	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.069	0.115	0.051	0.088	0.045	0.084	0.034
Panel D. Reading Test Scores (SDs)	0.016	0.017	0.011	0.005	0.055**	0.007	0.035
	(0.015)	(0.024)	(0.028)	(0.025)	(0.022)	(0.029)	(0.030)
N	3,229,736	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.115	-0.141	0.057	0.465	0.549	-0.294	-0.178
Panel E. Any Disc. Infr. (0/1)	-0.002	-0.030**	0.003	-0.003	0.006**	0.013	0.004
	(0.005)	(0.014)	(0.009)	(0.008)	(0.003)	(0.010)	(0.005)
N	3,229,736	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.074	0.197	0.070	0.056	0.014	0.068	0.016

Notes: Panels (A)-(E) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). The sample is restricted to students who are Black, Hispanic, and white. Binary variables related to suspension and disciplinary infractions indicate at least one out-of-school suspension (0/1), at least one in-school suspension (0/1), and at least one disciplinary infraction (0/1), respectively. Students are divided into subgroups by race/ethnicity and gender. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.

Table B10: Heterogeneous Treatment Effects by Race/Ethnicity and Gender

	(1) Black Males	(2) Black Females	(3) White Males	(4) White Females	(5) Hispanic Males	(6) Hispanic Females
Panel A. Aggressive Behavior (0/1)	-0.037***	-0.017***	-0.002	0.002	0.002	-0.002
	(0.010)	(0.005)	(0.005)	(0.002)	(0.005)	(0.002)
N	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.107	0.030	0.028	0.006	0.029	0.005
Panel B. Disruptive Behavior (0/1)	-0.014*	-0.002	-0.002	0.000	-0.001	0.001
	(0.008)	(0.005)	(0.003)	(0.001)	(0.004)	(0.002)
N	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.068	0.021	0.014	0.003	0.014	0.003
Panel C. Bus Misbehavior (0/1)	0.003	0.007*	-0.003	-0.000	0.002	0.003**
	(0.006)	(0.004)	(0.003)	(0.002)	(0.004)	(0.002)
N	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.038	0.014	0.010	0.002	0.015	0.003
Panel D. Insubordination (0/1)	-0.015***	-0.004	0.000	-0.001	0.002	-0.002
	(0.005)	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)
N	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.032	0.010	0.004	0.001	0.004	0.001
Panel E. Inappropriate Language (0/1)	-0.010*	-0.003	-0.003	-0.000	-0.000	-0.001
	(0.005)	(0.003)	(0.003)	(0.001)	(0.004)	(0.001)
N	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.023	0.007	0.006	0.001	0.007	0.001
Panel F. Fight (0/1)	-0.002	0.005	0.002	0.001	0.007**	0.003***
	(0.008)	(0.004)	(0.002)	(0.001)	(0.003)	(0.001)
N	414,029	398,467	790,168	737,144	304,723	292,933
Baseline Mean	0.032	0.007	0.004	0.000	0.006	0.000

Notes: Panels (A)-(F) report triple-differences estimates corresponding to Equation (1a). The sample contains all students enrolled in traditional public elementary schools in North Carolina between the 2013/14-2018/19 school years (excluding students enrolled in third grade). The sample is restricted to students who are Black, Hispanic, and white. Binary variables related to disciplinary infractions indicate at least one disciplinary infraction (0/1) of the type indicated. Students are divided into subgroups by race/ethnicity and gender. Heteroskedasticity-robust standard errors are clustered at the school level. Asterisks indicate statistical significance: * p<0.10, ** p<0.05, and *** p<0.01.