From Retributive to Restorative: An Alternative Approach to Justice*

Anjali Adukia
University of Chicago
and NBER

Benjamin Feigenberg
University of Illinois
at Chicago

Fatemeh Momeni
University of Chicago

January 2022

Abstract

Schools have traditionally employed punitive methods of discipline when trying to shape children's behavior, but more recently, districts have been experimenting with restorative approaches to engaging with students. We use variation in the timing of the introduction of restorative practices across Chicago Public Schools to evaluate the impact of restorative approaches on student outcomes. For high school students, we identify significant decreases in out-of-school suspensions and increases in perceived school climate in response to policy adoption. We also find evidence of a decrease in arrests that is consistent with spillovers in behavior outside of school.

*Contact: Anjali Adukia, adukia@uchicago.edu; Benjamin Feigenberg, bfeigenb@uic.edu; Fatemeh Momeni, fmomeni@uchicago.edu. We thank Roseanna Ander and Monica Bhatt for vital assistance making this work possible. For helpful feedback, we thank Hellen Antonopoulos, Monica Bhatt, Jon Guryan, and Jens Ludwig. For research support, we thank David Arbelaez, Emily He, Juan Miguel Jimenez, Farah Mallah, Eleni Packis, Puja Patel, Sarah Permut, and Jalisia Taylor-Singleton. For financial support, we thank AbbVie, the National Academy of Education, Spencer Foundation, and WT Grant Foundation. For data acquisition, we thank Chicago Public Schools and Chicago Police Department.

I Introduction

School disciplinary policies may reach beyond the creation of conditions for learning in the short term. They may also send signals to children about optimal ways to behave and how society should ideally work (Parsons 1959; Dreeben 1967; Jackson 1968; Bowles and Gintis 1976; Boyd and Richerson 1985). When school districts rely on primarily punitive responses to resolve minor conflicts, children may infer that the optimal approach to undesirable situations is one of retribution. However, if a school district instead emphasizes a positive, or restorative, approach to addressing behavior, children may adopt a more positive approach to situations in life. Indeed, social preferences may remain consistent and stable over time, such that habits formed early in life may influence the way people conduct themselves in later life (Chuang and Schaechter 2015).

The rate of school suspensions in the United States (U.S.) more than doubled between 1974 and 2006. In school year (SY) 2011-2012, approximately 3.5 million public school students were suspended from school, losing nearly 18 million days of instruction due to "zero-tolerance" policies. Suspensions and expulsions from school are associated with future academic difficulties, increased misconduct, and eventual involvement in the justice system (Fabelo et al., 2011; Shollenberger, 2015; Wolf and Kupchik, 2017). The rapid rise in suspension rates over recent decades paired with the negative long-run impacts of punitive disciplinary policies have prompted school officials to consider alternative ways to shape student behavior. Specifically, educators have begun experimenting with more "positive" techniques that emphasize community building and restitution or restoration, as an alternative to the traditional punitive approach (Losen, Hewitt and Toldson, 2014).

The restorative justice (RJ) approach used in the criminal justice system has become increasing popular in schools (Augustine et al., 2018). RJ as a philosophy emphasizes the reparation of harm between victims and offenders, engaging various stakeholders in the community through open dialogue and shared ownership of disciplinary justice (Karp and Breslin, 2001; Hopkins, 2003; González, 2012). While RJ has entered U.S. public education systems only recently in the form of restorative practices (RP), it has quickly increased in use despite the lack of quantitative evidence on the associated costs and benefits. Studies on the impacts of RP on educational and behavioral outcomes, within or outside of schools, are limited with most being correlational or descriptive. This study aims to fill a gap in the existing literature by providing novel evidence on the causal impacts of exposure to the RP approach (Augustine et al., 2018; Gregory et al., 2018; Acosta et al., 2019).

In this study, we examine how student-level short-run educational and behavioral outcomes and arrest rates respond to students' exposure to RP programs rolled out across high schools within the Chicago Public Schools (CPS) system beginning in school year (SY)

2013-2014.[1] To expand access to RP programming in schools, CPS provided training to school staff that emphasized less punitive and more positive strategies when engaging with students (for example, developing restorative mindsets and language in school staff, creating and implementing disciplinary protocols and processes, and strengthening student-teacher relationships.)

Using a within-district transition away from punitive approaches and towards restorative practices, we employ a difference-in-differences-style research design (based on the methodology developed in de Chaisemartin and D'Haultfoeuille (2020)) to examine how a positive approach to addressing perceived behavioral challenges impacts student disciplinary records, juvenile arrest rates, school climate, and academic outcomes. Given evidence of the disparate impact of traditional disciplinary approaches on male students, African American and Latinx students, and students with disabilities, we explicitly explore heterogeneity by student gender, race, ability status, and English Learner status.

We find that restorative practices decrease out-of-school suspensions for high school students. We do not find evidence of corresponding increases in in-school suspensions, suggesting that students are receiving more in-school instruction time in response to policy adoption. There are two potential explanations for these findings. First, it may be that RP is having a positive, productive impact on teacher behavior and/or student behavior. Teachers may be changing how they interact with students, better responding to students' individual needs, and avoiding escalation. This could then lead students to learn how to resolve conflicts appropriately and to feel more understood by adults and their peers. Second, the effects may be more mechanical such that teachers are told that they are not supposed to be suspending children as frequently and respond accordingly. If the latter explanation for our findings is correct, one concern is that those students who were previously identified as engaging in undesirable behaviors may now cause learning disruptions for their peers.

To distinguish between these alternative explanations for the measured declines in suspensions, we use person-level arrest data from the Chicago Police Department. Based on these data, we identify significant decreases in arrests among high students both in and outside of school. This evidence, in combination with a lack of any significant negative impacts on learning (as measured by GPA and test scores) suggests that the restorative practices intervention we study is generating genuine changes in underlying student behaviors. In subsequent heterogeneity analyses, we present evidence that the absolute gains from the introduction of restorative practices are particularly large among Black students, who experience the highest baseline rates of suspension and arrest.

[1]For brevity, we will refer to school years by the year in which the spring term occurs (*e.g.*, school year 2013-2014 is 2014 or SY14).

II Disciplinary Practices in Chicago Public Schools

We study the impact of restorative practices in the context of Chicago Public Schools (CPS), the third largest school district in the U.S., which serves over 340,000 students annually across more than 600 schools. The population of CPS is racially and economically diverse. Of the students attending CPS in SY21, 36% were identified as African American, 47% as Hispanic, and 11% as White and over 63% were eligible to receive free or reduced priced lunches.[2]

Like many other large, urban school districts, CPS employed punitive methods of student discipline in the past. In the 1980s and 1990s, the district implemented zero-tolerance policies mandating the use of suspensions and expulsions in response to student misconduct violations. These policies have come under scrutiny at the federal, state, and local level alike as they resulted in high suspension rates, especially among students of color (Stevens et al., 2015) and had important distributional consequences as students from the most vulnerable backgrounds, such as those living in poverty, those with disabilities and those with a history of abuse or neglect, were more likely to be suspended (Sartain, Allensworth and Porter, 2015).

In the past decade, school districts across the country have started to recognize the potential adverse effects of such zero-tolerance policies on student outcomes and introduced alternative approaches in response to misconduct violations. In 2014, CPS announced a disciplinary policy reform plan called the Suspensions and Expulsion Reduction Plan (SERP), with the goal of decreasing the number of out-of-school suspensions and expanding resources and training on school discipline to school staff across the district. Since then, CPS has continued to implement policy changes through the student code of conduct, by limiting the length of suspensions for substance use infractions, by requiring district administrator approval for suspending students for certain behaviors, and most recently by removing in-school suspension for first-time lower-level infractions. These efforts are specifically intended to reduce inequities in suspension rates by race and other student characteristics (Sartain, Allensworth and Porter, 2015).

The expansiveness of reforms undertaken in the past decade highlights the need for a rigorous research approach to isolating the impacts of any one policy related to disciplinary practices within CPS. Since pre-post analyses will confound the effects of these various reforms, we propose to isolate the impact of restorative practices programs using a difference-in-differences research design, which we will describe extensively in Section V. Our analysis focuses on understanding the impact of restorative practices in CPS high schools.

[2] https://www.cps.edu/about/district-data/

II.A Rollout of Restorative Practices Programs at CPS

In SY14, as a part of the SERP reform and as the district transitioned away from zero tolerance policies, CPS's Office of Social and Emotional Learning (OSEL) began to roll out district-wide Restorative Practices (RP) programs. The district started by working with 56 schools (including 22 high schools) in SY14 and has expanded their programs to reach 279 schools (including 74 high schools) by SY19. Appendix A1 presents this expansion across CPS schools between SY14 and SY19.

The district offers three different programs to support high schools in the adoption of restorative practices. These programs include RP Leadership, RP Coaching, and RP Peer Council. All these programs are based on the same fundamental RP practices including a restorative mindset, restorative language, restorative conversation, talking and peace circles, and peer conferences. These practices all promote the development of student Socio-Emotional Learning (SEL) skills (such as self-reflection and empathetic listening) through collaboration between teachers, students, and staff and creation of non-judgmental safe spaces for communication.

The most intensive of these programs is RP Coaching, a 10-20 week training through which a restorative practices Coach trains administrators and designated staff to model and implement restorative practices within their school. This flexible model is designed to meet schools' needs and abilities in developing a menu of restorative practices that is most appropriate for the context of their school. The second program is RP Leadership, whereby OSEL supports schools to strengthen internal leadership capacity for building sustainable school-wide systems that restoratively develop community and address behavioral concerns. During this process, designated school staff participate in a cohort to receive intensive training and ongoing coaching, and at the end of the program receive a Restorative Practices Leadership Certification and become Classroom and Discipline Leads. The third program –RP Peer Council– is a student-led process in which a small group of members work with referred students (who were involved in misconduct incidents or conflicts) to understand the impact of their actions on other individuals and school culture.

Our evaluation focuses on understanding the impact of restorative practices in CPS high schools as a whole and our analysis, for the most part, does not distinguish between these three programs.

Schools are selected to receive restorative practices programs based on a variety of factors including a school's interest and readiness to strengthen their restorative practices, a school's out-of-school suspension rate, a school's suspension rate for "priority" student groups, a school's climate indicators on the "My Voice My School" (MVMS) survey (also known as

the CPS 5Essentials survey), and input from those directly working with the schools.[3] The criteria used to allocate restorative practices programs motivates our difference-in-differences research design. Specifically, since these criteria indicate that schools receiving RP programming are likely to differ on various dimensions (for example, suspension rates) when compared to schools not receiving RP programming, our research design relies on a weaker conditional exogeneity assumption that requires only that *changes* in outcomes of interest over time are independent of RP programming assignment.

III Data Sources

Our analysis uses four main sources of data: Programming data from CPS's Office of Social and Emotional Learning (OSEL), CPS administrative data, CPS data on student responses to the MVMS survey, and Chicago Police Department (CPS) arrests data.

III.A Restorative Practices Programming Data

To identify the timing of treatment for students enrolled in a given school, we use programming data provided to us by the OSEL at CPS. This data includes records on which schools received restorative practices training in each school-year between SY14 and SY19. For schools that received restorative practices training in this time-frame, this data also includes information about which of the programs was implemented in a given school-year.

III.B CPS Administrative Data

Our analysis also uses information on CPS's student-level administrative data between SY09 and SY19. These data files include information on student-level outcome and demographic information. The outcome variables are constructed based on information on academic indicators such as semester-level grades, test scores, attendance records, we well as disciplinary records such as incidents of in-school and out-of-school suspensions. The administrative records include student-level demographic information such as race, gender, socio-economic indicators, disability status, and English learner status for students enrolled in any CPS schools. The CPS administrative data files also include information on student-level enrollment history, which we link to the programming data files from the OSEL to determine the timing of treatment at the student level.

The district requires all its neighborhood and charter schools to report data according to the same guidelines and the administrative datasets are compiled by CPS from data reported by each school.[4] CPS maintains a general student database in which each student

[3] "Priority" student groups have included students with IEPs and African American students since these are the student groups suspended at the highest rates.

[4] "Neighborhood" schools refer to those schools under the direct purview of CPS. They are overseen by the Chicago School Board and follow a universal curriculum (Chicago Public Schools, 2020). "Charter" schools

is identified by a unique student ID. All distinct administrative files are linked together by student ID.

III.C School Climate Data

Since 2011, CPS has administered annual surveys called "My Voice, My School" (MVMS) to understand the experiences of key stakeholders (students, teachers, and parents/guardians) in the school environment. To understand the impact of restorative practices training on school climate, we use data from student responses to the MVMS survey. The student survey is administered to students enrolled in grades six to twelve and is composed of 21 constructs. We construct a climate index using data from student responses to 12 out of the 21 constructs that most directly speak to a student's perception of school climate.[5]

III.D CPD Arrest Data

To explore the effects of restorative practices programs on student arrests, we use data from the Chicago Police Department (CPD), which includes individual-level arrest records from July 1, 2008 through June 30, 2019. The arrest data includes information on the type (violent or non-violent), the location, and the time of arrest but does not include any information about convictions. We use this information to explore the impact of restorative practices on different types of arrests and to separately analyze impacts on the arrests that take place in school versus outside of school. Prior research has demonstrated that student arrests have particularly negative long-term impacts on a range of student outcomes, highlighting the importance of including this outcome measure in our analyses. The arrests-based analysis also allows us to probe the possibility that any measured changes in student disciplinary outcomes (i.e., suspensions) are driven by changes in teacher and administrator responses to misconduct rather than by changes in student behavior. To the extent this is the case, we would expect changes in arrests (and out-of-school arrests, in particular) to be muted in comparison to changes in disciplinary outcomes initiated by school staff.

The CPD and CPS data files are joined using probabilistic matching over a child's name, date of birth, gender, and home address.

IV Study Sample

Our analysis includes observations from students who were enrolled in 298 CPS neighborhood and charter high schools between SY09 and SY19 for at least one day.

are independent, free neighborhood schools that develop innovative curriculum and are governed by the Illinois State Charter School Commission (Illinois Network of Charter Schools, 2020).

[5]These 12 constructs are Peer Support for Academic Work, Emotional Health, Academic Engagement, Human and Social Resources in the Community, Student Classroom Behavior, Academic Personalization, Parent Supportiveness, Psychological Sense of School Membership, Safety, School-Wide Future Orientation, School Safety, and Student-Teacher Trust (University of Chicago Impact, 2020).

We focus our analysis on high school students for two main reasons. First, the main goal of the restorative practices programs in CPS is to reduce instances of in-school disciplinary responses, which are far more prevalent in high schools compared to elementary schools. In SY14, 16% of CPS high school students received an out of school suspension compared to 9% of 6th-8th graders, 5% of 4th-5th graders and only 2% of 1st-3rd graders. Similarly, over 15% of CPS high school students received in-school suspensions in SY14, compared to only 3% of 6th-8th graders, 2% of 4th-5th graders and less than 1% of 1st-3rd graders. High school students are also more likely to experience both in-school and out-of-school arrest incidents. For example, our analysis suggests that in SY14, 2% (5%) of high school students were arrested in (outside) CPS schools, compared to 0.4% (0.6%) of elementary school students. The low baseline rate of school disciplinary incidents and arrests in elementary schools poses a measurement challenge limiting our power to detect potential impacts on these margins. Second, our identification strategy relies on the assumption that conditional on baseline observables, students enrolled in schools that did and did not adopt restorative practices would have exhibited parallel trends in their outcomes in the absence of the roll-out of restorative practices. Our analysis suggests that this key identifying assumption does not generally hold for the elementary school sample, preventing us from causally estimating impact.[6]

Before turning to the research design, we next present descriptive data on the characteristics of the CPS high schools included in our study sample. Specifically, Table 1 presents average characteristics for students enrolled in the 298 CPS high schools in our sample in the school year prior to the roll-out of restorative practices (SY13), separately for schools that did and did not receive any restorative practices programming at some point between SY14 and SY19.[7] This table shows that high schools that received restorative practices training differ from never treated high schools in several ways. These high schools were significantly larger schools (with about twice as many students enrolled) and enrolled a slightly lower share of African American students. The treated high schools also used suspensions as disciplinary tools more intensely as students who enrolled in these schools had on average twice as many in-school suspension days (0.49 versus 0.25) and slightly more out-of-school suspension days (1.07 versus 0.73) than those enrolled in never-treated schools. Finally, students in the treated high schools had more absent days and lower GPAs at baseline, on average. The average differences we identify between subsequently treated and untreated high schools motivate our choice to employ a research design that relies on parallel trends-type (rather

[6]See Section V for a detailed description of our identification strategy. Elementary school event studies are presented in Appendix Figures A2 and A3.

[7]Appendix Tables A1 and A2 present average characteristics by demographic groups.

than strict exogeneity-based) identifying assumptions.

V Research Design

We study the impact of the introduction of restorative practices on a host of student-level disciplinary and academic outcomes, as well as measures of student arrest incidents. We also investigate how student perceptions of school climate respond to the roll out of restorative practices.

To identify treatment effects associated with exposure to restorative practices, we rely on variation in exposure induced by the rollout of restorative practices over time and across schools. Since student enrollment choices may respond endogenously to restorative practices exposure, we identify student-level treatment exposure based on the first school that each student attended within the CPS system in a given grade range, as well as the year and grade level in which that student enrolled in CPS.[8] To guide thinking, if student i attended high school j from 2010 to 2012, and then moved to high school k, the student's treatment exposure remains a function of the timing of RP rollout in school j, regardless of whether the first intervention in school j occurred before or after the student had transferred. The subsequent analysis includes one observation per year per student for every student who was enrolled for at least one day in any CPS school in the corresponding year, according to the enrollment history files.[9]

Our identification strategy is premised on the assumption that students enrolling in schools that did and did not adopt restorative practices would have exhibited parallel trends in relevant outcomes in the absence of the rollout of the restorative practices treatment, conditional on controls for the set of student socio-demographic characteristics described below. An extensive recent literature has highlighted that estimators derived from standard two-way fixed effects models employed to identify treatment effects in settings with multiple treated groups and staggered rollout of treatment are unbiased only if treatment effects are homogeneous across time and group (Sun and Abraham, 2020; Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfoeuille, 2020). In practice, however, there are a number of reasons to hypothesize that the effect of exposure to restorative practices may vary with intensity (i.e., number of years) of exposure as well as the timing of introduction. First, student outcomes may be a function of cumulative exposure to restorative practices to the

[8] As such, elementary school students are assigned to the first CPS elementary school in which they enrolled, while high school students are assigned to the first CPS high school in which they enrolled. Since enrollment records are unavailable prior to SY09, we assign students enrolled in CPS prior to SY09 to schools based on their SY09 enrollment record.

[9] We exclude the following observations from this sample: students in grades K-2, students who have progressed to grade levels not offered by their initial schools, students past their expected school exit year; and any observations beyond our event study window (-4 to +3 years since treatment) from students who attended treatment schools.

extent that behavioral changes take time to manifest. Second, teachers' disciplinary practices and school climate more generally may evolve over time as the core principles of restorative practices become more ingrained. Third, the refinement of restorative practices programming over time may generate treatment effect heterogeneity as a function of the timing of its introduction. This anticipated treatment effect heterogeneity (which is ultimately borne out in the data) implies that standard two-way fixed effects models are inappropriate for our study setting. The resultant bias arises at least in part from the fact that standard two-way fixed effects models rely on already-treated groups when constructing counterfactuals; to the extent that changes in outcomes in these already-treated groups are themselves partly driven by the dynamic effects of the treatment, this comparison introduces bias. As shown in Sun and Abraham (2020), even event study models that separately estimate the effects of treatment as a function of treatment timing will be biased in the presence of cross-group treatment effect heterogeneity.

To test our identifying assumptions and estimate the causal effect of restorative practices in the presence of heterogeneous treatment effects, we rely on an estimator derived in de Chaisemartin and D'Haultfoeuille (2020), which is designed to produce unbiased estimates of the average effect of treatment on the treated (both averaged across post-treatment periods and separately by treatment timing) when such heterogeneity is present. This estimator uses only not yet-treated groups (students assigned to not-yet-treated schools, in our study setting) to predict counterfactual outcomes and so ensures that treatment effect estimates are not contaminated by treatment-induced changes in outcomes in already-treated groups.

To formally characterize the de Chaisemartin and D'Haultfoeuille (2020) estimator in the context of our study setting, we define $D_{i,g,t}$ as an indicator for restorative practices exposure of student i with assigned school g in school year t. We classify each school as exposed to the restorative practices treatment in all years after its introduction; in practice, we cannot measure whether restorative practices continued to be employed in subsequent years.[10] Following the notation from the authors' derivation, we define $N_{g,t}$ as the number of observations corresponding to school g in school year t and we define $N_{d,d',t} = \sum_{g:D_{g,t}=d,D_{g,t-1}=d'} N_{g,t}$ as the total number of students assigned to schools in school year t that had treatment value d' in school year $t-1$ and treatment value d in school year t. Next, we define:

$$(1) \quad DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

This expression returns a weighted average of the difference between the change in

[10] To the extent that a subset of schools transitioned away from restorative practices, our treatment effect estimates will consequently represent lower bounds on the true causal impact of persistent RP exposure.

outcomes between school year $t-1$ and t in schools first treated in school year t and the change in outcomes between $t-1$ and t in schools untreated through school year t. As shown in de Chaisemartin and D'Haultfoeuille (2020), we can then take a weighted average of $DID_{+,t}$ across all school years from $t=2$ to $t=T$ (where T is the final school year in the study sample) to produce an unbiased estimator of the average treatment effect in the first post-treatment school year of all schools that become treated during the sample period. Specifically, the weighted average is constructed as follows:

$$DID_M = \sum_{t=2}^{T}\left(\frac{N_{1,0,t}}{N_S} DID_{+,t}\right) \tag{2}$$

where N_S is the total number of observations corresponding to students in the year that their assigned school is first treated. Finally, we employ this same approach to construct treatment effect estimates specific to the number of school years since initial treatment exposure and, alternatively, as a function of the number of school years until initial exposure. These latter placebo estimates can then be used to evaluate the parallel trends assumption, as in the standard event study framework.

Turning back to the study setting, one key challenge is that we are interested in analyzing changes for a wide range of outcomes in response to the introduction of restorative practices. Since the parallel trends assumption must be evaluated for each outcome of interest, we present event study plots for all outcomes subsequently analyzed in our main tables. Following the notation used above, $Y_{i,g,t}$ is the outcome of interest for student i who was first enrolled in school g and is being observed in school year t. The restorative practices curriculum was provided across grade levels within adopting schools. Consequently, $D_{g,t,l}$, our treatment measure, is an indicator defined by whether restorative practices were introduced in school s exactly l years after school year t (or $|l|$ years before for negative-valued l). Based on the structure of our data, we can observe high school students who move up one grade level each year for a maximum of four school years prior to and three school years subsequent to initial treatment exposure. All event studies consequently present estimates constructed over this horizon.

Across analyses, our benchmark models also include the following student-level covariates: age fixed effects,[11] cohort fixed effects,[12] gender fixed effects, race fixed effects (students who identify as African American, Latinx, White, Asian, or other races), an indicator for

[11] Age is defined as the student's age by June 20 of the last calendar year of the school year (the last possible end date for a school year).

[12] Cohort defines a set of grade levels and school years corresponding to the same set of students in the absence of entry/exit or grade retention (i.e., one cohort includes first grade students in 2010-2011, second grade students in 2011-2012, etc.).

homelessness, an indicator for whether the student is enrolled as an English Learner, indicators for student disability classification (students with a 504 plan,[13] cognitive disability, or physical disability), an indicator for being on an individualized education plan, and an indicator for whether the student is on a free or reduced lunch plan.[14] In practice, the inclusion of these covariates improves the precision of estimates, but does not alter the basic pattern of findings or our conclusions regarding the validity of the parallel trends assumption that underpins the research design. To account for the school-level nature of treatment assignment, standard errors are clustered at the level of the school in which each student first enrolled across analyses.

The event study plots presented in Figures 1 and 2 provide support for the parallel trends assumption with respect to key outcomes of interest for the high school subsample. In subsequent analyses, we combine estimates of the instantaneous and dynamic effects of restorative practices exposure to produce a single estimate of the causal effect of treatment on the treated for each outcome of interest. To do so, we construct the following estimator of the average cumulative effect of restorative practices over $k+1$ treatment periods (where $k = 3$, as discussed above):

$$\hat{\delta}_{+,0:k} = \sum_{l=0}^{k} \omega_{+,k,l} DID_{M,l} \tag{3}$$

Here, $DID_{M,l}$ is defined analogously to DID_M and captures the weighted average effect of treatment l periods after initial treatment exposure. $\omega_{+,k,l}$, the weight assigned to the treatment effect l periods after initial treatment exposure, is defined as $\frac{N_l^1}{\sum_{l=0}^{k} N_l^1}$, where N_l^1 is the number of students reaching l school years after initial treatment exposure by the end of the study period (year T, corresponding to school year 2019).

VI Main Results: Behavior Changes

We seek to understand the role that school behavioral policies may play in shaping child behavior. Specifically, we examine the shift from more punitive practices to more restorative practices in response to perceived undesired student conduct.

Changing behavior inside of school? First, we examine the impact of the introduction of restorative practices on in-school behavioral outcomes. Figure 1 shows a trend for out-

[13] Having a 504 plan is not disability, but reflects the need for accommodations in the absence of diagnosed learning disabilities and special instruction requirements. For the purposes of selective enrollment, students with a 504 plan are considered disabled.

[14] In specifications that employ absent days as the outcome of interest, we also include yearly total "member days" as a control. Member days represent the total number of days that a student was present in a school and days that the student was absent from the school.

of-school suspensions that is statistically indistinguishable from zero before treatment and then indicates a pattern of decreases in out-of-school suspensions over time after treatment. This translates to an average decrease of out-of-school suspensions by 2.1 percentage points (Table 2, column 1).[15]

There are not corresponding changes that are statistically distinguishable from zero in in-school suspensions and days absent, which suggests that students are not getting less in-school instruction time on average (Figure 1; Table 2, columns 2 and 3).[16]

Changing behavior outside of school? We are interested in understanding whether being exposed to restorative justice affects conflict resolution. If being exposed to restorative practices is changing student behavior, then one should expect to see changes in behavioral outcomes outside of school. One proxy of measuring this is through effects on arrests. To explore this, we draw from Chicago Police Department data which has information on person-level arrests.[17] On average, arrests decreased by 0.7 percentage points, or by 10 percent relative to the baseline mean (Table 3, column 1). In Figure 1, we show pre-trends that are not statistically different from zero for child arrests followed by a declining pattern in arrests after the introduction of RP.[18]

In school, an adult determines the referrals to law enforcement, so decreases in arrests could correspond to dictates to reduce overall punitiveness corresponding to the introduction of RP. For example, if a teacher or principal is discouraged from referring kids for arrests, the arrest numbers would mechanically go down. However, if the arrests are decreasing outside of school, then that increases confidence that student behavior is actually changing. To understand whether this reduction in arrests was driven by those arrests that are made on school property and during school hours or outside of schools, we analyzed the impacts on arrests by geographic location and time. We see that arrests conducted both in school (i.e. incidents happening both inside school location and during school hours) and outside of school (i.e. incidents happening either outside school location or outside school hours) decreased (Table 3, columns 2 and 3, respectively), with much larger declines among in-school arrests (24 percent versus eight percent based on baseline means). This suggests that behavior changed beyond the school context into the neighborhood context, but it changed

[15] We see a decline in out-of-school suspensions for elementary school students using our main specification (Appendix Table A3), but the statistical significance is sensitive to specification choice.

[16] We see a similar effect for in-school suspensions among elementary school students (Appendix Table A3, Appendix Figure A2).

[17] Drawing from a data source independent from Chicago Public Schools increases our confidence that the results are not driven by explicit instruction of teachers to be less punitive. Police officers in the neighborhood context are not likely to be under the same authority as teachers in the local school.

[18] We see a decline in the likelihood of being arrested for elementary school students using our main specification (Appendix Table A4), but the direction and statistical significance of this estimate are sensitive to specification choice.

more than within the school context.[19]

A broader question is whether a restorative justice approach to conflict can decrease violence.[20] To explore this question, we examined changes in arrests for violent and non-violent offenses. We see that arrests decreased for both types of offenses (Table 3, columns 4 and 5), but they decreased more for non-violent offenses. This suggests that while they may aid in the reduction of violence, restorative practices may have more efficacy in ameliorating non-violent conflict.

VII Potential Mechanisms

School climate and student learning. We see that the introduction of restorative practices resulted in a decrease in out-of-school suspensions (Table 2, column 1). This could be happening for a few reasons. On the one hand, RP could be having a positive, productive impact on adult behavior (how they interact with students, maybe making them understand the whole student, not jump to escalation) and/or student behavior (they are learning how to resolve conflicts, they feel more understood by the adults and their peers, etc). On the other hand, the effect could be mechanical where teachers are under explicit instruction not to suspend students. If it were the former, we may expect to see increases in perceptions of school climate. If it were the latter, a concern/criticism could be that these students who were previously deemed as conducting undesirable behavior by adults in the school will cause disruptions and we should then see decreases in achievement, in terms of learning measures.

In Table 4, we show results on student-reported measures of school climate and on student learning. We see increased perceptions of school climate by 0.021 standard deviations. This includes measures of students reporting whether they feel that they belong at school, that their teachers trust them, that the school is a safe place, and that they have emotional health.

However, we find no consistent, statistically significant impacts on GPA or reading or math scores (Table 4), which suggests that the shift away from punitive, incapacitation-focused disciplinary responses does not seem to have been detrimental to the learning outcomes of the broader student body. The lack of consistent impacts on student academic outcomes suggests that this concern may not be of first-order importance in the study setting.[21]

[19] This in-school effect could be the net impact if school officials within the school were discouraged from arresting students and if students themselves changed their behavior.

[20] Such reforms are being experimented within criminally-accused situations. Results from our setting could inform practices in contexts separate from schools.

[21] We also see null effects among elementary school students (Appendix Table A5) with similar impacts on GPA. The actual estimates for the reading and math test effects are sensitive to specification choice however.

Implementation matters. Restorative practices comprise a wide range of implementation approaches, which can then make it harder to replicate and scale. To understand what specific set of practices were most effective, we explored differential impacts for the two predominant practices implemented: RP Coaching and RP Leadership.

RP Coaching consists of 10-20 week programs where an RP Coach trains administrators and designated staff to model and implement restorative practices within their school. This flexible model is designed to meet schools' needs and abilities in developing a menu of restorative practices, creating a restorative mindset and language among school staff, implementing improved disciplinary protocols using proactive strategy, and increasing staff-to-staff and staff-to-student trust. This is the most intensive implementation model.

Through the RP Leadership Schools Cohort, OSEL supports schools in developing their internal leadership capacity for building a sustainable school-wide system that systematically and restoratively builds community and addresses behavior. Over the course of the year, school-based Classroom and Discipline Leads participating in the cohort receive intensive training and on-going coaching. By meeting program requirements, Classroom and Discipline Leads receive Restorative Practices Leadership Certification.

In Appendix Table A6, we show suggestive evidence that the results are mainly driven by RP Coaching approach, the more intensive of the RP practices.

VIII Treatment Heterogeneity

We see suggestive evidence that some of the impacts may be driven by improvements in outcomes for particular students.

Suspensions. The absolute reductions in out-of-school suspensions may be particularly driven by declines in out-of-school suspensions of Black females (-4.0 percentage points), Black males (-3.8 percentage points), White males (-2.5 percentage points), and White females (-2.0 percentage points) (Appendix Table A7, column 1). The absolute gains are also concentrated among native English speakers (Appendix Table A8, suggesting that RP implementation, which can be nuanced and requires clear communication between parties, may have been better translated to those who were fluent in the instructional medium. The impacts are greater for those students who recently made the transition into the high school context, with suggestive absolute declines in out-of-school suspensions by 2.8 percentage points for 9th and 10th graders as compared to absolute declines by 1.7 percentage points for 11th and 12th graders (Appendix Table A9). This could be because students new to a context are less likely to have preset defaults of how to act and are more likely to adopt the new regime, or in other words, it is harder for students who have already conformed to existing norms to adopt new norms. The results are driven by students without documented

disabilities (Appendix Table A10).

There are suggestive decreases in absent days, netting out in-school suspension days, of Black males and of students in their early high school years (ninth and tenth graders), which would indicate that these students were likely receiving more instructional time following the introduction of RP (Appendix Tables A7 and A9).

Arrests. We wish to disentangle whether the underlying mechanism is a change in how youth resolve conflicts or a change in how adults perceive youth behavior (and therefore may be more or less likely to assume misconduct). In Appendix Table A7, we show suggestive evidence that these policing results are being driven by absolute reductions in arrests of young Black males (-1.4 percentage points) and young Hispanic males (-0.9 percentage points), with smaller but significant reductions for Black females (-0.8 percentage points) and White females (-0.6 percentage points). The absolute declines in arrests are driven by native English speakers (Appendix Table A8, suggesting that improved conflict resolution may be facilitated by linguistic homophily, and by younger high school students (Appendix Table A9), suggesting that it is easier to change the behavior of children when they enter a new social context as opposed to those who have already established norms in a context. The absolute decreases in arrests are most prevalent for students with documented disabilities (-1.2 percentage points), but they are also present for those without documented disabilities (-0.6 percentage points) (Appendix Table A10). Future work would examine how the introduction of RP influenced adult attitudes and behavior towards children.

IX Sensitivity Analysis: Alternative Specifications

We check the sensitivity of our results to using different empirical approaches and specifications. Overall, results remain similar to our main estimations.

Standard difference-in-differences empirical approach. Instead of using the de Chaisemartin and D'Haultfoeuille (2020) empirical approach, we use a standard difference-in-differences design. The results remain similar to the effects estimated in our benchmark specifications that use de Chaisemartin and D'Haultfoeuille (2020) (Appendix Tables A11 and A12, Panel A).

School-by-cohort fixed effects. In our main specification that follows de Chaisemartin and D'Haultfoeuille (2020), we use school fixed effects and cohort fixed effects separately. We change this to look at school-by-cohort fixed effects instead, and the results on behavioral outcomes and perception of school climate are similar to our main specifications. However, our findings on testing outcomes are not as robust: while the effects on GPA and reading scores remain insignificant, we find a small but significant negative effect on math test scores (Appendix Tables A11 and A12, Panel B).

Excluding charter schools. For our main specification, we include all observations of students who were enrolled in both neighborhood and charter schools in the corresponding school year. To restrict the sample to students who stayed in traditional neighborhood schools, we exclude observations of students who ever attended a charter school. The results on behavioral outcomes remain largely unchanged. Impacts on perceptions of climate are positive but become imprecise, but the findings on learning outcomes remain similar (Appendix Tables A11 and A12, Panel C).

X Conclusion

Historically, parents have sent their children to school with an implicit trust that the policies and practices of a school, if implemented properly, would necessarily result in the best outcomes, not only for their children but also for society. School officials themselves, however, struggle with the decision as to which policies are necessarily optimal, particularly when establishing safety and disciplinary systems. Schools tend to be risk-averse, and the inherently "safe" option is to have no tolerance for any breaches of what is considered to be appropriate conduct. On the other hand, by enforcing an overly retributive system, schools may be inadvertently cultivating a less tolerant society and exacerbating already stark disparities for students from disadvantaged backgrounds. The lack of clarity regarding the costs and benefits of a more or less punitive system necessitates a rigorous evaluation of different school policies and practices that are implemented with the intention of improving behavior and increasing safety of the school.

A rigorous examination of educational and behavioral impacts from changes in restorative practices and socioemotional learning programming can shed light on this issue. By understanding the immediate and future impacts of school disciplinary practices, we can more deeply understand the role schools play in influencing children's present and future behavior and preferences.

We study the causal impact of the rollout of restorative practices in Chicago Public Schools. We use cross-school variation in the timing of the introduction of RP to understand how adoption of a restorative approach affects students' academic and disciplinary outcomes, as well as their involvement with the criminal justice system. Our evidence suggests that the introduction of RP in CPS high schools reduced the likelihood of out-of-school suspensions by 13 percent and reduced the likelihood of arrest by 10 percent. We identify sizable declines in both in-school and out-of-school arrests, indicating that the changes in disciplinary outcomes we identify reflect genuine changes in underlying behavior rather than solely changes in how teachers and school administrators respond to behavioral challenges. Consistent with this hypothesis, we identify significant improvements in perceived school climate in response to

the introduction of RP. We do not find any robust evidence that RP significantly impacts student grades or test scores.

Turning to treatment effect heterogeneity, we find that absolute declines in the likelihood of out-of-school suspensions and arrests are particularly large among Black students, who face the highest suspension and arrest rates at baseline. Though less precise, point estimates also indicate that Black students experience the most positive absolute changes in their perceptions of school climate in response to the introduction of RP. Taken together, our findings indicate that RP interventions like those we evaluate have the potential to meaningfully impact those students most exposed to punitive disciplinary practices at baseline. Future research should examine the longer-term implications of these changes in disciplinary outcomes with regards to high school completion, post-secondary enrollment and future criminal justice system involvement.

References

Acosta, Joie, Matthew Chinman, Patricia Ebener, Patrick S. Malone, Andrea Phillips, and Asa Wilks. 2019. "Evaluation of a Whole-School Change Intervention: Findings from a Two-Year Cluster-Randomized Trial of the Restorative Practices Intervention." *Journal of Youth and Adolescence*, 48(5): 876–890.

Augustine, Catherine, John Engberg, Geoffrey Grimm, Emma Lee, Elaine Wang, Karen Christianson, and Andrea Joseph. 2018. *Can Restorative Practices Improve School Climate and Curb Suspensions? An Evaluation of the Impact of Restorative Practices in a Mid-Sized Urban School District.* RAND Corporation.

Callaway, Brantly, and Pedro Sant'Anna. 2020. "Difference-in-Differences with multiple time periods." *Journal of Econometrics, forthcoming.*

Chicago Public Schools. 2020. "Curriculum, Instruction, and Digital Learning."

de Chaisemartin, Clément, and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 1688–1699.

Fabelo, Tony, Michael D Thompson, Martha Plotkin, Dottie Carmichael, Miner P Marchbanks, and Eric A Booth. 2011. "Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement." *New York: Council of State Governments Justice Center.*

González, Thalia. 2012. "Keeping Kids in Schools: Restorative Justice, Punitive Discipline, and the School to Prison Pipeline." *Journal of Law and Education*, 41(2): 56.

Gregory, Anne, Francis L. Huang, Yolanda Anyon, Eldridge Greer, and Barbara Downing. 2018. "An Examination of Restorative Interventions and Racial Equity in Out-of-School Suspensions." *School Psychology Review*, 47(2): 167–182.

Hopkins, Belinda. 2003. *Just schools: A whole school approach to restorative justice.* Jessica Kingsley Publishers.

Illinois Network of Charter Schools. 2020. "Charter Public Schools Facts & Info."

Karp, David R, and Beau Breslin. 2001. "Restorative justice in school communities." *Youth & Society*, 33(2): 249–272.

Losen, Daniel, Damon Hewitt, and Ivory Toldson. 2014. "Eliminating Excessive and Unfair Exclusionary Discipline in Schools Policy Recommendations for Reducing Disparities." *Policy Recommendations for Reducing Disparities*, 16.

Sartain, Lauren, Elaine M Allensworth, and Shanette Porter. 2015. "Suspending Chicago's Students." *The University of Chicago Consortium on Chicago School Research*, 74.

Shollenberger, Tracey L. 2015. "Racial disparities in school suspension and subsequent outcomes." *Closing the school discipline gap: Equitable remedies for excessive exclusion*, 31–44.

Stevens, W David, Lauren Sartain, Elaine M Allensworth, and Rachel Levenstein. 2015. "Discipline Practices in Chicago Schools." *The University of Chicago Consortium on Chicago School Research*, 52.

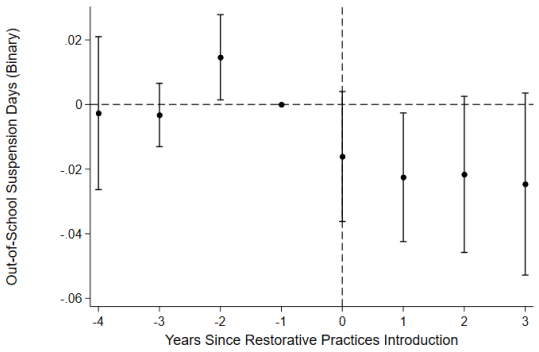
Sun, Liyang, and Sarah Abraham. 2020. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics, forthcoming*.

Wolf, Kerrin C., and Aaron Kupchik. 2017. "School Suspensions and Adverse Experiences in Adulthood." *Justice Quarterly*, 34(3): 407–430.

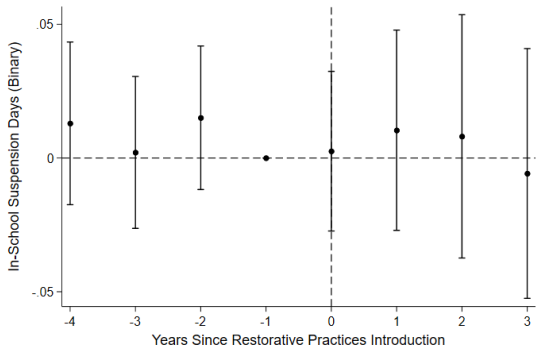
XI Main Figures

Figure 1: Event Studies: Behavioral Outcomes

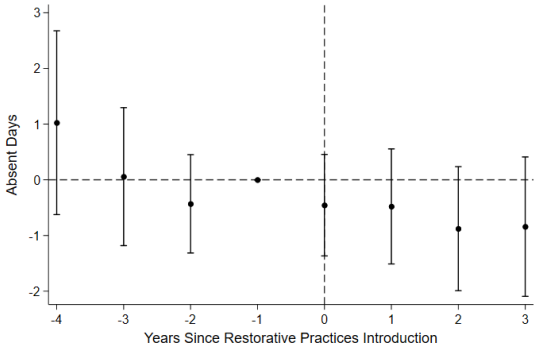
(a) Out-of-School Suspensions



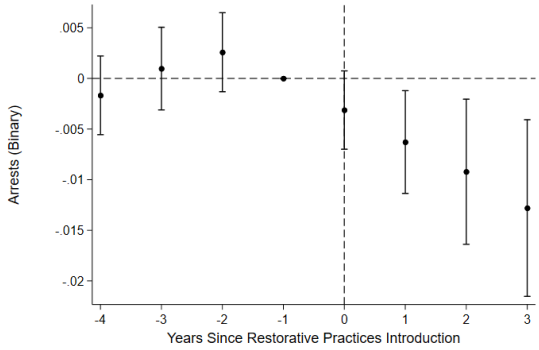
(b) In-School Suspensions



(c) Absent Days

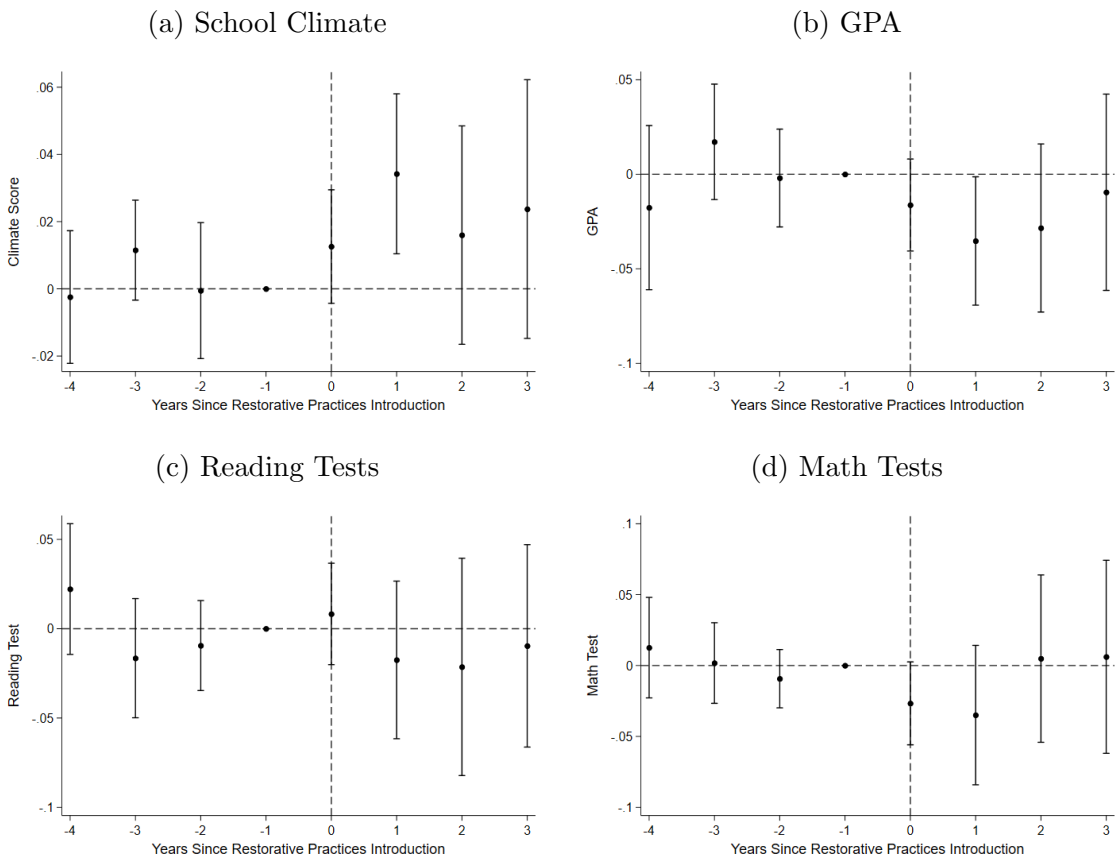


(d) Arrests



Notes: These figures show the event studies around the introduction of RP on students' in-school behavioral outcomes (out-of-school suspensions and in-school suspensions) over time. The effects are shown for overall student samples, as well as by grade and by race/gender. This includes students observed in SY09-SY19 as 9th-12th graders. Suspension data are collected by Chicago Public Schools. An out-of-school suspension is defined as the removal of a student from class attendance or school attendance. An in-school suspension is defined as the removal of a student from their regular educational schedule for more than 60 minutes of the school day to an alternative supervised setting inside the school building. Arrest data are collected by the Chicago Police Department. The main arrest outcome is defined as any number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest. Absence is defined as the total number of days absent, minus the total number of out-of-school suspension days that a student had in the school year, regardless of school.

Figure 2: Event Studies: School Climate and Learning



Notes: Panel A shows the impact of the introduction of RP on students' perceptions of school climate over time. These include students observed in SY09-SY19 as 9th-12th graders. The School Climate index measures student SEL levels and supportive environments. It includes the following constructs: Peer Support for Academic Work, Emotional Health, Academic Engagement, Human and Social Resources in the Community, Student Classroom Behavior, Academic Personalism, Parent Supportiveness, Psychological Sense of School Membership, Safety, School-Wide Future Orientation, School Safety, and Student-Teacher Trust. The GPA outcome is calculated using exclusively semester final grades from SY09 up to SY19. The GPA outcome in this study is the mean of the numeric grades (equivalent to the letter grades) registered in the data for all for-credit courses. Numeric grades are calculated as follows: A is equivalent to 4; B, to 3; C, to 2; D, to 1; and F, to 0. The data does not allow us to differentiate a F grade from a pass/fail or a for-credit course, so all F grades are counted as grades of for-credit courses and included in our data. It is important to note that the GPA used in our analyses is not the same GPA the students see on their transcripts because the schools use a different procedure to calculate the GPA.

Math and reading scores are standardized by test, school year, and grade.

XII Main Tables

Table 1: Baseline Characteristics: Chicago Public Schools High Schools

Variable	Non-Treated (1)	Treated (2)	Difference (3)
Number of Students	538.5814 (425.8919)	1003.6923 (774.8702)	465.1109*** (106.3646)
Out-of-School Suspension Days	0.7263 (2.6132)	1.0740 (3.2822)	0.3477** (0.1757)
In-School Suspension Days	0.2481 (1.2416)	0.4929 (1.7242)	0.2448** (0.1199)
Absent Days	16.1901 (18.6785)	23.2671 (23.0871)	7.0770*** (1.7019)
Ever Arrested	0.0738 (0.2614)	0.0943 (0.2923)	0.0206 (0.0170)
GPA	2.6150 (0.9870)	2.3380 (1.0159)	-0.2770** (0.1148)
English Learner	0.0530 (0.2241)	0.0706 (0.2561)	0.0175 (0.0136)
Students in Temporary Living Situations	0.0518 (0.2216)	0.0608 (0.2390)	0.0090 (0.0120)
Individualized Education Plan	0.1379 (0.3447)	0.1512 (0.3582)	0.0133 (0.0132)
Economically Disadvantaged	0.7857 (0.4104)	0.8217 (0.3828)	0.0360 (0.0373)
Gender: Female	0.5145 (0.4998)	0.4926 (0.4999)	-0.0219 (0.0153)
Race: African American	0.4765 (0.4995)	0.4205 (0.4936)	-0.0561 (0.0750)
Race: White	0.0830 (0.2758)	0.0989 (0.2985)	0.0159 (0.0333)
Race: Latinx	0.4020 (0.4903)	0.4392 (0.4963)	0.0373 (0.0644)
Race: Asian	0.0049 (0.0697)	0.0083 (0.0909)	0.0035 (0.0032)
Disability: Cognitive	0.1244 (0.3300)	0.1372 (0.3440)	0.0128 (0.0133)
Disability: None	0.8362 (0.3701)	0.8262 (0.3789)	-0.0100 (0.0110)
Disability: Physical	0.0106 (0.1025)	0.0113 (0.1055)	0.0006 (0.0018)
Disability: 504	0.0288 (0.1672)	0.0254 (0.1573)	-0.0034 (0.0048)

Notes: This table presents the baseline mean values of characteristics of non-treated high schools (column 1) to treated high schools (column 2), prior to the introduction of RP in SY14. Their associated differences (column 3) are based on regression with treatment indicator variable, and the standard errors are clustered at the school-level. The number of students is based on CPS student enrollment data in SY13. The rest of the variables in this table are based on CPS student-level data. Robust standard errors are reported with *** denoting statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 2: Restorative Practices and In-School Behavioral Outcomes

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)
RP	-0.021∗	0.004	-0.639
	(0.010)	(0.018)	(0.470)
Baseline Mean	0.158	0.134	19.250
Observations	1,176,280	1,176,280	1,176,280

Notes: The sample includes observations at the student-school year level. We exclude observations from this simple which are in one of the following criteria: students in K-2 grades; students who were enrolled for at least one day in a charter school in the corresponding school year; students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The elementary school sample includes 3rd-8th graders, and the high school sample includes 9th-12th graders. In columns 1 and 2, the out-of-school suspension (OSS) days and in-school suspension (ISS) days outcomes are the total number of ISS or OSS days that the student received in the corresponding school year, regardless of the school. In column 3, absent days is adjusted to total absent days minus out-of-school suspension days. And regressions for the absence days outcome include student member days of the corresponding school as a control. In column 4, GPA is calculated using only semester grades. In columns 4 and 5, math and reading scores are standardized by test, school year, and grade. In column 7, misconduct is the number of misconduct incidents in which a given student was involved in a given year. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, treatment school, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: ELL indicator, homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). The estimates are drawn from CPS data from SY09-SY19. Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table 3: Restorative Practices and Policing Outcomes

	Arrests Overall	In-School Arrests	Out-of-School Arrests	Violent Arrests	Non-Violent Arrests
	(1)	(2)	(3)	(4)	(5)
RP	-0.007**	-0.006**	-0.004*	-0.002*	-0.007**
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Baseline Mean	0.069	0.025	0.053	0.024	0.056
Observations	1,197,382	1,197,382	1,197,382	1,197,382	1,197,382

Notes: The sample includes observations at the student-school year level. We exclude observations from this sample which are in one of the following criteria: students in K-2 grades; students who were enrolled for at least one day in a charter school in the corresponding school year; students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The elementary school sample includes 3rd-8th graders, and the high school sample includes 9th-12th graders. Columns 1 and 2 report the average effect on students' arrests outcomes. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, treatment school, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: ELL indicator, homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). The estimates are drawn from CPD data from July 1st, 2008 to June 30th, 2019. Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table 4: Restorative Practices: School Climate and Learning Outcomes

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
RP	0.021+	-0.023	-0.009	-0.012
	(0.011)	(0.016)	(0.019)	(0.021)
Baseline Mean	-0.032	2.467	-0.070	-0.071
Observations	709,862	775,816	707,766	700,262

Notes: The sample includes observations at the student-school year level. We exclude observations from this simple which are in one of the following criteria: students in K-2 grades; students who were enrolled for at least one day in a charter school in the corresponding school year; students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The elementary school sample includes 3rd-8th graders, and the high school sample includes 9th-12th graders. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, treatment school, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: ELL indicator, homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). The estimates are drawn from MVMS scores from SY11-SY18 for 6th-12th graders, and are standardized by year and grade. Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Appendices

A Appendix Figures

Figure A1: Roll-out of Restorative Practices Programs in CPS

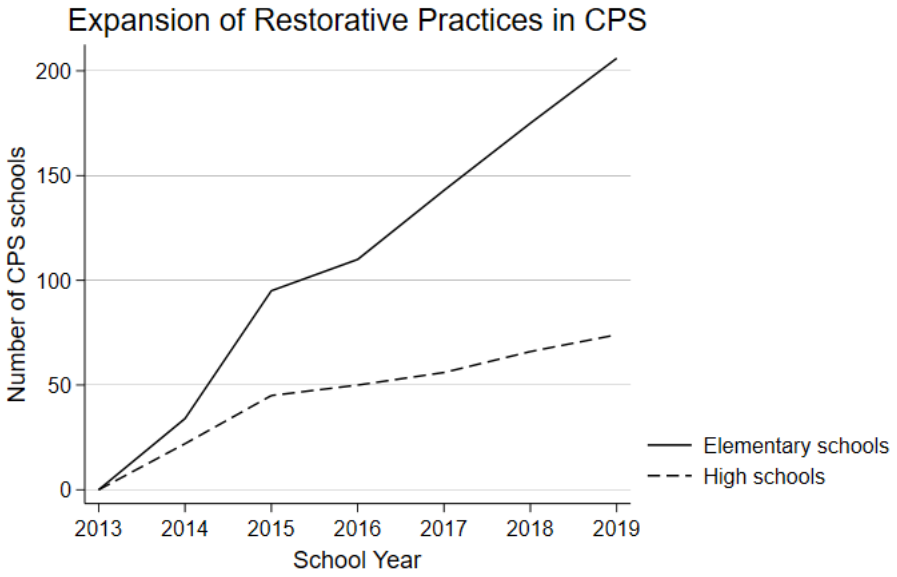
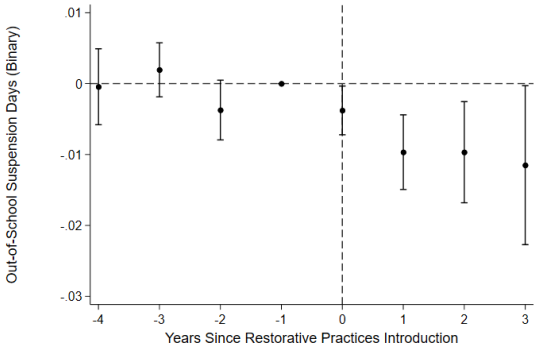
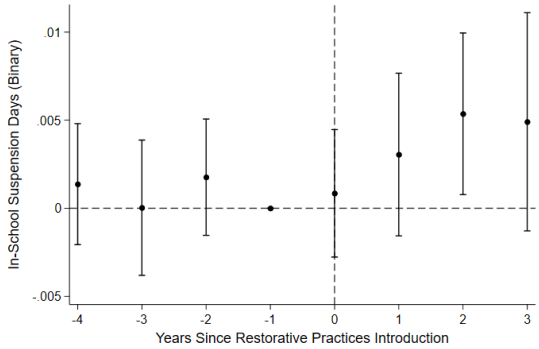


Figure A2: Elementary School Event Studies: Behavioral Outcomes

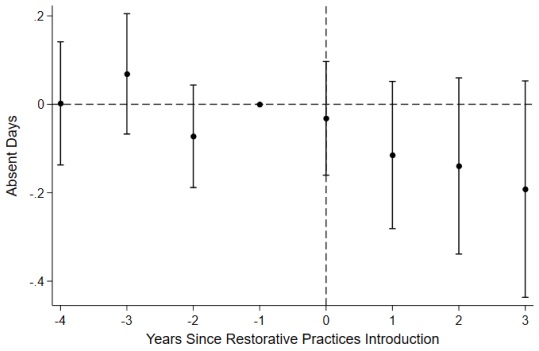
(a) Out-of-School Suspensions



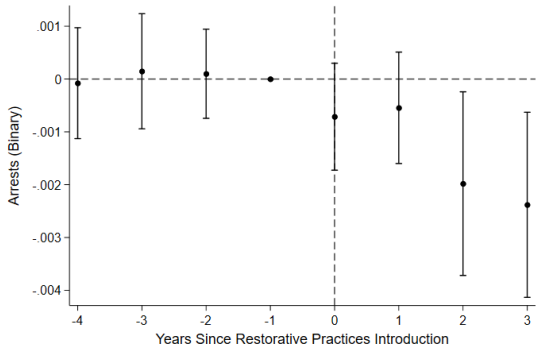
(b) In-School Suspensions



(c) Absent Days

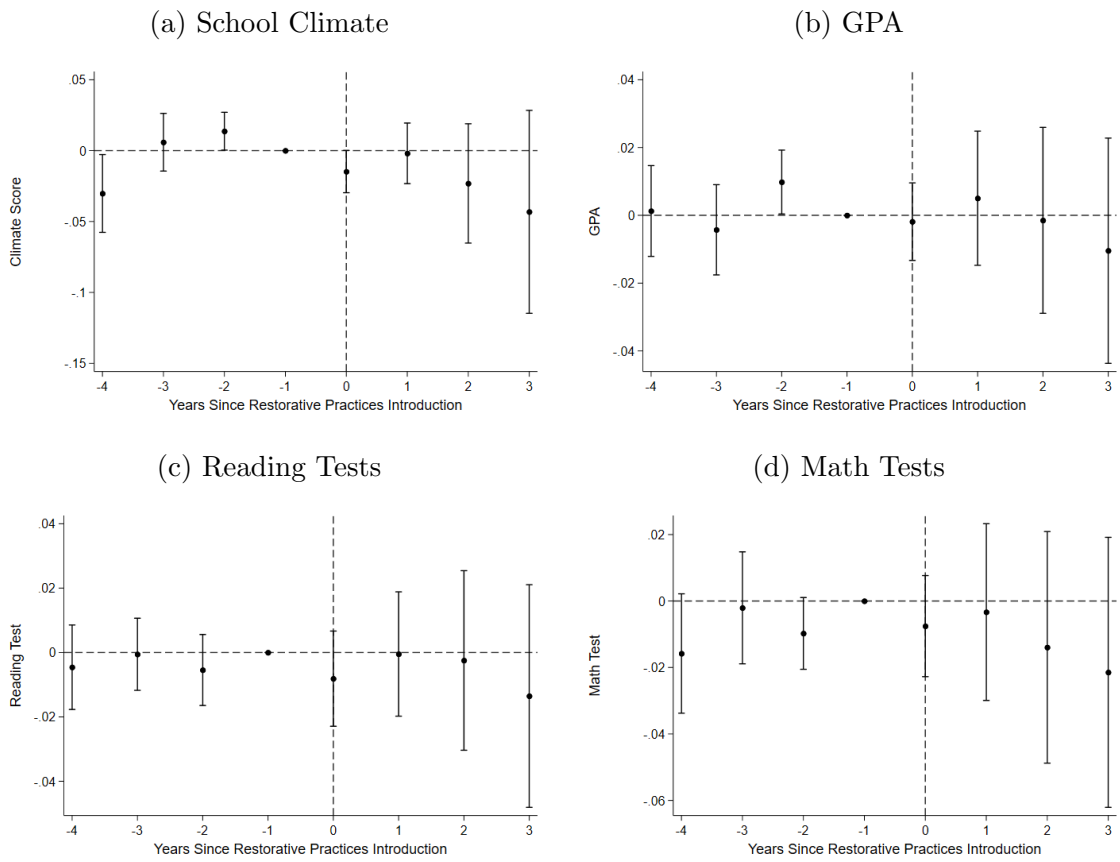


(d) Arrests



Notes: These figures show the event studies around the introduction of RP on students' in-school behavioral outcomes (out-of-school suspensions and in-school suspensions) over time. The effects are shown for overall student samples, as well as by grade and by race/gender. This includes students observed in SY09-SY19 as 9th-12th graders. Suspension data are collected by Chicago Public Schools. An out-of-school suspension is defined as the removal of a student from class attendance or school attendance. An in-school suspension is defined as the removal of a student from their regular educational schedule for more than 60 minutes of the school day to an alternative supervised setting inside the school building. Arrest data are collected by the Chicago Police Department. The main arrest outcome is defined as any number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest. Absence is defined as the total number of days absent, minus the total number of out-of-school suspension days that a student had in the school year, regardless of school.

Figure A3: Elementary School: Climate and Learning



Notes: Panel A shows the impact of the introduction of RP on students' perceptions of school climate over time. These include students observed in SY09-SY19 as 9th-12th graders. The School Climate index measures student SEL levels and supportive environments. It includes the following constructs: Peer Support for Academic Work, Emotional Health, Academic Engagement, Human and Social Resources in the Community, Student Classroom Behavior, Academic Personalism, Parent Supportiveness, Psychological Sense of School Membership, Safety, School-Wide Future Orientation, School Safety, and Student-Teacher Trust. The GPA outcome is calculated using exclusively semester final grades from SY09 up to SY19. The GPA outcome in this study is the mean of the numeric grades (equivalent to the letter grades) registered in the data for all for-credit courses. Numeric grades are calculated as follows: A is equivalent to 4; B, to 3; C, to 2; D, to 1; and F, to 0. The data does not allow us to differentiate a F grade from a pass/fail or a for-credit course, so all F grades are counted as grades of for-credit courses and included in our data. It is important to note that the GPA used in our analyses is not the same GPA the students see on their transcripts because the schools use a different procedure to calculate the GPA.

Math and reading scores are standardized by test, school year, and grade.

B Appendix Tables

Table A1: Baseline Characteristics: by Student and School Demographics

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	GPA (4)	Arrests (5)	Climate Index (6)
Native English Speakers						
Baseline Mean	0.180	0.135	18.453	2.463	0.073	-0.003
Observations	1,097,251	1,097,251	1,097,251	724,841	1,118,309	663,523
English Learners						
Baseline Mean	0.132	0.093	17.603	2.377	0.034	0.032
Observations	78,080	78,080	78,080	50,088	78,113	45,678
Grades 9 and 10						
Baseline Mean	0.200	0.145	16.808	2.393	0.078	-0.001
Observations	629,202	629,202	629,202	406,017	646,927	386,580
Grades 11 and 12						
Baseline Mean	0.148	0.117	20.344	2.531	0.062	-0.001
Observations	521,196	521,196	521,196	358,165	524,330	307,198
Any Disability						
Baseline Mean	0.226	0.160	21.981	2.259	0.098	0.001
Observations	214,344	214,344	214,344	131,061	220,428	115,024
504 Disability						
Baseline Mean	0.155	0.102	20.007	2.491	0.042	0.016
Observations	51,782	51,782	51,782	36,580	52,497	31,125
Physical Disability						
Baseline Mean	0.178	0.119	24.228	2.400	0.075	0.047
Observations	15,249	15,249	15,249	9,575	16,036	6,414
Cognitive Disability						
Baseline Mean	0.245	0.177	22.235	2.193	0.112	-0.006
Observations	146,660	146,660	146,660	83,808	151,254	76,412
No Disability						
Baseline Mean	0.168	0.127	17.750	2.494	0.065	-0.001
Observations	961,105	961,105	961,105	644,096	975,999	594,405

Table A2: Baseline Characteristics: by Student Race and Gender

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	GPA (4)	Arrests (5)	Climate Index (6)
Black Female						
Baseline Mean	0.245	0.183	21.813	2.422	0.063	-0.055
Observations	256,763	256,763	256,763	152,363	261,688	148,186
Hispanic Female						
Baseline Mean	0.083	0.061	17.084	2.657	0.012	0.027
Observations	263,189	263,189	263,189	176,571	266,259	169,068
White Female						
Baseline Mean	0.047	0.040	15.650	3.031	0.012	0.126
Observations	57,494	57,494	57,494	51,693	58,723	36,985
Asian Female						
Baseline Mean	0.014	0.014	9.013	3.288	0.003	0.122
Observations	6,442	6,442	6,442	5,523	6,546	3,243
Other Female						
Baseline Mean	0.017	0.017	9.668	3.317	0.006	0.117
Observations	17,187	17,187	17,187	15,378	17,342	11,464
Black Male						
Baseline Mean	0.304	0.241	21.014	2.022	0.183	-0.042
Observations	231,997	231,997	231,997	126,897	237,882	125,166
Hispanic Male						
Baseline Mean	0.146	0.094	16.466	2.299	0.053	-0.009
Observations	258,263	258,263	258,263	172,364	261,951	161,007
White Male						
Baseline Mean	0.097	0.075	15.370	2.689	0.039	0.093
Observations	53,817	53,817	53,817	47,997	55,583	33,591
Asian Male						
Baseline Mean	0.052	0.044	10.956	2.866	0.003	0.052
Observations	5,261	5,261	5,261	4,491	5,337	3,097
Other Male						
Baseline Mean	0.027	0.025	9.647	3.056	0.013	0.081
Observations	17,034	17,034	17,034	14,763	17,205	11,398

Table A3: Elementary Schools: Restorative Practices and In-School Behavioral Outcomes

	Out-of-School Suspensions	In-School Suspensions	Absent Days
	(1)	(2)	(3)
RP Any	-0.008**	0.003+	-0.105
	(0.003)	(0.002)	(0.075)
Baseline Mean	0.065	0.030	7.902
Observations	2,308,111	2,308,111	2,308,111

Notes: The sample includes observations at the student-school year level. We exclude observations from this simple which are in one of the following criteria: students in K-2 grades; students who were enrolled for at least one day in a charter school in the corresponding school year; students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The elementary school sample includes 3rd-8th graders, and the high school sample includes 9th-12th graders. In columns 1 and 2, the out-of-school suspension (OSS) days and in-school suspension (ISS) days outcomes are the total number of ISS or OSS days that the student received in the corresponding school year, regardless of the school. In column 3, absent days is adjusted to total absent days minus out-of-school suspension days. And regressions for the absence days outcome include student member days of the corresponding school as a control. In column 4, GPA is calculated using only semester grades. In columns 4 and 5, math and reading scores are standardized by test, school year, and grade. In column 7, misconduct is the number of misconduct incidents in which a given student was involved in a given year. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, treatment school, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: ELL indicator, homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). The estimates are drawn from CPS data from SY09-SY19. Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A4: Elementary School: Restorative Practices and Policing Outcomes

	Arrests Overall	In-School Arrests	Out-of-School Arrests	Violent Arrests	Non-Violent Arrests
	(1)	(2)	(3)	(4)	(5)
RP Any	-0.001∗	0	-0.001+	0	-0.001∗
	(0.001)	(0)	(0)	(0)	(0)
Baseline Mean	0.008	0.003	0.006	0.004	0.006
Observations	2,318,003	2,318,003	2,318,003	2,318,003	2,318,003

Notes: The sample includes observations at the student-school year level. We exclude observations from this sample which are in one of the following criteria: students in K-2 grades; students who were enrolled for at least one day in a charter school in the corresponding school year; students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The elementary school sample includes 3rd-8th graders, and the high school sample includes 9th-12th graders. Columns 1 and 2 report the average effect on students' arrests outcomes. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, treatment school, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: ELL indicator, homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). The estimates are drawn from CPD data from July 1st, 2008 to June 30th, 2019. Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A5: Elementary School: Restorative Practices: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
RP Any	-0.017	-0.001	-0.006	-0.01
	(0.013)	(0.009)	(0.01)	(0.012)
Baseline Mean	0.001	3.060	-0.028	-0.032
Observations	716,161	1,986,128	2,166,683	2,175,587

Notes: The sample includes observations at the student-school year level. We exclude observations from this simple which are in one of the following criteria: students in K-2 grades; students who were enrolled for at least one day in a charter school in the corresponding school year; students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The elementary school sample includes 3rd-8th graders, and the high school sample includes 9th-12th graders. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, treatment school, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: ELL indicator, homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). The estimates are drawn from MVMS scores from SY11-SY18 for 6th-12th graders, and are standardized by year and grade. Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A6: Implementation Type of Restorative Practices

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	Arrests (4)	Climate Index (5)
RP Coaching	-0.023∗	0.022	-0.826	-0.006+	0.024+
	(0.011)	(0.014)	(0.515)	(0.004)	(0.013)
Baseline Mean	0.186	0.148	20.304	0.081	
Num. Obs.	988,703	988,703	988,703	1,006,660	575,954
RP Leadership	0.019	0.004	-2.020	0.005	0.064+
	(0.022)	(0.056)	(4.288)	(0.038)	(0.038)
Baseline Mean	0.067	0.009	19.534	0.025	
Num. Obs.	192,956	192,956	192,956	196,298	103,413
RP Peer Council	0.003	-0.006	1.111	0.003	
	(0.021)	(0.055)	(3.763)	(0.004)	
Baseline Mean	0.075	0.056	23.842	0.025	
Num. Obs.	49,718	49,718	49,718	50,520	

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A7: Restorative Practices: Race-by-Gender Treatment Heterogeneity

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	Arrests (4)	Climate Index (5)
Black Female	-0.040**	-0.020	-0.758	-0.011	-0.008+
0.032+					
	(0.013)	(0.016)	(0.606)	(0.026)	(0.004)
(0.019)					
Hispanic Female	-0.006	0.016	-0.395	-0.024	-0.002
0.020					
	(0.006)	(0.024)	(0.709)	(0.025)	(0.002)
(0.014)					
White Female	-0.02*	-0.003	-0.738	0.002	-0.006+
0.001					
	(0.008)	(0.017)	(0.814)	(0.031)	(0.003)
(0.026)					
Asian Female	-0.009	0.003	0.161	-0.084	-0.005
0.037					
	(0.01)	(0.018)	(1.649)	(0.097)	(0.007)
(0.074)					
Other Female	-0.007	0.020+	-0.041	0.005	0.001
-0.026					
	(0.008)	(0.012)	(0.985)	(0.064)	(0.003)
(0.040)					
Black Male	-0.038*	-0.010	-1.769**	-0.014**	0.026
	(0.016)	(0.018)	(0.590)	(0.005)	(0.02)
Hispanic Male	-0.007	0.032	-0.106	-0.009*	0.006
	(0.011)	(0.026)	(0.513)	(0.004)	(0.011)
White Male	-0.025**	0.001	-0.508	-0.007	0.008
	(0.008)	(0.022)	(0.695)	(0.007)	(0.029)
Asian Male	-0.020	-0.015	-1.107	0.008	-0.022
	(0.033)	(0.025)	(1.959)	(0.009)	(0.139)
Other Male	0.000	0.021	0.635	-0.002	-0.042
	(0.011)	(0.018)	(0.920)	(0.008)	(0.039)

Table A8: Restorative Practices: English Learner Treatment Heterogeneity

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	Arrests (4)	Climate Index (5)
Native English Speakers	-0.023∗	0.001	-0.639	-0.008∗∗	0.024∗
	(0.009)	(0.02)	(0.466)	(0.003)	(0.011)
English Learners	0.014	0.041	-0.094	-0.006	-0.012
	(0.015)	(0.027)	(0.769)	(0.005)	(0.023)

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A9: Restorative Practices: Grade-Level Treatment Heterogeneity

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	Arrests (4)	Climate Index (5)
Grades 9 and 10	-0.028**	0.002	-0.835*	-0.009**	0.013
	(0.01)	(0.021)	(0.363)	(0.003)	(0.014)
Grades 11 and 12	-0.017*	0.011	-0.343	-0.004	0.029*
	(0.008)	(0.018)	(0.632)	(0.003)	(0.013)

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A10: Restorative Practices: Treatment Heterogeneity by Ability Status

	Out-of-School Suspensions (1)	In-School Suspensions (2)	Absent Days (3)	Arrests (4)	Climate Index (5)
Any Disability	-0.019	0.014	-0.672	-0.012**	0.002
	(0.013)	(0.022)	(0.531)	(0.005)	(0.019)
504 Disability	-0.019	0.012	-0.461	-0.013+	0.002
	(0.013)	(0.023)	(0.758)	(0.008)	(0.024)
Physical Disability	-0.018	0.027	-1.941	-0.022	0.012
	(0.021)	(0.026)	(1.787)	(0.019)	(0.068)
Cognitive Disability	-0.015	0.017	-1.08*	-0.012*	0.007
	(0.016)	(0.019)	(0.482)	(0.005)	(0.017)
No Disability	-0.023*	0.003	-0.619	-0.006**	0.023*
	(0.01)	(0.02)	(0.399)	(0.002)	(0.011)

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A11: Robustness: Behavioral Outcomes

	Out-of-School Suspensions	In-School Suspensions	Absent Days	Arrests Overall
	(1)	(2)	(3)	(4)
Panel A: Difference-in-Differences				
RP Any	-0.021∗	0.011	-2.17∗∗∗	-0.006∗∗∗
	(0.011)	(0.016)	(0.774)	(0.002)
Baseline Mean	0.158	0.134	19.250	0.069
Observations	1,121,056	1,121,056	1,121,056	1,148,890
Panel B: School-by-Cohort FEs				
RP Any	-0.020+	-0.005	-0.023	-0.009∗∗
	(0.011)	(0.018)	(0.403)	(0.003)
Baseline Mean	0.158	0.134	19.250	0.069
Observations	509,543	509,543	509,543	512,652
Panel C: Dropping Charter Schools				
RP Any	-0.02∗	0.016	-0.334	-0.008∗∗
	(0.01)	(0.019)	(0.532)	(0.003)
Baseline Mean	0.156	0.140	19.400	0.065
Observations	755,527	755,527	755,527	772,577

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.

Table A12: Robustness: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
Panel A: Difference-in-Differences				
RP Any	0.028**	-0.013	-0.010	-0.015
	(0.012)	(0.021)	(0.022)	(0.022)
Baseline Mean	-0.032	2.467	-0.070	-0.071
Observations	578,675	893,154	722,226	719,099
Panel B: School-by-Cohort FEs				
RP Any	0.018*	0.014	-0.01	-0.065**
	(0.009)	(0.02)	(0.011)	(0.023)
Baseline Mean	-0.032	2.467	-0.070	-0.071
Observations	310,054	343,817	199,188	203,070
Panel C: Dropping Charter Schools				
RP Any	0.012	-0.02	-0.037	-0.035+
	(0.012)	(0.016)	(0.023)	(0.021)
Baseline Mean	-0.030	2.477	-0.058	-0.054
Observations	482,123	735,989	442,507	436,709

Notes: Robust standard errors clustered by school are reported with ** denoting statistical significance at the 1 percent level, * at the 5 percent level, and + at the 10 percent level.