colorlinks=true, linktoc=all, linkcolor=gray,

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

May 2022
PRELIMINARY DRAFT

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. 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, both violent and non-violent, 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. For helpful feedback, we thank Roseanna Ander, Hellen Antonopoulos, Monica Bhatt, Jon Guryan, Ariel Kalil, Jens Ludwig, Martha Minow, Dick Murnane, Anita Wadhwa, and seminar participants at NBER, the National Academy of Education, the Spencer Foundation, and the William T. Grant Foundation. For excellent 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 William T. Grant Foundation. For data acquisition, we thank Chicago Public Schools and Chicago Police Department.

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; Bowles and Gintis, 1976). 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 Schechter, 2015).

Over the last five decades, school officials increased their use of exclusionary discipline, with the rate of school suspensions more than doubling for Black and Latinx children since 1974. 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 (Losen et al., 2015). Being in a stricter school can lead to long-term negative consequences such as decreased educational attainment, increased misconduct, and increased likelihood of engaging with the criminal justice system (Fabelo et al., 2011; Shollenberger, 2015; Wolf and Kupchik, 2017; Bacher-Hicks, Billings and Deming, 2019).

School officials report that classroom management and discipline represent the hardest parts of their jobs (Evertson and Weinstein, 2006; Kauffman et al., 2011). Thus, despite a growing awareness of the potential harms of suspensions, educators seek concrete responses to undesirable behavior, particularly in a context where 80 percent of schools report having incidents of violence, theft, or other crimes (Griffith and Tyner, 2019; Wang et al., 2020). 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 increasingly popular in schools. 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 (McCold and Wachtel, 1998; Fulkerson, 2001; Karp and Breslin, 2001; McGarrell, 2001; Hopkins, 2003; González, 2012; Angel et al., 2014; Wadhwa, 2015; Augustine et al., 2018; Gregory et al., 2018; Acosta et al., 2019; Shem-Tov, Raphael and Skog, 2021). 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 restorative justice in schools.

In this study, we examine how student-level short-run educational and behavioral outcomes 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, the effects may be mechanical because teachers were instructed to reduce the frequency of suspensions. Alternatively, 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. RP may teach students how to resolve conflicts more effectively, to understand their roles in conflicts, and to feel more understood by adults and 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. We identify significant decreases in child arrests both during school hours and on school grounds and outside of school. This evidence suggests that the introduction of restorative practices gener-

[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).

ated meaningful changes in underlying student behaviors and provides evidence that school practices may meaningfully shape socializing behaviors. Consistent with the theory that RP may shift school culture, we find an increase in student-reported measures of school climate perceptions, which include survey measures of perceived student classroom behavior, psychological sense of school membership, student-teacher trust, and school safety. These perceived improvements may also contribute to improvements in perceived behavior.[2]

A common concern is that reduced punitiveness in the absence of behavioral change may lead to increased classroom disruption. While we do not identify GPA or test score gains in response to the introduction of RP, negative but small and statistically insignificant average impacts suggest the shift toward restorative practices does not seem to have been overly detrimental to the learning outcomes of the broader student body, on average. The lack of evidence that reduced punitiveness leads to increased classroom disruption and worsened academic performance is reinforced by students' perceptions of improvements in classroom behavior.

In subsequent heterogeneity analyses, we present evidence that Black students, who experience the highest baseline rates of suspension and arrest, are those who benefit most from the introduction of restorative practices.

I Conceptual Framework: Shaping Student Behavior in Schools

School officials view classroom management and discipline as an important but difficult aspect of their roles. Their goal is to create an environment that is conducive for learning (Evertson and Weinstein, 2006; Kauffman et al., 2011). This involves responding to what they perceive as being undesirable behavior.

Consider a simple example involving two periods where there is an incident with three main student actors. In period 1, a student exhibits undesirable behavior ("offender") towards another individual ("victim"), where there are other students who passively or actively observe the incident ("bystanders").

In period 2, the school officials respond to the undesirable behavior. There are different goals for each student actor. First, there is the offender. The goal is that they are held accountable and that they learn the appropriate behavior for the future. Second, there is the victim where the goal of any response is that they feel "whole" again and that justice has been served. Third, there are the bystanders where the goal is for them to learn the appropriate behavior and be deterred from exhibiting the undesirable behavior in the future.

A common response by school officials involves exclusionary disciplinary practices, typically in the form of suspensions. At best, this removes the offender from a situation but

[2]We speak of "perceived" behavior because it may be that students are actually behaving in an undesirable way or it may be that adults are perceiving them to be behaving in an undesirable way.

neglects to impart desired behavior. It may give the victim a reprieve from interacting with the offender, but it remains unclear whether they feel justice. Justice itself necessitates a sense of accountability. Victims often report that offenders need to understand the harm that they caused in order for that offender to truly feel accountable for their actions. In the case of suspensions, if the offender is simply removed from a situation without understanding the harm they caused or how they made the victim (or their loved ones) feel, it is more difficult for them to take proper responsibility for their actions. The best-case scenario for the bystanders is that there is a deterrence effect and possibly a reprieve if the offender was causing disruptions to the learning environment.

At worst, the exclusionary response could be counterproductive to the long-term goals of school officials and perpetuate long-term harm through negative impacts on educational attainment or criminal legal system involvement (Fabelo et al., 2011; Shollenberger, 2015; Wolf and Kupchik, 2017; Bacher-Hicks, Billings and Deming, 2019).

School officials are increasingly aware of the negative consequences associated with a stricter school environment. Teachers, however, report needing concrete tools to meaningfully achieve justice and accountability in response to disciplinary situations without generating the potential harms related to exclusion.

This has led to the introduction of "restorative justice" (RJ). RJ is a philosophy rather than a specified set of practices and involves repairing harms between victims and offenders and restoring relationships, or transforming them in cases where there was not a pre-existing relationship. In RJ, the different stakeholders are being engaged through open dialogue with the goal of increased perspective taking and shared ownership of disciplinary justice. The concept originated in the criminal legal system and was first adapted to the school context in Australia. Increasingly, school districts across the U.S. have been introducing the RJ approach in order to purposively shift away from a more punitive atmosphere.

RJ can constitute a range of practices so it is typically referred to as restorative practices (RP) in the school context. It can manifest in different ways: restorative circles, peer juries, or peace rooms. Each agent has to agree to whatever the process is. For example, a victim will not be forced to participate if they feel that the process will re-traumatize them or if they do not want to discuss their experiences. It should mean a sense of justice for each party involved. It can be a conference between the offender and the victim, it could involve bringing two victims together who went through a similar experience, it could be an interaction between an offender and a victim's family or friend circle, or it could involve bringing together two people who committed similar offenses. In sum, it can look different in different settings and situations (McCold and Wachtel, 1998; Fulkerson, 2001; Karp and Breslin, 2001; McGarrell, 2001; Hopkins, 2003; González, 2012; Angel et al., 2014; Wadhwa,

2015; Augustine et al., 2018; Gregory et al., 2018; Acosta et al., 2019; Shem-Tov, Raphael and Skog, 2021).

II Policy Setting: Chicago Public Schools

We study the impact of restorative practices in partnership with the 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 (Chicago Public Schools, 2020).

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 levels due to 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. This spurred various policy changes through the student code of conduct which included removing suspensions as a disciplinary response for a certain tier of infractions,[3] limiting the length of suspensions for substance use infractions, 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; Lai, n.d).

[3]For grades three through twelve, out-of-school suspensions are now only permitted if a student's attendance endangers others, causes chronic/extreme interruption to others' participation in school, and prior interventions have been used. For students in kindergarten through grade two, central administration approval is required for any suspension.

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. This initiative was meant to not only give teachers alternative tools to suspension but also to improve the school environment itself. The district started by working with 22 high schools and 34 elementary schools in SY14. They expanded their RP programs to reach 279 schools (74 high schools and 205 elementary schools) by SY19.

The district offers three different programs to support high schools in the adoption of restorative practices. These programs include RP Coaching, RP Leadership, 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 each promote the development of student socioemotional learning (SEL) skills (self-reflection, empathetic listening, etc.) through collaboration between teachers, students, and staff and creation of non-judgmental safe spaces for communication. CPS had a grant from the U.S. Department of Justice (DOJ) to help with their initial rollout of RP programming.

The most intensive of these programs is RP Coaching, in which an RP coach trained administrators and designated staff to model and implement restorative practices within their school. Coaches were initially drawn from 15 different vendors with specialists who really understood restorative justice and how to adapt to different and dynamic school situations.[4] These coaches came to schools and met with teachers, school administrators, and other designated school staff two to three times a week every week for the academic year. 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 and that could adapt to evolving situations. Once the DOJ funding ran out in SY16, CPS reduced the number of vendors from which they drew and also reduced the frequency of coach engagement in schools to one day per week.[5] They also had to reduce the number of schools to which they could roll out RP programming.

The second program is RP Leadership, which entails a lighter touch intervention in schools. In RP Leadership, similar to RP Coaching, OSEL aims strengthen internal leadership capacity for building sustainable school-wide systems that restoratively develop community and address behavioral concerns. In these schools, they typically focused on training

[4]The longer that coaches stayed involved with a school, the more likely they would be incorporated as regular CPS school staff.

[5]Currently, half of the coaches come from CPS staff, many of whom originated from the original vendors from the initial RP rollout, and the other half come from a local vendor called Alternatives.

a smaller number of school administrators for a much shorter amount of time. 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, although we also examine heterogeneity by program intensity.

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), school size, and input from those directly working with the schools (network specialists).[6] 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 that expected *changes* over time in outcomes absent treatment 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 (CPD) 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 CPS's student-level administrative data from SY09 to SY19. These data files include information on student-level outcomes and demographic information. The outcome variables include test score and GPA measures, attendance records, and

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

records of in-school and out-of-school suspensions. In addition, these files include information on student race, gender, a proxy for economic disadvantage, disability status, and English learner status for those enrolled in CPS. The administrative data files include information on student-level enrollment history, which we link to the programming data files from the OSEL to construct a student-level measure of treatment exposure.[7]

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 investigate the impact of restorative practices training on school climate, we leverage 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, community environment, and parental supports.[8]

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. We separately investigate the impact of restorative practices by arrest type and by whether arrests took 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.[9] The CPD and

[7]CPS maintains a general student database in which each student is identified by a unique student ID. The distinct CPS administrative files are linked together by student ID. Using CPS administrative data, we construct a dataset for analysis which includes one row per year per student if the student was enrolled in any CPS school for at least one day in the corresponding year, according to the enrollment history data.

[8]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).

[9]It is important to note that the arrests data have no information about convictions, so included individuals may not have actually committed the criminal offenses for which they are arrested.

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

III.E Study Sample

Our analysis includes observations from students who were enrolled in any of 171 CPS neighborhood and charter high schools between SY09 and SY19 for at least one day.

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 and to distinguish behavioral from more mechanical responses to the introduction of RP. The second reason we focus our analysis on high school students is that student survey data on school climate, which permits us to investigate potential mechanisms driving estimated impacts on administrative outcomes, has limited elementary school coverage.

Before turning to the research design, we next present descriptive data on the characteristics of the CPS high schools and students included in our study sample. Specifically, Table 1 presents average characteristics for students enrolled in the 171 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.[10] This table shows that high schools that received restorative practices training differed from never treated high schools in several ways. These high schools are significantly larger schools (with about twice as many students enrolled). The treated high schools also used suspensions as disciplinary tools more intensely. Students who enrolled in subsequently treated schools had on average twice as many in-school suspension days (0.49 versus 0.25) and nearly 50% more more out-of-school suspension days (1.07 versus 0.73) than those enrolled in never-treated schools. Finally, students in treated high schools had more absent days and lower GPAs at baseline, as well as more negative perceptions of their school

[10] Appendix Tables A1 and A2 present average characteristics by demographic group and based on alternative sample partitions.

climates.[11] As noted, 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 than strict exogeneity-based) identifying assumptions.

IV 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 high school that each student attended within the CPS system, as well as the year and grade level in which that student enrolled in CPS.[12] To guide thinking, if student i attended high school g from 2010 to 2012, and then moved to high school g', the student's treatment exposure remains a function of the timing of RP rollout in school g, regardless of whether the first intervention in school g 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 high school in the corresponding year, according to the enrollment history files.[13]

Our identification strategy is premised on the assumption that students enrolling in schools that did and did not adopt restorative practices over a given period would have exhibited parallel trends in relevant outcomes in the absence of the rollout of the restorative practices treatment. 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.

[11] To ensure that our attendance measure is not mechanically correlated with our measure of OSS, we subtract the number of OSS days from total number of absences. ISS is not considered an absence because the student is still in a supervised setting inside their school.

[12] 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.

[13] We exclude the following observations from this sample: 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 assigned to treatment schools.

First, student outcomes may be a function of cumulative exposure to restorative practices to the 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. The fact that nearly half of CPS high schools are ultimately treated indicates that accounting for treatment effect heterogeneity is particularly important in our study setting.

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.[14] 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

[14]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.

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 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:

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

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, and $Y_{g,t}$ is the corresponding average outcome value for students assigned to school g 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 g exactly l years after school year t (or $|l|$ years before for negative-valued l).

Across analyses, our benchmark models also include the following student-level covari-

ates: age fixed effects,[15] cohort fixed effects,[16] gender fixed effects, race fixed effects (students who identify as African American, Latinx, White, Asian, or other races), an indicator for homelessness, an indicator for whether the student is enrolled as an English Learner, indicators for student disability classification (students with a 504 plan,[17] cognitive disability, or physical disability), an indicator for being on an individualized education plan, and an indicator for whether the student is eligible to receive free or reduced-price lunch.[18] In practice, the inclusion of these covariates improves the precision of estimates in some instances, but does not alter the basic pattern of findings nor 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, across analyses we cluster standard errors at the level of the school in which each student first enrolled.

The event study plots presented in Figures 2 and 3 provide support for the parallel trends assumption with respect to key outcomes of interest. 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 is set to 3 to avoid small cell sizes):

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

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 SY19).

[15] 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).

[16] 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 SY11, second grade students in SY12, etc.).

[17] Having a 504 plan is not indicative of a 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.

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

V 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 student misconduct.

Changing behavior inside of school? First, we examine the impact of the introduction of restorative practices on in-school behavioral outcomes. Figure 2 shows an event study plot for out-of-school suspensions that is indicative of growing declines in out-of-school suspensions in the years after initial treatment exposure. Aggregating instantaneous and dynamic estimates, we identify a decrease in out-of-school suspensions of 0.14 days, or 18 percent (Table 2, column 1). In contrast, estimated impacts on in-school suspensions and days absent are statistically indistinguishable from zero, with negative point estimates for days absent and estimates that are inconsistent in sign for in-school suspensions (Figure 2; Table 2, columns 3, 4, and 5). Taken together, these findings suggest that students are receiving weakly more in-school instruction time, on average.

Although our analysis is focused on high school students, we find parallel evidence of declining out-of-school suspension days for elementary school students with statistically insignificant impacts on in-school suspension days and days absent for these grade levels (Appendix Table A4, Appendix Figure A1).

Changing behavior outside of school? We are interested in understanding whether being exposed to restorative practices affects conflict resolution. To examine whether being exposed to restorative practices is changing student behavior rather than simply changing how adults in schools respond to student behavior, we draw on CPD arrest data. Police officers operating outside of schools are not under the same authority as teachers. Consequently, arrest records can be used to produce an independent measure of perceived student behavior.

In Figure 2, we show an event-study plot for number of arrests, which exhibits a relatively flat pre-trend followed by a decline in arrests that increases in magnitude with time since the introduction of restorative practices. The estimated aggregate impact is an average decrease of 0.02 arrests, which represents a 16 percent decline relative to the baseline mean (Table 3, column 1).[19]

While the estimated decline in arrests in response to the introduction of restorative practices is consistent with improved student behavior, school staff are ultimately tasked with making referrals to law enforcement. Consequently, decreases in arrests could still

[19]We also estimate a decline in the likelihood of being arrested for elementary school students (Appendix Table A5, column 1), though the coefficient is small in magnitude and not statistically distinguishable from zero.

reflect the fact that adults in schools are induced to reduce overall punitiveness in response to the introduction of RP. To distinguish between alternative explanations for the aggregate decline in student arrests, we next examine whether reductions in arrests were driven entirely by arrests made on school property and during school hours, or whether we also identify declines in arrests outside of school grounds or school hours that cannot be explained by changes in school staff referral behavior. In columns 2 and 3 of Table 3, we provide evidence that aggregate arrest declines reflect decreases in both in-school and out-of-school arrests (by 26.7 percent and 12.6 percent, respectively).[20] These findings provide evidence in support of the hypothesis that student behavior is responding to the introduction of restorative practices.

A broader question is whether a restorative justice approach to conflict can decrease violence.[21] To explore this question, we examined changes in arrests for violent and non-violent offenses. We see that percentagewise arrest declines are similar for both types of offenses (Table 3, columns 4 and 5), indicating that the protective effects of restorative practices may extend to relatively more serious violent offenses.

VI Potential Mechanisms

School climate and student learning. We saw that the introduction of restorative practices resulted in a decrease in out-of-school suspensions (Table 2, columns 1 and 2), and the evidence presented above on falling out-of-school arrests suggests that this effect is not simply the mechanical result of teachers being under explicit instruction not to suspend students. As such, estimated RP impacts likely reflect some combination of changes in adult behavior (for instance, how they interact with and understand students) and student behavior (for example, how students respond to conflict or to feeling more understood by adults in school and their peers). Consistent with the hypothesis that restorative practices engender genuine changes in staff and student attitudes and behaviors, we find significant improvements in student-reported measures of school climate (Table 4). Specifically, we identify a 0.021 standard deviation improvement in perceived school climate. This aggregate impact is driven by particularly large increases in students' perceptions of their peers' classroom behavior, in their psychological sense of school membership, and in levels of student-teacher trust.

Despite these improvements in school climate, we find negative, statistically insignificant impacts on a range of academic outcome measures, including GPA as well as reading and math test scores (Table 4). A common concern is that reduced suspensions of stu-

[20] In-school arrests are classified as incidents happening both inside the school location and during school hours while out-of-school arrests are incidents happening either outside the school location or outside school hours.

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

dents engaged in undesirable behaviors may disrupt student learning. While we do not identify any improvements in academic performance in response to the introduction of RP, the shift away from punitive, incapacitation-focused disciplinary responses does not seem to have been overly detrimental to the learning outcomes of the broader student body, on average. This basic conclusion is reinforced by student self-reports indicative of improved student classroom behavior.[22] Before examining treatment effect heterogeneity by student background to shed light on potential explanations for these insignificant average impacts on academic performance, we briefly consider the importance of program intensity in explaining our findings.

Implementation matters. Restorative practices comprise a wide range of implementation approaches, which can make it hard to replicate and scale successful models. To understand what specific set of practices was most effective, we explore differential impacts for the two predominant practices implemented in high schools: RP Coaching and RP Leadership.[23]

As discussed, RP Coaching is the more intensive of the two programs and involves an RP coach who trains administrators and designated staff to model and implement restorative practices and then meets regularly with staff throughout the school year. RP Leadership is a lighter touch intervention in which a smaller number of school administrators are trained for a much shorter amount of time.

In Appendix Tables A7 and A8, we show suggestive evidence that pooled RP impacts are mainly driven by the RP Coaching approach, the more intensive of the RP practices. However, it is important note that the relative infrequency with which schools have participated in the RP Leadership program means that Leadership program-specific treatment effect estimates are generally imprecise.

VII Treatment Heterogeneity

To understand the distributional implications of the aggregate impacts presented above and to shed light on potential mechanisms, we consider treatment effect heterogeneity with a particular emphasis on differential impacts by student race and gender.

Heterogeneity by Race and Gender The aggregate reductions in the number of out-of-school suspensions we estimate are driven by declines in out-of-school suspensions among Black male and female students, who experience declines of 0.307 and 0.276 suspension days, respectively (Appendix Table A9, column 1). In Appendix Table A9 (column 4), we show that Black male and female students similarly experience the largest absolute reductions

[22] We also find null effects among elementary school students (Appendix Table A6).

[23] We do not separately analyze impacts for RP Peer Council, which was introduced only in the last year of our study period (SY19).

in arrests (with estimated declines of 0.067 and 0.014 arrests, respectively). While Black students are most frequently suspended and arrested at baseline, these large absolute declines suggest that they may differentially benefit from the introduction of restorative practices on other dimensions as well. Indeed, we see a significant (1.77) decline in absent days among Black males, above and beyond the identified reduction in out-of-school suspension days.

Turning to academic outcomes, this increase in instruction time for Black males (and perhaps improvement in Black male students' sense of school membership resulting from less punitive punishment exposure) translates into significant math test score gains. In contrast, we find that Hispanic male students experience significant declines in both math and reading test scores (test score impacts are positive but insignificant for Black females and negative but insignificant for all other subgroups aside from "Other"). Given self-reported improvements in classroom behavior, we do not believe that these negative impacts on non-Black students reflect disruption effects. Moreover, we find no clear evidence of treatment effect heterogeneity for Hispanic students as a function of school's Black enrollment share, which might be indicative of a role for heterogeneity in the quality of RP implementation explaining these estimates of heterogeneous test score impacts. Though admittedly speculative, we posit that, by inducing teachers to treat students more empathetically and to engage more holistically with students, RP may lead to a modest reduction in the emphasis placed on academic preparation as measured by the standardized test scores to which we have access.

Heterogeneity by English learner status, disability status, and grade level For the sake of comprehensiveness, we also examine heterogeneity on alternative margins. We find that reductions in out-of-school suspensions and arrests are concentrated among native English speakers (Appendix Table A11), 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. We also find larger absolute declines in the number of out-of-school suspension days and arrests for 9th and 10th graders, who are suspended and arrested more frequently at baseline, as compared to 11th and 12th graders (Appendix Table A13). Lastly, we find that declines in out-of-school suspensions do not vary significantly with disability status, while estimated declines in arrests are notably larger for disabled students (Appendix Table A15).

VIII Sensitivity Analysis: Alternative Specifications

We investigate the sensitivity of results to a range of alternative empirical approaches and specifications. Overall, results remain similar to our benchmark findings.

Standard difference-in-differences empirical approach. Instead of using the de Chaisemartin and D'Haultfoeuille (2020) estimator, we employ a standard difference-in-differences

design. The results remain qualitatively similar to the effects estimated in our benchmark specifications, with a notably larger estimated decline in absent days driven by the differential pre-trends apparent for this outcome (Appendix Tables A17 and A18, Panel A).

Excluding charter schools. For our main specifications, we include all observations for students who were enrolled in either neighborhood or charter schools in a given school year. Here, we restrict the sample to students who remained in traditional neighborhood schools and so exclude all observations for 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, and the modest estimated decline in math test scores is significant at the 5% level (Appendix Tables A17 and A18, Panel B).

Only Age and Cohort Fixed Effects as Controls We verify that results are not overly sensitive to the exclusion of covariates by estimating models that include only age and cohort fixed effects. We arrive at qualitatively similar conclusions, although the estimated impact on school climate is no longer significant at conventional levels (Appendix Tables A17 and A18, Panel C).

IX 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 number of out-of-school suspension days by 18 percent and reduced the number of student arrests by 16 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 student 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 in the aggregate.

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. We find corresponding evidence of math test score gains among Black males, who experience particularly large declines in suspensions and arrests. 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.

Angel, Caroline M, Lawrence W Sherman, Heather Strang, Barak Ariel, Sarah Bennett, Nova Inkpen, Anne Keane, and Therese S Richmond. 2014. "Short-term effects of restorative justice conferences on post-traumatic stress symptoms among robbery and burglary victims: a randomized controlled trial." *Journal of Experimental Criminology*, 10(3): 291–307.

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.

Bacher-Hicks, Andrew, Stephen B Billings, and David J Deming. 2019. "The school to prison pipeline: Long-run impacts of school suspensions on adult crime." *National Bureau of Economic Research.*

Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life.* Basic Books.

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

Chicago Public Schools. 2020. "District Data | Chicago Public Schools."

Chuang, Yating, and Laura Schechter. 2015. "Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results." *Journal of development economics*, 117: 151–170.

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

Dreeben, Robert. 1967. "The contribution of schooling to the learning of norms." *Harvard Educational Review*, 37(2): 211–237.

Evertson, Carolyn M, and Carol S Weinstein. 2006. *Handbook of classroom management: Research, practice, and contemporary issues.* Routledge.

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.*

Fulkerson, Andrew. 2001. "The use of victim impact panels in domestic violence justice approach." *Contemporary Justice Review*, 4: 355–368.

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.

Griffith, David, and Adam Tyner. 2019. "Discipline Reform through the Eyes of Teachers." *Thomas B. Fordham Institute*.

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

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

Kauffman, James M, Patricia L. Pullen, Mark P. Mostert, and Stanley C. Trent. 2011. *Managing classroom behavior: A reflective case-based approach.* Pearson.

Lai, Ijun. n.d. "Short-term impacts of Chicago's Suspensions and Expulsions Reduction Plan (SERP)." *Dissertation Paper*.

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.

Losen, Daniel J, Cheri L Hodson, Michael A Keith II, Katrina Morrison, and Shakti Belway. 2015. "Are we closing the school discipline gap?"

McCold, Paul, and Benjamin Wachtel. 1998. "Community is not a place: A new look at community justice initiatives." *Contemporary Justice Review*, 1(1): 71–85.

McGarrell, Edmund F. 2001. *Restorative Justice Conferences as an Early Response to Young Offenders.* US Department of Justice, Office of Justice Programs, Office of Juvenile ....

Parsons, Talcott. 1959. "The school class as a social system: Some of its functions in American society." *Harvard Educational Review*, 29: 297–318.

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

Shem-Tov, Yotam, Steven Raphael, and Alissa Skog. 2021. "Can Restorative Justice Conferencing Reduce Recidivism? Evidence From the Make-it-Right Program." *National Bureau of Economic Research*.

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*.

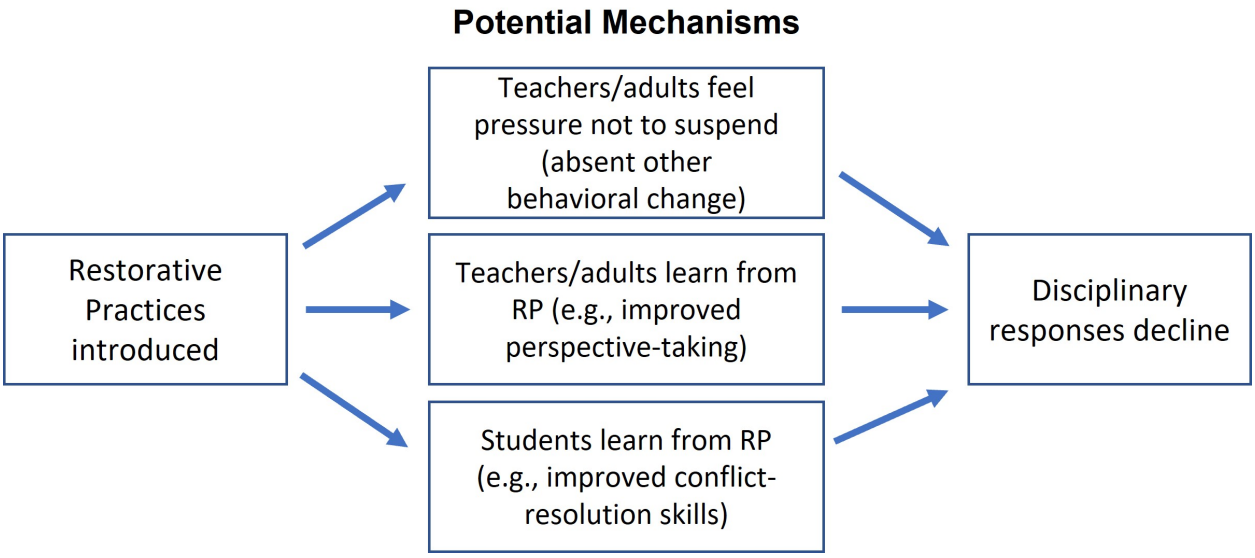
Wadhwa, Anita. 2015. *Restorative justice in urban schools: Disrupting the school-to-prison pipeline.* Routledge.

Wang, Ke, Yongqiu Chen, Jizhi Zhang, and Barbara A Oudekerk. 2020. "Indicators of School Crime and Safety: 2019. NCES 2020-063/NCJ 254485." *National Center for Education Statistics*.

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

X Main Figures

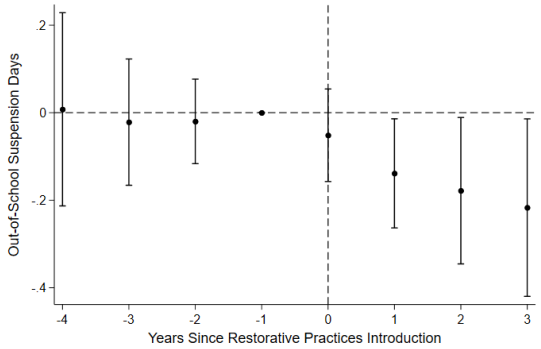
Figure 1: RP: Potential Mechanisms



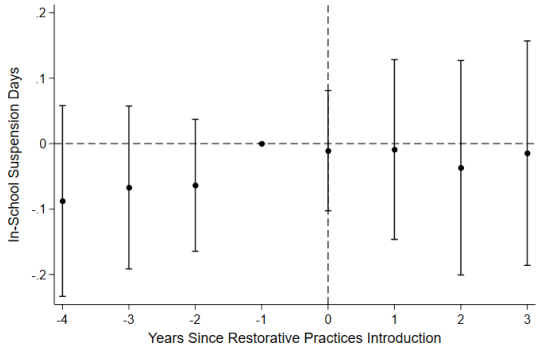
Notes: This schematic shows the simple channels through which RP may influence changes in disciplinary responses.

Figure 2: Event Studies: Behavioral Outcomes

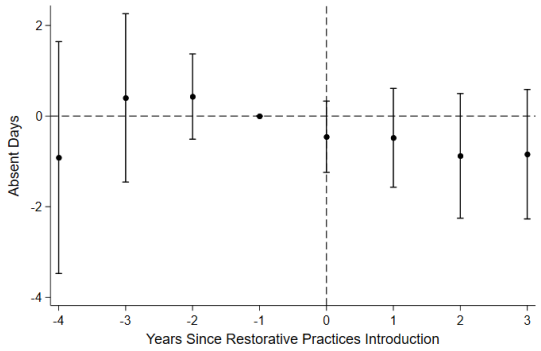
(a) Out-of-School Suspensions



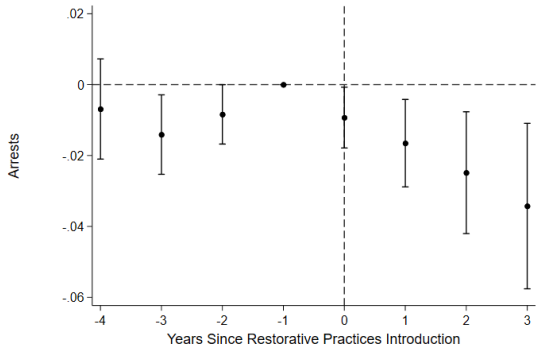
(b) In-School Suspensions



(c) Absent Days

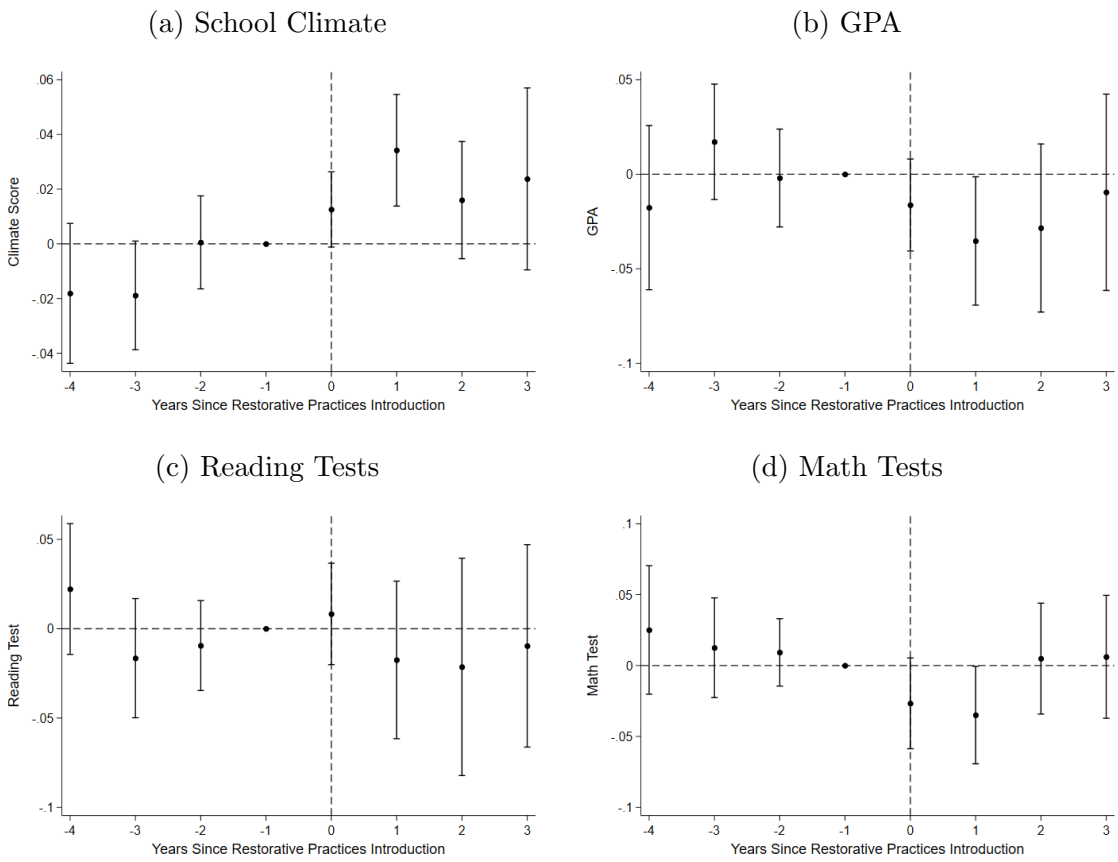


(d) Arrests



Notes: These figures show the event studies around the introduction of RP on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Suspension and absence 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. 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. Arrest data are collected by the Chicago Police Department. The main arrest outcome is defined as the number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest. Estimates are based on the methodology developed in de Chaisemartin and D'Haultfoeuille (2020) and described in text. Bars represent 95% confidence intervals based on standard errors clustered by school.

Figure 3: Event Studies: School Climate and Learning



Notes: These figures show the event studies around the introduction of RP on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The index 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. Estimates are based on the methodology developed in de Chaisemartin and D'Haultfoeuille (2020) and described in text. Bars represent 95% confidence intervals based on standard errors clustered by school.

XI Main Tables

Table 1: Baseline Characteristics: Chicago Public Schools High Schools

Variable	Non-Treated (1)	Treated (2)	Difference (3)
Number of Students	538.6 (425.9)	1003.7 (774.9)	465.1** (106.4)
Out-of-School Suspension Days	0.73 (2.61)	1.07 (3.28)	0.35* (0.18)
In-School Suspension Days	0.25 (1.24)	0.49 (1.72)	0.25* (0.12)
Absent Days	16.19 (18.68)	23.27 (23.09)	7.08** (1.70)
Number of Arrests	0.09 (0.55)	0.13 (0.62)	0.04 (0.03)
Ever Arrested	0.07 (0.26)	0.09 (0.29)	0.02 (0.02)
GPA	2.62 (0.99)	2.34 (1.02)	-0.28* (0.11)
Math Test Score	0.17 (1.08)	-0.076 (0.92)	-0.24 (0.15)
Reading Test Score	0.15 (1.06)	-0.066 (0.94)	-0.21 (0.15)
Climate Score	0.060 (0.53)	-0.037 (0.52)	-0.097** (0.027)
English Learner	0.05 (0.22)	0.07 (0.26)	0.02 (0.01)
Students in Temporary Living Situations	0.05 (0.22)	0.06 (0.24)	0.01 (0.01)
Individualized Education Plan	0.14 (0.35)	0.15 (0.36)	0.01 (0.01)
Economically Disadvantaged	0.79 (0.41)	0.82 (0.38)	0.04 (0.04)
Gender: Female	0.51 (0.50)	0.49 (0.50)	-0.02 (0.02)
Race: African American	0.48 (0.50)	0.42 (0.49)	-0.06 (0.08)
Race: White	0.08 (0.28)	0.10 (0.30)	0.02 (0.03)
Race: Hispanic/Latino	0.40 (0.49)	0.44 (0.50)	0.04 (0.06)
Disability: Cognitive	0.12 (0.33)	0.14 (0.34)	0.01 (0.01)
Disability: None	0.84 (0.37)	0.83 (0.38)	-0.01 (0.01)
Disability: Physical	0.01 (0.10)	0.01 (0.11)	0.00 (0.00)
Disability: 504	0.03 (0.17)	0.03 (0.16)	-0.00 (0.00)

Notes: This table presents the baseline mean values of characteristics of non-treated high schools (column 1) and treated high schools (column 2), prior to the introduction of RP in SY14. The associated differences (column 3) are derived from regressions of the given outcome on a treatment indicator variable, with the standard errors 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 from SY13. ** denotes 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 Suspension		In-School Suspension		Absent Days
	Days	Binary	Days	Binary	
	(1)	(2)	(3)	(4)	(5)
RP	-0.139*	-0.021*	-0.017	0.004	-0.639
	(0.066)	(0.010)	(0.060)	(0.018)	(0.505)
Baseline Mean	0.781	0.158	0.419	0.134	19.25
Observations	1,176,280	1,176,280	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 sample which are in one of the following criteria: 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. In columns 1 and 3, 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 columns 2 and 4, the OSS and ISS binary outcomes indicate whether a student ever received either of these types of suspensions in the corresponing school year, regardless of the school. 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. In column 5, absent days is adjusted to equal total absent days minus out-of-school suspension days. Regressions for the absence days outcome include student member days in the corresponding school year as a control. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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. Estimates are based on the methodology developed in de Chaisemartin and D'Haultfoeuille (2020) and described in text.

Table 3: Restorative Practices and Policing Outcomes

	Number of Arrests Overall	Number of In-School Arrests	Number of Out-of-School Arrests	Number of Violent Arrests	Number of Non-Violent Arrests
	(1)	(2)	(3)	(4)	(5)
RP	-0.02**	-0.008**	-0.012*	-0.004**	-0.016**
	(0.007)	(0.002)	(0.005)	(0.001)	(0.004)
Baseline Mean	0.125	0.030	0.095	0.029	0.096
Num. Obs.	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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. This table reports the average effect of restorative practices on students' arrests outcomes. Arrest data are collected by the Chicago Police Department. The arrest data includes information on the type (violent or non-violent), the location, and the time of arrest. The main arrest outcome is defined as the number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest. In-school arrests are defined as incidents that happened both inside the school location and during school hours, and out-of-school arrests are defined as incidents that happened either outside the school location or outside school hours. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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. Estimates are based on the methodology developed in de Chaisemartin and D'Haultfoeuille (2020) and described in text.

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
	(0.008)	(0.018)	(0.013)	(0.014)
Baseline Mean	-0.032	2.467	-0.070	-0.071
Observations	709,862	775,816	707,766	700,262

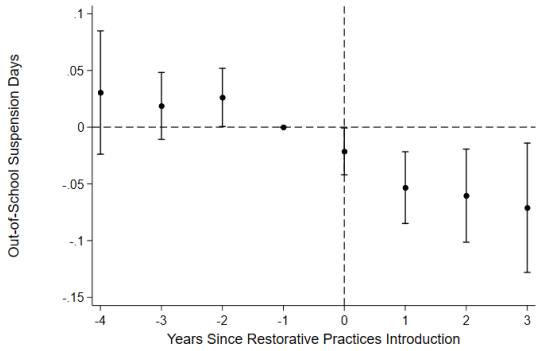
Notes: The sample includes observations at the student-school year level. We exclude observations from this sample which are in one of the following criteria: students in grades that are not offered by their first schools; students past their expected school exit year; and any observations beyond the our event study scope (-4 to +3 years since treatment). Student treatment assignment is determined by the first school enrolled since SY09. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 9th-12th graders, and are standardized by year and grade. The index 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. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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). 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. Estimates are based on the methodology developed in de Chaisemartin and D'Haultfoeuille (2020) and described in text.

Appendices

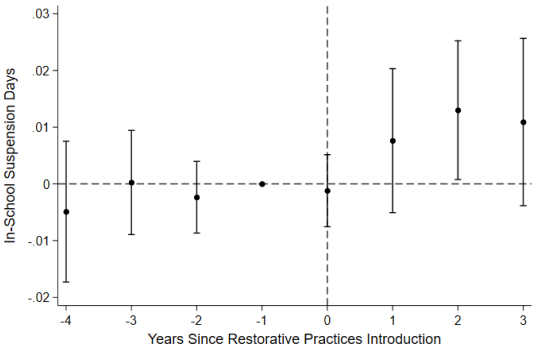
A Appendix Figures

Figure A1: Elementary School Event Studies: Behavioral Outcomes

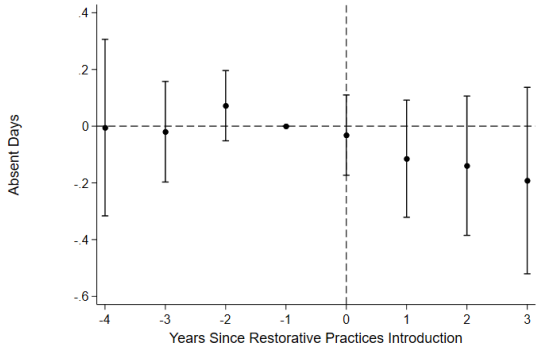
(a) Out-of-School Suspensions



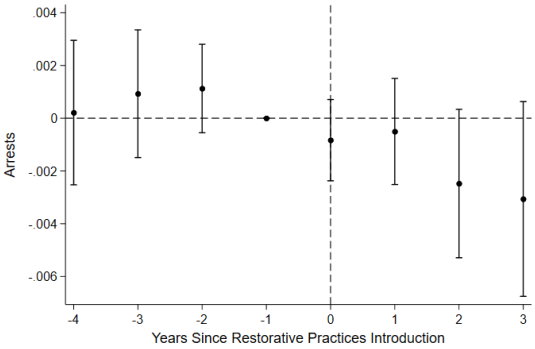
(b) In-School Suspensions



(c) Absent Days

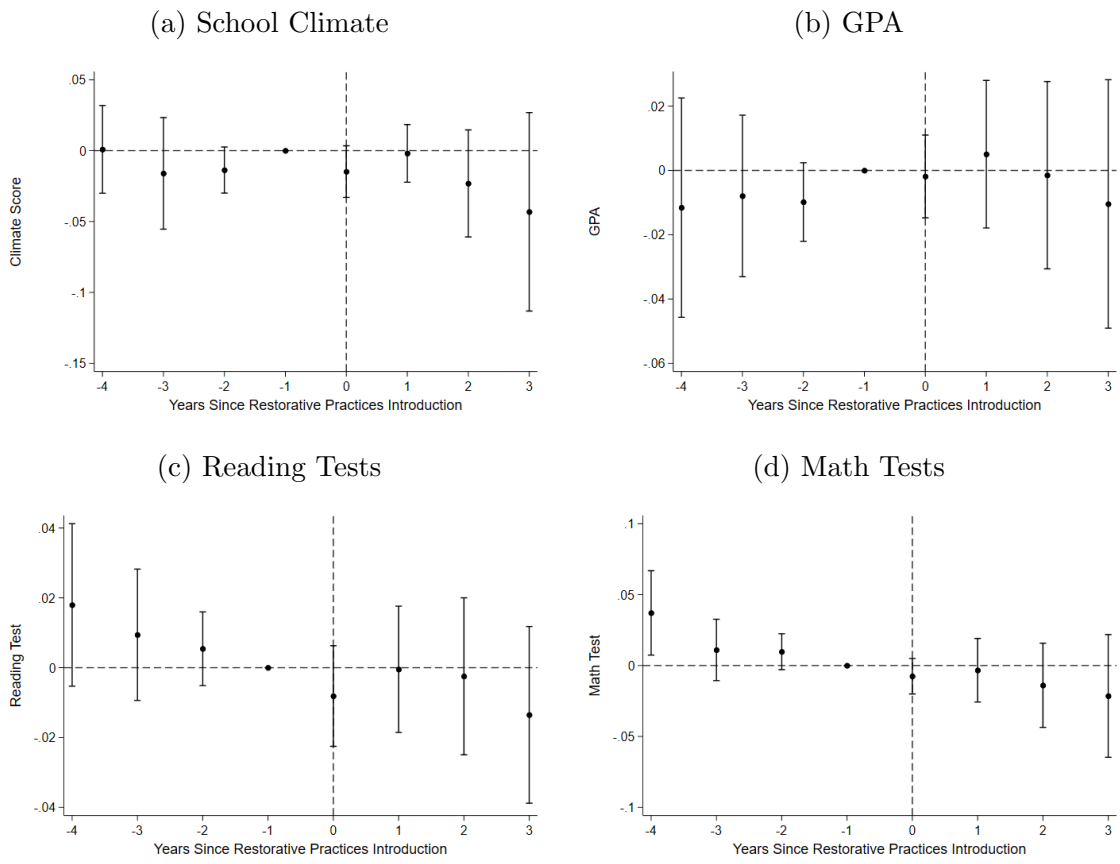


(d) Arrests



Notes: These figures show the event studies around the introduction of RP on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 3 and 8. Suspension and absence 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. 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. Arrest data are collected by the Chicago Police Department. The main arrest outcome is defined as the number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest.

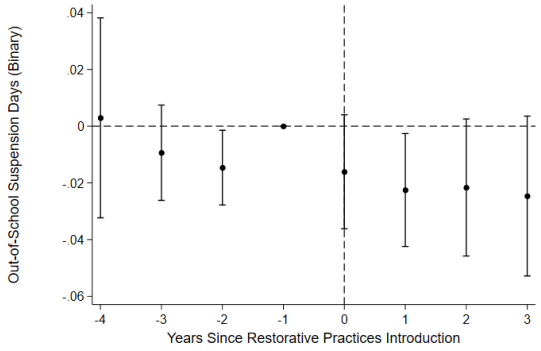
Figure A2: Elementary School: Climate and Learning



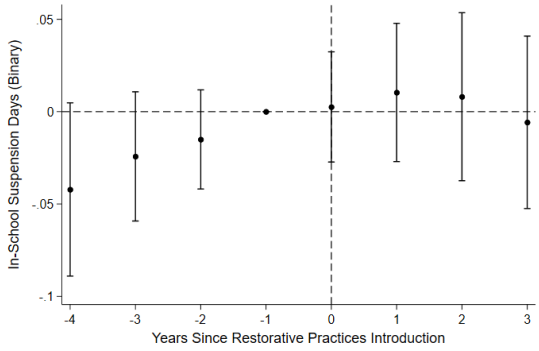
Notes: These figures show the event studies around the introduction of RP on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 3 and 8. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The index 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.

Figure A3: High School: Behavioral Outcomes, Binary Versions

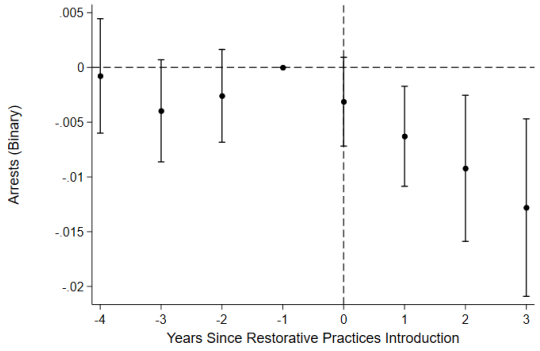
(a) Out-of-School Suspensions



(b) In-School Suspensions



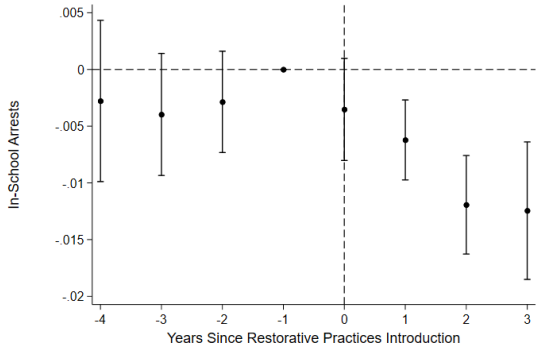
(c) Arrests



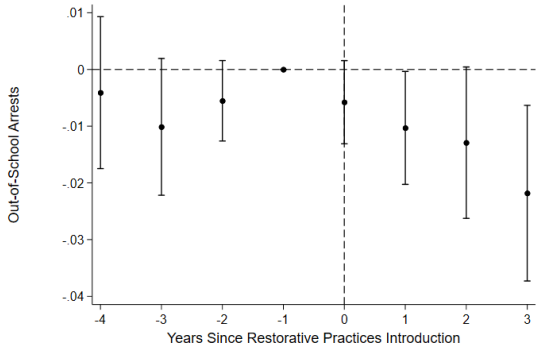
Notes: These figures show the event studies around the introduction of RP on binary in-school behavioral outcomes (out-of-school suspensions and in-school suspensions) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Suspension and absence data are collected by Chicago Public Schools. An out-of-school suspension is defined as whether a student was ever removed from class attendance or school attendance. An in-school suspension is defined as whether a student was ever removed 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 whether a student was ever arrested in a given year, regardless of the type of arrest or the location of the arrest.

Figure A4: High School Event Studies: Policing Outcomes

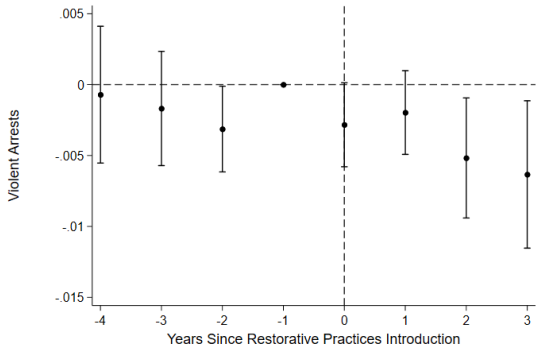
(a) Number of In-School Arrests



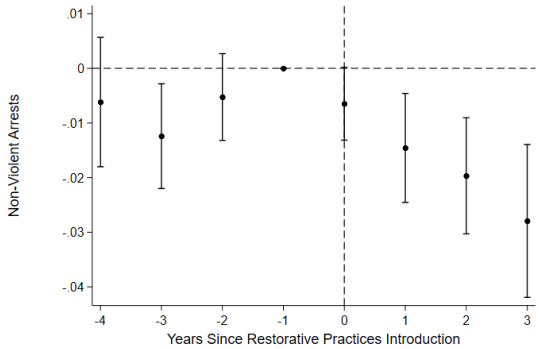
(b) Number of Out-of-School Arrests



(c) Number of Violent Arrests



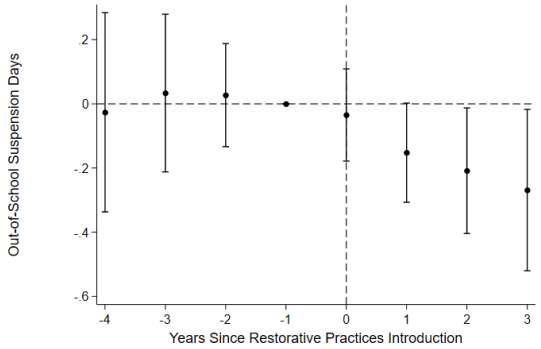
(d) Number of Non-Violent Arrests



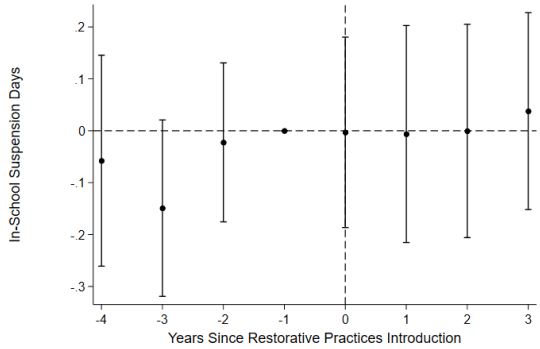
Notes: These figures show the event studies around the introduction of RP on students' arrest outcomes (out-of-school vs in-school, and violent vs non-violent) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Arrest data are collected by the Chicago Police Department. The arrest data includes information on the type (violent or non-violent), the location, and the time of arrest. The main arrest outcome is defined as the number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest. In-school arrests are defined as incidents that happened both inside the school location and during school hours, and out-of-school arrests are defined as incidents that happened either outside the school location or outside school hours.

Figure A5: High School Event Studies: RP Coaching: Behavioral and Policing Outcomes

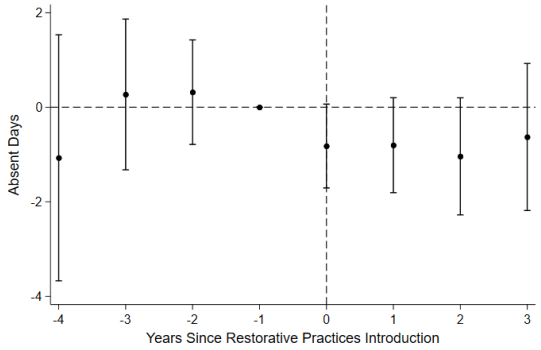
(a) Number of Out-of-School Suspensions



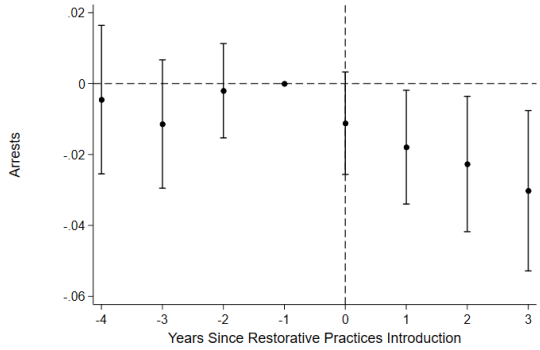
(b) Number of In-School Suspensions



(c) Absent Days



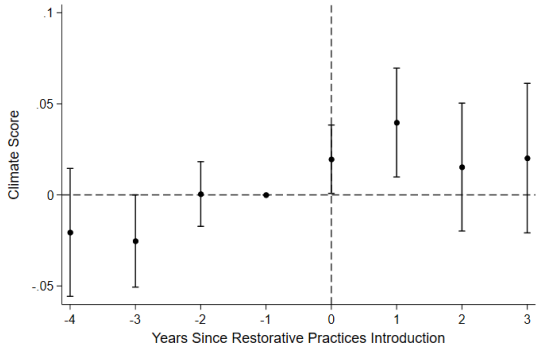
(d) Number Arrests



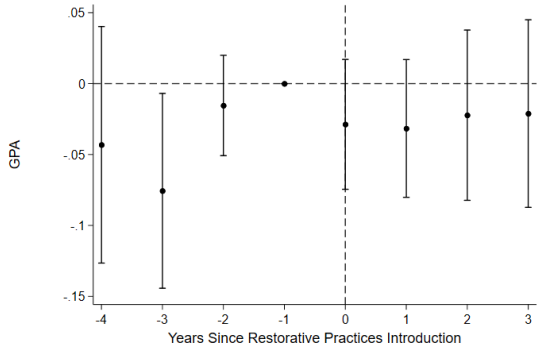
Notes: These figures show the event studies around the introduction of RP Coaching on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. 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 the 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 A6: High School Event Studies: RP Coaching: School Climate and Learning

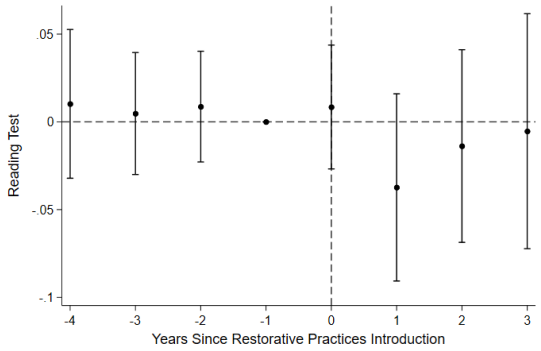
(a) Climate Index



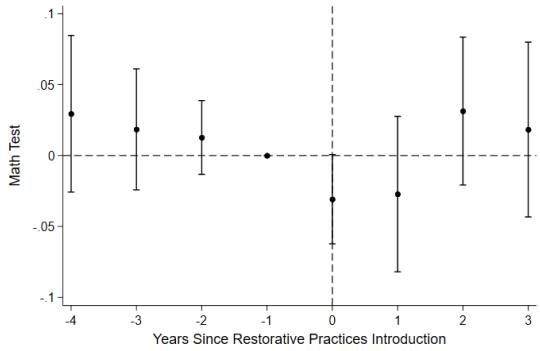
(b) GPA



(c) Reading Scores



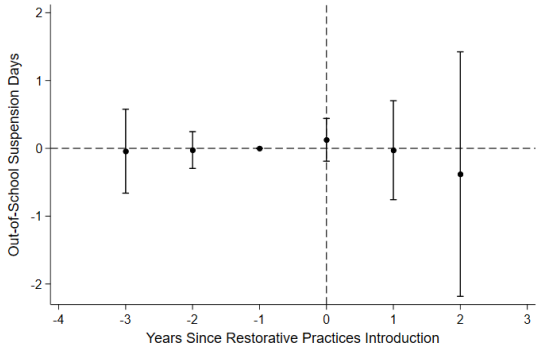
(d) Math Scores



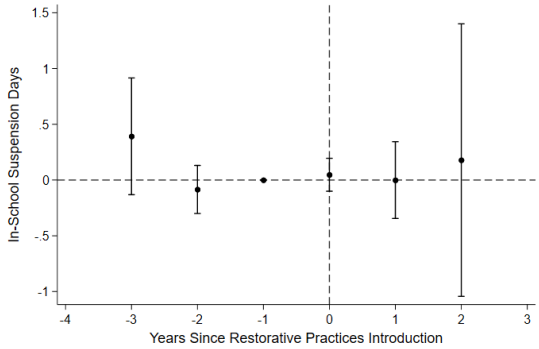
Notes: These figures show the event studies around the introduction of RP Coaching on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 9th-12th graders, and are standardized by year and grade. The index 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.

Figure A7: High School Event Studies: RP Leadership: Behavioral and Policing Outcomes

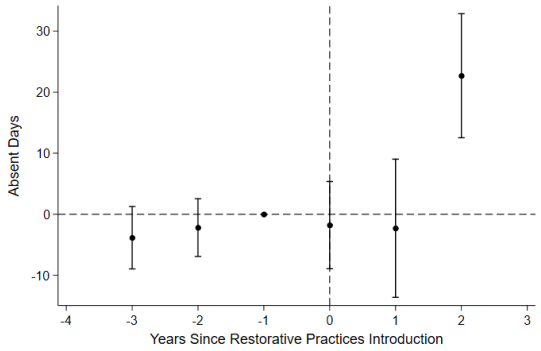
(a) Number of Out-of-School Suspensions



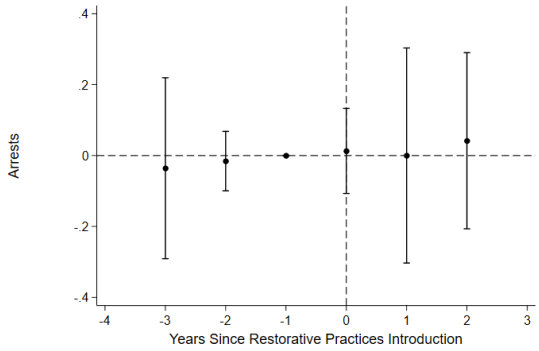
(b) Number of In-School Suspensions



(c) Absent Days



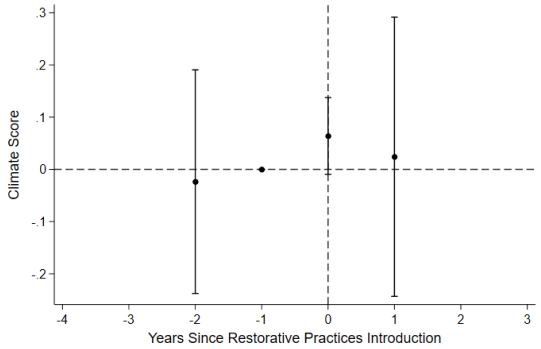
(d) Number Arrests



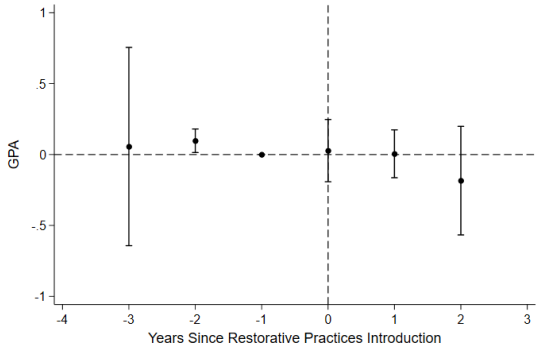
Notes: These figures show the event studies around the introduction of RP Leadership on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. 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 the 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 A8: High School Event Studies: RP Leadership: School Climate and Learning

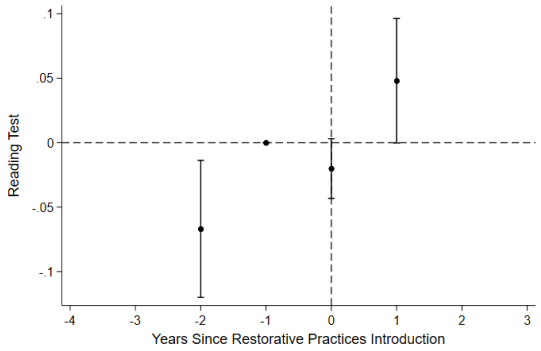
(a) Climate Index



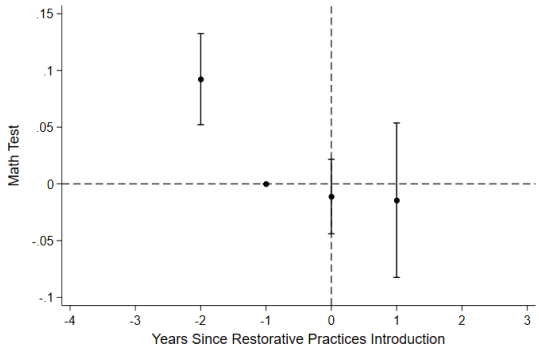
(b) GPA



(c) Reading Scores



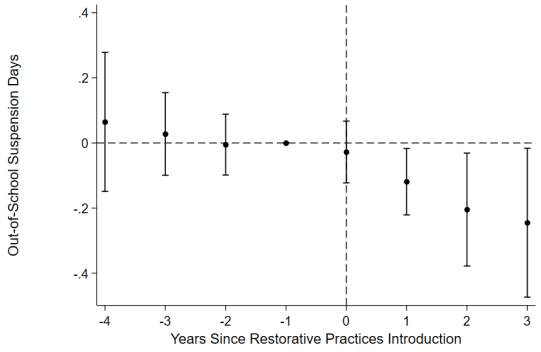
(d) Math Scores



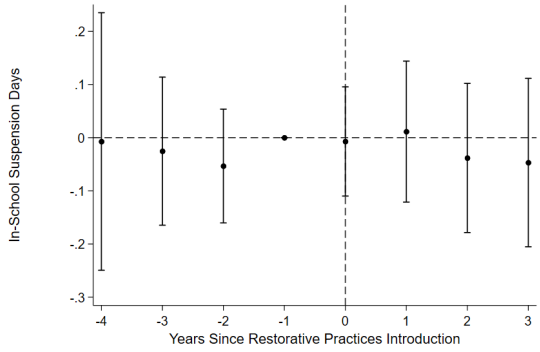
Notes: These figures show the event studies around the introduction of RP Leadership on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 9th-12th graders, and are standardized by year and grade. The index 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.

Figure A9: High School Event Studies: Difference-in-Differences: Behavioral and Policing Outcomes

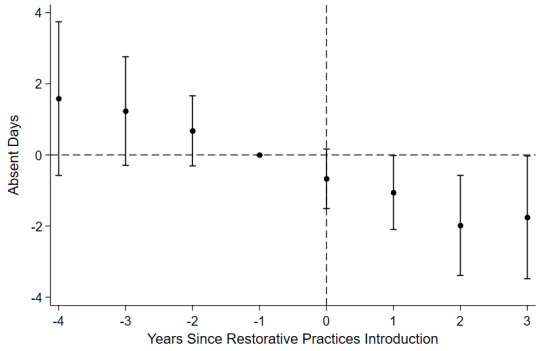
(a) Number of Out-of-School Suspensions



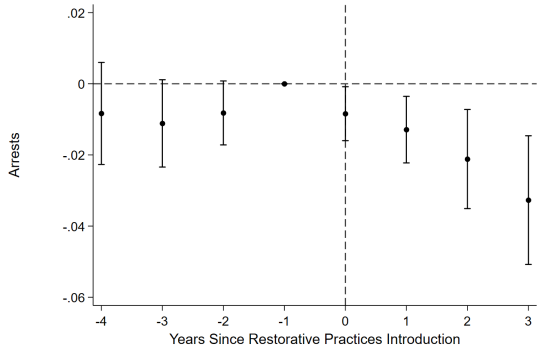
(b) Number of In-School Suspensions



(c) Absent Days

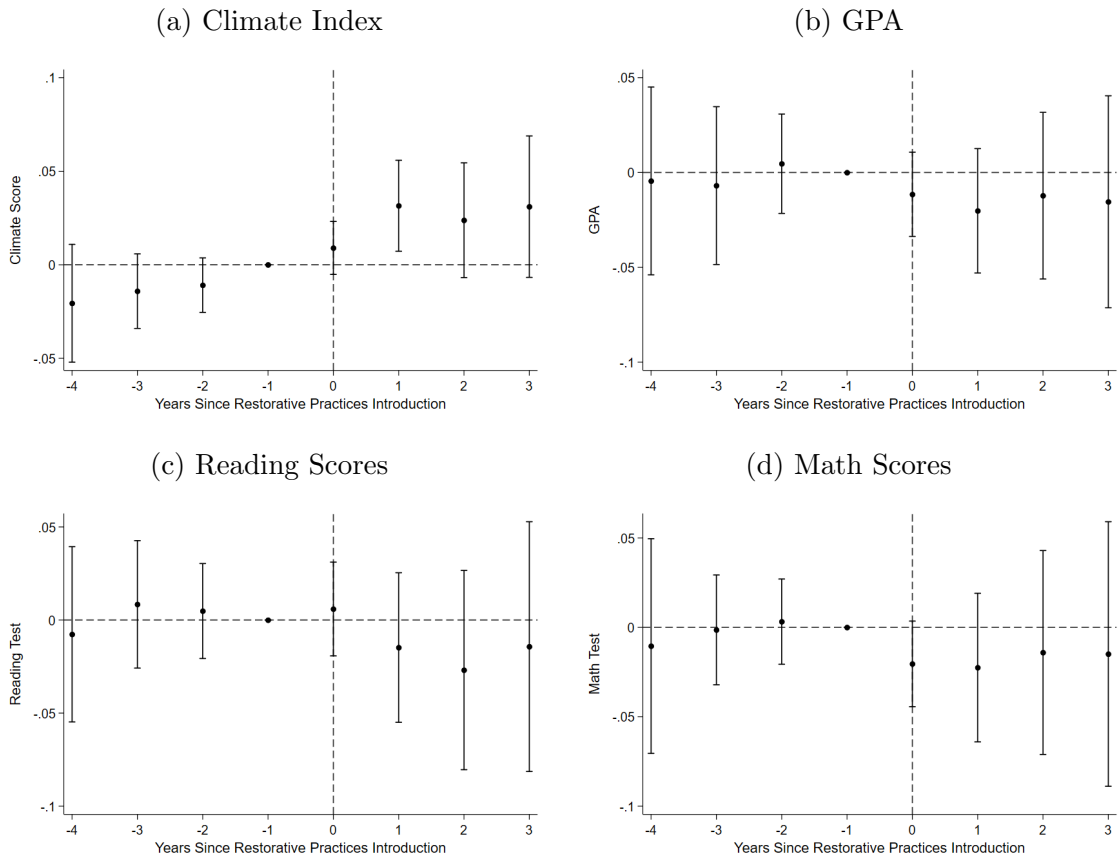


(d) Number Arrests



Notes: These figures show the event studies using a difference-in-differences model around the introduction of RP on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. 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 the 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.

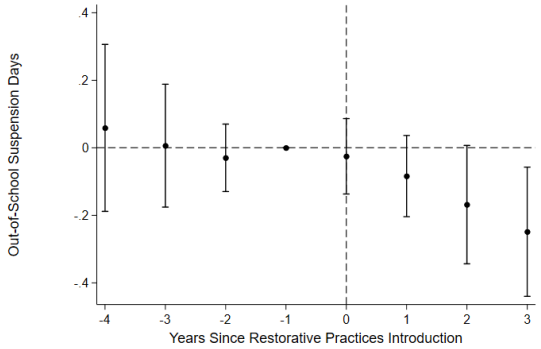
Figure A10: High School Event Studies: Difference-in-Differences: School Climate and Learning



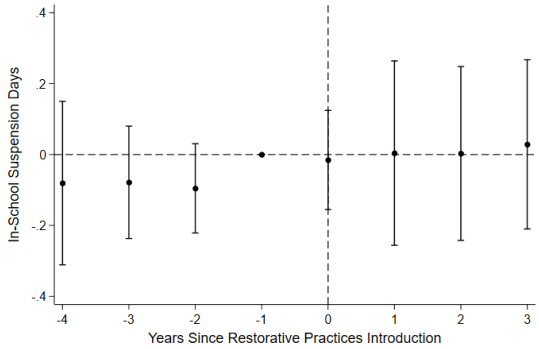
Notes: These figures show the event studies using a difference-in-differences model around the introduction of RP on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 9th-12th graders, and are standardized by year and grade. The index 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.

Figure A11: High School Event Studies: Drop Charter Schools: Behavioral and Policing Outcomes

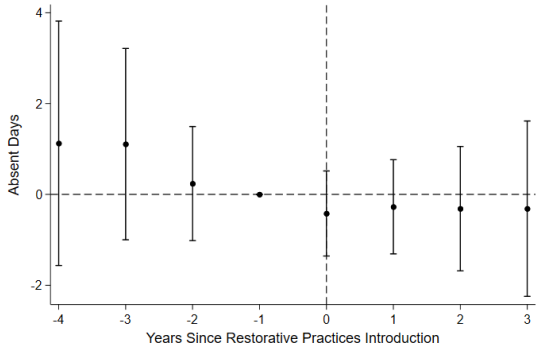
(a) Number of Out-of-School Suspensions



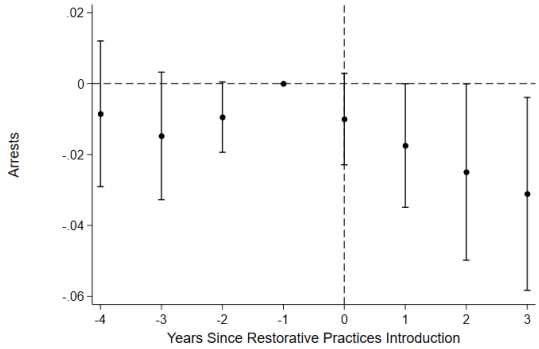
(b) Number of In-School Suspensions



(c) Absent Days

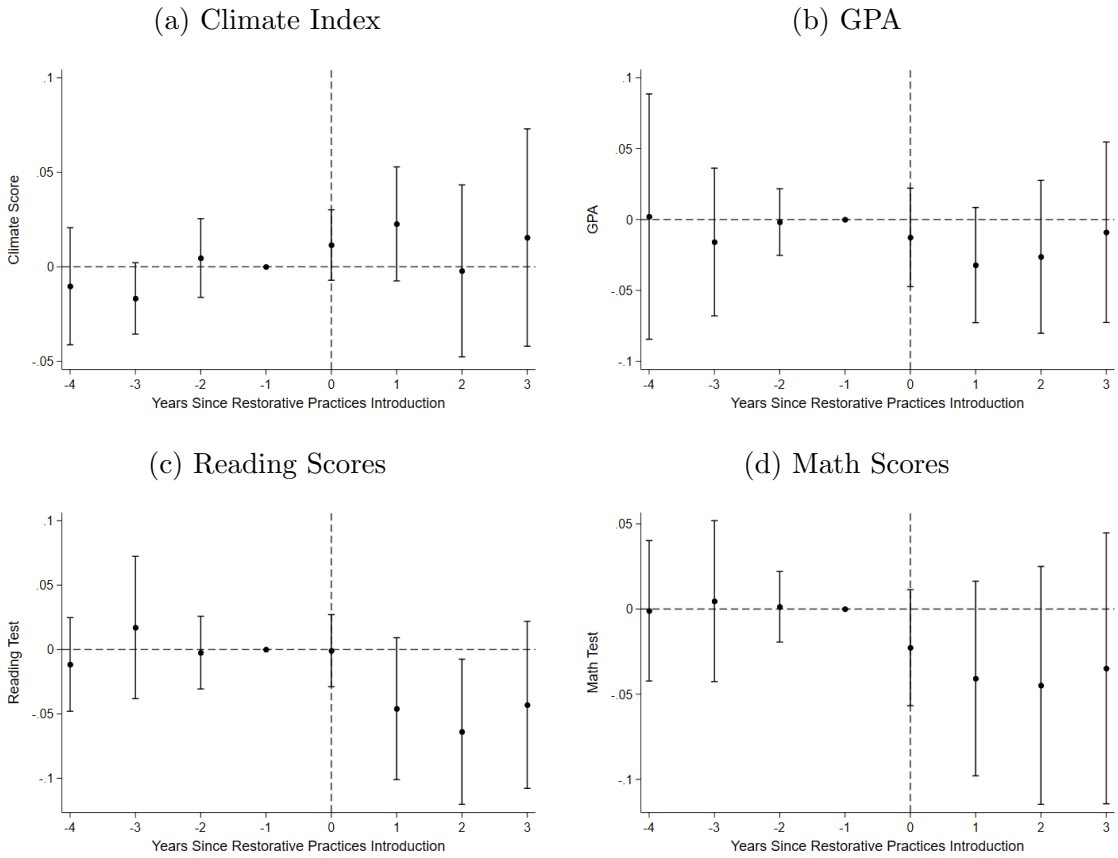


(d) Number Arrests



Notes: These figures show the event studies when dropping students in charter schools around the introduction of RP on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. 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 the 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.

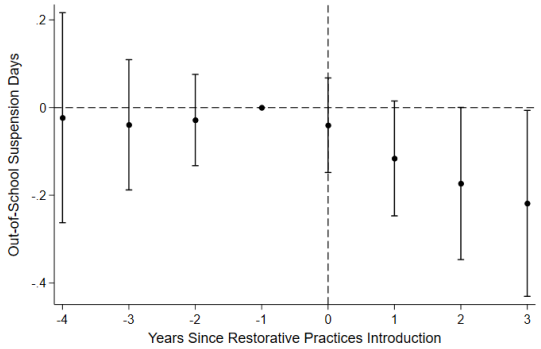
Figure A12: High School Event Studies: Drop Charter Schools: School Climate and Learning



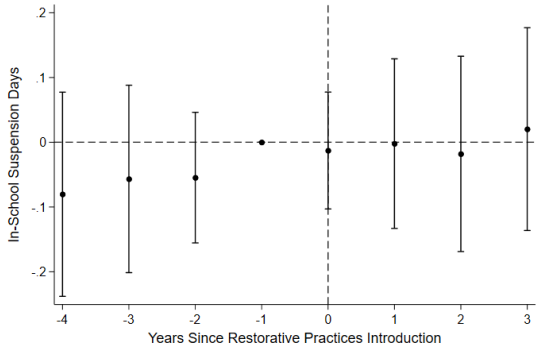
Notes: These figures show the event studies when dropping students in charter schools around the introduction of RP on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 9th-12th graders, and are standardized by year and grade. The index 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.

Figure A13: High School Event Studies: Only Age and Cohort Fixed Effects as Controls: Behavioral and Policing Outcomes

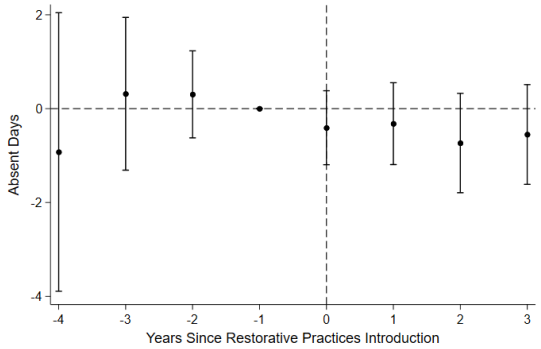
(a) Number of Out-of-School Suspensions



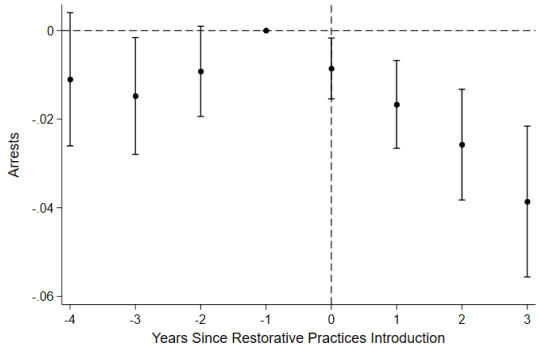
(b) Number of In-School Suspensions



(c) Absent Days

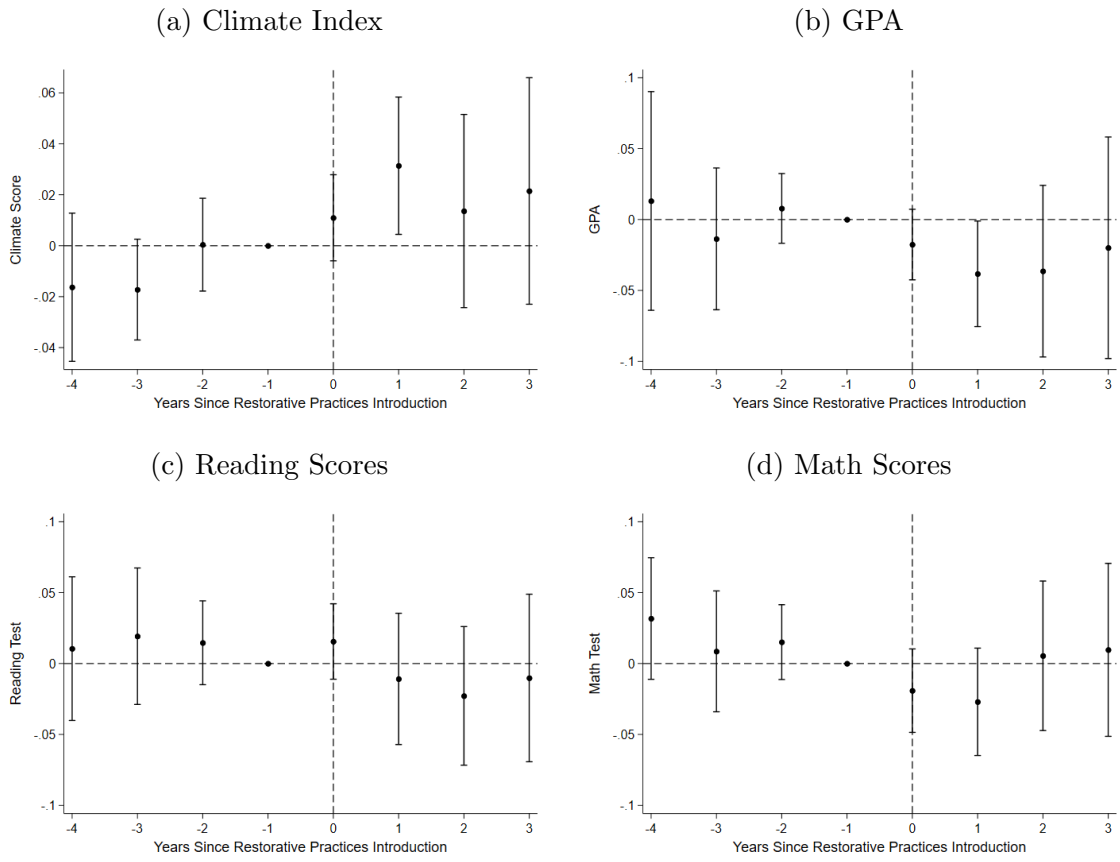


(d) Number Arrests



Notes: These figures show the event studies when only including age and cohort fixed effects as controls around the introduction of RP on in-school behavioral outcomes (out-of-school suspensions, in-school suspensions, and absent days) and policing outcomes (overall arrests) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. 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 the 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 A14: High School Event Studies: Only Age and Cohort Fixed Effects as Controls: School Climate and Learning



Notes: These figures show the event studies when only including age and cohort fixed effects as controls around the introduction of RP on students' perceptions of school climate and academic outcomes (GPA, reading test score, and math test score) over time. The sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 9th-12th graders, and are standardized by year and grade. The index 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

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	GPA (4)	Number of Arrests (5)	Climate Index (6)
Native English Speakers						
Baseline Mean	0.965	0.424	18.453	2.463	0.132	-0.003
Observations	1,097,251	1,097,251	1,097,251	724,841	1,118,309	663,523
English Learners						
Baseline Mean	0.561	0.236	17.603	2.377	0.059	0.032
Observations	78,080	78,080	78,080	50,088	78,113	45,678
Grades 9 and 10						
Baseline Mean	1.140	0.490	16.808	2.393	0.153	-0.001
Observations	629,202	629,202	629,202	406,017	646,927	386,580
Grades 11 and 12						
Baseline Mean	0.696	0.319	20.344	2.531	0.096	-0.001
Observations	521,196	521,196	521,196	358,165	524,330	307,198
Any Disability						
Baseline Mean	1.049	0.628	21.981	2.259	0.195	0.001
Observations	214,344	214,344	214,344	131,061	220,428	115,024
504 Disability						
Baseline Mean	0.727	0.302	20.007	2.491	0.066	0.016
Observations	51,782	51,782	51,782	36,580	52,497	31,125
Physical Disability						
Baseline Mean	0.920	0.398	24.228	2.400	0.198	0.047
Observations	15,249	15,249	15,249	9,575	16,036	6,414
Cognitive Disability						
Baseline Mean	1.130	0.719	22.235	2.193	0.223	-0.006
Observations	146,660	146,660	146,660	83,808	151,254	76,412
No Disability						
Baseline Mean	0.922	0.372	17.750	2.494	0.114	-0.001
Observations	961,105	961,105	961,105	644,096	975,999	594,405

Notes: This displays the mean for each outcome variable 1 year before treatment for students at any grade level between 9 and 12, broken out by English Language Learner status, grade grouping, and disability status.

Table A2: Baseline Characteristics: by Student Race and Gender

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	GPA (4)	Number of Arrests (5)	Climate Index (6)
Black Female						
Baseline Mean	1.322	0.542	21.813	2.422	0.085	-0.055
Observations	256,763	256,763	256,763	152,363	261,688	148,186
Hispanic Female						
Baseline Mean	0.304	0.134	17.084	2.657	0.015	0.027
Observations	263,189	263,189	263,189	176,571	266,259	169,068
White Female						
Baseline Mean	0.180	0.100	15.650	3.031	0.013	0.126
Observations	57,494	57,494	57,494	51,693	58,723	36,985
Asian Female						
Baseline Mean	0.093	0.020	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	0.061	0.036	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	1.880	0.906	21.014	2.022	0.387	-0.042
Observations	231,997	231,997	231,997	126,897	237,882	125,166
Hispanic Male						
Baseline Mean	0.670	0.246	16.466	2.299	0.086	-0.009
Observations	258,263	258,263	258,263	172,364	261,951	161,007
White Male						
Baseline Mean	0.417	0.199	15.370	2.689	0.059	0.093
Observations	53,817	53,817	53,817	47,997	55,583	33,591
Asian Male						
Baseline Mean	0.126	0.060	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	0.115	0.053	9.647	3.056	0.016	0.081
Observations	17,034	17,034	17,034	14,763	17,205	11,398

Notes: This displays the mean for each outcome variable 1 year before treatment for students at any grade level between 9 and 12, broken out by race-by-gender groups.

Table A3: High Schools: Restorative Practices and Policing Outcomes

	Arrests Overall (Binary) (1)	In-School Arrests (Binary) (2)	Out-of-School Arrests (Binary) (3)	Violent Arrests (Binary) (4)	Non-Violent Arrests (Binary) (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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. This table reports the average effect of restorative practices on students' binary arrest outcomes. Arrest data are collected by the Chicago Police Department. The arrest data includes information on the type (violent or non-violent), the location, and the time of arrest. The main arrest outcome is defined as whether a student was arrested in a given year, regardless of the type of arrest or the location of the arrest. In-school arrests are defined by whether a student had an incident that happened both inside the school location and during school hours, and out-of-school arrests are defined as incidents that happened either outside the school location or outside school hours. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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 A4: Elementary Schools: Restorative Practices and In-School Behavioral Outcomes

	Out-of-School Suspension		In-School Suspension		Absent Days
	Days	Binary	Days	Binary	
	(1)	(2)	(3)	(4)	(5)
RP	-0.047**	-0.008**	0.006	0.003+	-0.105
	(0.017)	(0.003)	(0.005)	(0.002)	(0.093)
Baseline Mean	0.246	0.065	0.055	0.030	7.902
Num. Obs.	2,308,111	2,308,111	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 sample which are in one of the following criteria: students in K-2 grades; 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 sample includes all students observed between SY09 and SY19 at any grade level between 3 and 8. In columns 1 and 3, 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 columns 2 and 4, the OSS and ISS binary outcomes indicate whether a student ever received either of these types of suspensions in the corresponing school year, regardless of the school. In column 5, absent days is adjusted to equal total absent days minus out-of-school suspension days. Regressions for the absence days outcome include student member days in the corresponding school year as a control. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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 A5: Elementary School: Restorative Practices and Policing Outcomes

	Number of Arrests Overall	Number of In-School Arrests	Number of Out-of-School Arrests	Number of Violent Arrests	Number of Non-Violent Arrests
	(1)	(2)	(3)	(4)	(5)
RP	-0.001	-0.001	-0.001	0	-0.001+
	(0.001)	(0)	(0.001)	(0.001)	(0.001)
Baseline Mean	0.012	0.004	0.009	0.004	0.008
Num. Obs.	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 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 sample includes all students observed between SY09 and SY19 at any grade level between 3 and 8. Arrest data are collected by the Chicago Police Department. The arrest data includes information on the type (violent or non-violent), the location, and the time of arrest. The main arrest outcome is defined as the number of arrests experienced by students in a given year, regardless of the type of arrest or the location of the arrest. In-school arrests are defined as incidents that happened both inside the school location and during school hours, and out-of-school arrests are defined as incidents that happened either outside the school location or outside school hours. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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 A6: Elementary School: Restorative Practices: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
RP	-0.017	-0.001	-0.006	-0.01
	(0.013)	(0.011)	(0.009)	(0.010)
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 sample which are in one of the following criteria: students in K-2 grades; 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 sample includes all students observed between SY09 and SY19 at any grade level between 3 and 8. The School Climate index measures student SEL levels and perceptions regarding the supportiveness of school environments. The estimates are drawn from MVMS scores from SY11-SY18 for 6th-8th graders, and are standardized by year and grade. The index 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. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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). 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: Implementation Type of Restorative Practices: Behavioral and Policing Outcomes

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	Number of Arrests (4)
RP Coaching	-0.16∗	0.006	-0.826+	-0.02∗
	(0.081)	(0.084)	(0.497)	(0.008)
Baseline Mean	0.966	0.517	20.304	0.148
Num. Obs.	988,703	988,703	988,703	1,006,660
RP Leadership	0.052	0.024	-2.02	0.007
	(0.254)	(0.112)	(4.64)	(0.106)
Baseline Mean	0.148	0.020	19.534	0.044
Num. Obs.	239,074	239,074	239,074	243,649

Table A8: Implementation Type of Restorative Practices: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
RP Coaching	0.024+	-0.026	-0.011	0.001
	(0.012)	(0.023)	(0.022)	(0.02)
Baseline Mean	-0.048	2.383	-0.155	-0.176
Num. Obs.	575,954	606,147	603,546	597,157
RP Leadership	0.064+	0.017	0.014	-0.013
	(0.038)	(0.094)	(0.017)	(0.024)
Baseline Mean	-0.075	2.553	0.032	0.006
Num. Obs.	103,413	131,561	69,101	69,131

Notes: These tables show results by the type of Restorative Practices program that was implemented in high schools. The district offers three different programs to support high schools in the adoption of restorative practices: RP Coaching, RP Leadership, and RP Peer Council. RP Peer Council results are not shown here because it was only delivered in SY19, so we are only able to estimate instantaneous effects for it. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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). 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: Race-by-Gender Treatment Heterogeneity: Behavioral and Policing Outcomes

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	Number of Arrests (4)
Black Female	-0.276∗	-0.133∗	-0.758	-0.014∗
	(0.115)	(0.067)	(0.538)	(0.007)
Hispanic Female	-0.042	0.059	-0.395	-0.004+
	(0.031)	(0.059)	(0.749)	(0.002)
White Female	-0.064∗	-0.025	-0.738	-0.005
	(0.032)	(0.053)	(0.684)	(0.004)
Asian Female	0.002	0.004	0.161	-0.005
	(0.053)	(0.033)	(1.906)	(0.01)
Other Female	-0.038	0.026	-0.041	0.002
	(0.048)	(0.031)	(0.847)	(0.003)
Black Male	-0.307∗	-0.144	-1.769∗∗	-0.067∗∗
	(0.12)	(0.104)	(0.46)	(0.02)
Hispanic Male	0.008	0.129	-0.106	-0.012
	(0.065)	(0.08)	(0.724)	(0.01)
White Male	-0.039	-0.014	-0.508	-0.008
	(0.065)	(0.061)	(0.625)	(0.013)
Asian Male	-0.051	0.004	-1.107	0.022
	(0.072)	(0.044)	(1.931)	(0.015)
Other Male	-0.089	0.009	0.635	-0.02
	(0.056)	(0.038)	(0.979)	(0.016)

Notes: This table shows results by race-by-gender. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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, and disability (having a 504 plan, physical, cognitive, none). 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: Race-by-Gender Treatment Heterogeneity: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
Black Female	0.032	-0.011	0.001	0.012
	(0.02)	(0.019)	(0.02)	(0.017)
Hispanic Female	0.02	-0.024	-0.03	-0.035
	(0.014)	(0.024)	(0.027)	(0.039)
White Female	0.001	0.002	-0.035	-0.053
	(0.03)	(0.04)	(0.034)	(0.036)
Asian Female	0.037	-0.084	-0.133	-0.082
	(0.091)	(0.093)	(0.118)	(0.141)
Other Female	-0.026	0.005	0.086+	0.041
	(0.035)	(0.044)	(0.048)	(0.047)
Black Male	0.026	-0.013	0.012	0.044*
	(0.022)	(0.03)	(0.022)	(0.018)
Hispanic Male	0.006	-0.041	-0.044**	-0.051+
	(0.014)	(0.031)	(0.017)	(0.026)
White Male	0.008	-0.032	-0.006	-0.032
	(0.031)	(0.029)	(0.038)	(0.049)
Asian Male	-0.022	-0.124	-0.072	-0.096
	(0.123)	(0.096)	(0.257)	(0.269)
Other Male	-0.042	-0.078	0.047	-0.022
	(0.044)	(0.053)	(0.058)	(0.059)

Notes: This table shows results by race-by-gender. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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, and disability (having a 504 plan, physical, cognitive, none). 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: Restorative Practices: English Learner Treatment Heterogeneity: Behavioral and Policing Outcomes

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	Number of Arrests (4)
Native English Speakers	-0.15∗	-0.027	-0.639	-0.021∗∗
	(0.059)	(0.067)	(0.426)	(0.006)
English Learners	0.114	0.211∗	-0.094	-0.015
	(0.072)	(0.094)	(0.617)	(0.011)

Table A12: Restorative Practices: English Learner Treatment Heterogeneity: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
Native English Speakers	0.024∗	-0.02	-0.012	-0.016
	(0.011)	(0.018)	(0.015)	(0.02)
English Learners	-0.012	-0.03	0.03	0.035
	(0.028)	(0.038)	(0.02)	(0.033)

Notes: These tables show results by English Language Learner status. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, and cohort defined as the difference between grade of entry and school year of entry. Each specification includes the following individual-level covariates: homelessness indicator, IEP indicator, free or reduced lunch indicator, gender, race, and disability (having a 504 plan, physical, cognitive, none). 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 A13: Restorative Practices: Grade-Level Treatment Heterogeneity: Behavioral and Policing Outcomes

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	Number of Arrests (4)
Grades 9 and 10	-0.184**	-0.028	-0.835*	-0.026**
	(0.066)	(0.059)	(0.388)	(0.008)
Grades 11 and 12	-0.100*	0.031	-0.343	-0.008*
	(0.046)	(0.056)	(0.662)	(0.004)

Table A14: Restorative Practices: Grade-Level Treatment Heterogeneity: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
Grades 9 and 10	0.013	-0.011	0.001	0.006
	(0.013)	(0.025)	(0.027)	(0.024)
Grades 11 and 12	0.029*	-0.038	-0.044*	-0.031
	(0.013)	(0.023)	(0.018)	(0.024)

Notes: These tables show results by grade-level. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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). 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 A15: Restorative Practices: Treatment Heterogeneity by Ability Status: Behavioral and Policing Outcomes

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	Number of Arrests (4)
Any Disability	-0.112∗	-0.012	-0.672	-0.04∗∗
	(0.05)	(0.076)	(0.497)	(0.013)
504 Disability	-0.095	-0.004	-0.461	-0.017
	(0.091)	(0.083)	(0.853)	(0.014)
Physical Disability	-0.185	-0.061	-1.941	-0.074
	(0.156)	(0.2)	(1.963)	(0.048)
Cognitive Disability	-0.081	0.004	-1.08∗	-0.045∗∗
	(0.082)	(0.11)	(0.55)	(0.014)
No Disability	-0.147∗	-0.027	-0.619	-0.016∗∗
	(0.063)	(0.054)	(0.477)	(0.006)

Notes: This table shows results by ability status. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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, and race. 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 A16: Restorative Practices: Treatment Heterogeneity by Ability Status: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
Any Disability	0.002	-0.045∗	0.03	0.04
	(0.018)	(0.022)	(0.021)	(0.025)
504 Disability	0.002	-0.019	-0.058	-0.036
	(0.022)	(0.037)	(0.042)	(0.037)
Physical Disability	0.012	-0.102	0.01	0.001
	(0.059)	(0.065)	(0.067)	(0.084)
Cognitive Disability	0.007	-0.045	0.041∗	0.052∗
	(0.022)	(0.028)	(0.018)	(0.025)
No Disability	0.023+	-0.018	-0.013	-0.021
	(0.012)	(0.022)	(0.015)	(0.014)

Notes: This table shows results by ability status. 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 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 sample includes all students observed between SY09 and SY19 at any grade level between 9 and 12. Each specification includes the following fixed effects: school year, student's age by June 20th of the school year, 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, and race. 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 A17: Robustness: Behavioral Outcomes

	Number of Out-of-School Suspensions (1)	Number of In-School Suspensions (2)	Absent Days (3)	Number of Arrests Overall (4)
Panel A: Difference-in-Differences				
RP	-0.154+	0.00127	-2.127**	-0.0113+
	(0.0865)	(0.0703)	(0.772)	(0.00600)
Baseline Mean	0.781	0.419	19.250	0.125
Num. Obs.	1,126,001	1,126,001	1,126,001	1,174,233
Panel B: Dropping Charter Schools				
RP	-0.12+	0.003	-0.334	-0.02*
	(0.065)	(0.097)	(0.552)	(0.009)
Baseline Mean	0.775	0.433	19.400	0.116
Num. Obs.	755,527	755,527	755,527	772,577
Panel C: Only Age and Cohort Fixed Effects as Controls				
RP	-0.128+	-0.004	-0.488	-0.021**
	(0.069)	(0.057)	(0.396)	(0.005)
Baseline Mean	0.781	0.419	19.250	0.125
Num. Obs.	1,178,613	1,178,613	1,178,613	1,199,850

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 A18: Robustness: School Climate and Learning

	Climate Index (1)	GPA (2)	Reading Scores (3)	Math Scores (4)
Panel A: Difference-in-Differences				
RP	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
Num. Obs.	578,678	893,541	722,596	719,473
Panel B: Dropping Charter Schools				
RP	0.012	-0.02	-0.037*	-0.035
	(0.014)	(0.02)	(0.019)	(0.026)
Baseline Mean	-0.030	2.477	-0.058	-0.054
Num. Obs.	482,123	735,989	442,507	436,709
Panel C: Only Age and Cohort Fixed Effects as Controls				
RP	0.019	-0.028	-0.006	-0.008
	(0.014)	(0.021)	(0.017)	(0.018)
Baseline Mean	-0.032	2.467	-0.070	-0.071
Num. Obs.	710,651	777,447	709,056	701,525

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.