UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Evaluating the Effects of School Based Restorative Practices

Permalink
https://escholarship.org/uc/item/631028b4

Author
Darling-Hammond, Sean

Publication Date
2022

Peer reviewed|Thesis/dissertation

Evaluating the Effects of School Based Restorative Practices

By

Sean Darling-Hammond

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Public Policy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Chancellor's Professor Rucker Johnson, Chair
Professor Steven Raphael
Professor Jesse Rothstein
Assistant Professor Jason Okonofua

Spring 2022

Abstract

Evaluating the Effects of School Based Restorative Practices

by

Sean Darling-Hammond

Doctor of Philosophy in Public Policy

University of California, Berkeley

Professor Rucker Johnson, Chair

Schools recruit exclusionary discipline—such as suspensions and expulsions—to deter students from misbehaving, and to protect students from the harms associated with exposure to student misbehavior. While often implemented with good intentions, recent research indicates that exposure to discipline increases (rather than deters) misbehavior; increases risks of dropout, juvenile involvement, and adult incarceration; and exerts secondary harms, reducing school climate among those whose peers are suspended.

Black students bear the brunt of school discipline. Across student subgroups and school contexts, Black students are far more likely to be suspended or expelled. Whether one looks at female students, economically disadvantaged students, preschool students, or students in charter schools, Black students are overrepresented among those disciplined. Research demonstrates that stark racial disparities in discipline are not merely a function of racial disparities in student misbehavior, nor of how students sort into schools. Instead, disparities are largely driven by school practices. The harmfulness and unevenness of school discipline thus present a pressing equity issue: how can schools ameliorate racial disparities in discipline?

In reply, scholars have suggested, and schools have implemented, restorative practices, which include proactive practices to inculcate conflict resolution skills and strengthen community bonds (e.g., community building circles) and responsive practices to resolve conflicts and repair relationships (e.g., harm repair circles). Proponents argue that because these practices address a root cause of student discipline, they have the potential to ameliorate racial disparities while enhancing school climates, academic engagement, and academic performance.

However, a review of extant quantitative research surfaces a critical distinction between restorative programs and restorative practices. Restorative *programs* are systems of hiring, training, and support that are designed to encourage school community members to engage in restorative practices. Restorative *practices* are the specific actions that community members might engage in in a restorative school, and that theoretically can yield benefits for students, staff, and community members. Research has focused almost exclusively on the impact of restorative *programs* and has shown mixed results. This research also suffers from a core challenge: programs often do not accrue to students being exposed to restorative practices. Prior research thus leaves unclear whether restorative *practices* can have intended impacts.

This study merges and leverages multiple forms of California administrative data to track school practices and student outcomes over space and time, and finds that students who gain exposure to restorative practices see declines in discipline and increases in academic achievement, and declines in related racial disparities. It also finds that schools that increase their use of restorative practices see declines in misbehavior, gang membership, victimization, depressive symptoms, and substance abuse; and improvements in GPA and school climate. Taken together, these results present a strong case for the effectiveness of restorative practices at improving outcomes for students and schools. However, if these analyses answer one question about restorative practices, then they suggest another: *how* can schools increase student exposure to these potent practices?

A review of implementation guides surfaces insights for employing restorative programming that can engender student exposure to restorative practices. Districts and states can empower schools to persevere in their implementation of restorative practices by providing sustained funding. Schools, meanwhile, can take three approaches. First, they can commit to culture change by reducing reliance on exclusionary discipline. Second, they can preempt caregiver concerns via proactive communication about the psychological benefits of restorative approaches (and the potential harms of punitive ones). And, finally, schools can ensure widespread utilization (and exposure) by providing staff *throughout* the school with training regarding how to implement restorative practices in varied situations and when interacting with students of all backgrounds— and by empowering and incentivizing staff to use these practices.

This dissertation is dedicated…

to my dear son, Kofi. Each day, you remind me of what it means to love someone deeply and unequivocally; to be awed by who they are and who they are becoming; to be hopeful that the world will appreciate and nurture their brilliant mind and loving spirit; and to feel motivated to do *my* part to create such a world…

to the millions of amazing children throughout the world with unimaginable potential who deserve to be treated with dignity and afforded opportunities to thrive…

to their parents and caregivers, yearning for a just society where they can sleep soundly with the knowledge that when their children leave their care, their children will be embraced by a safe and affirming community…

and to the many teachers, administrators, coordinators, counselors, psychologists, nurses, and staff striving against incredible odds to cultivate such beautifully interconnected communities; and to create schools that celebrate the wonder of each child.

May this dissertation be a ripple with constructive interferant potential. May it join the work of many others. And may we foster a wave of equitable and effective school practices.

Acknowledgments

Some people can't believe in themselves until someone else believes in them first.	Better than a thousand days of diligent study is one day with a great teacher.
– Sean Maguire, Good Will Hunting	– Japanese proverb

A social science PhD is a leap of faith—a gamble in which one bets on their ability to expand knowledge in the service of equity. Doubts abound. "Am I *smart* enough? Am I *knowledgeable* enough? Am I *driven* enough? And even if I am, am I *persuasive* enough to encourage the many gatekeepers (of funding and influence) to give me a chance to make a difference?" It is far easier to turn away from these questions than to surge forward with the brash confidence that seems required to *begin* (let alone complete) the endeavor. And so, I nearly did turn away. Countless times.

But then, Valentina, you believed in me. You believed in me so unflinchingly. You told me, "of *course* you can ace the GRE. Of *course* you can get into a great school. Of *course* you can conduct incisive research. You just have to go for it, babe, and you'll achieve it." It was the first time someone besides my parents or sisters had been so clear in their confidence in me. It shifted the ground under my feet. It lifted me up. It fortified my heart. Because of you, I *did* apply. I *did* persist. And now, I am ecstatic to be on the precipice of doing the work I feel I am called to do. Thank you, Valentina, for believing in me.

I've come to believe that a career in social science is a marathon. And that when you remember your "why," each step can feel as natural as breathing. Mom, Dad, thank you for instilling a clear sense of my "why." Dad, you taught me to believe that every person is part of one human family, and that there is no greater purpose than empowering empathy, unity, and equity. The knowledge that there is a difference worth making has been a guiding light during long nights. Mom, thank you for showing me not only why we run the marathon, but also *how* to do it. You have provided me with the most unbelievable example of what it means to bring a brilliant mind *and* caring heart to this critical work; to identify important questions by truly listening to those we serve; and to glean insights that can expand thriving for students, teachers, and communities. And you have modeled an unfathomable, unshakable commitment to progress. When I face setbacks, I think of the thousands of obstacles you have overcome, take a breath, and discover the strength to take another step.

At times, this PhD has felt less like jogging and more like chess or basketball. In those moments, I have been deeply grateful to the "coach" in my corner. Rucker, you have poured your soul into nurturing my curiosity, building my capabilities, honing my instincts, and empowering my success. I am unbelievably lucky to have had you as my advisor, and I look forward to bothering you with requests for advice and collaboration for years to come.

Midway through my PhD, my life changed. Valentina and I welcomed Kofi into the world, and into our lives. They say "kids come with their own personality!" Kofi definitely did. He was curious. talkative. playful. mobile. and *active*. There were times when I wondered if I could keep up with my PhD *and* my toddler. In those moments, Auntie, Uncle, Elena, and Terry—*you were my superheroes.*

Management, and UC Berkeley Graduate Division for supporting and celebrating my work. And thank you to Anne Gregory, and to the wonderful folks at Oakland Unified School District (Curtiss Sarikey, Barb McClung, and Andrea Bustamante) for being willing to partner with me on this work as it evolves.

Even further beyond the academic bubble, thank you to the "Travengers"—Alex, Andy, Anton, David, Grant, Guido, Hsin, Milo, Noah, Sean, Troy, and Victor— and the rest of my ninja warrior family— Adam, Barclay, Brian, Brittney, Calle, Cara, Chris, Colt, Dalton, Daniel, David, Deren, Donovan, Dougie, Drew, Ethan, Eric, Francisco, Gabe, Greg, Glenn, Hunter, Jamie, Jeri, Jesse, Jessie, JJ, Jody, Josh, Justin, Karen, Kevin, Kim, Kyle, Lindsay, Maggi, Mark, Meagan, Megan, Michelle, Mike, Morgan, Naj, Nick, Noah, Olya, Pauline, Rachael, Rian, Robbie, Strat, Tajma, Tee, Tiana, Tim, Tristan, Tyler, Verdale, Zhanique, and so many others. No matter what is happening in my professional life, there is such deep solace in knowing I always have a home with my ninjas—and that we can always come together and train together to overcome the world's toughest obstacles.

And thank you to my life homies who have been by my side from long before, and throughout, this recent chapter. Ajith, Akin, Anjana, App, Con-Bon, Danesh, Forkin, Fung, Jake, Jerome, Joester, Julie, Krystal, Malia, Moha, Nathaniel, Nayad, Nico, Richard, Sherrie, Suganthi, Vasu, Vijay, Yee, Yoyo — I know *for a fact* that if I'm ever in trouble, I can lean on you, and that, despite the fact that I can be immeasurably silly, you *get* me. That knowledge gives me the strength to take the kinds of risks necessary to chase one's dreams.

Ultimately, no words can fully capture the gratitude I feel for the many people who have made this PhD possible. In the hopes that a song by the smoothest 90's R&B group can do it better, I want to close with these lyrics.

> 'Cause even though when times got rough
> You never turned away
> You were right there
> And I thank you
> When I felt I had enough
> You never turned away
> You were right there
> And I thank you

With all of my heart,

- Sean

Table of Contents

List of Tables

<h1 style="text-align: center;">List of Figures</h1>

Introduction: Exclusion and Restoration

At first glance, Darius Robinson's[1] experience of school might seem extraordinary. Facing harsh punishment since preschool, he became disillusioned and, after being expelled for a minor act of defiance, he chose to drop out of school entirely. As Darius put it,

> I dropped out of school – actually they kicked me out – because I didn't want to give them my hat. It was real zero tolerance! I was expelled for defiance for putting a hat in my backpack instead of giving it to them. And I had had bad experiences since preschool so it was easy for me to be like "[forget] this." As a teenager, I was thinking "you don't care about us anyway. You just get paid checks per student in a seat. And they think we don't know, but we know.

The "bad experiences" Darius alluded to included suspensions throughout his k-12 experiences. His trajectory – of early and frequent exclusion followed by detachment and drop out – is alarmingly common, particularly among Black students. While Black youth represent 19% of preschool students, they represent 47% of students suspended from preschool (Government Accountability Office [GAO], 2018). Research has documented the relationship between exclusionary discipline and negative behavioral, academic, and carceral outcomes (Bacher-Hicks et al., 2019; LiCalsi et al., 2021). When compared to Black students who are not suspended, Black students who are suspended are two times less likely to graduate from high school, two times more likely to experience an arrest; and *seven* times more likely to experience juvenile confinement and adult confinement (Shollenberger, 2015). And Black students are overrepresented among those suspended at every grade level and in every educational context—from preschool to high school; in charter schools and traditional public schools (GAO, 2018). Research has suggested that racial disparities in discipline are also related to staggering and persistent racial disparities in academic achievement (Pearman et al., 2019).

Research thus suggests that discipline may be harmful. But that is not to say schools employ discipline without purpose. Instead, administrators and practitioners have long relied on suspensions and expulsions to deter misbehavior and to avoid the harms that can flow from students engaging in, or experiencing, bullying and violence (Griffith & Tyner, 2019). Many studies have revealed the adverse educational and behavioral effects of sustained exposure to misbehaving students (Deming, 2011; Imberman et al., 2012; Kinsler, 2013). For their part, educators have exhibited concern about negative spillover effects for over a century (e.g., Bagley, 1914) and thus have relied on suspensions to curb misbehavior (Adams, 2014). However, more recent research calls into question whether suspensions actually deter misbehavior (LiCalsi et al., 2021; Pesta, 2021), and even suggests that suspensions may exert *secondary* harms, leaving unsuspended students feeling anxious and disconnected from school environments, and harming their academic performance (Lacoe & Steinberg, 2019; Perry & Morris, 2014).

Thus, the question arises—are there alternatives to exclusionary discipline that might do more to reduce misbehavior while also avoiding the potential harms and apparent racial injustices occasioned by reliance on exclusionary discipline?

---

[1] All student and school names have been replaced with pseudonyms to protect student privacy.

Darius certainly found one. After dropping out of school, Darius spent time homeless and in gangs. However, after a run-in with the police, he was given two options: return to school, or go to jail. He chose school, but assumed he would eventually drop out of school again. Things did not go as he expected. The new school, Alice Walker Academy,[2] had recently adopted restorative practices. In a restorative paradigm, schools focus on strengthening relationships, proactively teaching students the skills needed to manage conflict, and guiding students through conflict resolution. Darius remembers attending community building circles in which students would share their emotions and deepen their connection to one another; and surface and resolve conflicts in a healthy manner, with the support of the community. At first, he thought it was "some kumbaya bullshit." But, he reveals, it worked—not just for him, but for the school overall:

> [After two weeks of circles at Alice Walker Academy], I realized it was the first time in my life I ever wanted to be at a school! Like we got circle today, I gotta go! I wanted to be in class, do projects, interact, be one of the first students called on. I felt good being up here! And the school had kids from West Oakland, East Oakland, Richmond, and yet there were two fights in the entire school the whole year. We had kids from south and north Richmond on the same basketball team and the team went undefeated while beating primary schools. We were kids that the system said couldn't function in the same environment. That's wild! All my friends [from before I went to Alice Walker Academy] are dead or in jail. Without [restorative practices], I'd probably be dead or in jail too. After I graduated, I realized I could bring this to homies to change my community. I was like, "this is what I want to do." I had already lost four friends to the justice system—four sentenced to 10-15 years under the age of 20. I had seen four murdered in the same year. I wanted to save my friends' lives.

Darius eventually became a restorative practitioner and trainer. Another restorative practitioner named Nia experienced a similar trajectory:

> Restorative practices saved my life. It's a lifestyle, not a practice or a program or none of that bullshit. It's not something you turn on or turn off. Once you start doing it, you will start having restorative conversations and learn to be a good listener. And you make really lasting relationships because [restorative practices] teach you not to be afraid of opening up to people.

According to Darius and Nia, restorative practices proved transformative for their life trajectories, and for the trajectories of many of the students they touched through their trainings. Might these practices have the potential to improve outcomes for students across the country? Some are skeptical, and research on restorative practices has sadly not kept pace with social change. While thousands of schools now implement restorative practices, few researchers have attempted to identify their impacts. Those that have attempted to ascertain the impacts of these practices have faced myriad challenges, including uncertainty regarding how to categorize restorative practices, difficulty finding programming that can reliably induce increases in implementation of these practices, and selection effects at the level of the student and at the level of the school.

---

[2] All student and school names have been replaced with pseudonyms to protect student privacy.

While daunting, the task of identifying the impacts of restorative practices is critical. From a policy perspective, it presents a first-order question that precedes inquiry regarding the kinds of programming that might shift practices, or the kinds of school conditions that might empower sustained implementation. Put another way, if restorative practices are truly effective, evidence of their potential impacts could empower schools and districts who adopt these practices to weather caregivers' concerns and pundits' critiques (Pollack et al., 2019); and to explore innovative solutions to natural implementation challenges (Garnett et al., 2020; Gregory, Ward-Seidel, & Carter, 2021). However, if, on the other hand, restorative practices are ineffective, thousands of schools across the country could leverage new insights to shift away from these practices, and to seek other practices better suited to meet students' needs. This dissertation thus seeks to answer a paramount, principal question: does student exposure to restorative practices drive improvements in academic, disciplinary, mental health, and school climate measures; and can it reduce racial disparities in discipline and academic achievement?

Section I of this dissertation summarizes research regarding the causes and consequences of exclusionary discipline. Research demonstrates that school practices play a role in generating racial disparities, and that school practices can be marshaled to reduce disparities.

Section II explains what restorative practices are, and what we know about their effectiveness. Rooted in relationships, these practices seek to ensure students are empowered to avoid, navigate, and repair conflict—all in the service of building a healthy school climate.

Section III summarizes quantitative research regarding the effectiveness of restorative practices, which largely supports the potential of these practices to improve school climate, reduce misbehavior, and reduce discipline. However, research has thus far been impeded by what could be termed the program / practice gap. Research has identified effects of various restorative *programs* (e.g., trainings designed to encourage teachers to use restorative practices). However, programs often do not result in teachers using restorative practices, and teachers can arrive at these practices via many other means. In short, research on programs may fail to detect the effectiveness of restorative practices—the true target of interest. What is needed, then, is research designed to identify the impacts of student exposure to restorative *practices*.

Section IV discusses challenges for identifying the causal impact of exposure to restorative practices. It then summarizes novel analyses suggesting that student exposure to restorative practices causes student benefits related to a suite of academic, disciplinary, behavioral, school climate, and mental health outcomes; and reduces racial disparities.

Finally, section V presents policy implications of this research, and suggests means of ensuring restorative practices can achieve their potential.

I. Discipline, Disparities, and School Practices

Schools recruit exclusionary discipline—such as suspensions and expulsions—in an effort to manage student behavior. For decades, researchers have theorized, and educators have assumed, that school exclusion can generate a deterrence effect that improves student behavior (Bagley, 1914; Casella, 2003; Ewing, 2000; Kafka, 2011; Matthews & Agnew, 2008). The theory follows that by removing misbehaving students from school environments, schools can convince all students not to misbehave lest they receive the sanction of exclusion. Moreover, by removing misbehaving students, educators seek to avoid the negative spillover effects that occur when students are exposed to misbehaving students (Deming, 2011; Imberman et al., 2012; Kinsler, 2013). However, other research suggests that exclusionary discipline may have perverse effects, with studies linking both direct and indirect exposure to lower academic achievement, less school connectedness, and more misbehavior, juvenile involvement, and adult incarceration (Bacher-Hicks et al., 2019; Lacoe & Steinberg, 2019; LiCalsi et al., 2021; Perry & Morris, 2014; Pesta, 2021; Way, 2011). Importantly, these correlations, while troubling, do not demonstrate that discipline *causes* these harms, and researchers have leveraged varied approaches to generate causal evidence that exposure to discipline is harmful. Concerningly, the harms of discipline are not experienced equally by all students. Black students are much more likely to be disciplined than their White peers (GAO, 2018). These disparities cannot be explained by student behavior, and differential treatment by teachers plays a role in driving discipline disparities (Barrett et al., 2021; Gregory et al., 2016; Huang & Cornell, 2017; Okonofua & Eberhardt, 2015; Owens & McLanahan, 2020). Given the role schools play in generating racial disparities in discipline, schools must shift their practices if they hope to reduce Black-With discipline disparities.

Why Do We Discipline?

In his seminal and aptly titled book regarding school discipline, Bagley (1914) laid out the justification for suspending students. "The group," he explained, "must be protected against the individual, [and] ... the individual must be protected against himself – against the impulses and desires that would interfere with his growth and his development" (p. 216). This view was not only popular among Bagley's contemporaries; it remained the prevailing perspective for over a century (Adams, 2014; Casella, 2003; Ewing, 2000; Kafka, 2011; Matthews & Agnew, 2008). Mindful of the need to "protect" students from the negative impacts of exposure to individual misbehaving students (Deming, 2011; Imberman et al., 2012; Kinsler, 2013), educators have long turned to suspensions to deter, and limit students' exposure to, misbehavior. But why are suspensions theorized to reduce misbehavior? Bagley's (1914) views on the topic are illuminating:

> Suspension . . . is the "natural" punishment for school offenses—a loss of "privilege" which ought to have an effective "sting." ... In handling a critical situation, it is often necessary for the welfare of the majority to send pupils from school and to keep them from school until reasonable obedience, order, and industry are assured. Where the compulsory attendance laws are well enforced, the pupil of school age will either have to remain in school or be committed to a reformatory, and where the right kind of cooperation exists between the school authorities and the juvenile courts, it is often a simple matter to settle troublesome cases by the expedient of suspension, depending upon a wholesome fear of the serious consequences to wheel the recalcitrant into line. (p. 208)

Historically, then, suspensions were believed to be effective in part because students were expected to understand that, if suspended, they would run the risk of commitment to a "reformatory" (juvenile hall). Given the conditions present in juvenile halls (both in the early twentieth century and today; Macallair, 2015), we can surmise that suspensions were meant to operate much like criminal sanctions, generating a deterrence effect that improves student behavior via coercion (e.g., Casella, 2003; Ewing, 2000; Matthews & Agnew, 2008). Bagley (1914) believed that the threat of punishment was a potent driver of student behavior, writing:

> The principle underlying the employment of rewards and penalties is very simple. If a certain kind of behavior invariably brings an unpleasant consequence, this type of behavior will tend to be repressed or inhibited. … From the point of view of psychological theory, there can be little doubt ...[that] the discipline of the "unpleasant" or the "disagreeable" [nature] undoubtedly teaches its lesson with greater certainty and celerity than the discipline of the "pleasant" or the "agreeable" [variety]. This can be fairly well demonstrated in the experimental study of animal psychology. (pp. 164–166)

Bagley then went on to reason that just as animals can be more rapidly induced to learn and perform tricks via punishments, children can be more readily encouraged to follow school rules via suspensions than via rewards. Bagley's argument evidences a straightforward view of child and adolescent psychology—misbehavior is merely a byproduct of incentive structures that do not sufficiently deter it. It is perhaps not surprising, then, that Bagley (1914) advocated for teachers to be able to leverage suspensions frequently and forcefully:

> The iron hand may be needed to initiate order and to teach the very basic lessons of respect for the authority of the law and for the rights of others. … Coercive measures must be swift, certain, and unerring. … Obedience must be secured, and there must be no halt in the proceedings until obedience is forthcoming. (pp. 132–139)

This view—that sanctions swiftly and universally applied can deter misbehavior—persisted for decades after Bagley wrote *School Discipline*. Then, in the 1980s when adolescent crime rates unexpectedly skyrocketed, this view combined with widespread concern about school violence to catalyze a movement toward zero tolerance policies (American Psychological Association Zero Tolerance Task Force, 2008; Michels et al., 2016). Under zero tolerance policies, students who break school rules face mandatory penalties which can include suspension, expulsion, and referral to law enforcement. These policies have been widely blamed for a huge surge in both discipline rates and in Black-White discipline disparities (Hoffman, 2014; Losen & Skiba, 2010). Because zero tolerance policies sometimes *require* that misbehaving youth be referred to law enforcement, these policies have also been implicated in the genesis of a "school to prison pipeline" (Skiba et al., 2016). In an influential report, Fabelo et al. (2011) reviewed Texas administrative data that revealed that students expelled or suspended had a significantly higher likelihood of being involved in the juvenile justice system. Still, as recently as 2014, surveys of school staff demonstrated that many teachers and principals support both zero tolerance policies and the widespread use of suspensions to address misbehavior (Adams, 2014).

Zero tolerance policies are particularly prominent in urban charter schools, and parents of students in these schools (including minority parents) often indicate that they value this disciplinary approach (Golann et al., 2019). Charter CEOs perceive their punitive approaches as being drivers not only of the academic success of the students they serve, but also of the success

of Charter schools at persuading parents to send their children to charter schools. In an *Education Week* opinion piece, Golann and Debs (2019) write:

> When some have challenged this model as unethical and racist, supporters point to parent demand. Eva Moskowitz, the CEO of Success Academy, which now runs 47 no-excuses charter schools with long waiting lists in New York City, has argued that parents who "believe in stricter discipline" are "voting with their feet" by enrolling their children in these schools.

In some respects, the approach to discipline in charter schools mirrors that endorsed over a century ago by Bagley (1914). Golann and Debs (2019) continue:

> No-excuses students are typically required to wear uniforms, sit straight, with their hands folded on the table, and their eyes continuously on the teacher. At breaks, they walk silently through the halls in single-file lines. Students who follow these stringent expectations are rewarded with privileges, while violators are punished with demerits, detentions, and suspensions.

In sum, educators utilize discipline, and parents select into punitive schools, because of persistent and widespread beliefs that discipline deters students from engaging in misbehavior (Bagley, 1914; Casella, 2003; Ewing, 2000; Kafka, 2011; Matthews & Agnew, 2008) and "protects" students from the negative effects of exposure to misbehaving students (Deming, 2011; Imberman et al., 2012; Kinsler, 2013). Thus, discipline is perceived by many as having both a primary benefit for students who experience it, and secondary benefits for students whose peers experience it.

However, researchers have long cautioned that discipline may fail to realize intended impacts on both primary and secondary dimensions. At the primary level, researchers warn that direct exposure to discipline may lead students to distrust and feel defiant toward adults in school (Pesta, 2021; Way, 2011); may drive *increases* in both misbehavior (LiCalsi et al., 2021) and youth involvement in the juvenile justice system (Bacher-Hicks et al., 2019; Fabelo et al., 2011; Hoffman, 2014; Losen & Skiba, 2010; Skiba et al., 2016); may exert psychological harms (Eyllon et al., 2020); and may reduce academic performance (Perry & Morris, 2014). And at the secondary level, researchers express concern that indirect exposure to discipline can lead youth to feel anxious and disconnected from their school environment, and may harm (rather than buttress) academic performance (Hinze-Pifer & Sartain, 2018; Lacoe & Steinberg, 2019; Perry & Morris, 2014).

Who is right? And how can we know? In this section, I will briefly explore some of the methodological challenges researchers face in identifying the primary and secondary impacts of school discipline. Thereafter, I will explore the methodological tools researchers have employed to overcome these challenges, and summarize the research base regarding the impacts of discipline. Finally, I will review research regarding racial disparities in discipline and elevate related insights into the scope, impacts, origins, and means of combatting these disparities. I begin this inquiry by exploring some of the correlates of exposure to discipline.

Correlates of Discipline

Scholarship demonstrates that exclusionary discipline is associated with negative academic, juvenile, and carceral outcomes. Balfanz et al. (2015) followed nearly 200,000 ninth graders and found that, compared to students who were not suspended, students who were suspended just *one* time in ninth grade were twice as likely to drop out of school; and the dropout rate increased steadily with each successive suspension. High school graduation rates showed an inverted pattern, declining steadily with each successive suspension, as did college enrollment rates, and post-secondary attainment. Also exploring correlates of discipline, Shollenberger (2015) reviewed data from the National Longitudinal Survey of Youth, tracking 9,000 students from 1997 through 2010. As depicted in Figure 1, she found that, during this thirteen-year period, compared to Black students who did not experience suspensions, Black students who experienced suspensions were 2.6 times less likely to graduate from high school, 2.0 times more likely to be arrested at some point in their lives, 7.0 times more likely to suffer juvenile confinement, and 7.5 times more likely to experience adult incarceration. Most jarringly, she found that 67% of Black students who experienced suspensions went on to experience arrests. Among White students, the relationship between suspensions and negative outcomes was similar, but Black students who experienced suspensions saw the highest rates of negative outcomes.

Figure 1

Risk of Negative Outcomes for Black and White Students Who Were, or Were Not, Suspended
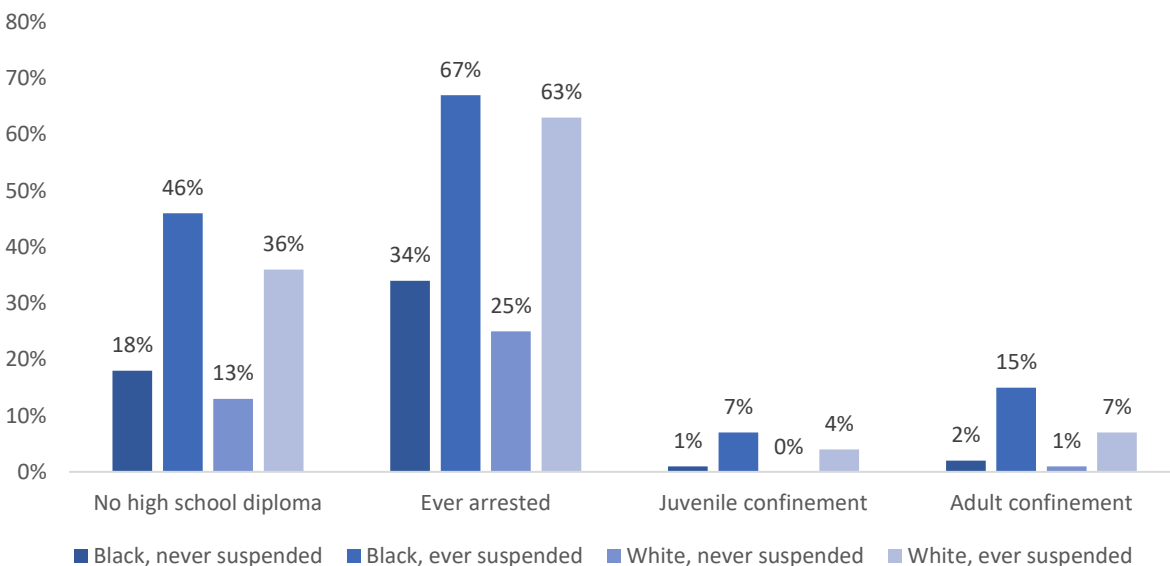
Note. Adapted from Adapted from "Racial disparities in school suspension and subsequent outcomes: Evidence from the National Longitudinal Survey of Youth," by T. Shollenberger, 2015. In D. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 37). Teachers College Press.

More recent work by Eyllon et al. (2020) suggests that early exposure to discipline is related to adolescent mental health challenges. Why might this be? One possible answer is that exposure to discipline depletes students' sense of connection to their school community—a key predictor of mental health (McChesney & Aldridge, 2018) and academic engagement (Del Toro & Wang, 2020). Relatedly, I recently reviewed California Healthy Kids Survey (CHKS) data and discovered a troubling link between exclusion and students' psychological wellbeing. Multivariate regressions based on data from 6,550 Black students from 256 schools who participated in CHKS surveys in the 2017-18 and 2018-19 school years (Table 1) suggest that experiencing a suspension in the last twelve months is negatively related to feelings of belonging in school, controlling for student sex, Hispanic ethnicity, parental education, and free-or-reduced-price lunch status. In similarly adjusted models, being in a school with a higher discipline rate among Black students was related to lower feelings of belongingness. Finally, the two phenomena (personal exposure to discipline and a high Black discipline rate) were each negatively related to feelings of belongingness after controlling for the other. And the Black discipline rate was negatively related to feelings of belongingness even among Black students who were not, themselves, subjected to discipline in the prior twelve months. These regressions suggest that, for Black students, there is a negative relationship between personally experiencing discipline and feeling welcome; and between seeing Black *peers* experience discipline and feeling welcome.

Table 1

Regression Models Predicting Black Students' Sense of Belonging (1–5) as a Function of Their Own Disciplinary Experiences and the Black Discipline Rate in Their Schools

	Model			
Variable	1	2	3	4 [a]
Suspended in last 12 months	-0.37^{***} (0.05)		-0.33^{***} (0.05)	––
Black discipline rate		-0.89^{***} (0.17)	-0.57^{**} (0.18)	-0.54^{**} (0.20)
Constant	3.09	3.15	3.16	3.19
n	6,550	6,550	6,550	5,775

Note. Models generated by author using California Healthy Kids Survey data for 2017–2018 and 2018–2019. All models control for student sex, whether students identify as Hispanic, parental education, and free- or reduced price-lunch status.
[a] Model 4 is limited to Black students who were not suspended in the last 12 months.
$^{**}p < .01.$ $^{***}p < .001.$

That the Black discipline rate is related to Black students' sense of connection to their school foreshadows ways that racial disparities in discipline might be related to other educational disparities. For example, reviewing district-level data from around the country, Pearman et al. (2019) have demonstrated that districts with larger Black-White disparities in discipline also tend to have larger Black-White disparities in achievement, even after controlling for many district-

level characteristics. Taken together, correlational research indicates that exposure to discipline and exposure to discipline disparities are both related to negative educational, psychological, juvenile, and carceral outcomes.

Correlation, of course, is not causation. In the foregoing section, I will explore the challenges researchers face when attempting to identify the causal impact of exposure to discipline, and the tools they leverage to overcome challenges and identify causal effects.

Impacts Obscured

Children vary. Caregivers vary. Schools vary. These statements may appear banal, but viewed through the lens of efforts to glean the causal impacts of exposure to discipline, each statement carries enough punch to knock a research team out in the first round.

Let's begin with the first statement, and recall research demonstrating that exposure to discipline is correlated with myriad negative outcomes. Is the association between discipline and outcomes a function of student exposure to discipline, or of the kinds of students who tend to experience discipline? An obvious example of this paradox is the link between student behavioral tendencies and discipline. Research indicates, for example, that children who grow up in neighborhoods with a higher degree of structural disadvantage are more likely to develop externalizing behaviors (Beyers et al., 2003), and that externalizing behaviors are related to exposure to school discipline (Kalu et al., 2020). Neighborhood structural disadvantage predicts not only misbehavior, but also adult incarceration (Alvarado, 2020). Given this, we might wonder whether the association between discipline and adult incarceration is really a reflection of how neighborhood structural disadvantage generates disparities in misbehavior which redound to both discipline and adult incarceration. There are, of course, more direct pathways linking student characteristics to both discipline and outcomes. For example, research indicates that low-income students are more likely to attend punitive schools (Kupchik & Ward, 2014), and that lower socioeconomic status is related to lower health and educational outcomes (Adler et al., 1994; von Stumm et al., 2020; Yao & Robert, 2008). Thus, any observed relationship between discipline and outcomes could merely reflect the way that affluence is negatively associated with discipline and positively associated with myriad positive outcomes.

Even if children did not vary at all, we would face challenges due to links between caregiver characteristics and both discipline and student outcomes. Earlier, I quoted Eva Moskowitz, the CEO of Success Academy, as saying that Black and Hispanic parents "believe in stricter discipline" and are "voting with their feet" (i.e., showing their support for strict disciplinary practices) by enrolling their children in charter schools (Golann & Debs, 2019). While bombastic, Moskowitz's claims (about charter school discipline and student compositions) are well supported. Charter schools have the highest discipline rate among all categories of schools (GAO, 2018). And, in the 2019-20 school year, charter schools served far more Black and Hispanic students (per capita) than traditional public schools, with 52% of charter school students being Black or Hispanic (as compared to 38% of traditional public school students; de Brey et al., 2019). The discrepancy was even starker in urban settings, where 50% of public school students were Black or Hispanic (de Brey et al., 2019) as compared to about 74% of charter school students were Black or Hispanic (Xu, 2022). Black and Hispanic parents are also more likely to live in communities that are heavily policed (Beckett et al., 2006; Fagan et al., 2010; Kirk, 2008), and individuals (particularly boys of color) who live in heavily policed neighborhoods also exhibit higher likelihoods of low educational achievement (Legewie & Fagan, 2019) and mental health challenges (Bennett, 2020; Hatzenbuehler et al., 2015). Taken

together, Black and Hispanic caregivers are more likely to send their children to more disciplinarian schools (Owens & McLanahan, 2020) such as charter schools. They are also more likely to live in heavily policed neighborhoods (which subject children to police contact risks that are associated with low educational attainment and mental health challenges). Might it be possible that the relationships between discipline and academic performance, and between discipline and mental health, are both truly driven by the way lower wealth, minority families are confined in housing choices to certain residential areas and sort their children into high discipline schools and high police contact neighborhoods? A similar theme emerges when one considers the role that parental education and wealth play. Research by Jackson (2009) demonstrates that more affluent and better-educated caregivers more often sort their children into schools that are better resourced, have smaller class sizes, and employ higher-quality teachers (i.e., teachers who score more highly on academic valued-added measures). These schools often have lower discipline rates (e.g., Finn et al., 2003; Williams & Wiggan, 2016) and higher academic achievement (Shin & Raudenbush, 2011). And research has often linked parental education and affluence with myriad positive academic, behavioral, and occupational outcomes (e.g., Dubow et al., 2009). As such, it is feasible that the link between discipline and academic performance derives from the relationship between caregiver affluence and school discipline, on the one hand, and between caregiver affluence and student outcomes, on the other.

A final potential source of spurious relationships between discipline and student outcomes stems from how schools that use disciplinary practices might also have other characteristics that can drive negative student outcomes. For example, these schools tend to have higher percentages of low-income and minority students (GAO, 2018), and this kind of concentrated disadvantage is related to negative peer effects, which are, in turn, related to lower academic engagement and performance (Hoxby, 2006). These schools also tend to have lower percentages of teachers of color (Roch et al., 2010), and this lack of teacher diversity is related to worse student-teacher relationships for Black students (Papageorge et al., 2020), and lower academic performance for Black students (Redding, 2019). Thus, relationships between high discipline rates and negative outcomes could conceivably be a function of the ways that more disciplinarian schools also feature concentrated disadvantage, negative peer effects, and lower teacher diversity—rather than some meaningful relationship between discipline and outcomes.

In sum, researchers struggle to identify the relationship between discipline and outcomes because children vary, caregivers vary, and schools vary—each in ways that are correlated with school disciplinary policies and student outcomes. So how can researchers overcome these challenges and identify and isolate the causal effects of school discipline policy (independent of these other important factors)?

Impacts Uncovered

Researchers seeking to glean the causal effects of discipline (or any social phenomena) generally leverage one of three approaches. Before describing these approaches, we must review two key terms. When seeking to understand the causal impact of any phenomenon (e.g., school discipline, placement in juvenile confinement, or exposure to restorative practices), researchers will be faced with one of two situations. The first, and substantially more common, situation is that exposure to the phenomenon is endogenous, meaning it is related to other unmeasured phenomena (such as student traits, caregiver sorting preferences, or school practices) which are, themselves, related to outcomes of interest. Above, I described many reasons to suspect that exposure to discipline is endogenous, and why this endogeneity muddles (or confounds) our

ability to understand the impact of discipline. The second situation is that exposure to the phenomenon is exogenous (meaning unrelated to unmeasured factors that also predict the outcome). In these rare cases, ascertaining the impact of the phenomenon is as simple as determining whether certain expected outcomes follow the occurrence of the phenomenon. For example, if suspensions were exogenous, then to measure the impact of suspensions on academic achievement, we would merely ascertain how academic performance shifts after a suspension. If only!

Researchers recruit strategies that are responsive to the degree of endogeneity they face, and attempt to either 1) adjust for endogenous variation, 2) identify and exploit naturally occurring exogenous variation, or 3) generate and exploit artificial exogenous variation. Research on juvenile courts provides prime examples of each strategy.

Strategy 1: Adjust for Endogeneity

Increasingly, researchers face situations where variation is unquestionably endogenous, but where available data are so thorough that the data may be recruited to attempt to model, and adjust for, sources of endogeneity. For example, Jordan (2012) sought to ascertain the impact of juvenile confinement, and had relatively detailed data regarding 300 youth in three Pennsylvania counties, including whether they were detained (the exposure of interest) and a number of variables that could be used to predict each juvenile's probability of being detained, including their age at referral, their gender, their race, the number of prior referrals to juvenile court they had received, whether they had been incarcerated before, whether they received a predisposition release, the offense they were charged with, their role in the offense, the number of victims, whether a weapon was used during the crime (and if so, what kind), and type of defense attorney (public versus private). The author used these variables to construct a logistic regression model and predict each juvenile's unique estimated probability of receiving a juvenile confinement. They then matched each juvenile who received a juvenile confinement to a "virtual twin," or a juvenile with a remarkably similar probability of being confined who was *not* confined. Finally, they evaluated whether juveniles who were confined had, on net, a higher recidivism rate than their virtual twins, and found that they did. This process (known as "propensity score matching") has been employed in a number of studies and contexts to leverage data to adjust for endogeneity. More recently, Walker and Herting (2020) used a similar approach, applying a propensity score matching approach to review data from 46,000 juveniles and determine that pretrial detention increased recidivism.

These approaches have been leveraged to study a range of questions related to racial bias. For example, when I sought to understand the impact of workplace diversity on racial bias, I recruited eleven variables in the General Social Survey, including workplace sector, residential diversity, and commuting zone. Leveraging these variables, I constructed a logistic model to predict each White individual's probability of having a Black coworker, then matched each White individual who had a Black coworker with the White individual who had the most similar probability of having a Black coworker but who did not have one. Finally, I compared mean bias levels among those who had a Black coworker to mean bias levels among their "virtual twins" (those who had similar probabilities but did not have a Black coworker), and found evidence to suggest that having a Black coworker was related to lower levels of anti-Black bias (Darling-Hammond, Lee, & Mendoza-Denton, 2021).

Each of these research designs relied on a relatively heroic assumption: that the logistic regression models used to predict exposure to the phenomena of interest (juvenile confinement

or workplace diversity) included *all* sources of endogeneity. One can imagine many important variables could have been omitted from these models. If so, the resulting causal estimate would be biased. Are there approaches that sidestep this process of attempting to adjust for the seemingly endless sea of sources of endogeneity?

Strategy 2: Exploit Naturally Occurring Exogenous Variation

Like Jordan (2012), Aizer and Doyle (2015) also sought to evaluate the impact of juvenile detention. However, they leveraged two structural aspects of juvenile courts. First, juvenile defendants are assigned to judges randomly. Secondly, as I discussed in research regarding the structure of juvenile courts (Darling-Hammond, 2017), juvenile judges enjoy essentially unfettered discretion. Thus, the approaches that judges take to similar cases can vary wildly. As a result, one can estimate the probability that a given judge will send a child to juvenile detention based on how often they have sent previous, randomly selected defendants to juvenile detention. Because assignment to judges is random, the probability that a juvenile defendant will receive juvenile detention can be seen as having a random component. The structure of juvenile courts therefore creates a situation where exposure to juvenile incarceration can be seen as partially exogenous. Aizer and Doyle exploited this partial exogeneity, using an instrumental variables approach to estimate that juvenile incarceration decreases high school graduation by thirteen percentage points and increases adult incarceration by twenty-three percentage points.

In some cases, institutional structures do not present opportunities to exploit exogenous variation. What can researchers do to create exogenous variation?

Strategy 3: Generate Artificial Exogenous Variation

Courts around the country have recently implemented restorative justice proceedings to help guide juvenile defendants through a process of understanding and repairing the harm they caused through their illegal actions. While initial research suggests that participation in restorative processes is associated with lower rates of recidivism, restorative proceedings are lengthy and resource-intensive. Thus, counties must make difficult choices regarding which juvenile defendants will, and will not, gain access to these promising practices.

Shem-Tov et al. (2021) leveraged this situation, working with the San Francisco District Attorney's office to randomly assign certain juveniles to receive restorative justice conferencing. By creating artificial, exogenous variation in who gained exposure to these conferences, the authors were able to ascertain that participation in the program markedly reduced recidivism. It is worth noting that this kind of randomized controlled field experiment design was availing here, but would not have been in the case of evaluating the impact of juvenile detention, for at least one major reason: ethics. While it is arguably ethically permissible to use random processes to determine which juveniles gain access to a finite, potentially positive resource, it is unquestionably unethical to use random processes to determine which juveniles will be forced to go to juvenile detention. Thus, where ethical considerations permit, researchers can generate and exploit exogenous variation to glean causal effects.

We have now reviewed three approaches for gleaning causal effects. What have researchers found when employing these three strategies to identify the causal impacts of discipline on student outcomes?

Causal Impacts of School Discipline

LiCalsi et al. (2021) recently leveraged the first approach (adjusting for endogeneity) to determine the causal impacts of discipline. Leveraging extraordinarily detailed data from the New York City Department of Education, the authors were able to operationalize a propensity score matching approach, using a machine learning approach (recruiting *eighty* variables) to estimate each student's unique propensity of being suspended, matching students who had similar propensities of suspension but who had distinct disciplinary experiences, and then regressing students' outcomes on their suspension experiences with additional adjustment for student and school demographics, and student and school standardized test performance. Via this process, they found that receiving an out-of-school suspension was related to higher rates of student misbehavior in each of the subsequent four years after the suspension.

Bacher-Hicks et al. (2019) used the second approach (exploiting naturally occurring exogenous variation). When Charlotte-Mecklenburg schools redrew its district boundaries, many students who had previously been in the same schools were suddenly (and quasi-randomly) assigned to new schools. In a manner similar to Aizer and Doyle (2015), the authors used schools' prior disciplinary histories to determine the probability that a student who attended any given school would receive a suspension. By exploiting the sliver of exogeneity created by the boundary redrawing, the authors were able to determine that exposure to punitive schools depresses academic performance and increases the odds of arrest and adult incarceration for *all* evaluated sub-categories of students. Thus, whereas prior literature and theory had assumed that suspending some students would improve long-run outcomes for the majority of remaining students, strict schools exerted negative long-run effects for students regardless of the demographic subgroup to which they belonged. Finally, the authors found that exposure to "strict schools" exerted a uniquely pernicious impact for Black students.

Research thus indicates that discipline can have perverse effects on those who experience *direct* exposure to discipline. But what of those who experience discipline indirectly? As noted, educators often leverage discipline to ensure students can avoid exposure to misbehaving students, which, research suggests, can have negative educational and behavioral consequences (Deming, 2011; Imberman et al., 2012; Kinsler, 2013). However, even if we assume that exposure to *misbehavior* is harmful, it does not necessarily follow that indirect exposure to *discipline* is beneficial. Put another way, it may be the case that while indirect exposure to suspensions limits students' levels of exposure to misbehavior, they also exert other harms. Perry and Morris (2014) theorized, for example, that indirect exposure to discipline could erode students' sense of trust in the fairness of a school which could lead otherwise compliant students to misbehave.

Perhaps not surprisingly, then, some research has identified *negative* spillover effects, indicating that discipline may harm school climate for students and teachers (Lacoe & Steinberg, 2019); may lead to negative academic outcomes (Perry & Morris, 2014); or may simply not have an effect on unsuspended students at all (LiCalsi et al., 2021). In a sophisticated analysis that used the first causal research approach (adjusting for endogeneity), Perry and Morris (2014) reviewed data regarding over 14,000 students from a school district in Kentucky. They leveraged a within-student estimator to overcome many sources of endogeneity. In short, they first evaluated how students' outcomes shifted over time, and then evaluated how students' exposure to discipline shifted over time, and finally explored whether students whose exposure to discipline grew over time also saw their outcomes decline over time. By "fixing" the student (and evaluating how changes within a student over time in the level of exposure are related to

changes within a student over time in the outcome), this approach adjusts for all time-invariant student-level endogeneity. Using this approach, the authors found that students in schools that grew more reliant on discipline saw a slowing in math and reading achievement growth. Hinze-Pifer and Sartain (2018) leveraged a similar approach to identify the impact of exposure to schools that shifted *away* from discipline, and found that reductions in suspensions were related to small, but significant, increases in student test scores and attendance.

Taken together, research on the effects of discipline complicates the notion that discipline generates a deterrence effect, or has positive externalities. Instead, research suggests that discipline might lead to increases in misbehavior for those disciplined, and *negative* externalities for those not disciplined (including declines in academic achievement). Given research regarding the harms of discipline, one might worry that higher discipline rates among Black students present an equity issue. This begs the question, what is the scope of Black-White discipline disparities, from where do they emerge, and how can they be addressed? I answer those questions in turn below.

Widespread and Enduring Black-White Disparities in Discipline

The most recent federal education data demonstrates that, without question, Black students are disciplined more harshly than their White peers (GAO, 2018). As detailed in Figures 2 and 3, Black students were 3.9 times more likely than White students to receive an out-of-school suspension. And Black-White disparities appeared in all evaluated school contexts and among all evaluated student sub-populations. For example, Black boys are disciplined more harshly than White boys; Black and disabled students are disciplined more harshly than White and disabled students; Black students in poor schools are disciplined more harshly than White students in poor schools; and on. Most jarringly, Black preschool students are 3.7 times more likely to receive and out-of-school suspension than White preschool students. Finally, perhaps unsurprisingly given their stated focus on using discipline to regulate behavior, Black-White discipline disparities are more severe in charter schools than in any other school context, as Black charter school students are nearly *five times* more likely to have received an out-of-school suspension than White charter school students.

Figure 2

Black–White Disparities in Various Exclusionary Outcomes
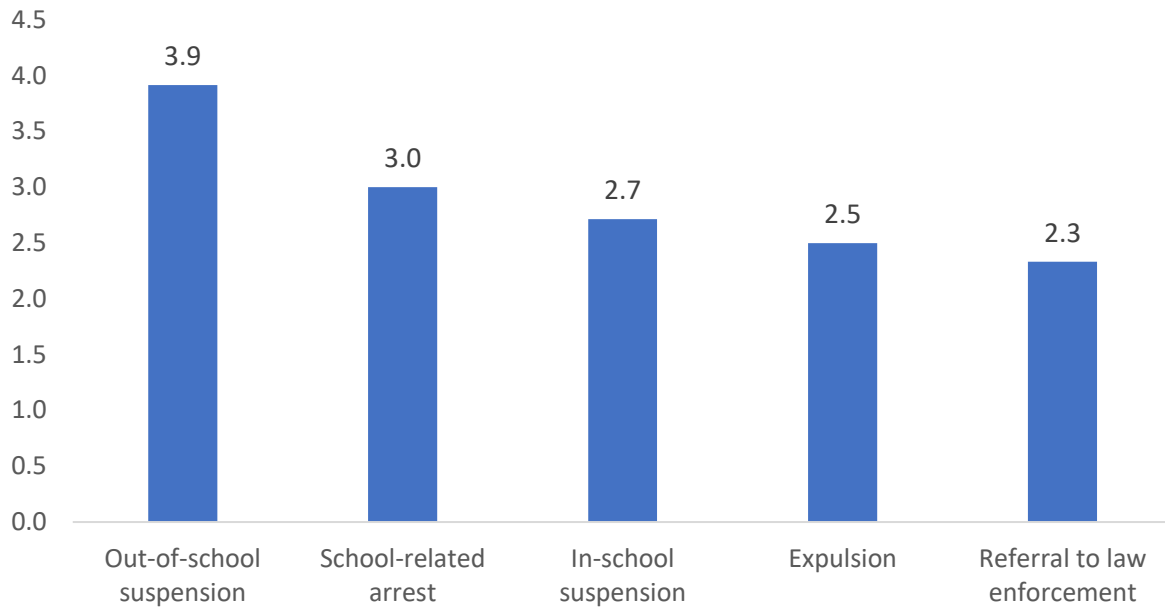
Note. The chart indicates how much more likely Black students were than White students to receive a given exclusionary outcome. For example, Black students were 3.9 times more likely to receive an out-of-school suspension than White students. Adapted from "Discipline disparities for Black students, boys, and students with disabilities," by Government Accountability Office, 2018, p. 71 (https://www.gao.gov/assets/gao-18-258.pdf).

Figure 3

Black–White Disparities in Out-of-School Suspensions, by Student Demographics
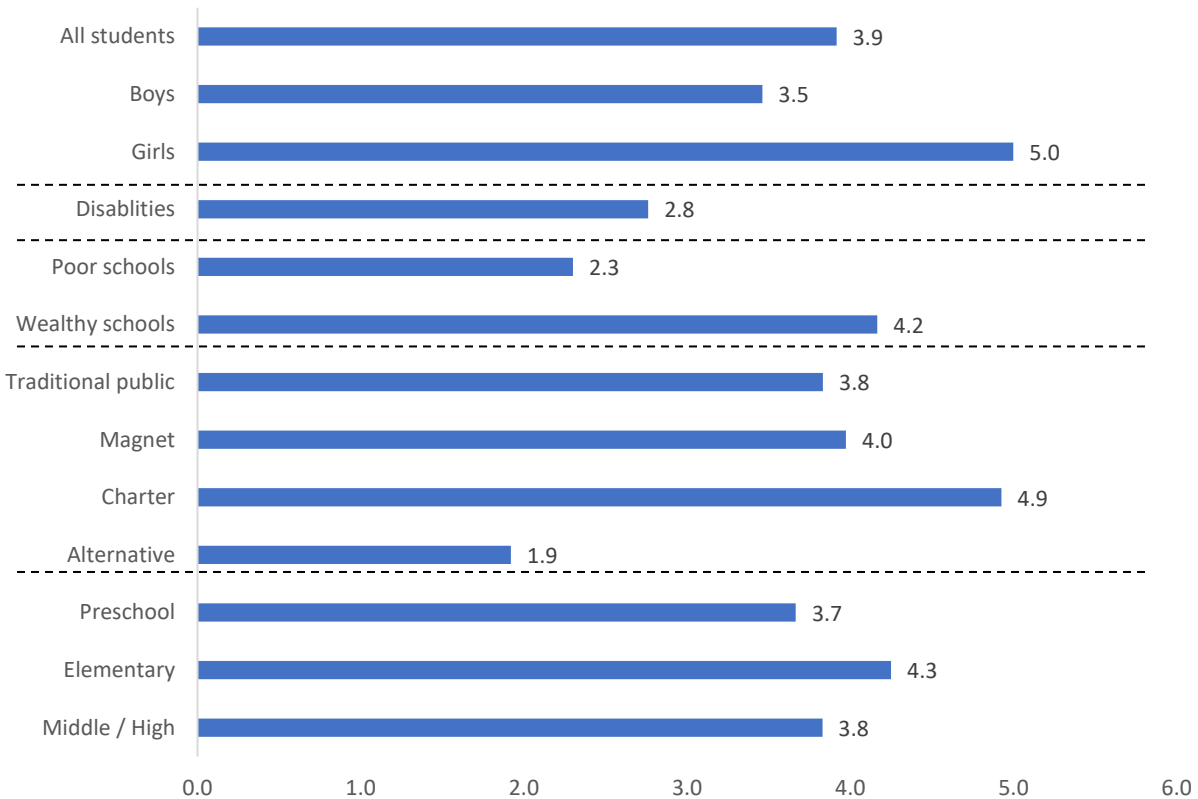
Note. The chart indicates, for various student subpopulations, how much more likely Black students were than White students to receive an out-of-school suspension. For example, among charter school students, Black students were 4.9 times more likely to receive an out-of-school suspension than their White peers. Dashed lines separate types of student and school subdivisions. Adapted from "Discipline disparities for Black students, boys, and students with disabilities," by Government Accountability Office, 2018, p. 71-81 (https://www.gao.gov/assets/gao-18-258.pdf).

These disparities also appear in California schools. My review of 2018-19 data from the California Longitudinal Pupil Achievement Data System (CALPADS) revealed that Black students are more likely to receive out of school suspensions than any other category of students, and that this phenomenon persists for students at each grade level; and for students who are, or are not, economically disadvantaged (Figure 4). Put another way, regardless of grade level or economic status, Black students were more than twice as likely to receive an out-of-school suspension as White students, Hispanic students, or Asian students.

Figure 4

Percentage of California Students Who Received Out-of-School Suspensions in 2018–2019, by School and Student Type, and by Student Racial Identity
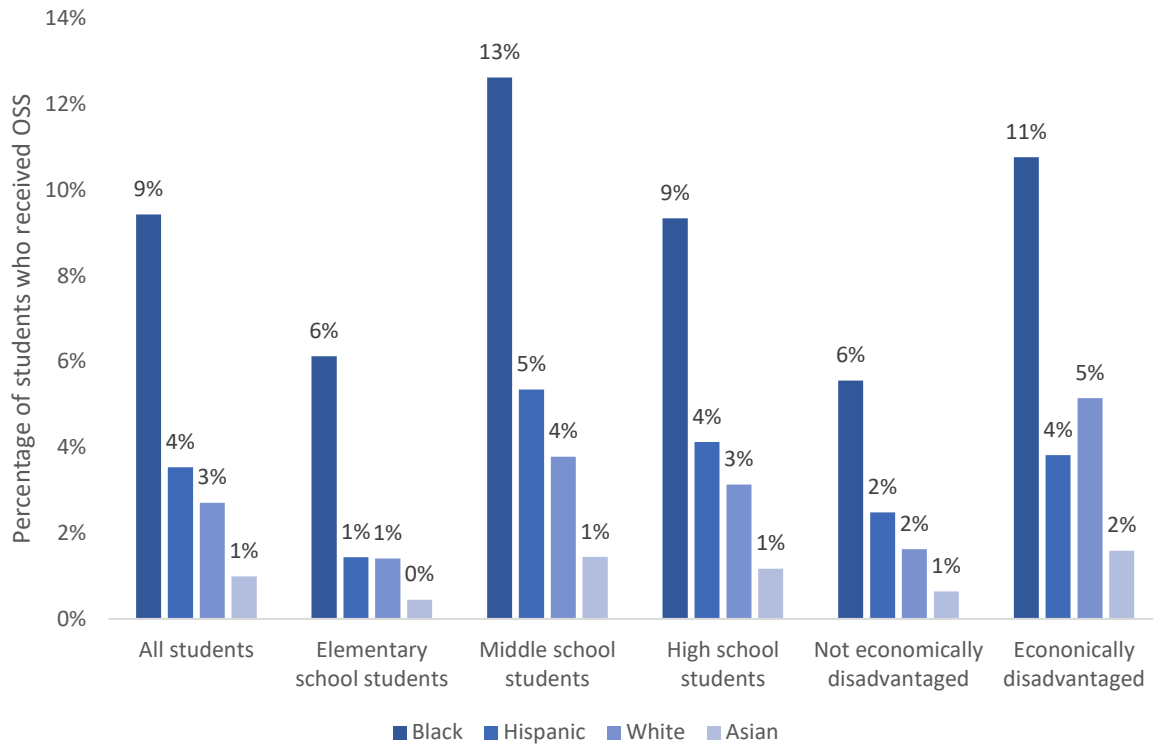
Note. Chart developed by author based on data from $N = 3,256,134$ students in the CALPADS 2018–2019 data system.

Because Black students are uniquely vulnerable to school exclusion, and exclusion is related to academic achievement, this dissertation focuses primarily on Black-White disparities in discipline and academic achievement. However, other student characteristics are also associated with educational disadvantage. For example, data from the National Assessment of Educational Progress (2019) demonstrates a persistent Hispanic-White gap in standardized test performance. Asian students represent a large and growing proportion of California's youth, and the impacts of school practice reforms on Asian students is remarkably understudied. In this dissertation, I thus review outcomes for Black, White, Asian, and Hispanic students separately. The central question of this section, however, is why are Black students so uniquely vulnerable to school exclusion, and what role, if any, do school practices have in ameliorating Black-White disparities? I review these questions below.

The Role of School Practices in Engendering Racial Disparities in Discipline

Research strongly indicates that differential treatment may be a driver of racial disparities in discipline. In a seminal, randomized controlled trial, Okonofua and Eberhardt (2015) asked teachers to read vignettes describing a student who engaged in two consecutive misbehaviors. Some teachers were randomly assigned to read a vignette about a student with a stereotypically White sounding name, and others were assigned to read a vignette about a student with a stereotypically Black sounding name. Aside from the students' names, the vignettes were identical in the two conditions. After reading about the two incidents of misbehavior, teachers were asked to indicate how troubled they felt about the student's behavior, and how harshly they felt the student should be disciplined. As depicted in Figure 5, compared to those assigned to read about a White student, teachers randomly assigned to read about a Black student felt significantly more troubled by the student's behavior, and suggested harsher discipline. The authors deem this the "two strikes" effect: even when Black and White students engage in identical behavior, after just *two* misbehaviors, Black students are more often deemed "troublemakers" and subjected to harsh discipline.

Figure 5

Racially Differential Responses to Identical Student Behavior
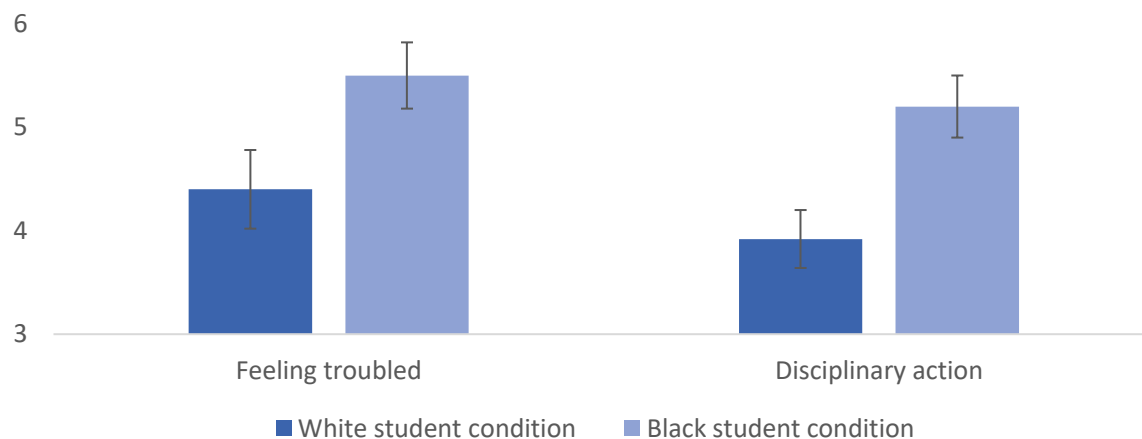
Note. Error bars represent 95% confidence intervals around mean values. Adapted from "Two Strikes: Race and the Disciplining of Young Students," by J. Okonofua and J. Eberhardt, 2015, *Psychological Science*, *26*(5), p. 619 (https://doi.org/10.1177/0956797615570365). Copyright 2015 by Sage Publishing. Adapted with permission.

The GAO (2018) report suggested that discipline disparities emerge even in preschool. Might differential treatment play a role even in early childhood contexts? Research with preschool teachers by Gilliam et al. (2016) suggests it might. In an eye-tracking study, the authors asked teachers to watch six minutes of video content of young children playing and scan for any "problem behaviors." In reality, all of the children were playing nicely, and no problem

behavior was present. However, when asked to find misbehavior, teachers focused significantly more of their attention on Black children, and on Black boys in particular.

Of course, the vignette and eye-tracking studies are one step removed from real-world conditions. However, research using real-world data has found that racial disparities in discipline cannot be fully accounted for by racial disparities in student behavior (Huang & Cornell, 2017), leaving open the possibility that differential treatment plays a role. In a recent article, Owens and McLanahan (2020) reviewed 5,000 student records from the Fragile Families and Child Wellbeing Study—a population-based birth cohort study of children born in large U.S. cities between 1998 and 2000—to explore how much of the Black-White discipline disparity stems from Black students sorting into more punitive schools, how much stems from Black students misbehaving more, and how much stems from Black students being treated more harshly than White students in the same school. Using parent and teacher surveys to identify students' behavioral tendencies, and using school-level data to identify school demographic composition, they find that Black students were more likely to receive suspensions than White student with similar behavioral tendencies and in schools with similar racial compositions. They conclude that racial differences in misbehavior accounted for only *nine percent* of Black-White disparities in discipline; that between-school sorting accounted for 21 percent of the gap; and that differential treatment accounted for a whopping 46 percent of the Black-White discipline gap.

Specific evidence of differential treatment has also been observed in real-world data. Gregory et al. (2016) report that, compared to White students, Black students are 26.2% more likely to receive an out-of-school suspension for their first recorded disciplinary incident. This work is bolstered by that of Barrett et al. (2021), who find Black-White disparities in discipline across districts, across schools, and within schools, and find that when a Black and White student fight, the Black student is punished more harshly even after controlling for students' economic status, prior disciplinary histories, prior behavioral histories, sex, special education status, and academic achievement. Also leveraging a within-incident, Shi and Zhu (2022) review K-12 data from 2008 through 2018 from the North Carolina Education Research Data Center. They find that even when a Black and White student are collectively involved in an incident of misbehavior, and when the school codes the Black and White student as having engaged in the same misconduct, the Black student nonetheless generally receives a more severe disciplinary sanction.

While research has not precisely pinpointed the causes of racially differential treatment, it has provided some clues. One source may be racial bias. Using data from the Project Implicit data system, and from the Office of Civil Rights Data Collection, Riddle and Sinclair (2019) found that school districts that have a higher degree of pro-White / anti-Black subconscious bias also tend to evidence larger Black-White disparities in school discipline. But how might subconscious bias engender disparities? Ample research demonstrates that subconscious racial bias more readily impacts decisions when actors have more discretion or are faced with more uncertainty (Dovidio & Gaertner, 2000; Johnson et al., 1995; Pearson et al., 2009; Saucier et al., 2005). Absent a zero tolerance policy or discipline ban, teachers always face both discretion and uncertainty when choosing whether, and if so how, to discipline a student. Psychological research thus suggests that teachers' subconscious biases could drive these decisions, and may partially explain differential treatment. Smolkowski et al.'s (2016) recent work lends credence. They find that racial disparities in discipline emerge most in contexts where discretion is most pronounced, providing further support for the notion that implicit bias may act via discretion to engender racial disparities in discipline.

A related, recent explanation for racial disparities is teacher-student race matching. Those contributing to this burgeoning body of work posit that—whether due to harboring lower levels of anti-Black bias or enjoying higher degrees of cultural congruence with Black students—Black educators generally hold more positive views of, and are better able to engender positive behavior from, Black students. Work by Papageorge et al. (2020) demonstrates how Black educators do indeed hold more positive expectations of Black students: compared to White teachers, Black teachers are 12% more likely to predict that a given Black student will graduate high school, and 30% more likely to predict that the student will complete a four-year college degree. In addition, they found that White teachers held higher expectations of White students than they did of Black students. So, what does research suggest happens when Black students are matched with Black educators? In their review of evidence regarding student-teacher race matching, Redding (2019) found that studies generally find that across educational contexts, Black students are described as being less likely to exhibit externalizing behaviors (e.g., physical aggression, or disobeying rules) when rated by a Black teacher (as compared to when rated by a teacher of another race). They note that in some studies (e.g., Downey & Pribesh, 2004; Wright et al., 2017), Black-White disparities in misbehavior vanished when Black students were reviewed by Black educators. Relatedly, many scholars (Grissom & Jones, 2020; Meier & Stewart, 2003) have found that schools with a more diverse teaching workforce exhibit lower discipline rates for Black students; and work by Rocha and Hawes (2009) finds that such schools also exhibit smaller Black-White discipline disparities. Why do diverse teacher workforces correlate with lower discipline rates for Black students, and lower Black-White discipline disparities? Roch et al.'s (2010) study provides one answer: schools with more diverse teaching workforces are more likely to employ discipline policies that are "learning-oriented" rather than "sanction-oriented."

Taken together, available research suggests that school policies and practices (such as hiring practices and discipline policies) contribute to racial disparities in discipline rates. In short, extant research provides evidence that, in many cases, schools are treating Black students more harshly. However, under Title VI of the Civil Rights Act, schools, as government actors, may not treat students of different racial groups distinctly (Department of Education [DOE] & Department of Justice [DOJ], 2014). As such, evidence that differential treatment in schools is driving racial disparities in discipline could give government actors reason to intervene. How, then, did the federal government react to evidence of a link between racial disparities in discipline and racially disparate treatment?

The Restorative Reaction

The DOE and DOJ (2014) issued a joint "Dear Colleague" letter in which they clarified that "[f]ederal law prohibits public school districts from discriminating in the administration of student discipline." The letter warned that the departments could claw back Title I funds from districts if they violated federal law. Finally, the letter suggested that to avoid violations, schools should consider various alternatives to exclusionary discipline, "including conflict resolution [and] restorative practices" (DOE & DOJ, 2014). Concurrently, the DOE provided districts with grant funding (DOE, 2014) and technical guidance (Alfred, 2013) for implementing restorative practice programs. As demonstrated in Figure 6, data from the California School Staff Survey suggests that the combination of the Dear Colleague letter, grant funding, and technical supports preceded a large and discontinuous increase in schools' use of restorative practices. It also suggests that restorative practices uptake continued steadily after 2014-15.

Figure 6

Mean Staff Responses Regarding Whether "This School Helps Students Resolve Conflicts With One Another"
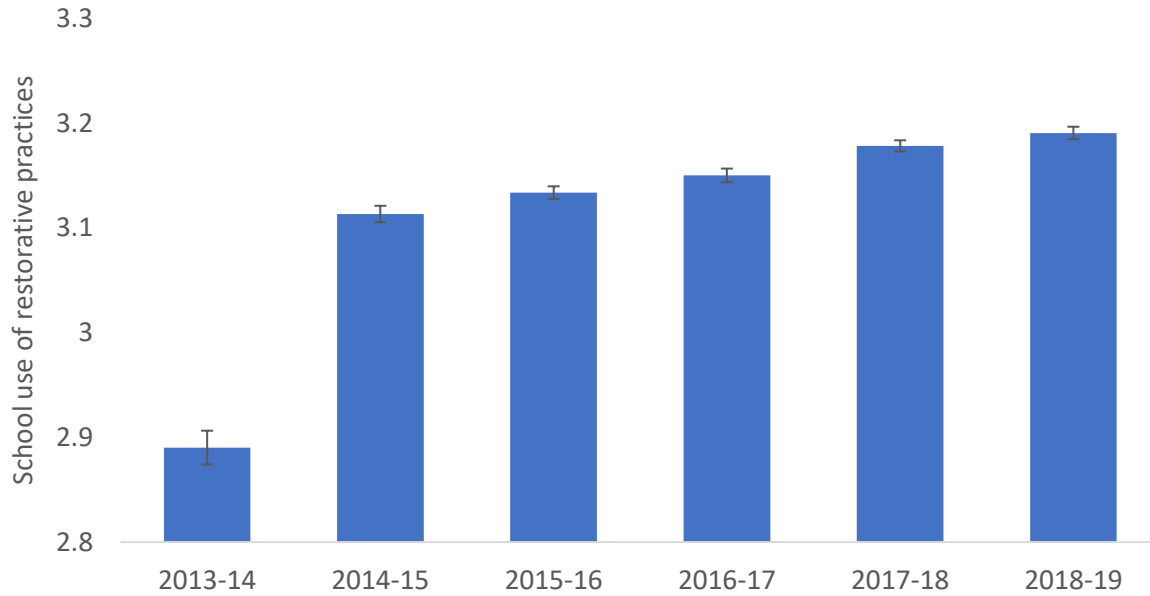
Note. Responses ranged from 1 (*strongly disagree*) to 4 (*strongly agree*). Chart generated by the author using 227,579 surveys from the California School Staff Survey database. Bars depict average staff responses in each year. Error bars represent 95% confidence intervals around mean values.

In the next section, I explore what restorative practices are and why they are theorized to reduce discipline rates and ameliorate discipline disparities.

II. What Are Restorative Practices?

Restorative practices encompass a wide array of practices designed to repair harm when conflict occurs; and to proactively improve relationships so punishable conduct is less common—all in the service of improving outcomes for students, school staff, and communities. Critically, restorative *practices* are distinct from restorative *programs* (see Figure 7).

Figure 7

Relationships Between, and Typologies of, Restorative Programs, Practices, and Outcomes

Programs, Practices, and Outcomes

Restorative *programs* are systems of hiring, training, and support that are designed to encourage school community members (staff and students alike) to engage in restorative practices. These can include hiring a restorative coordinator to manage conflict repair sessions; training teachers in relationship-building dialogue techniques; providing ongoing coaching and professional development to improve and expand practices; and incentive structures designed to shift practitioners away from exclusionary practices and towards conflict remediation. There is no established definition regarding which program model must be present for a school to be categorized as "restorative." Due to the lack of clear criteria, schools identified as using "restorative programming" comprise a diverse tapestry. Still, at their core, these schools have invested at least some amount of time and energy into encouraging community members to use restorative practices, ostensibly to improve relationships and eschew exclusion.

Restorative programs fall loosely into two models: "add-on" programs and "whole school" programs. In the former, schools add limited restorative functions to their existing disciplinary arrangements. This can take the form of diverting some students who would otherwise be suspended to restorative proceedings, or of hiring a single restorative coordinator to oversee selective restorative activities within the school. The latter, whole school, model involves providing instruction in restorative concepts and skills to all school personnel (both staff and students) so restorative concepts and approaches are infused in as many school interactions as possible. The "whole school" model can be augmented with continuous professional development in the form of coaching and/or professional learning communities (or

PLCs) dedicated to expanding and improving the use of restorative practices. Both "add-on" and "whole-school" models are often embedded within other schoolwide initiatives that are designed to improve school climate, such as Social and Emotional Learning (SEL) and Positive Behavioral Interventions and Supports (PBIS).

Restorative *practices* are the specific actions that community members might engage in in a restorative school, and that theoretically can produce certain benefits for students, staff, and community members. Restorative practices can be roughly subdivided into two types: *repair practices* and *community building practices*.

The first restorative practices that were formally introduced into schools were repair practices, and often were described by the related term "restorative justice." As theorists (e.g., González, 2012; Karp & Breslin, 2001; Zehr, 2002) explain, in the K-12 setting, repair practices are meant to bring together all stakeholders to resolve issues rather than control student misbehavior through punitive exclusionary approaches. Programs for fostering repair practices range from training for teachers in conflict-responsive dialogue techniques for the classroom to hiring professional restorative coordinators to guide restorative conferences with students, staff, and other stakeholders. Formal conferences can include victims, misbehaving students (often described as "respondents" as they are asked to respond to, or repair, the harm they've caused), and facilitators, but may also include community members (e.g., witnesses, friends, and family members). The term "victim" is often used broadly and can include school community members who speak to the general harm caused by respondents' actions (e.g., in the case of vandalism). Together, all of the conference participants (including the respondent) aim to determine a reasonable and restorative response to the harm done. These can include community service, restitution, apologies, or agreements to change specific behaviors, such as the respondent agreeing to comply with certain conditions, sometimes in exchange for incentives (Stinchcomb et al., 2006).

The second body of restorative practices are "community building" practices. These practices are designed to foster an interconnected school community and healthy school climate in which punishable transgressions are less common (Brown, 2017). The best-known community building practices are community-building circles, which are semi-regular convenings (e.g., "each Monday morning in homeroom") structured to help students and staff deepen relationships and trust so that misbehavior becomes less common. Another common community-building practice is the re-entry circle. In these circles, community members gather to help students who have been removed from the school community (for example, due to out-of-school suspensions) to feel re-integrated into the community. These circles are designed to ensure returning community members have the social support needed to thrive (and to avoid misbehaving). A final body of community building practices are practices designed to help students develop their social and emotional capacities to manage conflict when it occurs. These include role-playing conflict situations, reflecting on past conflicts, and discussing sources of stress and anxiety in students' lives. Capacity development activities often occur during community-building circles.

Programs designed to catalyze community-building practices include widespread training in affective communication techniques to bolster social connections, hiring restorative coordinators to lead community-building activities, and providing teachers with training and coaching regarding how to lead community-building and re-entry circles.

The Theory of Restoration

> To me, an ideal justice system ... would be a problem-solving and a healing system rather than a punitive system... Think about when you're a kid and you throw a baseball through your neighbor's window and if you're so lucky to have the kind of parents who would take you by the ear to your neighbor, have you apologize, find out how much it costs, and if you're so lucky to have an allowance, redact it until you have paid them back for how they paid to have that window repaired; right? You've learned something and you've redeemed yourself; right? ... Whatever it is, those are wake-up moments for us, and I think our justice system should be about those things. I think that would cause the moral change within us. It would be driven by notions of empathy, compassion, repair, atonement, these types of things. That really is what restorative justice is about.

> —Sujatha Baliga, *Interview with Awakin, November 8, 2014*

> Criminal justice harms people who harm people to show that harming people is wrong… Restorative justice invites us to be present to one another in ways that bring about healing and wholeness rather than in ways that deepen harm and hostility. And importantly, it gives us the tools to do so. RJ is effective…because it responds to human need. It is attuned to peoples' yearning to be in good relationship with one another.

> —Fania Davis, *Comments at the Harvard Divinity School, October 3, 2017*

Proponents of restorative practices (e.g., Baliga, 2021; Davis, 2019; Tyler, 2006; Zehr, 2002) argue that restorative practices can mitigate reliance on exclusionary discipline by addressing the root causes of misbehavior, all while improving school climate and academic engagement. They argue that while traditional discipline approaches merely *manage* student behavior, restorative approaches *develop* students' social and emotional capacities and nurture school relationships so students are less likely to misbehave. They argue further that RP can help students view institutional power as more just by giving students agency and by creating a clearer tie between student behavior and scholastic responses. In this way, restorative practices differ from exclusionary discipline which, theory and research suggest, may lead students to feel school rules are unfair, may fracture student-teacher relationships, and may catalyze an attitude of defiance (Pesta, 2021; Way, 2011).

Restorative practices have also gained popularity as a means of addressing disproportionalities in school discipline. As discussed above, psychologists (e.g., Okonofua & Eberhardt, 2015; Okonofua et al., 2016, 2020) have identified that one cause of racial disparities in discipline is that teachers are more likely to perceive an act of misbehavior by a Black student as indicating that the student is a "troublemaker"; but that enhancing student-teacher relationships can stem this tendency and reduce disparities. Accordingly, RP advocates (e.g., Gregory et al., 2016) argue that RP can address disproportionalities by facilitating positive student-teacher relations regardless of student demographics.

Perhaps because restorative practices represent a striking and multifaceted departure from typical disciplinary regimes, some find it difficult to imagine a restorative paradigm. Below, I thus describe what might be a "typical day" for a student in a restorative school.

Aaliyah is a 7th grader at a diverse middle school. On a Monday morning, she walks into her homeroom classroom and sees the chairs arranged in a circle. She sits down next to her peers. After doing a brief check-in, her teacher asks the students what emotional skills they would like to practice that week. Her classmates offer suggestions—one says he wants to practice being empathetic; another says she wants to work on listening without judging. Aaliyah says she likes both of those. After everybody has expressed a commitment, the circle begins. One by one, the students discuss their experiences over the weekend—what they did, how they were challenged, how they responded, how they grew, what they regretted, what they would like to do differently, what they were proud of… The students share and explore one another's emotional worlds. They offer perspectives and ideas to one another. The students even role-play difficult conversations they had or want to have. By the end of the circle, which takes about an hour, Aaliyah knows new things about her peers and about herself. She's practiced useful communication skills. And she feels more fully seen by her classmates and her teacher.

The next day, as Aaliyah is walking through the hall, she accidentally bumps into Walden, a student from another classroom. She is surprised when Walden yells at her and curses her out. She feels herself get flush, and yells back, calling Walden a name. As the volume rises, Mr. Macky, a teacher who happened to be nearby, calmly walks over and quietly asks Aaliyah and Walden to take a deep breath and walk outside together. It's a sunny day, so the brief walk gives both students a chance to calm down. As they walk, both students try to imagine the situation from the other students' perspective and start to feel a little guilty about how they acted. Once outside, Mr. Macky reassures the students that conflict happens sometimes, reminds them that they can make things right, and asks the students to take turns trying to describe what happened. He also asks them to share any feelings they were having. Walden goes first, and Aaliyah listens as Walden says that he was already having a really bad day because his brother is in the hospital, so when Aaliyah bumped into him, it really set him off. He says he knew it might have been an accident, but in the moment, it really felt like Aaliyah bumped into him on purpose. And then he was really hurt by the name Aaliyah called him. Aaliyah apologizes for calling Walden a name, and tells Walden he didn't deserve that. Then, consistent with her Monday commitment, she tries to empathize. She admits that when she's worried about something, it's easier for her to get into misunderstandings. Walden calms. Aaliyah continues that it hurt her feelings when Walden yelled at her when she didn't bump into him on purpose. Walden thinks for a second, and says, "Yeah, I don't like when people assume I did bad stuff on purpose either, so I get why you reacted the way you did. Sorry I made an assumption." Both students say they feel better about the interaction, and promise to try to do things differently if they bump into each other again, and then go to class. But the interaction sticks with Aaliyah. She realizes she's feeling worried about Walden because she really didn't realize he was going through so much. So, at lunch, she takes a second to go over to Walden and ask if there is anything she can do to help. Walden admits that he'd love it if she sat with him for lunch since his sibling usually would join him, and the two of them have a great conversation.

This depiction stems from reviews of dozens of practitioner guides (see, e.g., Darling-Hammond et al., 2020); from my own experiences working in restorative contexts; and from interviews with restorative practitioners and students in restorative schools. It features two restorative practices that often emerge in schools using a "whole-school model": a community-

building circle in which students shared deeply and practiced conflict resolution skills; and a teacher-guided, impromptu conflict repair conversation. It also shows the virtuous cycle that exposure to restorative practices can elicit. The community building circle allowed Aaliyah to practice conflict resolution skills and, with the guidance of a teacher, she was able to recruit these skills to resolve a conflict with Walden. Thereafter, the same skills encouraged Aaliyah to deepen her connection with Walden, leading to more community building, and potentially reducing the likelihood that either Aaliyah or Walden will have future conflict. All of this was possible, however, because students and teachers throughout the school were empowered (via training and practice) to recruit restorative practices when the moment presented.

As noted previously, restorative practices are theorized to reduce discipline by enhancing school climates (Brown, 2017). Given ample research documenting the positive psychological, behavioral, and academic correlates of positive school climates (Cohen et al., 2009; McChesney & Aldridge, 2018; Thapa et al., 2013; Wang & Degol, 2016), one might expect restorative practices to not only reduce discipline, but also improve student mental health and academic performance. What does extant research say about the impacts of restorative practices across these dimensions?

III. Quantitative Research on Impacts of Restorative Practices

As I have written in three reviews of quantitative research on restorative practices (Darling-Hammond et al., 2020, 2022; Fronius et al., 2019), research on the impacts of school-based restorative practices is in a relatively nascent state, but generally favors the potential of these practices to improve school climates; and to reduce student misbehavior, exclusionary discipline, and racial disparities in discipline. Research on these practices can be loosely grouped into three types:

1. Psychological research regarding the impact of interventions that encourage teachers to adopt restorative mindsets (i.e., to believe that improving student-teacher relationships can improve student behavior, and that relationships can improve with time);
2. Pre-post studies which measure how student outcomes shift after schools introduce restorative programming; and
3. Randomized controlled trials that evaluate whether students in schools randomly assigned to receive restorative programming see more improvement in student outcomes than students in schools randomly assigned *not* to receive restorative programming.

While each type of research augments our understanding of the impact of restorative practices (and the conditions under which they achieve intended outcomes), each faces a common challenge. All three focus on how student outcomes shift after the introduction of restorative *programming*. As noted previously, while restorative *practices* are theorized to improve student outcomes, restorative programs are distinct from restorative practices. Programs are systems of training and support designed to encourage school community members to learn, and engage in, restorative practices. As many researchers have documented (see, e.g., Blood & Thorsborne, 2005; Gregory & Evans, 2020; Gregory, Ward-Seidel, & Carter, 2021), districts often face significant challenges when trying to help staff proceed from receiving training to engaging in practices. Because practices are the drivers of outcomes, and because restorative programming may not always succeed at catalyzing restorative practice utilization, most research, therefore, relies on the assumption that programming shifts practices.

Some research designs (such as randomized controlled trials, or observational difference-in-difference designs) also rely on comparing schools that receive restorative programming to schools that do not. These research designs not only assume that programming shifts practices, but also assume that schools that do not receive formal restorative programming are not utilizing restorative practices more than those that do.

But are these assumptions warranted? As depicted in Figure 8, there are many reasons we might worry that implementing restorative programs may not lead to student exposure to restorative practices; or that schools that do not receive formal restorative programming might, nonetheless, have high levels of restorative practice utilization. Each of the rectangles below represents a moderating factor that can partially determine whether implementation of a restorative program accrues to downstream student exposure to restorative practices.

Our first potential moderator is program quality. My reviews of research and practitioner guides related to restorative practices (Darling-Hammond et al., 2020; Fronius et al., 2019) revealed staggering heterogeneity not only in what constitutes "restorative programming," but in the extent to which programs provided actual instruction, opportunities for practice, coaching, and peer learning. Some restorative programs did not appear designed in such a way that one could reasonably expect that teacher participation in the program would shift teacher practices.

In my interviews with teachers and restorative practitioners, interviewees often maligned restorative trainings that simply provided "a binder" to teachers and oriented them to the contents, expecting that teachers' perusal of complex material might change the way teachers relate to students. One can imagine that these kinds of programs would be unlikely to empower teachers to adopt relational mindsets and abandon punitive ones; or enhance teachers' capacities to build a strong school community, inculcate conflict resolution skills, and facilitate conflict resolution when misbehavior occurs. Interviewees also indicated that these low-touch training approaches are actually quite commonly employed because they cost less money and require less time to implement than more robust alternatives. Thus, when attempting to understand the impact of exposure to restorative practices, we may worry that research that assumes programs shift practices could overstate the extent to which "treated" schools use restorative practices, and therefore underestimate the impact of these practices.

Figure 8

Pathways to Student Exposure to Restorative Practices
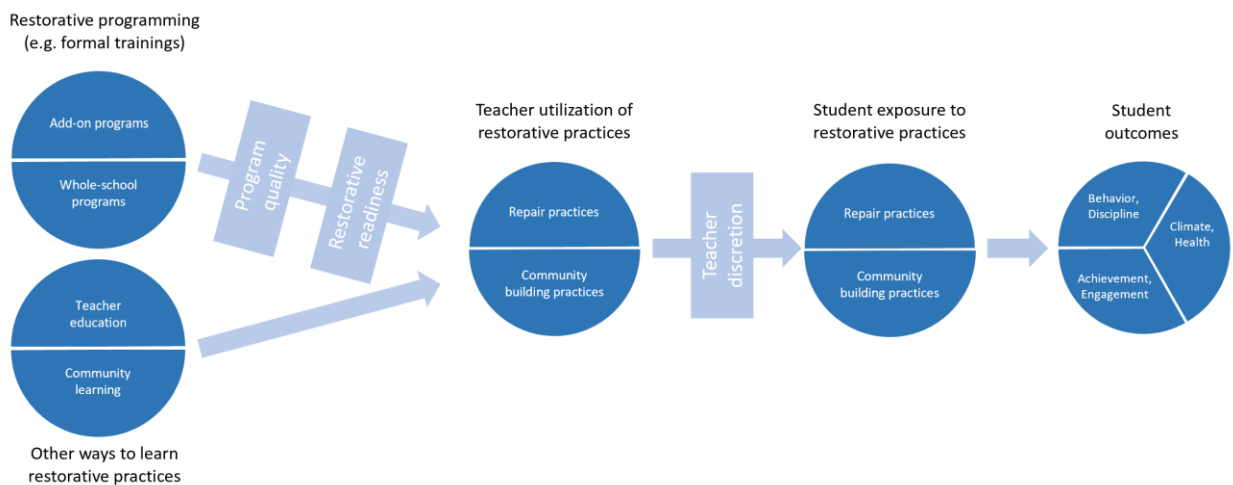
The second potential moderator is that even when schools select restorative programming that *can* shift teacher practices, school staff may not be sufficiently receptive to the programming to shift their practices. Staff may be less receptive to restorative programming if they personally adhere to the notion that discipline is necessary to manage student behavior. This notion of the cultural fit between the mores of a school and the ethical pillars of restorative practices (e.g., that teachers can elicit pro-social behavior by appealing to students' intrinsic desire for positive relationships rather than relying on exclusion) is often described as "restorative readiness" (Garnett et al., 2020; Gregory & Evans, 2020). Researchers have theorized that schools that are low on restorative readiness will struggle to shift teacher practices (and, relatedly, student exposure to restorative practices). This presents another reason to worry that research identifying impacts of restorative practices by seeing student outcomes that follow the implementation of even well-run restorative programs might anticipate higher levels of restorative practice implementation than ultimately occur, and might, therefore, underestimate the effects of restorative practices.

A final moderator is teacher discretion in when, and with whom, to employ restorative practices. Research has previously identified the extent to which teachers not only enjoy, but employ, discretion in when to employ punitive approaches (Skiba et al., 2011), and that—even when student conduct is held constant—this discretion can encourage teachers to leverage more punitive approaches when Black students misbehave than when White students misbehave (Okonofua & Eberhardt, 2015). Clearly, individuals may employ discretion in when to recruit harmful or harsh practices. But research also indicates what are known as "boosting effects"— being more likely to employ helpful practices when interacting with White individuals than with individuals of other races (e.g., Kang et al., 2009; Smith et al., 2015). This could indicate that teachers may be more likely to employ restorative practices when interacting with White students, and employ exclusionary practices when interacting with Black students. Moreover, as discussed above, research indicates that White teachers expect less of Black students than they do of White students (Papageorge et al., 2020). To the extent that teachers' level of motivation to employ relational practices is a function of how much they believe in the capacity of a given student, White teachers may be less likely to leverage restorative practices when interacting with Black students. Taken together, research suggests that teachers (and particularly White teachers) may be more likely to engage in restorative practices when engaging with White students than when engaging with Black students. If so, we might worry that even when a school implements an effective restorative program, and even when teachers are culturally receptive to the programming, they may be more likely to utilize newfound restorative practices when engaging with White students. This can lead to unevenness in student exposure to these practices. If researchers had assumed that the program would lead teachers to use restorative practices uniformly in all interactions with students, then researchers' failure to appreciate this phenomenon (whereby Black students may experience very little exposure) could again lead researchers to overstate teachers' use of restorative practices and therefore underestimate their impact. In addition, if restorative practices prove effective at reducing discipline and improving other student outcomes, then if teachers use these practices with White students but not with Black students, it could increase racial disparities in discipline and academic achievement, among other outcomes.

The three moderators discussed above (program quality, restorative readiness, and teacher discretion) each present challenges in situations where researchers hope to identify how the introduction of a restorative program might impact student outcomes. However, as noted, in many cases, researchers also hope to compare schools that receive programming ("treated" schools) to schools that do not receive programming ("control" schools). In these cases, researchers also must content with the possibility that control schools utilize restorative practices at a high level. This is quite possible. First, as discussed in the prior section, a large and increasing number of schools implement restorative practices. Thus, randomly selected control schools may already be implementing programming. Second, even if control schools do not employ formal programming, they may hire teachers who are well versed in restorative practices. Today, many teacher-preparation and credentialing programs provide or require coursework related to restorative practices (e.g., School of Teacher Education, 2020). It is therefore becoming increasingly feasible for teachers at control schools to have exposure to restorative practices, even if those schools do not have restorative programming. Third, research has documented how teachers share new practices with one another (Rutkowski et al., 2013). Thus, even without formal programming, teachers in control schools could catalyze higher than expected levels of restorative practice utilization.

The foregoing discussion of research challenges related to using restorative *program* research to understand the impacts of restorative *practices* is mainly intended to demonstrate the need for research on restorative practices. However, it should not detract from the value of understanding the impact of restorative programs, which provide important insights into the impacts of restorative practices. Below, we summarize finding from the three major types of restorative program research: studies on interventions that can catalyze restorative mindsets, pre-post studies of restorative program implementation, and randomized controlled trials of restorative programs.

Psychological Research on Restorative Mindset Programs

A new body of psychological evidence indicates that encouraging teachers to adopt what could be termed "restorative mindsets" can help reduce racial disparities in discipline. Restorative practices encompass a broad umbrella of practices, but all are designed to proactively improve school relationships, and repair relational harm when conflict occurs. To implement these practices, teachers must adopt what could be termed a restorative mindset—one that accepts that relationships can improve with effort, that prioritizes relational repair, and that engenders empathy with student perspectives (Brown, 2017; Cavanagh et al., 2014; Evans & Lester, 2013; González, 2012; Gregory, Ward-Seidel, & Carter, 2021; Lustick, 2020; Mirsky & Wachtel, 2007). What does psychological evidence indicate about how the adoption of a restorative mindset impacts teachers' treatment of students of various racial backgrounds?

Okonofua et al. (2016) randomly assigned 31 teachers to one of two conditions. "Treated" teachers received an intervention designed to encourage the adoption of what they termed an "empathic mindset"—one that "prioritizes valuing and understanding students' experiences and negative feelings that give rise to misbehavior, sustaining positive relationships with misbehaving students, and working with students within trusting relationships to improve behavior" (p. 5221). Teachers read content and answered questions designed to encourage them to believe that student-teacher relationships are major drivers of student thriving, and that working to understand students' perspectives can improve student-teacher relationships. They then read about some of the stresses students might be facing, and read statements from students providing these students' perspectives. An example of the student perspectives received by the treated teachers appears below:

> In middle school, I didn't feel like I belonged. It seemed like the teachers always called on the other students. So, I didn't pay attention in class and sometimes I got in trouble. One day I got detention, and instead of just sitting there, my teacher talked with me about what happened. He really listened to me. And then he told me that he had trouble sometimes in middle school but that it gets better. It felt good to know I had someone I could trust in school. (Okonofua et al., 2016, p. s1)

Finally, they read statements by teachers who had worked to adopt relationship-oriented practices into their teaching. In sum, teachers in the "treated" condition were encouraged to adopt the kind of mindset required for implementing restorative practices—one that prioritizes and seeks to nurture relationships with students. "Control" teachers did not receive the empathic discipline intervention, and read placebo statements from students and teachers regarding classroom technology.

The authors reviewed outcomes for 1,682 students taught by the treated and control teachers. As depicted in Figure 9, students taught by treated teachers (those who were encouraged to adopt a restorative mindset) were significantly less likely to be suspended over the school year. This was true for Black and Latino boys, and particularly true for Black and Latino girls, who saw a markedly lower (70%) suspension rate in the treatment condition relative to the control condition. Students assigned to treated teachers also were significantly more likely to report that they felt respected by their teachers.

Figure 9

Student Discipline Rates by Teacher Condition
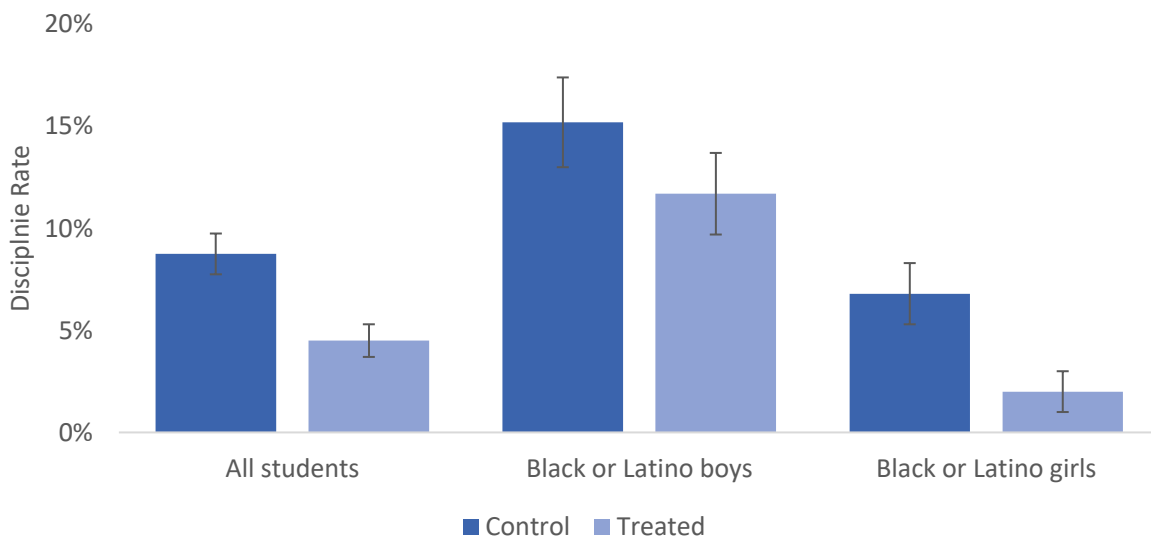
Note. "Treated" teachers were encouraged to adopt restorative mindsets (termed "empathic mindsets" by the authors). "Control" teachers were not encouraged to adopt restorative mindsets. Error bars represent 95% confidence intervals around mean values. Adapted from "Brief Intervention to Encourage Empathic Discipline Cuts Suspension Rates in Half Among Adolescents," by J. Okonofua, D. Paunesku, and D. Walton, 2016, *Proceedings of the National Academy of Sciences*, *113*(19), p. 5224 (https://doi.org/10.1073/pnas.1523698113). Adapted with permission.

In a recent study that I had the pleasure of collaborating on (Okonofua et al., 2020), we attempted to test whether adding another aspect of the restorative mindset, and simulating a restorative context, might boost the effects discovered in the prior study (Okonofua et al., 2016). We encouraged "treated" teachers to adopt the belief that students can grow, and that student-teacher relationships can grow with effort—two "growth mindsets" that are critical to restorative practices. We then asked teachers to read vignettes about a misbehaving student and answer questions about how troubled they were by the student's behavior and how harshly they believed the student should be disciplined. Some teachers read about a White student, and others read about a Black student. As noted, teachers in the treatment condition were encouraged to believe

that students can grow and that student-teacher relationships can improve (the "growth mindsets"). However, they were also asked to imagine having a brief conversation with the misbehaving student where they learned about the student's home life and aspirations. This latter manipulation was meant to simulate a critical facet of a restorative context—the opportunity to hear the perspectives of misbehaving students. Teachers in the control condition were not encouraged to make these mindset shifts, nor did they get students' perspectives. As anticipated based on prior work, we first found that teachers in the control condition exhibited racial disparities in how they responded to students. However, teachers in the treatment condition did not exhibit racial disparities in their responses. In addition, teachers in the treatment condition were significantly less likely to label the misbehaving student as a troublemaker, more likely to feel able to build a strong relationship with the student, and less likely to expect the student to get suspended in the future.

Together, these two studies suggest that encouraging teachers to adopt restorative mindsets, and placing them in restorative contexts, could enhance student-teacher relationships, reduce reliance on exclusionary discipline, catalyze more equitable treatment of students, and reduce racial disparities in discipline.

Pre–Post Studies Regarding Restorative Programming

Pre-post studies of restorative practices generally find that after the introduction of restorative practices, schools saw improvements in discipline, discipline disparities, misbehavior, and school climate (Figure 10). Results were more mixed regarding academic performance.

Figure 10

Results of Pre–Post Studies Regarding Restorative Programming
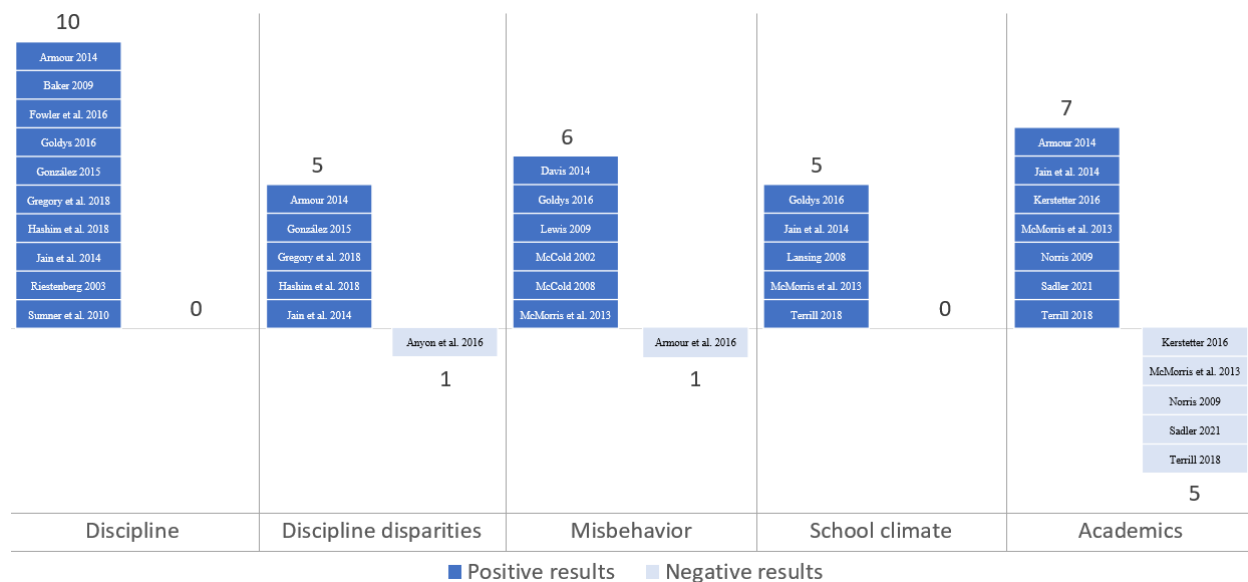
Nearly all pre-post studies evaluating rates of exclusionary discipline after restorative programming implementation report notable declines. Following restorative programming introduction, researchers reviewing school-level data have seen 87% drops in suspensions across two years of implementation (Sumner et al., 2010), 65% drops in suspensions among 6th graders in a middle school (Armour, 2014), 55% reductions in office referrals for students in an elementary school (Goldys, 2016); 57% drops in disciplinary referrals, 77% drops in suspensions, and 35% drops in time spend in in-school suspension (Riestenberg, 2003); and drops in out-of-school suspensions from 12% to 7% and in in-school suspensions from 19% to 7% over a five year period (Fowler et al., 2016). District-level analyses report similarly. Jain et al. (2014) report that Oakland schools that received *whole school* restorative programming saw students receive significantly fewer suspensions than students in the district overall. Three research teams reviewing outcomes in Denver Public Schools following districtwide implementation of restorative programming have noted marked and sustained declines in discipline rates, overall and for subcategories of students (Baker, 2009; González, 2015; Gregory et al., 2018). From 2006 to 2013, for example, González (2015) reports that overall suspension rates fell from 10.6% to 5.6%, rates for Black students fell from 17.6% to 10.4%, and rates for Latino students fell from 10.2% to 4.7%. Hashim et al. (2018) report a similar trend in Los Angeles Unified School District following the implementation of restorative programming in the 2014-15 school year—suspension rates for misconduct dropped for all measured categories of students.

Studies of the relationship between program implementation and discipline *disparities* have been largely encouraging. At the school level, Armour (2014) found that both the Black-White and Latino-White discipline gaps narrowed after RP implementation in a San Antonio middle school. And at the district level, Hashim et al. (2018) found that Black-White discipline disparities abated in Los Angeles schools after implementation; González (2015) and Gregory et al. (2018) reported that Black-White and Latino-White disparities diminished after Denver implemented restorative programming; and Jain et al. (2014) found that Oakland schools that used a whole-school restorative model saw the Black-White discipline gap decline from 12.6% to 9.2% over a three year period (all the more impressive given that Oakland schools that did *not* implement restorative programming during the same time period actually saw the Black-White discipline gap increase). In contrast to the studies above, in a review of administrative data from a large urban district, Anyon et al. (2016) found that, following the introduction of restorative programming, discipline rates abated overall, but racial discipline gaps persisted.

As with disciplinary outcomes, pre-post studies generally suggest program implementation precedes declines in misbehavior. Davis (2014) reports that Oakland schools implementing restorative programming saw a 77% decrease in referrals for violence after two years. Lewis (2009) reports that a West Philadelphia High School saw violent acts and serious incidents drop by 52% in the first year of implementation and drop by an additional 40% in year two. Youth participating in a Pennsylvania restorative program saw their offending drop by 58% over three months (McCold, 2002), and by 50% over two years (McCold, 2008), and youth who fully completed the program saw the greatest reduction in recidivism rates (McCold, 2002, 2008). McMorris et al. (2013) report similarly positive results from their study of the "Family Group Conferencing" model adopted in Minnesota, reporting decreases in self-reported incidents of physical fighting and skipping school among conference participants in a six-week follow-up. Goldys (2016) reviews data from an elementary school that saw a 55% decrease in physical aggression after implementing restorative programming. In contrast to most studies on the topic,

Armour (2014) found that offense frequencies grew over the course of implementation in a San Antonio middle school, but did note that the implementation period coincided with marked student mobility, with 68% of the student body having moved into or out of the school during the study year.

The introduction of restorative programming typically precedes improvements in school climate measures, including improvements in conflict resolution skills. McMorris et al. (2013) found that participants in a restorative program reported increased fondness towards school, an augmented sense of connection to school, and improved problem after the six-week program. The Lansing School District (2008) reported, similarly, that in a six-week follow up of their program, 91% of students and 89% of parents indicated students had learned conflict resolution skills, 92% of students and 85% of parents indicated conflict had been resolved through restorative processes, and 90% of students indicated using new skills to resolve future disputes (90%). Jain et al. (2014) found that 69% of staff in Oakland schools implementing restorative programming believed that the programming had improved school climate, 67% indicated it helped students improve their social and emotional skills, and 64% believed that it helped facilitate caring relationships between teachers and students. However, Jain et al. also found discrepancies between staff and parental opinions: whereas 100% of principals believed that the programming had improved school climate, only 40% of parents agreed; and whereas 92% of principals believed programming had improved teacher-student relationships, only 28% of parents did. Goldys (2016) reports that 97.7% of students in an elementary school implementing programming indicated feeling safe in school after implementation. Focusing on three diverse, rural, West Coast schools, Terrill (2018) reports that teachers felt that implementing the programming resulted in greater respect by students for other students.

On the topic of academic outcomes, pre-post studies have reported mixed results. Two studies report positive findings. Armour (2014) found that 6th grade students in a restorative program for a year saw 11% improvements in their statewide reading passage rates and a 13% improvement in math, and that Black, Hispanic, economically disadvantaged, and special education students all saw strong improvements. Jain et al. (2014) compared three-year academic growth in Oakland schools implementing restorative programming to growth in schools not implementing programming. They found that students in implementing schools saw reading levels increase by 128% (compared to 11% in non-implementing schools), saw four-year graduation rates increase by 60% (versus 7%), and saw high school dropout rates decreased by 56% (versus 17%). Meanwhile, five studies report what could be termed ambiguous findings. Kerstetter (2016) compared outcomes at a charter elementary school implementing restorative programming to a "comparable" charter school implementing "no excuses" policies, and found that, in the study year, whereas 60% of "restorative charter" students were proficient on statewide tests, 36% of students in the comparison charter were. However, in the following year, the proportion of "restorative charter" students who were proficient had dropped from 60% to 47%. McMorris et al. (2013) note that students who participated in restorative programming in Minnesota schools saw increases in GPA and credit attainment, but declines in chances of being on track to graduate (although on-track markers rebounded the year after initial implementation). Sadler (2021) finds that academic performance for Black students in a large charter network diminished in the first year after RP adoption, but rose again in subsequent years. Norris (2009) reports no significant change in grade point average for participants in a restorative program (compared to non-participants). Reviewing data from a school implementing programming, Terrill (2018) reports that while grade point averages of students overall fell after RP

implementation, grade point averages increased among students who had received office referrals and were most likely to interface with programming.

Pre-post studies generally exhibit the most impressive results when reviewing outcomes in schools that adopted a "whole-school" restorative approach. For example, perhaps the most promising study of restorative programming (Jain et al., 2014) found that students in Oakland schools that adopted a "whole school" model not only saw impressive gains in discipline and academic achievement, but that students in "whole school" environments saw significantly more growth than those in schools that adopted a partial, or add-on, model. Many researchers have argued that schools that adopt a "whole school" model are more likely to successfully encourage teachers to engage in restorative practices (e.g., Gregory, Ward-Seidel, & Carter, 2021). Thus, one reading of these whole school studies is that whole school models more successfully catalyze the use of restorative *practices*, and these practices, in turn, drive positive student outcomes.

While pre-post studies often evidence encouraging results related to misbehavior, discipline, discipline disparities, and school climate, it is important to reemphasize that these studies suffer from internal validity concerns. At best, they document the *co-occurrence* of restorative program uptake and positive outcomes—a co-occurrence that could be a function of a meaningful relationship or could simply reflect a spurious one. For example, if the kinds of schools that tend to implement restorative programming also tend to be on pre-existing positive trajectories, or tend to adopt other impactful practices, then the correlation between restorative programming uptake and outcomes could really be a reflection of other facets of these schools (i.e., "school-level selection effects"). Along the same lines, if certain kinds of caregivers are more likely to place their children in restorative schools, and these students are already on positive trajectories (or are experiencing other phenomena which tend to drive positive outcomes), then the relationship between uptake and student outcomes could really be a reflection of the characteristics of the kinds of students that tend to be in restorative schools (i.e., "student-level selection effects").

Randomized Controlled Trials Regarding Restorative Programming

One mechanism for overcoming these student and school-level selection effects is to leverage research methods designed for identifying causal effects. The best known of these causal research methods is the randomized controlled trial (RCT), wherein a subset of teachers or schools are randomly assigned to receive some treatment (here, restorative programming), and another set is assigned not to receive the treatment. In school contexts, the outcomes of interest are often at the student level, so researchers compare outcomes for students of "treated" teachers or schools to outcomes for students of "controls" teachers or schools.

There have been, to date, six randomized controlled trials in U.S. k-12 contexts that have reviewed the impact of restorative practices on student outcomes. As noted above, the treatment in these studies (restorative programming) is distinct from the treatment of interest (restorative practices); and providing *restorative programming* (to teachers or schools) does not necessarily result in students gaining exposure to *restorative practices* (Blood & Thorsborne, 2005; Darling-Hammond et al., 2020; Gregory & Evans, 2020; Gregory, Ward-Seidel, & Carter, 2021). Still, while these RCTs leave important holes in our understanding of the impact of restorative practices, they provide a useful lens into the feasibility of providing restorative *programming* that can yield student benefits.

C. R. Cook et al. (2018) and Duong et al. (2019) report on randomized controlled trials in which elementary (for C. R. Cook et al.) and middle (for Duong et al.) school teachers were

randomly assigned to receive training in the "Establish, Maintain, Restore" (EMR) program, and to receive ongoing coaching. EMR encourages teachers to, among other things, appreciate the importance of student-teacher relationships, actively take steps to establish and maintain positive relationships, and affirmatively restore relationships when conflict has occurred. The program thus seeks to encourage teachers to adopt a restorative mindset and to engage in certain restorative practices related to student-teacher relationships. In their studies, teachers randomly assigned to receive EMR training saw statistically significantly greater declines in student misbehavior, as well as statistically significantly greater improvements in teacher-reported student-teacher relationships and researcher-observed student engagement. However, their studies did not evaluate impacts on discipline, nor did they disaggregate student data to ascertain whether EMR might empower teachers to abridge racial disparities in discipline. In addition, outcome data were collected only three months after EMR training, leaving the long-run impacts of the EMR training unclear.

Gregory, Huang, and Ward-Seidel (2021), meanwhile, report on an RCT featuring 18 elementary, middle, and high schools; and with discipline data from 5,878 students. Their careful review of intermediary variables indicated that the nine schools randomly assigned to receive restorative programming implemented the program with fidelity. For example, all nine principals attended the required training, met with their restorative coaches, and scheduled time in the school day for teachers and students to engage in restorative practices; and, across schools, 95% of staff designated to hold restorative circles received appropriate training. They found that students in schools randomly assigned to receive whole-school restorative programming saw significantly larger declines in discipline rates, but did not find evidence that programming abridged racial disparities in discipline. However, due to the COVID-19 pandemic, researchers were unable to collect data after year one of implementation.

Grant et al. (2022) evaluated the impact of the synergy of restorative programming (the SaferSanerSchools program provided by the International Institute of Restorative Practices) and Diplomas Now (a whole-school reform model of curriculum reform, staff support, and early student detection designed to avoid early dropout). From a sample of 33 elementary, middle, and high schools drawn from various cities across the country, they randomly assigned 17 schools to receive the synergy of programs and 16 schools to serve as controls. After adjustments for differential attrition, they recovered an analytic sample of 25 schools—13 in treatment, and 12 in control. Results indicated that students in treatment schools saw significant gains in measures of school climate. Teachers, meanwhile, saw null effects on school climate and teacher turnover.

Augustine et al. (2018) reviewed data from a randomized controlled trial of 44 Pittsburgh middle schools, finding that RJ implementation caused a 16% reduction in days lost to suspensions. They also found that implementation led to a small but notable reduction in the racial discipline gap, and an improvement in school climate based on teacher surveys. They did not, however, find a statistically significant link between RJ implementation and either arrests or absences. Concerningly, Black students in treatment schools experienced lower academic performance than their counterparts in control schools.

Acosta et al. (2019) conducted a randomized controlled trial of fourteen Maine middle schools. Unlike in the Pittsburgh RCT, the Maine RCT team tracked information about the extent to which students were exposed to restorative *practices*, anticipating that students in treatment schools would indicate higher exposure, and would, as a function of this higher exposure to restorative practices, experience improvements in school climate and other outcomes. Surprisingly, as depicted in Figure 11, students in treatment and control schools evidenced nearly

identical levels of exposure to restorative practices. They found, further, that assignment to restorative *programming* was not related to improvements in behavior or school climate, but given the disconnect between programs and practices, this finding is not terribly instructive.

It is important to note that the disconnect that Acosta et al. (2019) noted between restorative programming and restorative practices should not be construed as suggesting that randomized controlled trials *cannot* evaluate restorative practices. Indeed, theoretically, randomizing which schools receive restorative programming can yield causal estimates of the impacts of restorative practices. In *expectation*, treated and control schools are equivalent in their baseline levels of restorative programming and restorative practice utilization. As such, with a few assumptions, one can attribute any comparative improvements experienced by students in treated schools to restorative programming.

Figure 11

Exposure to Restorative Practices Among Treatment Versus Control Students in a Randomized Controlled Trial of Restorative Programming
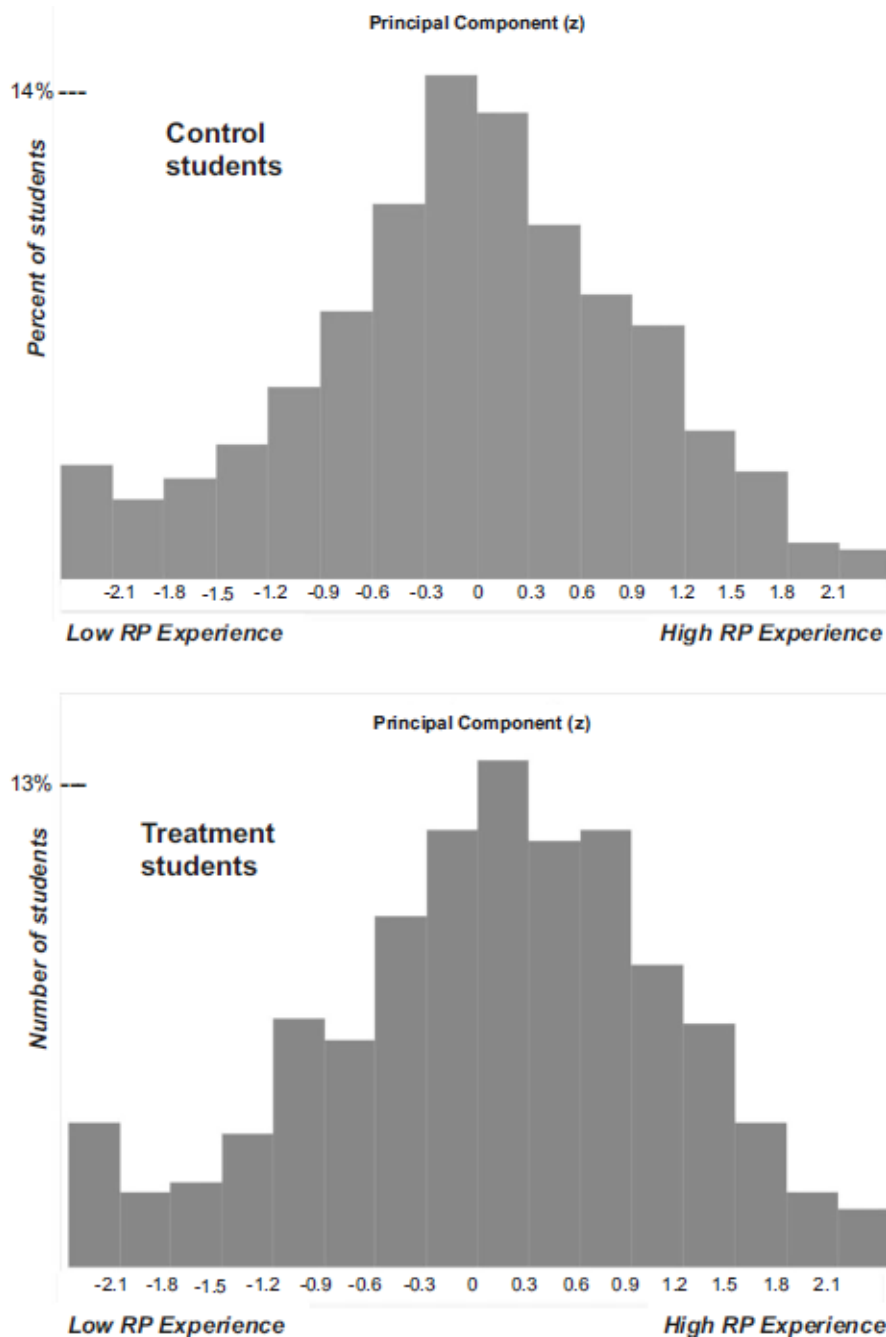
And if one assumes restorative programming drives restorative practices, one can attribute improvements to restorative practices. This assumption may be warranted when treated schools are randomized to receive immersive, high-quality programming. However, because randomized controlled trials typically involve small samples of schools, it is quite possible that treated and control schools do *not* have similar levels of restorative programming or restorative practice utilization at baseline; that control schools increase their use of restorative practices over the course of the study period; that treated schools utilize ineffective professional development, or have low levels of restorative readiness, and thus do not see practice shifts as a result of the training; or that treatment schools hire teachers that might leverage their discretion in ways that create racial disparities in exposure to restorative practices.

Can Randomized Controlled Trials Be Leveraged to Evaluate Restorative Practices?

How, then, could a research team leverage a randomized controlled trial framework to evaluate the impact of student exposure to restorative practices? That is the question driving a nascent research practice partnership between myself and Oakland Unified School District. Exploratory conversations have elevated a number of strategies that might allow our research to overcome some of the methodological challenges highlighted above. First, rather than randomizing at the level of the school, we plan to cluster randomize at the level of the teacher. Assuming the project moves forward, we will identify all incoming first-year teachers who have *not* previously received professional development in restorative practices. Next, we will randomly select half to receive restorative practice professional development (RPPD) in early Fall; and half to receive RPPD in early Spring. We will then use well-timed survey instruments to verify baseline comparability of treatment and control groups; determine the effectiveness of RPPD at shifting teaching practices; evaluate how RPPD impacts student exposure to restorative practices (e.g., does it shift exposure for all students, or only certain subcategories of students); and, finally (assuming RPPD shifts student exposure to restorative practices) to ascertain if utilization of / exposure to restorative practices improves outcomes for teachers and students.

We would survey all eligible teachers (and the students they serve) before the Fall semester and at the end of the Fall semester so we may compare outcome growth for teachers who received RPPD in the Fall to outcome growth for teachers who will not have received RPPD yet. These comparisons will allow us to answer two first-order questions:
- Do teachers who received RPPD exhibit increased restorative practice utilization?
- Do students served by these teachers exhibit more exposure to restorative practices?

Having vetted the core assumption underlying our RCT framework, we could then ascertain
- whether students served by teachers who receive RPPD engage in less misbehavior, receive less discipline, exhibit more academic engagement, demonstrate better academic performance, and indicate higher school climate; and
- whether teachers who receive RPPD indicate higher school climate and better job satisfaction.

Randomization at the level of the teacher addresses many concerns. First, the larger sample of teachers (compared to smaller samples achieved with school-level randomization) increases the probability that a random pull of treated and control units will be comparable at baseline. Second, teacher-level randomization provides a more effective pathway to a compelling control group. Given the proliferation of RPPD and the frequency with which teachers change schools (Daly et al., 2008; Nittler & Gerber, 2018), it is difficult to find a set of schools where *no*

teacher has received RPPD. It is far easier to find a set of incoming teachers who have not received RPPD.

Third, teacher-level randomization with staggered treatment creates a plausible means of overcoming differential attrition (should it emerge, as it has in other studies of restorative programming; Grant et al., 2022). In a recent randomized controlled trial that I fielded among Denver public employees, we staggered employee access to an anti-bias training (with "treated" employees receiving access to the training earlier than "control" employees). We had all employees complete two surveys—one before the treated group had access to the training and one after the treated group had access (but before the control group got access). We found that the treated group had lower survey participation on the second survey than the control group, suggesting that participation in the antibias training itself led to attrition. To address this issue, we restricted our analyses to treated and control individuals who completed the training when it became available to them. We saw no differential attrition among this subset of participants. A similar phenomenon could emerge in research regarding RPPD. Participation in RPPD often calls on teachers to examine their own teaching practices, and often invites teachers to consider the role that racial bias might have played in their decisions to discipline or cultivate relationships with students (e.g., Denver School-Based Restorative Practices Partnership, 2017; Yusem et al., 2016). One can imagine that participation in RPPD could lead to study attrition among teachers such that those who receive RPPD might be less willing to fill out end-line surveys if they feel fatigued by the training. Alternatively, participation could conceivably lead teachers to feel hopeful and become engaged than non-participation, leading to higher end-line survey participation rates. In a school-level framework, it might be difficult to discern drivers of differential attrition. However, in the teacher-level framework articulated above, if differential attrition emerges, we can restrict analyses to teachers who participate in RPPD (whether in the Fall or Spring) as a robustness check to ensure results are not an artifact of differential attrition.

Given its competitive advantages over a school-level randomization framework, a teacher-level RCT for evaluating the impacts of restorative practices is promising. Of note, such a framework respects both the distinction between restorative programming and restorative practices, and the need to use sophisticated research approaches to validate implementation assumptions. However promising, this framework does not yet exist, and will not exist for some time. Thus, the question becomes what, if anything, can we learn from *observational* data about the relationship between exposure to restorative practices and student outcomes? In some studies, researchers have tracked students' levels of exposure to restorative practices. What do these studies indicate?

Correlates of Exposure to Restorative Practices

In Acosta et al.'s (2019) randomized controlled trial, students in schools randomly assigned to receive restorative programming neither saw growth in their exposure to restorative practices nor saw improvements in measured outcomes. However, because Acosta et al. had student-level data on restorative practice *exposure*, they were able to ascertain correlates of exposure to restorative *practices*. And they found that—regardless of treatment condition— students who reported having more exposure to restorative practices reported higher school connectedness, better school climate, more positive peer relationships, better developmental outcomes, less physical victimization, and less cyberbullying. Gregory et al.'s (2016) review of student survey data from two high schools found, similarly, that students who indicated a high degree of restorative practice exposure received fewer defiance and misconduct referrals, and

this held for Black, Latino, White, and Asian students. Students who reported a high degree of exposure to restorative practices also experienced a smaller racial discipline gap, and were more likely to indicate feeling respected by their teachers. Finally, Darling-Hammond, Trout, et al. (2021) reviewed records from over 800,000 California middle and high school students and found that, across racial groups, students who indicated higher levels of exposure to restorative practices were less likely to have been suspended in the prior thirty days, and that the relationship held after controlling for a suite of student, parent, and district-level factors. They also found that students with higher levels of exposure to restorative practices evidenced markedly smaller Black-White discipline disparities. Specifically, as the authors detail, "students with the highest levels of exposure to restorative practices experienced Black–White discipline disparities that were five times smaller than those experienced by students with the lowest levels of exposure to restorative practices" (Darling-Hammond, Trout, et al., 2021, p. 3). While the magnitude was smaller, the authors' models also indicated smaller Hispanic-White discipline disparities at higher levels of restorative practice exposure. Finally, they found that, across racial groups, students with higher levels of exposure to restorative practices also had higher GPAs.

These three studies suggest that restorative practice exposure is correlated with positive outcomes. However, these studies suffer from the same student and school selection effects as the pre-post studies discussed above, and thus should only be viewed as documenting the co-occurrence of restorative practice exposure and positive outcomes. Even when exposure to restorative practices follows the implementation of a restorative program, it can still be correlated with potent confounders. For example, given research showing that White teachers hold lower expectations of Black students than White students (Papageorge et al., 2020) and that Black students exhibit more externalizing behavior when interacting with White teachers (Redding, 2019), White teachers may be less likely to achieve the rapport with Black students that encourages teachers to utilize restorative practices. If this is so, then student-level variation in exposure to restorative practices could be a function of student race which, ample research shows, is correlated with student discipline (e.g., GAO, 2018) and academic achievement (e.g., Pearman et al., 2019); and any observed relationship between exposure to restorative practices and positive outcomes might simply reflect the relative degree of structural advantage enjoyed by White students when compared to Black students. Another potential spurious pathway connecting student exposure to restorative practices and student outcomes is variation in how students sort themselves into education environments. Research demonstrates that Black students are more like than White peers to sort into more punitive schools (Owens & McLanahan, 2020). We may therefore worry that Black students might also be more likely to sort into schools that do not use restorative practices. Were this to be true, any observed relationship between student exposure to these practices and student outcomes might simply reflect student-level variation in how students sort into schools.

There is at least one more pathway that might lead to the observation of a spurious relationship between restorative practice exposure and student outcomes, and it stems from how students are placed in special education environments. When I represented students in juvenile proceedings, all of my clients were boys of color, and most had been diagnosed with emotional disturbance in early elementary school. These diagnoses mystified me: my interactions with these students did not reveal aggression, challenges forming relationships, or other common symptoms of emotional disturbance. In addition, all of my clients had long discipline histories stretching back to early elementary school. I found this equally perplexing. Home visits, interviews with family members, and lengthy intakes revealed students who were amiable and

agreeable. When I began my PhD studies, I quickly learned that my experience—of seeing boys of color with heightened probabilities of receiving special education designations and alarming exposure to discipline—was not atypical. Research by Sullivan and Bal (2013) demonstrates that, even after controlling for myriad student and school demographic factors, Black students are overrepresented among students receiving special education designations. And students with disabilities are disproportionately disciplined regardless of their grade level, the type of school they attend, or their schools' demographic composition (GAO, 2018). The situation is even more daunting for Black students with disabilities, who are overrepresented even compared to other students with disabilities or other Black students, and who have the *highest* degree of exposure to out-of-school suspensions of any two-way intersectional student group (e.g., Black and male or White and female). Specifically, among Black students with disabilities, *twenty-three percent* received an out-of-school suspension in the 2013-14 school year alone—as compared to 5.7% among all students, 12% among all students with disabilities, and 14% among Black students (GAO, 2018). So why might the relationship between special education designation and discipline confound our ability to understand the impact of exposure to restorative practices? Due to student tracking (Loveless, 2013), students with disabilities in a given school are often isolated in a unique school practice milieux, which may not include exposure to promising practices (such as restorative practices). Thus, even if analyses reveal that more restorative practice exposure is associated with positive outcomes, we might worry that the association merely reflects the extent to which students in special education classrooms have low levels of exposure to these practices and tend to experience negative student outcomes. This spurious link could be exacerbated in analyses focused on Black students, not merely because Black and disabled students experience such poor student outcomes, but also because *one in seven* Black students are designated as receiving special education services (GAO, 2018). All of this is to say that reviewing links between exposure to restorative practices and student outcomes without employing a strategy to overcome student-level selection effects almost certainly will yield a biased estimate.

The foregoing demonstrates how student-level variation in restorative exposure might be a function of student traits (and how the failure to account for student-level variation in exposure could lead to a misestimation of the impacts of restorative practices). The same logic applies to school-level confounders. For example, research by Roch et al. (2010) demonstrates that schools with diverse teacher workforces are more likely to adopt more "learning-oriented" (rather than "sanction-oriented") behavior management approaches. Restorative practices are certainly more "learning-oriented" than "sanction oriented," and we thus may worry that school-level variation in teacher adoption of restorative practices (and downstream student exposure to these practices) may be a function of teacher demographics. Given research (Rocha & Hawes, 2009) demonstrating that Black students experience less discipline in schools with more diverse teaching workforces, any observed relationship between teacher adoption of restorative practices and positive student outcomes (particularly for Black students) might simply reflect the relationship between teacher demographics and positive student outcomes.

Due to the potential for myriad student and school-level selection issues, studies documenting the co-occurrence of restorative practice exposure and positive student outcomes may be encouraging, but they are hardly convincing.

When reviewing *all* available quantitative evidence on restorative practices, we see that extant psychological, pre-post, and randomized controlled trial studies suggest that restorative *programming* may drive improvements in discipline, discipline disparities, and school climate

(among other outcomes); but we also see that programming may harm academic performance. However, myriad implementation guides (Garnett et al., 2020; Gregory, Ward-Seidel, & Carter, 2021) and Acosta et al.'s (2019) randomized controlled trial demonstrate that programs and practices are distinct. Extant literature thus leaves unclear the impact of restorative *practices* on critical student outcomes. And while three studies have reviewed correlates of restorative practice exposure, these studies merely demonstrate the co-occurrence of practices and positive outcomes, and thus fall far short of indicating that exposure to restorative practices *causes* positive outcomes. What is needed, then, is a means of estimating the causal impact of student exposure to restorative practices. The following sections present attempts at precisely this kind of causal estimation.

IV. Estimating the Effects of Exposure to Restorative Practices

In this section, I first discuss the challenges we must overcome in order to estimate the impact of student exposure to restorative practices. Next, I explore the strengths and limitations of the data on hand to address these challenges. Thereafter, I present three identification strategies designed to provide a sense of the relationship between exposure to restorative practices and outcomes, to overcome student-level selection effects, and to overcome school-level selection effects. And finally, I present the findings of my analyses.

Challenges to Causal Inference

In order to identify the impacts of exposure to restorative practices, one must overcome three core research challenges: (a) identifying students' levels of exposure to restorative practices, (b) tracking students' outcomes over time, and (c) ascertaining the causal relationship between exposure and outcomes. As discussed below, I will attempt to overcome the first challenge via approaches similar to those leveraged by prior researchers (Aizer & Doyle, 2015; Bacher-Hicks et al., 2019; Hinze-Pifer & Sartain, 2018; Perry & Morris, 2014) to use historical data to typify social institutions in terms of the practices they tend to use, and therefore ascertain the extent to which those who come into contact with that institution are likely to be exposed to these practices. I address the second challenge by linking information about school practices with California administrative data that tracks student attendance, academic achievement, and disciplinary experiences over time.

The third challenge is certainly the most complex because other phenomena (besides restorative programs) can drive student exposure to restorative practices, and because other phenomena (besides restorative practices) can drive student outcomes. Put another way, certain schools are more likely to implement restorative practices, and certain students are more likely to be exposed to these practices; thus, simply relating school-level restorative practice utilization or student-level restorative practice exposure to student outcomes will almost certainly yield biased estimates. While randomized controlled trials can isolate the unique impact of restorative programming on student outcomes, these programs do not always redound to restorative practices. Thus, as Darling-Hammond et al. (2020) have encouraged, I will turn to causal research methods that do not rely on program randomization, and instead recruit other established approaches (Angrist & Pischke, 2009; Gerber & Green, 2012; Glennerster & Takavarasha, 2013) to identify the relationship between exposure to restorative practices and student outcomes net of student and school selection effects. As discussed below, I will leverage within-student and within-school estimation approaches similar to those employed by Perry and Morris (2014) and Hinze-Pifer and Sartain (2018) to estimate the impacts of student exposure to exclusionary discipline. Our research journey begins in earnest with the first question: how can we identify students' levels of exposure to restorative practices over time?

Challenge 1—Identifying Student Exposure to Restorative Practices

In order to *compare* students' levels of exposure to restorative practices, one must first be able to *assess* a given student's level of exposure. This can be achieved by ascertaining whether students attend schools that use these practices, but to operationalize this approach, we must identify whether schools are using restorative practices. Faced with the similar challenge of identifying whether schools promoted diversity, Levine et al. (2019) reviewed the mission

statements for 100 schools, coding each for the extent to which they use language indicating a commitment to diversity (e.g., "a diverse student body," "promoting life-long learning and respect for diversity"). Comparing students of color who attended, or did not attend, schools that emphasized the value of diversity, they found that those who attended diversity emphasizing schools exhibited superior health outcomes. One can imagine that a similar approach (e.g., reviewing schools' policy handbooks or districts' Local Control Accountability Plans) might be leveraged to identify which schools or districts have invested in the use of restorative practices. However, in my review of quantitative evidence on restorative practices (Darling-Hammond et al., 2020), I found that policies on the books do not necessarily relate to practices in the halls, and that many schools attempt to sidestep scrutiny by indicating that they utilize some form of restorative practices (whether they do or not). We thus cannot rely on textual analysis methods to identify where students are being exposed to restorative practices.

Another means of identifying student exposure is to leverage student experiences, aggregated across the school, to typify schools. Aizer and Doyle (2015) used a similar approach to typify judges based on the decisions they handed down to juvenile defendants. By identifying the proportion of each judge's opinions that resulted in juvenile confinement, they were able to arrange judges along a continuum of punitiveness. Next, exploiting the fact that juvenile defendants are randomly assigned to judges, and instrumenting on the punitiveness of the judge to which a juvenile defendant was assigned, they ascertained the impact of juvenile incarceration on high school graduation and adult incarceration. Scholars (e.g., Bacher-Hicks et al., 2019; Hinze-Pifer & Sartain, 2018; Perry & Morris, 2014) have used similar approaches to typify schools, using historical data about how often schools issue suspensions to typify schools along a continuum of punitiveness, and use these scores to evaluate the impact of exclusionary discipline on student outcomes. One can imagine that a similar approach can be used to typify schools in terms of the extent to which students in those schools, in the aggregate, indicate that the school uses restorative practices. Put another way, just as a juvenile judge's past decisions can be used to typify judges and estimate the extent to which a juvenile defendant whose case is reviewed by that judge will experience a harsh sanction (Aizer & Doyle, 2015); and just as a school's past disciplinary actions can be used to typify schools on a scale of punitiveness to estimate the extent to which a student in that school will experience a harsh disciplinary sanction (Bacher-Hicks et al., 2019; Hinze-Pifer & Sartain, 2018; Perry & Morris, 2014); student reports regarding a school's use of restorative practices can be used to typify schools on a scale of restorative practice utilization and estimate the probability that a student in that school will be exposed to restorative practices.

To achieve (and leverage) this kind of typification, we must first develop a single, consistent measure of each student's level of "restorative practice exposure." Next, we must aggregate this measure across students in a school. Finally, by tracking the schools that students attend over time, we can estimate students' level of "restorative practice exposure" over time. Having constructed a measure of student exposure to our treatment, we can leverage various observational data analysis techniques to ascertain the relationship between exposure to restorative practices and student outcomes. Finally, to understand the *impact* of exposure to these practices, we must ascertain the portion of the relationship between exposure and outcome that is *not* a function of selection bias.

The first step to causal estimation, then, is to develop a standard means of measuring and tracking student exposure to restorative practices.

Available Data on Exposure to Restorative Practices. How can one measure students' levels of exposure to restorative practices? There are certainly many options one might consider, but they boil down to four basic questions:

1. Which survey questions best measure restorative practices (a multi-item scale capturing multiple facets of RP, or a single item scale capturing conflict resolution practices)?
2. What kinds of respondents are best positioned to speak to the existence of RP in an educational environment (students or school staff)?
3. At what level does exposure to restorative practices vary (school or district)?
4. During what time period do we measure restorative practices (a single-year period with specificity, or a multi-year period with precision)?

Which Survey Questions Reveal Exposure to Restorative Practices? As part of the aforementioned review of quantitative evidence on restorative practices (Darling-Hammond et al., 2020), I reviewed hundreds of pages of restorative practice implementation guides. In so doing, I identified a set of key restorative practices. These fell into three categories: practices designed to inculcate social and emotional skills necessary to resolve conflicts and deepen connections; practices designed to facilitate students' processes of conflict resolution; and practices designed to ensure a cohesive school community.

Seeking to quantify students' levels of exposure to restorative practices, I next looked to California Health Kids Survey (CHKS) student surveys. These surveys are completed by hundreds of thousands of California students across over one thousand schools each year. Schools participate bi-annually, meaning the set of schools that participate switches from year to year. However, in each bi-annual survey year, CHKS aims to survey 70% of all 5th, 6th, 7th, 9th, and 11th graders in each participating school. my review of CHKS data indicates they largely achieve this goal. CHKS includes a school climate module which is used annually to ask over one hundred thousand students in over three hundred schools to indicate the extent to which adults in their schools engage in a range of practices. As depicted in Table 2, many of the practices reviewed in the CHKS school climate module fall within the three core types of restorative practice that I identified in my prior review of all research on these practices.

Each of these measures ranges from 1 (low exposure) to 5 (high exposure). These measures can be used alone or can be scaled to create measures of restorative practice exposure. How can one decide which measure or measures to use?

Restorative practices are focused on helping students manage conflict, both proactively (through skill development) and responsively (through facilitated conflict resolution).

Some may argue that proactive restorative practices hue quite closely to social and emotional learning strategies. These strategies have been demonstrated to be effective in many contexts (Taylor et al., 2017), and thus a scale that includes social and emotional learning strategies could predict positive benefits for students not because of practices related to active conflict resolution, but because of practices related to social and emotional development. Thus, some may prefer a measure of restorative practice exposure that focuses solely on conflict resolution, such as the measure that stems from the question, "this school helps students solve conflicts with one another."

Alternatively, others may argue that for schools to be *truly* restorative, they must embrace both responsive *and proactive* restorative practices, and should feature elements from each of the three buckets I identified in my research review. These scholars might prefer a multi-item scale of restorative practices that captures elements sitting in each bucket. Thus, they might prefer a

scale using each of the eight items depicted above. Indeed, when I sought guidance from preeminent restorative practice researchers and practitioners, including A. Gregory (personal communication, November 23, 2020) and K. Hickman (personal communication, April 30, 2017), they opined that the scale was a strong mechanism for ascertaining student exposure to the kinds of restorative practices theorized to improve student outcomes. The eight-item exposure measure also had a high scale reliability coefficient (Cronbach's Alpha) with a score of 0.910, indicating excellent internal consistency. The average inter-item correlation of 0.636 (with correlations ranging from 0.361 to 0.762) further indicated that while items were related, they were not duplicative.

Table 2

California Health Kids Survey Items Utilized to Measure Restorative Practice Utilization in Schools, Subdivided by Practice Type

Practice type	Items
Repair	This school helps students solve conflicts with one another
	If I tell a teacher that someone is bullying me, the teacher will do something
Community building	This school encourages students to feel responsible for how they act
	This school encourages students to understand how others think and feel
	This school encourages students to care about how others feel
	Students are taught that they can control their own behavior
Breadth	Teachers show it is important for students of different races to get along
	The adults in this school respect differences in students

Both the single and eight-item scales proved related to real-world signals of restorative practice utilization. I reviewed practice guides for large districts which appear in the CHKS data and found that those that scored highest on the single-item and eight-item measures also had written documentation of their utilization of school-based restorative practices. Both approaches thus passed a litmus test of real-world validity.

Given the ostensible validity and value of both the single-item and eight-item scales, I ran all analyses based on both measures. Notably, reviewing middle school (grade 6-8) student surveys from 2013-14 through 2018-19, the single and eight-items measure proved strongly correlated at both the student level (r(234,575) = 0.82) and at the school level (r(1,189) = 0.91). Perhaps not surprisingly, then, when both measures were available, results converged regardless of whether I used the single or eight-item scale as the exposure measure. For the sake of brevity, I primarily present findings based on the eight-item scale, but use the single-item scale for one sensitivity analysis and for one analysis for which the eight-item scale was unavailable.

What Kinds of Respondents Are Best Positioned to Indicate Restorative Practice Utilization? Thus far, I have indicated my intention to use student survey responses to ascertain restorative practice utilization. But what of staff opinions? The California Survey of School Staff (CSSS) provides a rich set of data regarding school staff opinions, and includes each of the

questions used to construct the eight-item scale. Given this, one might wonder whether teacher survey responses might provide a better lens in to the extent to which schools utilize restorative practices.

I think this is unlikely for four reasons. First, CHKS's sampling frame is designed to ensure a representative group of approximately 70% of the students in each participating district take their surveys. Thus, when a school scores highly on a student-based measure of restorative practice utilization, we can be more certain that students of varying racial backgrounds are rating their school highly in its use of these practices. In contrast, CSSS staff data is not designed to be representative of staff in a given district, let alone a given *school*, and may capture only certain staff that serve certain kinds of students.

Next, whereas students have little incentive to rate their teachers highly in an anonymous survey, staff may overstate their use of restorative practices to buttress their own sense of self. This kind of self-concept-oriented "social desirability" effect has been observed in anonymous survey settings (Brenner & DeLamater, 2016; Larson, 2019) and thus could conceivably appear in CSSS data. Moreover, we may worry that teachers might provide rose-tinted assessments to ensure their schools perform more favorably on district or state accountability metrics. We see evidence of these effects when comparing school-level CSSS and CHKS measures of restorative practice utilization. Whereas only 3% of CHKS students (in all grades) give their school a perfect score on restorative practice utilization, a whopping 16% of staff give their school a perfect score. On the single-item score, whereas 10% of students give their school a perfect score, 29% of teachers do the same. We thus may trust students to present a more balanced view of their schools' use of restorative practices.

Third, while students do sometimes shift the school they attend, teachers shift school environments (and leave the teaching profession altogether) at alarming rates each year. Thus, to the extent that we hope to measure school utilization of restorative practices over time, we should be concerned that the composition of teachers (and therefore the composition of teacher survey takers) could change wildly over time. If it did, then any observed shifts in restorative practice utilization could really be a function of shifts in the kind of teachers hired or retained by a given school at a given point in time.

Finally, we may worry that teachers' ability to implement restorative practices with discretion could lead teachers to state that they are using these practices very often when they are truly using these practices in only selected situations (e.g., when interacting with White students). Leveraging these teacher responses could overstate the extent to which certain kinds of students (e.g., Black students) are being exposed to restorative practices.

However, one could make arguments in favor of the use of teacher surveys. For example, because some of the outcome data stems from student surveys, one might worry that examining the relationship between restorative practice exposure and these outcomes is tautological—students who are generally unhappy might rate their schools poorly and might also rate their lives poorly while students who are generally happy might do the opposite. This could create a spurious relationship between exposure and outcome. Relatedly, one might also worry that students' ratings of their level of exposure to restorative practice might partially reflect whether the student, themselves, has a proclivity towards misbehavior. This would suggest that school-level restorative practice utilization scores might actually detect cumulative misbehavior rates, which could be related to cumulative disadvantage, rather than actual teaching practices. If this is so, then relationships between estimated restorative practice utilization and student outcomes might simply reflect latent relationships between structural disadvantage and outcomes.

While teacher-survey-based measures have their drawbacks (as discussed above), school-level variation on teacher-based measures is arguably less likely to reflect differences in student composition, and is more likely to reflect differences in school practices. As such, one means of ascertaining whether student-survey-based measures of restorative practice utilization are valid (and don't merely reflect structural disadvantage) is to see if they align with teacher-survey-based measures of utilization. We can not only review whether the two measures align, generally, but also whether they align in varied contexts. For example, if student-survey-based measures simply reflected cumulative disadvantage, then we might expect that the relationship between student and teacher-based measures would be more positive in schools with a high degree of structural disadvantage (e.g., schools with more students who are economically disadvantaged or more students who are non-White), and could be flat or even negative in schools with a low degree of structural disadvantage. However, as depicted in Figure 12, we see that student-based scores and teacher-based scores align in all schools; in schools that are more or less economically disadvantaged; and in schools that have larger or smaller proportions of Black and Hispanic students. Together, this suggests that student responses are not reflections of underlying student characteristics, and can be used to identify school practices.

Figure 12

Relatedness of Measures of Restorative Practice (RP) Utilization Based on Student Surveys and Teacher Surveys
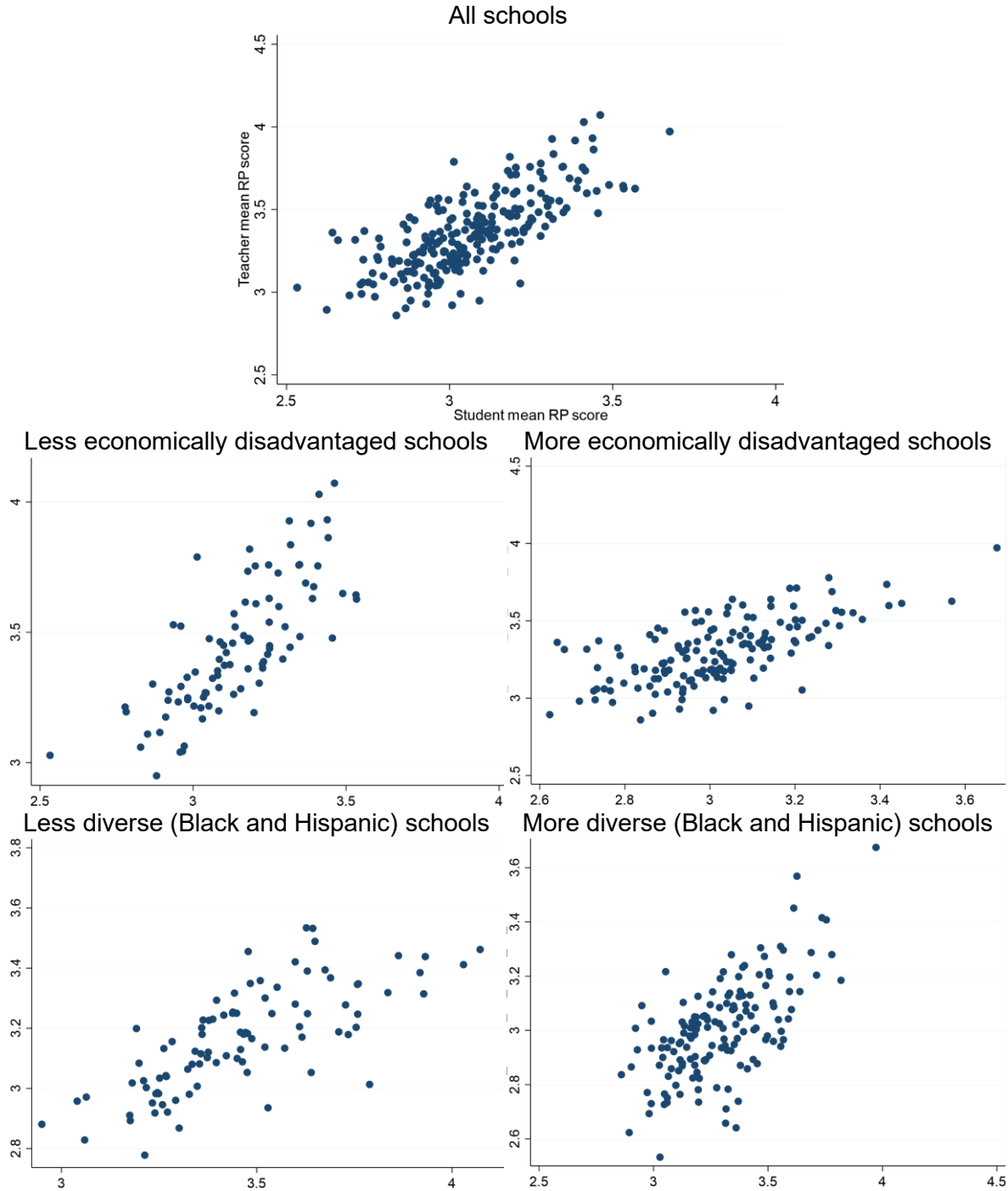
Note. Student mean RP scores were taken by averaging all student surveys from a given school from 2013–2014 through 2018–2019 and restricting to schools with 100 or more surveys. Teacher mean RP scores were taken using the same method and precision cutoff.

At What Organizational Level Should We Measure Restorative Practice Utilization?
Districts frequently issue policies regarding the practices to be implemented in schools. They also provide funding and support for the use of these practices. And district policies can be leveraged by students via legal proceedings to force schools to utilize district-enshrined practices. One might expect, then, that restorative practice utilization would vary primarily at the district level, rather than at the school level.

However, while districts often promulgate district-wide policies which have the force of law, they often use vague language that merely encourages schools to consider restorative practices from among a large number of alternatives to exclusionary discipline (such as positive behavioral interventions and supports). Districts can thus satisfy their legal obligations by considering and/or implementing one of any number of alternatives to exclusionary discipline. In addition, even when districts specifically express that schools should use restorative practices, districts often do not provide training or support to enhance schools' utilization of these practices. Many districts merely encourage schools to engage local restorative practitioners to provide professional development or services. In short, restorative practices are rarely a district-wide mandate, and even when they are, they are often a financially unsupported mandate. This can lead to substantial unevenness in the extent to which schools within the same district utilize restorative practices.

Reviewing CHKS data indicates that while there is variation between districts in terms of their utilization of restorative practices, there is far more variation between schools within districts. Indeed, when looking at school-level data nested in districts, we see that only 21% of the variation in schools' use of restorative practices (based on the eight-item scale score) occurs between districts, and 79% of variation occurs within districts (i.e., between schools). When reviewing the one-item measure, we see, similarly, that only 19% of variation occurs between districts; and 81% occurs between schools (i.e., within districts). I thus focus on variation in school-level utilization (and school-site related exposure) in my models.

During What Timeframe? To ascertain the extent to which each student was exposed to restorative practices, we must find a way of determining each school's level of restorative practice utilization. Theoretically, one could use student survey data to determine the extent to which each school used restorative practices in each year. However, because schools participate in CHKS biannually, the set of schools that participate in CHKS alternates year to year. This makes reviewing school-year level data for *each* year in the data impractical. Moreover, the sample of students who participate in CHKS in a given school in a given survey year could vary based on other factors, such as whether there were school or community events that might shift the composition of students available to take the survey during a given survey time frame. All of this can contribute to sampling error.

To maximize the precision with which I estimate restorative practice utilization, I therefore average restorative practice scores for each school over multi-year time periods. Most models utilize the six-year time period from 2013-14 through 2018-19 and restrict analyses to schools whose averages stem from scores from one hundred or more students. Other models utilize three-year time periods (2013-14 through 2015-16 and 2016-17 through 2018-19) and restrict the sample of schools to those whose averages stem from fifty or more students in *each* of these time periods.

For Which Grade Levels? Middle school presents a particularly delicate time for students. For example, P. J. Cook et al. (2008) have found that 6th grade students quasi-randomly placed in middle schools (rather than elementary schools) experience more discipline. This suggests that the transition from elementary to middle school marks a critical moment in the disciplinary trajectories of students. My review of 2018-19 California Longitudinal Pupil Achievement Data System (CALPADS) data echoes the critical nature of middle school. As depicted in Figure 13, the out-of-school suspension rate more than doubles between elementary and middle schools, and is higher in middle school than in high school.

Figure 13

Discipline Rate by School Type
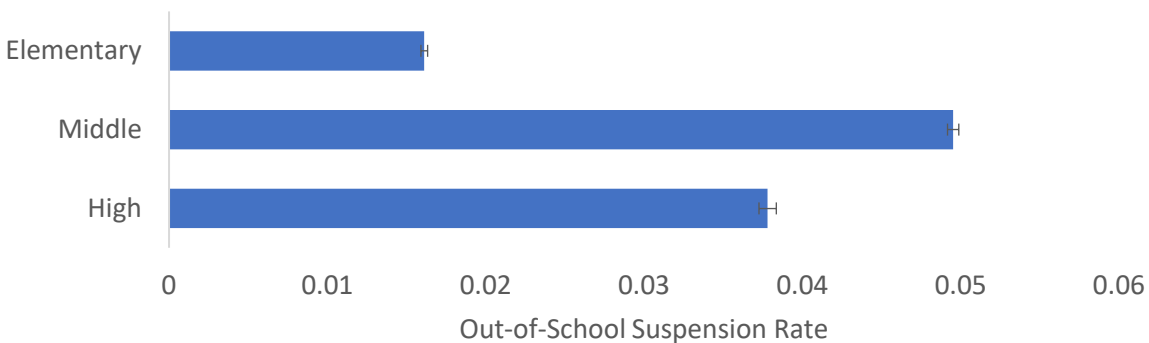
Note. Bars depict mean values (school type) for whether 3,254,662 students ever received an out-of-school suspension in the 2018–2019 school year. Error bars represent 95% confidence intervals around school type means. Because CALPADS has data for only certain grades, the elementary out-of-school suspension rate was calculated using data from Grades 3–5; the middle school rate was calculated using data from Grades 5–8; and the high school rate was calculated using data from Grade 11.

Researchers have also found that relational interventions implemented with middle school teachers can have outsized impacts on student discipline (e.g., Okonofua et al., 2016, 2020). These considerations strongly encourage reviews of the effects of exposure to restorative practices for middle school students.

Another more data-oriented consideration is that many of my models will rely on both CHKS data; and on two California administrative data sources:

1. CALPADS, which includes data regarding student economic disadvantage, migrant status, English language learner status, race, gender, grade level, school of attendance, whether received an out-of-school suspension, and number of days spent in out-of-school suspension; and

2. California Assessment of Student Performance and Progress (or CAASPP), which includes data regarding student special education status, and performance on two standardized tests (the smarter balanced English Language Arts assessment, and the smarter balanced math assessment)

CHKS data provide information about restorative practice exposure using the eight-item scale for primarily 7th, 9th, and 11th grade students. Thus, a key exposure measure is only available at the middle and high school level. The CAASPP data, meanwhile, provides information for 3rd through 8th, and 11th grades, but it is most useful when it includes data for prior years (i.e., 4th through 8th grades). Thus, the CALPADS and CAASPP data, when combined, are also limited to these grades. For the sake of economy, I will hereinafter refer to the merger of these two datasets as "the California administrative data."

Taken together, both CHKS and the California administrative data feature coverage at the middle school level. Finally, some of my models will attempt to ascertain how schools evolve in their use of restorative practices over time using two three-year periods. To avoid overlapping observations, we can focus on middle (6th through 8th) grade students.

Thus, partially due to the sensitive nature of middle school, and largely to enhance parsimony with the data sources leveraged to glean effects, I restrict all models to middle school students.

Schools' Restorative Practice Utilization Scores. Focusing on middle schools, utilizing the six-year time frame, and restricting to schools whose averages stem from one hundred or more student surveys, we can generate school-level restorative practice utilization scores for 485 schools. Figure 14 depicts the distribution of these school-level restorative practice utilization scores, and demonstrates that schools evidenced substantial variation in their use of restorative practices.

Figure 14

Histogram of School-Level Restorative Practice Utilization Scores
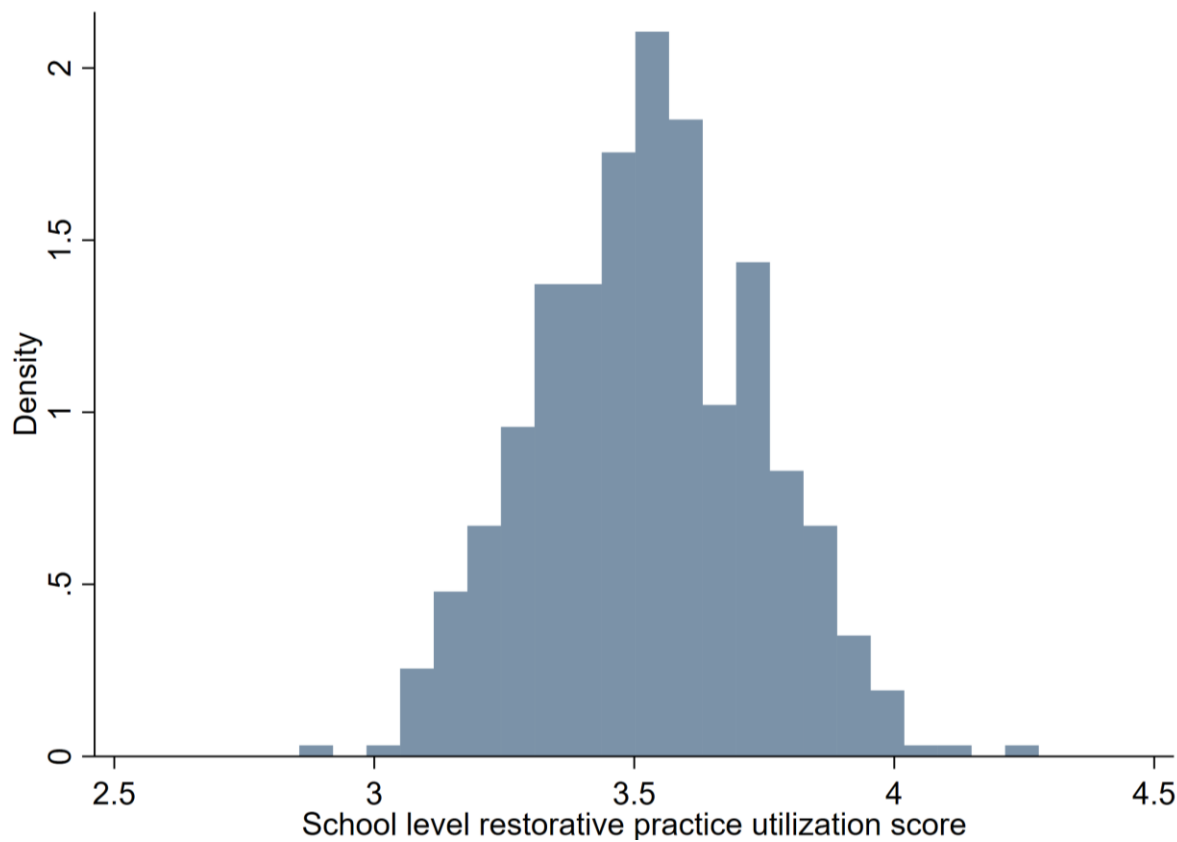
Note. The distribution shows school-level mean restorative practice utilization scores for the 485 schools for which there were 100 or more student-level restorative justice exposure scores to average over the 6-year period from 2013–2014 to 2018–2019. The 485 school-level scores are built on 219,627 student-level surveys over the 6-year timeframe.

In the same way that school punitiveness scores can be used to track student exposure to disciplinary environments (Bacher-Hicks et al., 2019; Hinze-Pifer & Sartain, 2018; Perry & Morris, 2014), the restorative practice utilization scores described above can be used to track students' levels of exposure to restorative practices over time. Having overcome our first challenge, we can move on to our second, and architect data that allows us to track student attendance and experience over time.

Challenge 2—Tracking Student and School Outcomes Over Time

Extant research has identified at least seven classes of outcomes that are potentially related to restorative practice exposure: misbehavior, victimization, discipline, school climate, academic achievement, attendance, and mental health (Darling-Hammond et al., 2020; Todic et al., 2020). As noted above, the CHKS data provide powerful data for identifying *school*-level restorative practice exposure. They also provide useful measures for certain outcomes at the

school level, including misbehavior, school climate, mental health, and academic achievement. And they provide means of capturing aggregate student body demographics over time. Critically, because CHKS does not follow individual students over time, it cannot be leveraged for certain casual inference techniques. However, because CHKS student-level data can be aggregated, this data can be used to track each school's use of restorative practices, and aggregate school-wide outcomes, over time, as well as keep track of how student body compositions might shift over time. The data can thus be used to explore a critical question: how do aggregate student body outcomes shift when schools increase (or decrease) their utilization of restorative practices?

  While CHKS data provide insights regarding *school*-level effects for many measures, other datasets provide useful measures of student-level outcomes and track student experiences over time. Two of the strongest such datasets are CAASPP and CALPADS, which are the state's administrative databases tracking students' school placements, demographics, academic growth, and disciplinary experiences. For each student in grades 3-8 and 11, the California administrative data captures the school they attended, their demographic information, their disciplinary experiences (whether they experienced an out-of-school suspension in a given school year, as well as how many days they were out-of-school suspended in a given year), and their smarter balanced scores in English and Math. Smarter balanced scores are standardized measures of achievement. Because they are both standardized and normalized within grade level, they are in many respects ideal for comparing students across school environments and for tracking students' growth over time. The California administrative data covers approximately 3.2 million students per year.

  Table 3 below summarizes the outcome variables available at the student level from California administrative data, and at the school level from CHKS.

Table 3

Outcome Measures Available at the Individual and School Levels

Outcome	Student level [a]	School level [b]
Discipline	Whether received an out-of-school suspension in prior year Days of out-of-school suspension in the prior year	
Academic achievement	Smarter Balanced Math Score (typically taken between February and May) Smarter Balanced English Language Arts Score (typically taken between February and May)	Student grade point average over last 12 months (self-report)
Attendance		Whether missed school for any reason in the past 30 days
Misbehavior		Whether engaged in various acts of misbehavior in the prior 12 months: fought, destroyed school property, carried a gun to school, carried another weapon to school
School climate		A scale score based on six school climate module responses: feel like part of school, feel close to people at school, feel happy at school, feel safe at school, feel an adult at school cares, feel an adult at school listens
Health		Whether missed school in last 30 days due to various health challenges: depressive symptoms, sleep deprivation, illness, substance use
Victimization		Whether experienced various kinds of victimization in the past 12 months: beat up; threatened harm; threatened or injured with weapon; stolen from; called names; had rumors told about; had sexual jokes told about; harassed based on race, religion, gender, orientation, disability, or anything else

[a] California administrative data. [b] California Healthy Kids Survey.

Merging Exposure and Outcome Data. As noted, CHKS provides useful data regarding (among other things) the extent to which schools utilized restorative practices; and California administrative data provide useful information regarding student outcomes, attendance, and demographics—all over time. We can thus merge the two datasets to answer a variety of important questions such as: Do certain kinds of students have more access to restorative practices? Does exposure to these practices yield improvements on disciplinary and academic dimensions? And might increasing schools' utilization of restorative practices help reduce racial disparities related to academic achievement and discipline?

To answer these questions, we must first merge the CHKS and California administrative data. We do this by leveraging student-level information regarding school attended in any given year (from California administrative data) and linking it with school-level information regarding restorative practice utilization (from CHKS). Because we do not have CHKS data for all schools, we cannot glean restorative practice exposure scores for certain students.

As noted earlier, CHKS and California administrative data cover different student grades at different frequencies. California administrative data include all 3-8 and 11th grade students in California – roughly 450,000 students per covered grade per year. Put another way, 42% of California administrative data cover elementary school students (1-5th grades), 42% covers middle school students (6-8th grades), and only 14% covers high school students (9-12th grades). The CHKS data related to restorative practice exposure comes from grades 7, 9, and 11, with approximately 35% stemming from middle school students, and the remaining 65% coming from high school students. Because my modeling strategies rely on having both year-on-year student-level data, and having overlap between California administrative data and CHKS data, I will focus analyses here on middle school data. However, the results I present in the foregoing sections are functionally identical to those that emerge when I include the entire universe of students in the California administrative data and the entire universe of schools for which I have CHKS data, suggesting that whatever benefits restorative practices confer are not limited to middle school students.

That said, limiting ourselves to middle school students, and utilizing the merging process described above, I recover a dataset that captures both longitudinal student experiences and school-level restorative practice exposure for approximately 350,000 students in each year. Analyses based on these students appear generalizable to students throughout the state. California administrative data include approximately 1.4 million middle school students each year. CHKS, meanwhile, can be used to generate restorative practice exposure scores for a subset of schools, and therefore for a subset of students. Our dataset for analyses regarding the effects of restorative practice exposure on academic and disciplinary outcomes is thus limited to the set of students who have scores on both the restorative practice exposure measure (from CHKS) and academic and disciplinary measures (from California administrative data)—a total of about 320,000 middle school students per year. One may thus worry that the set of students included in our analysis dataset (because they have CHKS data) are distinct from the set of students that are excluded from our analysis dataset (because they lack CHKS data). However, looking at 2018-19 data (Table 4), we see that the students for whom we have both CHKS and California administrative data look demographically quite similar to the students for whom we have California administrative data but lack CHKS data. This suggests that our sample is representative of the full universe of California middle school students.

Table 4

Comparison of 2018–2019 California Administrative Data for Which Restorative Practice (RP) Information Was and Was Not Available

Category	% with RP data available ($N \sim 320,000$)	% with RP data missing ($N \sim 1,100,000$)
Race		
White	22.4	23.1
Black	4.4	5.6
Hispanic	56.6	55.1
Asian	9.0	9.3
Male	51.1	51.2
Economically disadvantaged	60.0	61.2
English language learner	14.8	14.1
Special education	11.6	12.0
Migrant	1.1	0.7

Note. In 2018–2019, the subset of students for whom California administrative and RP data are considered to be available are those students who appear in California administrative data and who attended schools in 2018–2019 that administered 100 or more California Healthy Kids Surveys regarding restorative practice exposure (taken between 2013–2014 and 2018–2019). These students attended schools with sufficient California Healthy Kids Survey data to generate a precise estimate of their schools' levels of restorative practice utilization. Approximate sample sizes are provided as the exact sample size for the number of students for whom data is available or unavailable varies marginally for each student characteristic.

Having constructed a first-of-its-kind dataset tracking students' outcomes and levels of restorative practice exposure over time, we can now begin our journey into evaluating the impacts of exposure to these practices.

Challenge 3—Identifying the Impact of Exposure on Student Outcomes

How can we ascertain if student exposure to restorative practices causes improvements in student outcomes? This inquiry boils down to two questions – can we exploit exogenous variation? If not, can we limit the influence of endogenous variation?

The purest form of exogenous variation is random assignment. For example, to ascertain the impact of restorative conferencing in juvenile courts, Shem-Tov et al. (2021) randomly assigned juvenile defendants to either go through a restorative justice process, or to go through a typical juvenile proceeding. It is worth noting that juvenile courts represent a near-perfect venue for randomized controlled evaluations. For Shem-Tov et al., assignment to treatment all but guaranteed youth would be exposed to the restorative justice intervention; and assignment to the

control condition guaranteed youth would not be exposed. The ability to tightly control exposure to the treatment allowed the authors to generate exogenous variation in exposure to treatment which empowered them to ascertain a causal effect: restorative proceedings markedly reduced recidivism. In contrast, generating exogenous variation in student exposure to restorative practices in *schools* is substantially more complex. Unlike youth randomly assigned to restorative juvenile proceedings, youth in schools randomly assigned to teachers who receive training in restorative practices do not reliably experience exposure to restorative practices (e.g., Acosta et al., 2019), as many conditions must be met to ensure training accrues to exposure. These include the implementation of high-quality training that changes teaching practices; a high degree of training uptake among teachers; a cultural fit between the training and the school culture; structural conditions in the school that allow teachers to implement restorative practices; and teachers exercising their discretion in ways that do not engender inequities in exposure. In addition, while teachers in control schools do not receive professional development in restorative practices via the *experiment*, these teachers may nonetheless have prior training (e.g., received as part of their certification process, or received at another school before lateralling to their current school site). Thus, in the context of evaluating the impacts of restorative practice exposure, students in schools randomly assigned not to receive restorative programming may prove to be poor controls. And while teacher-level randomization may provide a means of evaluating the effectiveness of restorative practice professional development at shifting teaching practices and student outcomes, research leveraging this approach is in a nascent state and will take many years to produce research findings.

Given that we cannot generate artificial exogenous variation, how might we identify and exploit *naturally* occurring exogenous variation in student exposure to restorative practices, or at least take steps to ensure our estimates are not biased by a failure to account for naturally occurring endogenous variation? A clue can be found in work by Aizer and Doyle (2015), who rely on the random process that determines which juvenile judge will hear a given child's case to identify the impact of being placed in juvenile confinement. A similar approach could be leveraged in the context of evaluating school practices if assignment to various kinds of schools could be seen as random. Of course, under typical circumstances, students do not sort randomly into school environments. Indeed, research by Owens and McLanahan (2020) documents how Black students are more likely to sort into more punitive schools, and how this sorting explains 21 percent of the Black-White discipline gap. We thus cannot treat selection into schools that use varying degrees of restorative practices as a random process.

But can we imagine that student sorting is more random (or at least less intentional) in certain circumstances? Relatedly, to estimate the impacts of school suspension, Bacher-Hicks et al. (2019) first typified schools in terms of their punitiveness, but then exploited plausibly exogenous variation in exposure to more punitive schools—the re-zoning of schools within a district which forced students who lived in the same neighborhoods and previously attended the same school to suddenly and unexpectedly attend different schools. While we do not have such an exogenous source of variation in student exposure to restorative practices (such as a radical rezoning), we can focus on moments in students' educational trajectories where students shift educational environments not so much because they choose to sort, but more because they are forced to switch schools.

One example of such a moment is the situation in which a student is completing a given grade while attending a school that does not serve the next grade level (for example, a 5th grade student whose current school does not offer 6th grade). One strategy, then, is to evaluate how

students' exposure to restorative practices shifts when they move from 5th to 6th grade, and ascertain the relationship between changes in restorative practice exposure and changes in student outcomes. Because this approach explores changes over time within a given student, it is often described as "within-student" estimation. Within-student estimators have been recruited to identify the effects of student exposure to discipline (Hinze-Pifer & Sartain, 2018; Perry & Morris, 2014), and have the benefit of adjusting for all time-invariant student characteristics in one fell swoop, including any characteristics which drive students' sorting tendencies.

Importantly, within-student estimators do not adjust for time-*variant* student characteristics. For example, if students who sort into restorative schools tend to also be on pre-sorting trajectories of faster cognitive development (for example, due to lower levels of exposure to systematic disadvantage), then within-student estimators will be biased. Specifically, in this case, if we were to observe a positive relationship between increased exposure to restorative practices and increased student outcomes, this relationship could simply be a reflection of how students on faster growth trajectories both tend to select restorative schools and tend to see faster gains.

If we identify a strong effect using a within-student estimator, how can we bolster our confidence that what we are detecting is the impact of restorative practices, rather than time-variant student factors? One approach is to leverage a within-school estimator to ascertain if schools that grow more restorative over time also see aggregate student body outcomes improve. This approach adjusts for all time-invariant school characteristics, including the kinds of students who tend to sort into them. It thus provides a helpful compliment to the within-student estimator.

In the pages that follow, I will further describe and report the results of within-student and within-school estimates. Prior to reviewing these analyses, however, I will first explore whether certain kinds of students have more exposure to restorative practices than others. As noted above, there is reason to believe teachers will more readily use restorative practices when interacting with White, rather than Black, students (Kang et al., 2009; Okonofua & Eberhardt, 2015; Papageorge et al., 2020; Skiba et al., 2011; Smith et al., 2015). Thus, a review of predictors of student exposure to restorative practices will help fill a hole in restorative justice literature, which has only barely explored drivers of variation in student exposure to these potentially helpful practices (Fronius et al., 2019), and only in two studies using data from the 1997-98 school year (Payne & Welch, 2015, 2018). Having reviewed drivers of access to restorative practices, I will next review correlates of exposure to verify that exposure (as operationalized in this research) is related to positive student outcomes.

Access to Restorative Practices

We begin our inquiry by asking an implied question: who is gaining exposure to these theoretically beneficial practices? As depicted in Table 5, in the 2018-19 school year, student characteristics explained 12% of the variation in student exposure to restorative practices, and school characteristics explained 33% of the variation in restorative practice utilization. Concerningly, even after controlling for a range of other student demographics, Black students were markedly and significantly *less* likely to experience exposure to restorative practices. The same was true for Hispanic students and economically disadvantaged students. And schools with higher Black populations and higher economically disadvantaged populations had markedly lower levels of restorative practice utilization. School-level results were functionally identical when workforce experience and demographic variables were included (e.g., mean years of teaching experience; percent of teachers who were Black). Because no workforce variables

proved to be significant predictors of school-level restorative practice utilization, and because including them did not meaningfully shift predicted relationships between aggregate student demographic markers and restorative practice utilization, all workforce variables are omitted from the model presented below.

If exposure to restorative practices is theorized to reduce the likelihood that a student will experience exclusionary discipline, then in a sense, it is not surprising that the students who are most overrepresented among those disciplined also have the lowest levels of exposure to restorative practices. But the question remains—does exposure to restorative practices protect students against exclusionary discipline, improve other outcomes, and reduce racial disparities? To answer these questions, we turn to the next stage of our inquiry and review estimates of the relationship between exposure to restorative practices and student outcomes.

Table 5

Regression Models Predicting Restorative Practice Exposure With 2018–2019 Student and School Demographics

Predictor of restorative practice	Coefficient (robust *SE*)	
	Student level [a]	School level [b]
Grade (relative to Grade 6)		
7	−0.006 (0.007)	0.28[*] (0.12)
8	−0.008 (0.007)	−0.29[**] (0.09)
Race (White reference)		
Black	−0.098[***] (0.015)	−0.82[***] (0.19)
Hispanic	−0.044[***] (0.010)	0.20[*] (0.08)
Asian	0.057[*] (0.017)	0.23[*] (0.10)
Economically disadvantaged	−0.098[***] (0.009)	−0.44[***] (0.07)
Migrant	−0.019 (0.018)	−0.04 (0.58)
English language learner	−0.018[***] (0.005)	−0.08 (0.13)
Special education	0.003 (0.003)	−0.10 (0.23)
Female	0.000 (0.001)	−0.03 (0.27)
Intercept	3.640	3.75

Note. The first model was produced using student-level data and by regressing students' levels of restorative practice exposure (based on 6-year pooled averages from the California Healthy Kids Survey for schools with 100 or more student surveys) on student characteristics (from California administrative data). Restorative practice exposure scores ranged from 1 (*extremely low exposure*) to 5 (*extremely high exposure*). Thus, for example, being Black was associated with a 0.098-point lower restorative practice exposure score, and being economically disadvantaged was associated with a similar-sized, 0.098-point lower score. Standard errors for the first model were clustered at the school level. The second model was produced using school-level data and by regressing schools' restorative practice scores on school-wide student composition variables (e.g., proportion Black). It shows, for example, that a 1-proportion increase in the Black student population is related to a -.797-point lower restorative practice exposure score; thus, a 0.1 proportion (or 10%) increase is related to a 0.0797-point lower restorative practice exposure score. In Model 2, standard errors were clustered at the district level.
[a] $N = 318,831$ (students), and $r^2 = .12$. [b] $N = 482$ (schools), and $r^2 = .33$.
[*]$p < .05$. [**]$p < .01$. [***]$p < .001$.

Estimating Relationships via Adjusted Regression Analyses

Earlier, I noted that the main challenge of identifying the relationship between restorative practice exposure and outcomes is overcoming selection effects. California administrative data empower researchers to leverage many causal estimation techniques that can overcome, or at least mitigate, selection bias issues. One such approach is regression with adjustment including lagged outcome measures. By "lagged outcome measures," I mean the value of an outcome measure in a prior year. The benefit of these models is that they allow for a more precise estimate of the relationship between school experiences of interest in a given year and an outcome in a given year, net student characteristics that might have driven the outcome value in the prior year. So, for example, let's say we want to understand drivers of smarter balanced performance in the 2018-19 school year. A host of student characteristics might drive certain students to perform well on Smarter Balanced assessments year after year (e.g., family wealth or family education); and those same characteristics might encourage students to attend schools that utilize restorative practices. However, if we know how each student performed in 2017-18, we can regress a student's 2018-19 performance on their restorative practice exposure score in 2018-19, and then control for their 2017-18 performance. This helps ameliorate the confounding potential of their stable individual characteristics. We can further augment this approach by also controlling for student demographics.

Can this approach yield a causal estimate? Generally, in order for regression to return an unbiased causal estimate of the relationship between a treatment and outcome, the data must satisfy certain conditions. The most important of these (and the one most connected to treatment selection bias) is that the data must include all confounders, or all variables that predict both the treatment and the outcome (Angrist & Pischke, 2009).

As discussed above, race and other student demographics do indeed predict a great deal of the variation in restorative practice exposure, and they are also strong predictors of exposure to exclusionary discipline (GAO, 2018). They are thus confounders. Lagged, or prior year, outcome values are extremely strong predictors of later year outcomes. They are also unquestionably correlated with restorative practice exposure, if nothing else, due to their relationship to student demographics. They are thus also incredibly useful confounders to include in my models. Finally, given the relationship between individual student characteristics and restorative practice exposure, we can surmise that school-level mean values related to student characteristics (e.g., "percent of students who are Black") are related to school-level restorative practice utilization; and ample literature documents the relationship between school characteristics and student outcomes (e.g., Pearman et al., 2019). I thus include aggregate (school-wide) student characteristics in my models as well. In addition, research demonstrates that *staff* demographics are also drivers of school practice adoption (Roch et al., 2010), and that staff experience is also related to student discipline practices (Moore & Cooper, 1984; Williams et al., 2020). To further adjust for school-level endogeneity, I thus include terms, for teachers and principals alike, regarding mean years of experience, percent credentialed, percent female, percent Black, and percent White. In my models, I will thus adjust for the student and school-level variables listed in Table 6.

Because students in a given school may be more similar to one another than they are to students from the overall sample, I employ multivariate linear regression with clustered standard errors (at the school level). To leverage this approach, I regress 2018-19 student outcomes on each student's restorative practice exposure score (which is based on the level of restorative practice utilization in the school they attended in 2018-19), and adjust for the confounders listed

above. As noted, I measure 2018-19 restorative practice exposure by determining each school's level of utilization of restorative practices and by tracking students' attendance over time. I measure schools' restorative practice utilization by averaging all student survey responses regarding restorative practice exposure for a given school over a six-year time frame (2013-14 through 2018-19). This strategy presumes some degree of stability in schools' utilization of restorative practices over time. As I shall discuss later, when we evaluate how schools shift in their use of restorative practices over time, we see evidence of this kind of stability. However, to ensure that results are not an artifact of the timeframe used to develop the exposure, I re-ran all analyses using restorative practice utilization scores stemming from the most recent one, two, and three years of data. These models produce results that are concordant with what I present below.

Table 6

Student- and School-Level Characteristics Included in Multivariate Regressions

Student-level characteristics	School-level characteristics
Economic status	*Mean student characteristics*
Migrant status	Proportion of students with "low" economic status
English-language learner status	Proportion of students with migrant status
Special education status	Proportion of students designated as English
Race	language learners
Gender	Proportion of students receiving special education
Grade level	Proportion of students who are White, Black, Asian,
Prior year outcome values: whether	Hispanic
received an out-of-school suspension,	Proportion of students who are female
days of out-of-school suspension	Proportion of students in sixth grade, seventh grade,
received, Smarter Balanced ELA score,	eighth grade
Smarter Balanced Math score	Number of middle school students in a given school
	Mean teacher characteristics
	Mean years of experience for teachers
	Proportion of teachers with a credential
	Proportion of teachers who are female
	Proportion of teachers who are White
	Proportion of teachers who are Black
	Mean administrator characteristics
	Mean years of experience for administrators
	Proportion of administrators with a credential
	Proportion of administrators who are female
	Proportion of administrators who are White
	Proportion of administrators who are Black

Formally, my model is

$$\text{OUTCOME}_{18\text{-}19} = \alpha + \beta_1(\text{SRPE}_{18\text{-}19}) + X_i + \varepsilon, \tag{1}$$

where

- $\text{OUTCOME}_{18\text{-}19}$ is the 2018-19 outcome of interest in a given model (received an out-of-school suspension in 2018-19; smarter balanced English score in 2018-19; smarter balanced math score in 2018-19);
- $\text{SRPE}_{18\text{-}19}$ is the restorative practice exposure score for the school a student was in in 2018-19; and
- X_i is vector of covariates, including all 2017-18 outcomes, all student characteristics, and all school characteristics.

Here, β_1 is our coefficient of interest and, presuming the identifying assumptions related to regression are met, β_1 represents an unbiased estimate of the causal effect of exposure to restorative practices. As depicted in Table 7, adjusted regression models suggest that, overall and for each category of students reviewed, student exposure to restorative practices is related to higher English Language Arts scores, higher math scores, a lower probability of receiving an out-of-school suspension, and fewer days spent in out-of-school suspension. Before moving on to what these models indicate about racial disparities in discipline and achievement, it is important to note what they do *not* indicate. Some have expressed concern that by keeping unruly students in classrooms, restorative practices may actually harm outcomes for certain subcategories of students (Eden, 2020). And, as discussed above, some pre-post studies and one randomized controlled trial found that implementation of restorative programming is associated with or caused declines in academic achievement. Our models, which overcome the conflation of restorative programming and restorative practices, and focus on exposure to restorative *practices*, show no evidence of these negative externalities. Instead, students of all backgrounds (including White and Asian students) saw a positive association between restorative practice exposure and academic achievement.

Models also indicate that, relative to White students, the benefits are slightly more pronounced for Hispanic students, and substantially more pronounced for Black students. For example, a one-unit increase in restorative practice exposure is associated with a seven-unit increase in English Language Arts scores for Whites, but a seventeen-unit increase for Black students. Most notably, a one-unit increase in restorative practice exposure is related to 0.04 fewer days of out-of-school suspension for Whites, but 0.6 fewer days for Black students. The association is thus fifteen times stronger for Black students than for White students.

Because associations are stronger for Black and Hispanic students than for White students, as demonstrated graphically in Figures 15–18, all else being equal, these regression models indicate that at higher levels of restorative practice exposure, we see smaller Hispanic-White and Black-White achievement and discipline gaps. In these models, as we move from lowest to highest levels of restorative practice exposure, we estimate substantial declines in Black-White disparities in all four measures: a 9% decline for math scores, a 22% decline for English Language Arts scores, an 82% decline for out-of-school suspension rates; and a more than 100% decline for days suspended (meaning that at the highest levels of exposure, the Black-White disparity in days suspended disappears). While Hispanic-White disparities on these measures are generally smaller than Black-White disparities, results on Hispanic-White gaps are also notable and encouraging. Reviewing Hispanic-White disparities as we move from the lowest

to the highest levels of exposure to restorative practices, our models suggest a 7% decline for math score gaps, a 6% decline for English Language Arts, a 90% reduction for suspension rates, and a more than 100% reduction for days suspended. It is important to note that it is unlikely that any given student would move from the lowest to the highest levels of restorative practice exposure. Thus, the aforementioned percentages are not meant to suggest what is likely to occur if restorative practices are expanded, but merely to provide a lens into what may be possible.

Table 7

Relationship Between 2018–2019 Outcomes and 2018–2019 Restorative Practice Exposure, Based on Multivariate Regression With Adjustment for Student- and School-Level Factors

	Coefficient (Robust Standard Error)				
Outcome	All students	White	Black	Hispanic	Asian
ELA score	9.31***	8.24*	22.14***	10.86**	4.54
	(2.77)	(3.39)	(6.10)	(3.41)	(3.08)
n	266,223	56,339	11,892	155,211	23,276
Math score	8.57*	8.35	15.29*	12.63**	4.23
	(3.59)	(4.31)	(5.94)	(4.50)	(4.18)
n	265,816	56,224	11,856	154,983	23,266
Received OSS	−0.037***	−0.018*	−0.113***	−0.040***	−0.011
	(0.008)	(0.009)	(0.036)	(0.010)	(0.007)
n	269,210	57,354	12,116	156,594	23,384
Days in OSS	−0.17***	−0.04	−0.77***	−0.17***	−0.09
	(0.04)	(0.06)	(0.21)	(0.05)	(0.06)
n	269,210	57,354	12,116	156,594	23,384

Note. All models adjust for 2017–2018 outcomes (ELA score, math score, whether suspended, days suspended), student-level characteristics (economic status, migrant status, English-language-learner status, special-education status, race, gender, and grade level), school-level student body characteristics (percent economically disadvantaged, percent with migrant status, percent with English-language-learner status, percent with special-education status, percent female, percent in sixth grade, percent in seventh grade, percent in eighth grade, and middle school student population size), and school-level teacher and administrator characteristics (mean years of experience, percent with a credential, percent female, percent White, and percent Black). Models focusing on all students also adjust for percent Black, percent White, percent Asian, and percent Hispanic. Models focusing on racial subsamples do not include these student racial composition variables due to concerns regarding overcontrolling. Standard errors are clustered at the school level. ELA = English language arts; OSS = out-of-school suspension.
*$p < .05$. **$p < .01$. ***$p < .001$.

Figure 15

English Language Arts (ELA) Scores Estimated at Various Levels of Restorative Practice Exposure via Regression, by Race
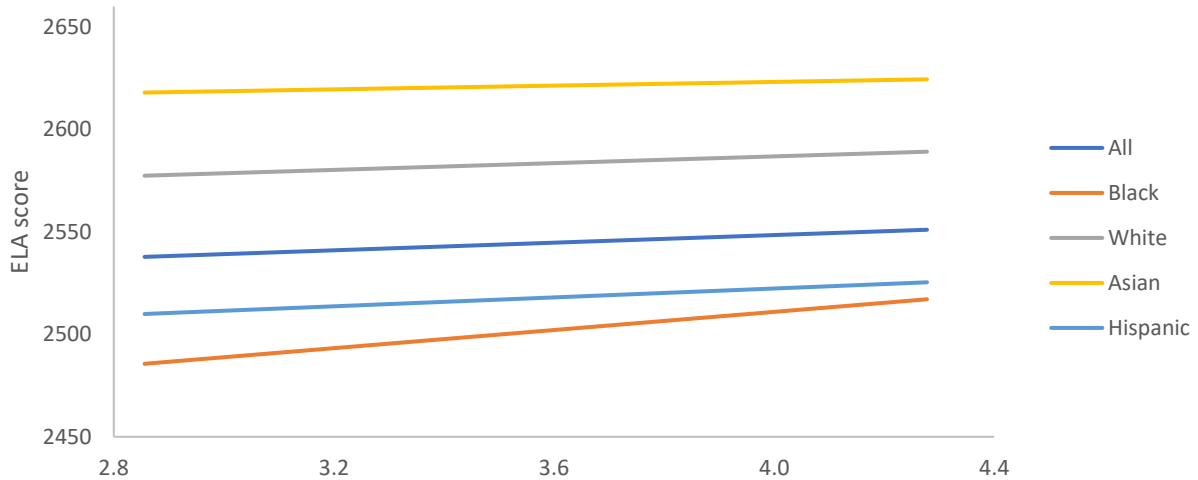
Note. Related models adjust for 2017–2018 outcomes, student characteristics, and school characteristics (see Table 7).

Figure 16

Math Scores Estimated at Various Levels of Restorative Practice Exposure via Regression, by Race
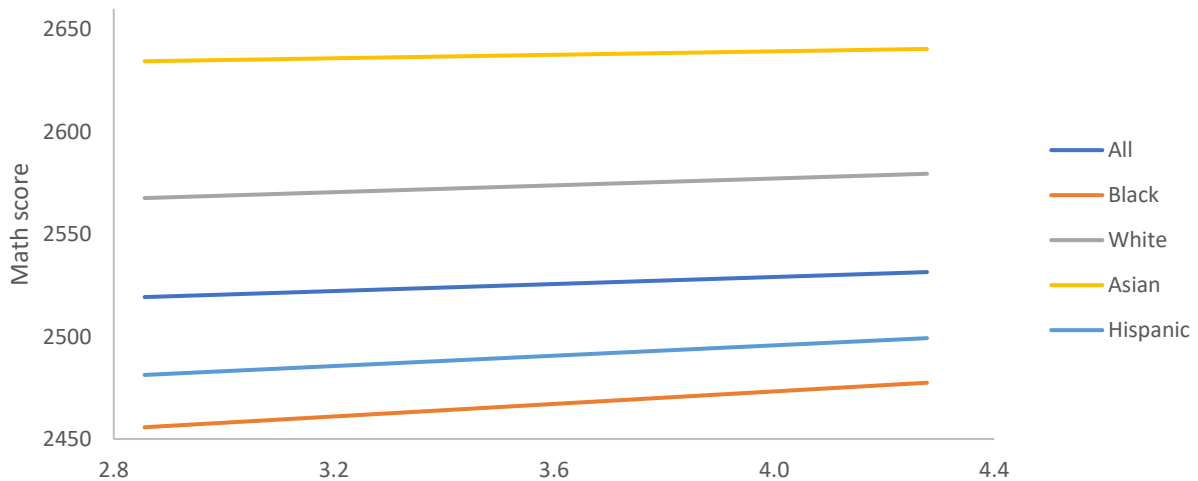
Note. Related models adjust for 2017–2018 outcomes, student characteristics, and school characteristics (see Table 7).

Figure 17

Out-of-School Suspension (OSS) Rates Estimated at Various Levels of Restorative Practice Exposure via Regression, by Race
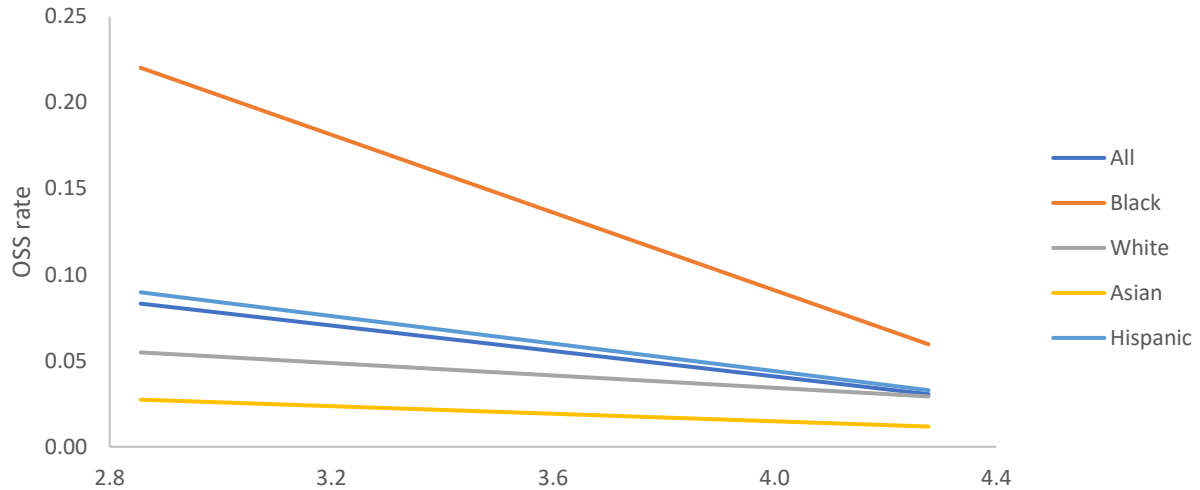
Note. Related models adjust for 2017–2018 outcomes, student characteristics, and school characteristics (see Table 7).

Figure 18

Mean Days Suspended Estimated at Various Levels of Restorative Practice Exposure via Regression, by Race
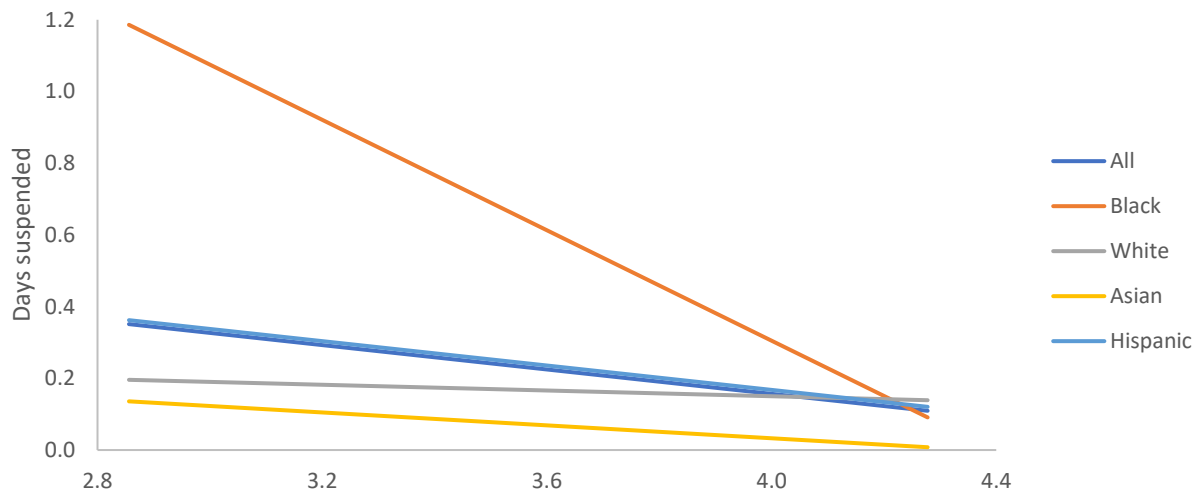
Note. Related models adjust for 2017–2018 outcomes, student characteristics, and school characteristics (see Table 7).

While these estimates are encouraging, they come with heavy caveats. First, these results hinge on the extent to which I have accurately ascertained school-level restorative practice utilization. As discussed above, some may feel that the eight-item scale captures aspects of school practices beyond what might be considered "restorative practices," and could be perceived as a proxy for a school being well run. Second, regression models only provide an unbiased estimate of the relationship between a predictor and outcome if they are properly specified. One could easily argue that any number of variables are non-linearly related to our outcome of interest, or our exposure; or that any number of variables (including our exposure) might reveal interaction effects if we conducted relevant tests. Finally, even assuming a well-classified exposure and properly specified model, the aforementioned regression results could still produce a biased estimate of the relationship between restorative practice exposure and student outcomes if the models do not include important student or school-level confounders. There are a seemingly infinite number of potential confounders one might include in these models, including markers of caregiver characteristics that relate to caregivers' tendencies to place their children in restorative schools, and school characteristics that relate to schools' tendencies to implement restorative practices.

In sum, our results hinge on 1) a properly classified exposure measure, 2) a properly identified functional form, and 3) the inclusion of all relevant student and school-level confounders. I attempt to address each of these issues in turn below.

To address the issue of exposure classification, I reproduce the regression results two ways:

- Using a single-item measure of exposure to conflict resolution practices. This measure is arguably less likely to proxy for other school practices (such as social and emotional learning).
- Using a measure of restorative practice utilization generated using staff surveys. Notably, due to teacher discretion in when (and with which students) to employ restorative practices, I would argue that this measure is less likely to accurately capture student exposure to restorative practices. Nonetheless, this measure captures dimensions of school practices that students may miss.

For the second issue (functional form), I reproduce results using a propensity score matching design as these designs are less sensitive to functional form specifications. Finally, to address student and school-level selection effects, I leverage within-student and within-school regression approaches.

Results Using a Single-Item Exposure Measure. The above results utilize an eight-item scale that is designed to capture three aspects of restorative practices: conflict resolution, community building, and scale. While the eight-item scale was validated in a number of ways, some may worry that the scale captures elements of what might be termed social and emotional learning (SEL). There is a broad and rich body of work demonstrating the benefits of SEL in schools (e.g., Taylor et al., 2017). Many have argued that effective implementation of restorative practices involves improving students' social and emotional capacities. However, given the relatedness of the eight-item scale and measures of SEL practice, it is worthwhile to evaluate whether using a single-item measure of exposure to conflict resolution practices generates similar results. I find that it does. As depicted in Table 8, while the effects are somewhat muted, using the single-item measure of exposure to conflict resolution practices, we see the same pattern: exposure to conflict resolution practices is associated with higher achievement and lower

rates of discipline for all categories of students; and associations are stronger for Black and Hispanic students than for White students.

Table 8

Relationship Between 2018–2019 Outcomes and Exposure to Conflict Resolution Practices, Based on Multivariate Regression With Adjustment for Student- and School-Level Factors

Outcome	Coefficient (Robust Standard Error)				
	All students	White	Black	Hispanic	Asian
ELA score	7.17^{**}	6.24^{*}	14.02^{*}	8.88^{***}	2.76
	(2.14)	(2.50)	(5.68)	(2.69)	(2.31)
n	279,013	60,100	12,142	162,470	24,074
Math score	6.22^{*}	6.87^{*}	9.41	8.98^{*}	−0.13
	(2.72)	(3.18)	(5.11)	(3.56)	(3.46)
n	278,572	59,975	12,106	162,220	24,063
Received an OSS	-0.031^{***}	-0.015^{*}	-0.095^{***}	-0.033^{***}	−0.009
	(0.006)	(0.006)	(0.028)	(0.008)	(0.006)
n	282,153	61,196	12,368	163,910	24,185
Days in OSS	-0.15^{***}	−0.03	-0.59^{***}	-0.16^{***}	−0.06
	(0.03)	(0.04)	(0.16)	(0.04)	(0.04)
n	282,153	61,196	12,368	163,910	24,185

Note. Unlike Table 7 (which shows relationships between exposure to restorative practices and student outcomes), this table displays relationships between exposure to conflict resolution practices and student outcomes. All models adjust for 2017–2018 outcomes (ELA score, math score, whether suspended, days suspended), student-level characteristics (economic status, migrant status, English-language-learner status, special-education status, race, gender, and grade level), school-level characteristics (percent economically disadvantaged, percent with migrant status, percent with English-language-learner status, percent with special-education status, percent female, percent in sixth grade, percent in seventh grade, percent in eighth grade, and middle school student population size), and school-level teacher and administrator characteristics (mean years of experience, percent with a credential, percent female, percent White, and percent Black). Models focusing on all students also adjust for percent Black, percent White, percent Asian, and percent Hispanic. Models focusing on racial subsamples do not include these student racial composition variables due to concerns regarding overcontrolling. Standard errors are clustered at the school level. ELA = English language arts; OSS = out-of-school suspension. $^{*}p < .05.$ $^{***}p < .001.$

Results Using an Exposure Measure Based on Staff Survey Responses. My main adjusted regression models leverage a school-level variable constructed by first generating a student restorative practice exposure scale score (based on eight student survey measures), and then averaging students' exposure scores across school environments to generate a restorative practice utilization score for each school. Notably, the eight student survey items used to construct my restorative practice utilization measure also appear on the California Survey of School Staff (CSSS). Theoretically, then, one can use the same process to develop school-specific restorative practice utilization scores based on *staff* surveys.

In the main models, I opted to use student surveys for many reasons. First, while the CHKS sampling is designed to capture representative subsets of students in each school, the CSSS sampling frame is not designed to capture representative subsets of teachers within each school. We thus may worry that variation in CSSS-based measures might reflect variation in the type of staff persons who participate in various school environments, rather than variation in school practices. Second, CHKS data include a variable which indicates the grade level of each respondent. CSSS data, meanwhile, does not provide information about the specific grade a given staff person serves. Thus, whereas one can aggregate CHKS data to estimate *middle-school-specific* conditions, one can only aggregate CSSS data to estimate conditions in the entire school, across all grades. Third, staff may be aware that, in some instances, CSSS data is used to evaluate schools, and may therefore censor their responses. Indeed, as discussed above, staff are much more likely than students to give their school a glowing score on restorative practice utilization. Finally, because teachers can exercise discretion regarding when (and with whom) to implement restorative practices, we may worry that staff scores might reflect a high degree of restorative practice utilization even in situations where whole swaths of the student body are enjoying very little exposure to restorative practices.

Still, some might prefer the use of staff surveys, or at least believe that these surveys capture a distinct and important aspect of school-level restorative practice utilization that is not captured by student surveys. One might also argue that teacher surveys avoid issues inherent in student surveys: that students' ratings of their level of exposure to restorative practice might partially reflect whether the student, themselves, has a proclivity towards misbehavior, and that a restorative practice utilization score based on student surveys might therefore reflect the degree of cumulative disadvantage in a school rather than actual teacher practices. Thus, below, I present results of adjusted regression models in which I predict the relationship between student outcomes (as measured by California administrative data) and restorative practice utilization (as measured by *staff* surveys).

As depicted in Table 9, estimates based on staff responses are directionally identical to those using student surveys, but are generally smaller. Still, estimates suggest strong effects on math scores across racial groups; and for Hispanic students across outcome measures.

Table 9

Relationship Between 2018–2019 Outcomes and Exposure to Restorative Practice Based on Multivariate Regression With Adjustment for Student- and School-Level factors and Using an Exposure Measure Constructed via Staff Survey Responses

	Coefficient (Robust Standard Error)				
Outcome	All students	White	Black	Hispanic	Asian
ELA score	11.32*	8.81	14.22	13.70***	7.59
	(5.55)	(5.44)	(8.44)	(6.51)	(6.84)
n	101,317	24,688	4,571	55,422	8,609
Math score	20.66**	26.45***	20.94*	22.53**	19.19*
	(6.86)	(6.64)	(8.14)	(8.47)	(8.51)
n	101,176	24,639	4,548	55,364	8,606
Received an OSS	−0.023	−0.016	−0.026	−0.035*	−0.011
	(0.014)	(0.019)	(0.037)	(0.015)	(0.013)
n	102,468	25,043	4,657	56,011	8,643
Days in OSS	−0.15	−0.05	−0.43	−0.22**	−0.18
	(0.09)	(0.10)	(0.28)	(0.08)	(0.13)
n	102,468	25,043	4,657	56,011	8,643

Note. Unlike Table 7 (which shows relationships between exposure to restorative practices and student outcomes in models that use a measure of restorative practice utilization based on student surveys), this table displays the relationship between exposure to restorative practices and student outcomes in models that use a measure of restorative practice utilization based on staff surveys. All models adjust for 2017–2018 outcomes (ELA score, math score, whether suspended, days suspended), student-level characteristics (economic status, migrant status, English-language learner status, special-education status, race, gender, and grade level), school-level student body characteristics (percent economically disadvantaged, percent with migrant status, percent with English-language-learner status, percent with special-education status, percent female, percent in sixth grade, percent in seventh grade, percent in eighth grade, and middle school student population size), and school-level teacher and administrator characteristics (mean years of experience, percent with a credential, percent female, percent White, and percent Black). Models focusing on all students also adjust for percent Black, percent White, percent Asian, and percent Hispanic. Models focusing on racial subsamples do not include these student racial composition variables due to concerns regarding overcontrolling. Standard errors are clustered at the school level. ELA = English language arts; OSS = out-of-school suspension.
*p < .05. **p < .01. ***p < .001.

Results Using a Propensity Score Matching Approach. As noted earlier, multivariate regression will only return an unbiased estimate of the relationship between exposure and outcome if certain conditions are met. Chief among these is a properly specified functional form. While there are many diagnostic tests to ensure a linear model is appropriate, one means of ascertaining whether functional form issues might be generating spurious relationships is to utilize propensity score matching. Propensity score matching designs are not reliant on functional form assumptions. Thus, where propensity score matching and multivariate regression approaches converge, we are less worried that results are an artifact of functional form decisions. As discussed below, results do converge. Below, I provide some explication of the propensity score matching approach, and present results generated using this approach.

To operationalize this propensity score matching approach, I first calculate a logistic regression model in which I use nearly all of the adjustment variables from the linear regression to predict whether or not a student was in a top quartile restorative school in 2018-19. Notably, the logistic regression model does *not* include workforce variables as these variables proved poor predictors of both restorative practice exposure and student outcomes. Also, as with the multivariate regressions, I include individual and school-level race variables in models predicting relations for all students, and I omit race variables in models predicting relationships for subgroups of students.

I then use the propensity score models to predict each student's unique probability of being in a top quartile school (they're "p-score" or "probability of treatment"). Finally, I match each student who was in a top quartile restorative school to the student with the closest p-score who was not in a top quartile restorative school. While there are other means of matching students, because there is a great deal of data to draw from here, I have chosen to execute matching "without replacement," meaning each student can be matched only one time. Unlike with matching with replacement, with this method, we need not worry that our model estimates an artifact of a single student being matched many times.

This approach has two benefits. First, it reduces our covariate matrix to a single dimension, allowing for easy pairing of "treated" and "control" cases. As such, this approach does not rely as heavily on functional form assumptions. Second, because the treated and control cases are more similar on covariates than the full sample of treated and control individuals, we also improve balance on covariates between treated and control groups. Critically, propensity score matching can only be executed when there is a sufficient "region of common support," meaning that for any given treated individual, there is a control individual with a sufficiently similar propensity score to find a match. As depicted in Figure 19, we have a strong region of common support in the full model and in sub-sample models.

Using a propensity-score matching approach, we generate relatively similar estimates to those generated via multivariate regression with adjustment (Table 10). Our PSM-based estimates suggest that exposure to restorative practices improves academic performance and reduces exposure to discipline for all students; and that effects on discipline measures are more pronounced for Hispanic and Black students, suggesting that increasing exposure to these practices could facilitate reductions in Black-White and Hispanic-White discipline disparities.

The two modeling strategies applied above may help address concerns that results are a result of either exposure misclassification or model misspecification. However, the most critical vulnerability of multivariate regression models is that they assume all relevant student and school-level variables are included in the model. To address concerns related to student and

school-level selection effects, below, I present findings from within-student and with-school models.

These models should not be considered mere sensitivity analyses. Within-student and within-school models allow for adjustment of all time-invariant student and school-level characteristics. In many cases, they are thus well positioned to glean causal effects. Because of their superior identification potential, I present these models as the best-identified estimates of the effect of exposure to restorative practices on student outcomes in this dissertation.

Figure 19

Regions of Common Support for Propensity Score Models
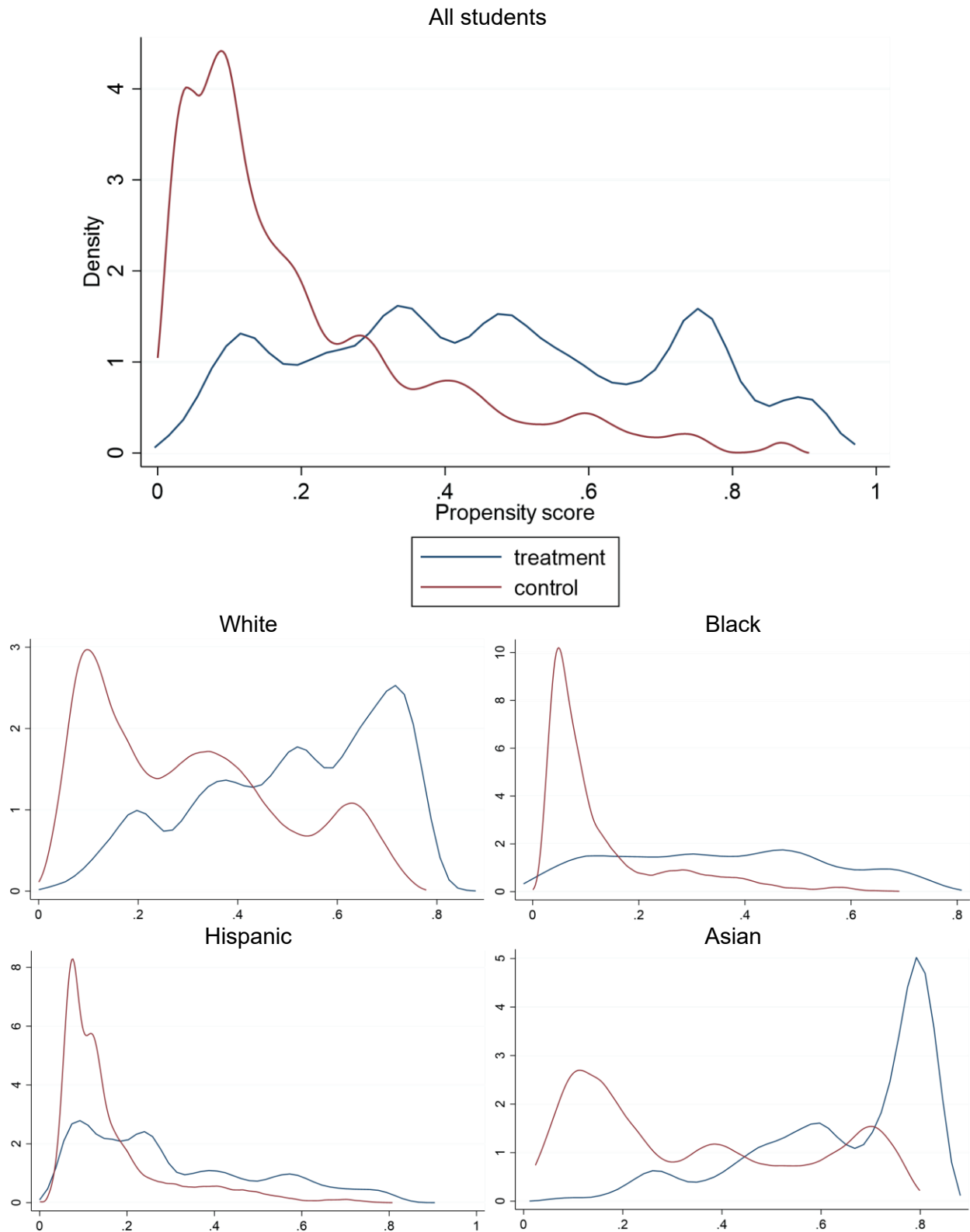
Table 10

Estimates, Based on Propensity Score Matching, of the Relationship Between 2018–2019 Outcomes and 2018–2019 Restorative Practice Exposure

	ATT (Standard Error)				
Outcome	All students	White	Black	Hispanic	Asian
ELA score	16.94***	16.01***	10.36***	4.95***	39.02***
	(0.50)	(0.81)	(3.05)	(0.77)	(1.12)
n	162,400	50,712	4,330	65,240	26,820
Math score	21.96***	16.39***	15.05***	6.25***	55.51***
	(0.59)	(0.93)	(3.38)	(0.85)	(1.34)
n	162,154	50,606	4,312	65,164	26,808
Whether suspended	−0.01***	−0.007***	−0.02*	−0.01***	−0.015***
	(0.001)	(0.002)	(0.009)	(0.002)	(0.002)
n	163,972	51,492	4,394	65,702	26,930
Days suspended	−0.04***	−0.02*	−0.07	−0.04***	−0.06***
	(0.01)	(0.01)	(0.06)	(0.01)	(0.01)
n	163,972	51,492	4,394	65,702	26,930

Note. Propensity scores are calculated based on estimated values for logistic regressions predicting exposure to restorative practices via a suite of student and school-level variables. ATT = Average Treatment Effect on the Treated.
*$p < .05$. ***$p < .001$.

Estimating Student-Level Causal Effects via Within-Student Regression

As noted previously, the regression and propensity-score-matching-based estimates presented above can only return an unbiased causal estimate of the student-level impact of exposure to restorative practices if they include *all* confounders. While I include a strong set of confounders in these models, it remains possible that certain unmeasured student characteristics predict both student exposure to restorative practices and downstream outcomes. Put another way, we may be concerned about student-level selection effects—that certain student characteristics might systemically drive certain students to attend more restorative schools. Thus, we may want to find a method that addresses what are known as time-invariant confounders—or student characteristics that do not vary over time but may nonetheless systematically drive students to attend or avoid restorative schools.

A brute force method for accounting for *all* time-invariant confounders simultaneously is differencing regression, or "within-student" regression. In differencing regression, we first calculate changes in outcome, then calculate changes in exposure, and finally ascertain whether changes in exposure are related to changes in outcome. Applying that approach to this data, we first calculate each student's change in outcome values between 2017-18 and 2018-19 ("delta

outcome"), and then calculate each student's change in restorative practice exposure during the same timeframe ("delta exposure"). We then regress delta outcome on delta exposure.

The resulting coefficient represents an estimate of the impact of changes in restorative practice exposure on changes in outcomes and is an unbiased causal estimate so long as there are no time-*varying* confounders omitted from the model.

While one can theoretically use this approach to measure the impact of changes in exposure to restorative practices occurring between any two grade levels, we may be wary about changes in restorative practice exposure that occur at "unnatural" times in a student's educational journey. For example, imagine a student who is finishing 7th grade and is currently attending a school that also offers 8th grade. If that student decides to switch schools between 7th and 8th grade, we may worry that the decision itself was motivated by a change in their feelings occasioned by exposure to school practices in use at their old school, as well as those in use at the new school. If there are many students for whom this pattern holds, then when we calculate the relationship between changes in outcomes and changes in restorative practice exposure, the detected relationship may be a function of the *kinds* of students who opt to switch schools (and therefore end up experiencing changes in restorative practice exposure) rather than a reflection of the impact that more exposure to restorative practices had on their outcomes.

One means of overcoming this issue is to restrict our analysis to students experiencing "natural" school switches. Imagine, instead, 5th grade students attending schools that do *not* offer 6th grade. They switch from one school to another in part due to natural necessity. Thus, if we measure the relationship between changes in outcomes and changes in restorative practice exposure for this subset of students, we will be less concerned that the detected relationship is a reflection of the kinds of students who make school switches as, in our case, all students will be making switches at least in part because they have to. We will thus be more persuaded that the detected relationship reflects the effect of exposure to restorative practices on student outcomes. One other benefit of focusing on the transition from 5th to 6th grade is that this transition is, research suggests, a precarious one for many students. P. J. Cook et al. (2008) have found that students who attend 6th grade in a middle school experience substantially more discipline than similarly situated students who attend 6th grade in an elementary school. One interpretation of these results is that middle schools are more disciplinarian environments than elementary schools—a finding that accords with my review of California administrative data (see Figure 4). However, another reading is that much can change for students as they enter 6th grade, and that 6th grade can reshape students' disciplinary trajectories. It is thus fitting that we analyze the impact of shifts in exposure to restorative practices as students traverse the delicate and potent transition from 5th to 6th grade.

One caveat with our within-student models is that whereas we can measure restorative practice exposure via our eight-item scale measure for students in 6th-8th grades, we cannot do so for students in 5th grades. This is because the CHKS survey for elementary school students does not include some of the eight items we used for the fuller scale. Fortunately, we *can* measure students' levels of exposure to arguably the core measure in our scale – their feelings about the extent to which their school helps students resolve conflicts. We will thus create a new measure of school-level utilization of conflict resolution practices that is based on averaging student means over appropriate timeframes in a given school. To estimate exposure to conflict resolution practices for 5th grade students in 2017-18, we will average CHKS surveys on the conflict resolution measure for all 5th grade students who attended a given school between 2014-15 and 2017-18 (a four-year period, and the largest period for which CHKS data is available). And to

estimate exposure to conflict resolution practices in 6th grade in 2018-19, we will average CHKS surveys on the measure for 6th – 8th grade students who attended a given school between 2015-16 and 2018-19 (a similarly-sized, four-year period). As before, we will limit our analyses to students who attended schools with adequate survey coverage in both periods—100 or more surveys. Results are consistent if we use other precision thresholds.

Formally, we run the following model:

$$\Delta \text{ OUTCOME}_{17\text{-}18\,/\,18\text{-}19} = \alpha + \beta_1(\Delta \text{ SRPE}_{17\text{-}18\,/\,18\text{-}19}) + \varepsilon \qquad (2)$$

Here again, OUTCOME represents our four outcome measures (smarter balanced math and English Language Arts scores; whether suspended and days of suspension), and β_1 represents our causal estimate of the relationship between changes in restorative practice exposure and changes in outcomes. As noted above, because this approach looks at variation within students over time, it "fixes" the student, and in so doing, deftly adjusts for all time-invariant student characteristics.

As with adjusted regression and propensity score matching, the differencing estimator suggested that, for students generally, increases in exposure to restorative practices (and specifically, conflict resolution practices) coincide with increases in English and math smarter balanced scores, and decreases in days suspended and the probability of experiencing a suspension (Table 11). The same was true for Black, White, and Hispanic students, with null effects for Asian students. Effects were stronger for Black and Hispanic students than for White students, again suggesting that exposure to these practices can help reduce racial disparities.

Figure 20 visually depicts the relationship between shifts in exposure to conflict resolution practices and shifts in student outcomes, for all students. As the figure illustrates, students that saw increases in their level of exposure to conflict resolution practices generally saw improvements in academic performance, and reductions in exposure to discipline. The visual demonstrates that students who saw year-on-year declines in exposure to conflict resolution practices generally saw worsening in academic and disciplinary outcomes; and students who saw year-on-year increases in exposure to conflict resolution practices generally saw improvements in these outcomes.

Table 11

Estimated Relationship Between Outcomes and Exposure to Conflict Resolution Practices via Within-Student Regression for the Transition From Fifth to Sixth Grade

	Coefficient (Robust Standard Error)				
Δ outcome	All students	White	Black	Hispanic	Asian
ELA score	13.50***	9.97***	8.85	15.67***	0.65
	(1.21)	(2.44)	(6.03)	(1.77)	(3.56)
n	34,015	8,155	1,208	17,996	3,737
Math score	22.52***	6.78**	22.63***	22.86***	2.42
	(1.19)	(2.31)	(5.96)	(1.77)	(3.29)
n	34,049	8,150	1,212	18,018	3,750
Days OSS	−0.10***	0.09	−0.54**	−0.11***	−0.02
	(0.02)	(0.06)	(0.20)	(0.03)	(0.03)
n	34,706	8,418	1,235	18,292	3,787
Received OSS	−0.029***	0.003	−0.09**	−0.031***	−0.005
	(0.004)	(0.008)	(0.03)	(0.007)	(0.008)
n	34,706	8,418	1,235	18,292	3,787

Note. Models are produced by regressing each student's shift in outcomes between 2017–2018 and 2018–2019 on that student's shift in their level of exposure to conflict resolution practices during the same time frame. Because data on the eight-item restorative practice utilization scale are only available for 7th grade, 9th grade, and 11th grade students, the models above also do not use that scale and instead focus solely on whether students were exposed to conflict resolution practices. Models are limited to students who switched schools between 2017–2018 and 2018–2019. In addition, models are limited to students who attended fifth grade in 2017–2018 at a school that did not offer sixth grade, suggesting that their school switches were more "natural" and less likely to be driven by responses to the school practices used in the school attended in 2017–2018. Finally, models are limited to students whose schools, in both 2017–2018 and 2018–2019, fielded at least 100 surveys capturing students' levels of exposure to conflict resolution practices. ELA = English language arts; OSS = out-of-school suspension.
**$p < .01$. ***$p < .001$.

Figure 20

Relationship Between Changes in Exposure to Conflict Resolution Practices and Changes in Academic and Disciplinary Outcomes for Students Transitioning From Fifth to Sixth Grade
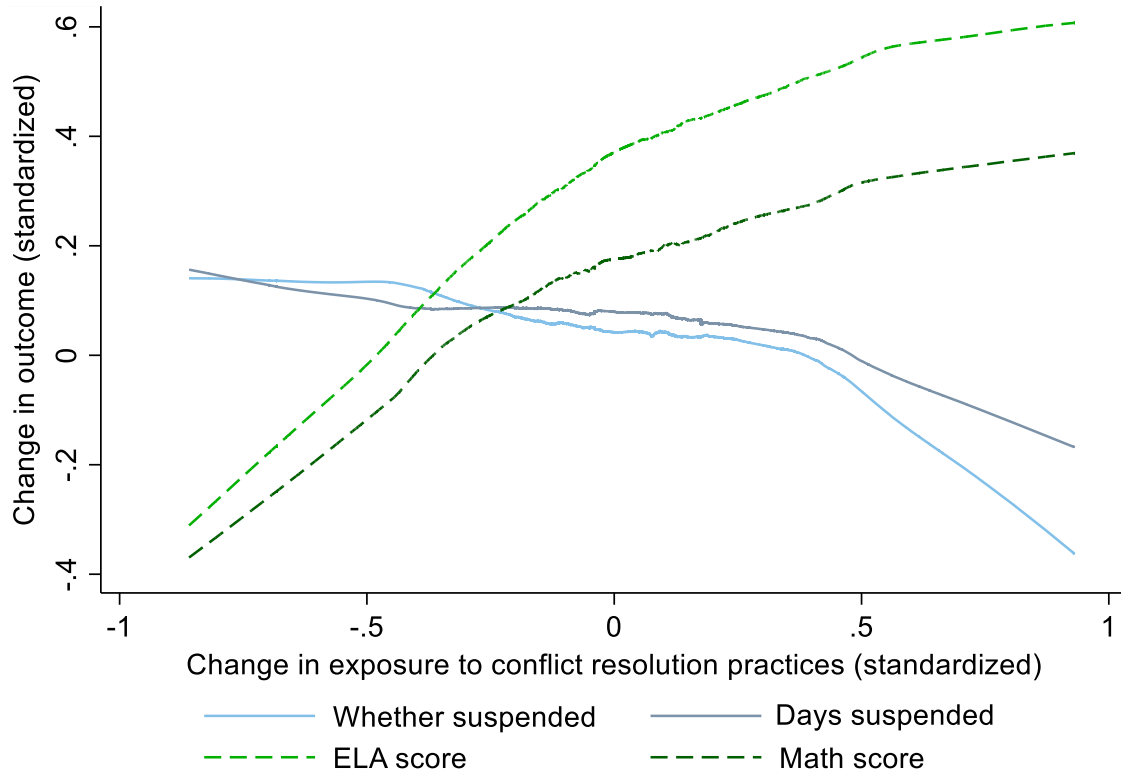
Note. Figure depicts locally weighted regressions predicting year-on-year changes in student outcomes based on year-on-year changes in student exposure to restorative practices. The measure related to year-on-year changes in conflict resolution practices is standardized by dividing by the standard deviation of the measure. The measures related to year-on-year changes in outcomes are standardized using the same approach.

Estimating School-Level Causal Effects via Within-School Regression

The within-student regression results discussed above are encouraging. They demonstrate that even when the student is held constant, shifts in restorative practice exposure predict improvements in academic and disciplinary outcomes. We can use the same approach to ask a related, and important, question: do schools that increase their level of restorative practice utilization also see improvements in academic and disciplinary dimensions?

As we shift to school-level analyses, it is worth noting that CHKS surveys ask students a range of questions related to misbehavior, victimization, school climate, mental health, attendance, and GPA. Moreover, in the same way that we can aggregate student-level restorative practice exposure scores to glean an estimate of school-level restorative practice utilization, we can pool student-level CHKS outcome scores to generate aggregate school-level outcome scores. And with sufficient data, we can estimate school-level restorative practice utilization and outcome scores *over time*, empowering various causal research methods.

We have CHKS data for six years—from the 2013-14 through 2018-19 school years. As discussed above, during this time period, many schools shifted in their use of restorative practices. To evaluate the impact of restorative practices on CHKS outcomes, we can thus use a "within-school" estimator. In this approach, we first calculate each school's change in restorative practice utilization between two, non-overlapping time waves. We then calculate their change in outcomes between the two waves. To estimate the impact of restorative practices, we calculate the *relationship* between *changes* in outcomes and *changes* in restorative practice utilization.

Here, we will fix the first wave to be the first three school years in our data: 2013-14 through 2015-16. We will fix the second wave to be the last three school years in our data: 2016-17 through 2018-19. As before, we will limit our analyses to sixth-eighth graders. This again respects the sensitivity of middle school years, but it also reflects another issue: double-counting. By limiting analyses to $6^{th} – 8^{th}$ graders, we limit the likelihood that a given student shows up in both our first wave and our second wave. To ensure our estimates of restorative practice utilization are precise for all schools in both time periods, we will limit analyses to schools with fifty or more surveys in each time wave.

The huge benefit of a "within-school" approach is that it accounts for all "time-invariant" endogeneity. By "time-invariant endogeneity," we mean stable school characteristics that systemically drive certain schools to adopt restorative practices *and* are also related to school-level outcomes of interest. These can also be termed "stable confounders." For example, given the relationship between student demographics and student exposure to restorative practices, we might expect, that school demographics (e.g., the percentage of students who are Black within a school) are related both to schools' levels of restorative practice utilization *and* to relevant school-level outcomes. In a typical regression framework, we would worry that failing to account for these stable characteristics might bias our estimate (which is why we adjusted for student demographics in our student-level models).

However, when we analyze the relationship between *changes* in restorative practice utilization and *changes* in outcomes, we fix our analyses to occur "within schools and over time." Because stable characteristics do not change over time, shifts in these characteristics over time are consistently "zero" and thus cannot be correlated with *changes* in restorative practice utilization, nor with *changes* in outcomes. With a within-school estimator, we thus no longer need worry about the possibility that stable characteristics operate as confounders. Put another way, with a "within-school" analysis, the failure to account for stable characteristics mathematically *cannot* bias our estimate (Angrist & Pischke, 2009).

However, a "within-school" estimator does not account for time-variant confounders. By time-variant confounders, we mean school characteristics that vary over time, and whose variation is correlated with variation in schools' utilization of restorative practices and with variation in schools' outcomes. Because student composition *can* change over time, and changes in student composition could theoretically be related to changes in restorative practice utilization and changes in outcomes, we may want to include terms in our regression that adjust for compositional changes schools experience over time. To operationalize this manner of within-school analysis (known as differencing regression with adjustment), our formal model is:

$$\Delta \text{ OUTCOME} = \alpha + \beta_1(\Delta \text{ RPU}) + \beta_i(\Delta \text{ X}_i) + \varepsilon \tag{3}$$

where
- Δ OUTCOME represents a given school's shift in mean outcome values between the two time points;
- Δ RPU represents a given school's shift in its level of utilization of restorative practices during the same time frame; and
- Δ X$_i$ represents a given school's shift in other school-level characteristics, specifically average racial demographics, average gender demographics, average parental education, and proportion of students receiving a free-or-reduced-price lunch.

Here again, β_1 is our measure of interest. It represents an unbiased estimate of the causal relationship between restorative practices and outcomes so long as there are no unmeasured time-variant confounders. It produces a conservative causal estimate because it detects only the impact of *changes* in restorative practice utilization (rather than the impact of differential exposure within a single time point).

Applying the data restrictions described above, we identify 220 schools for which we can identify their shift in restorative practice utilization and their shift in outcomes. As depicted in Figures 21 and 22, schools evidenced meaningful variation in terms of how much they shifted in their use of restorative practices, and how much they shifted in aggregate outcomes (e.g., the depressive symptom rate).

Figure 21

Distribution of School-Level Shifts in Restorative Practice Utilization
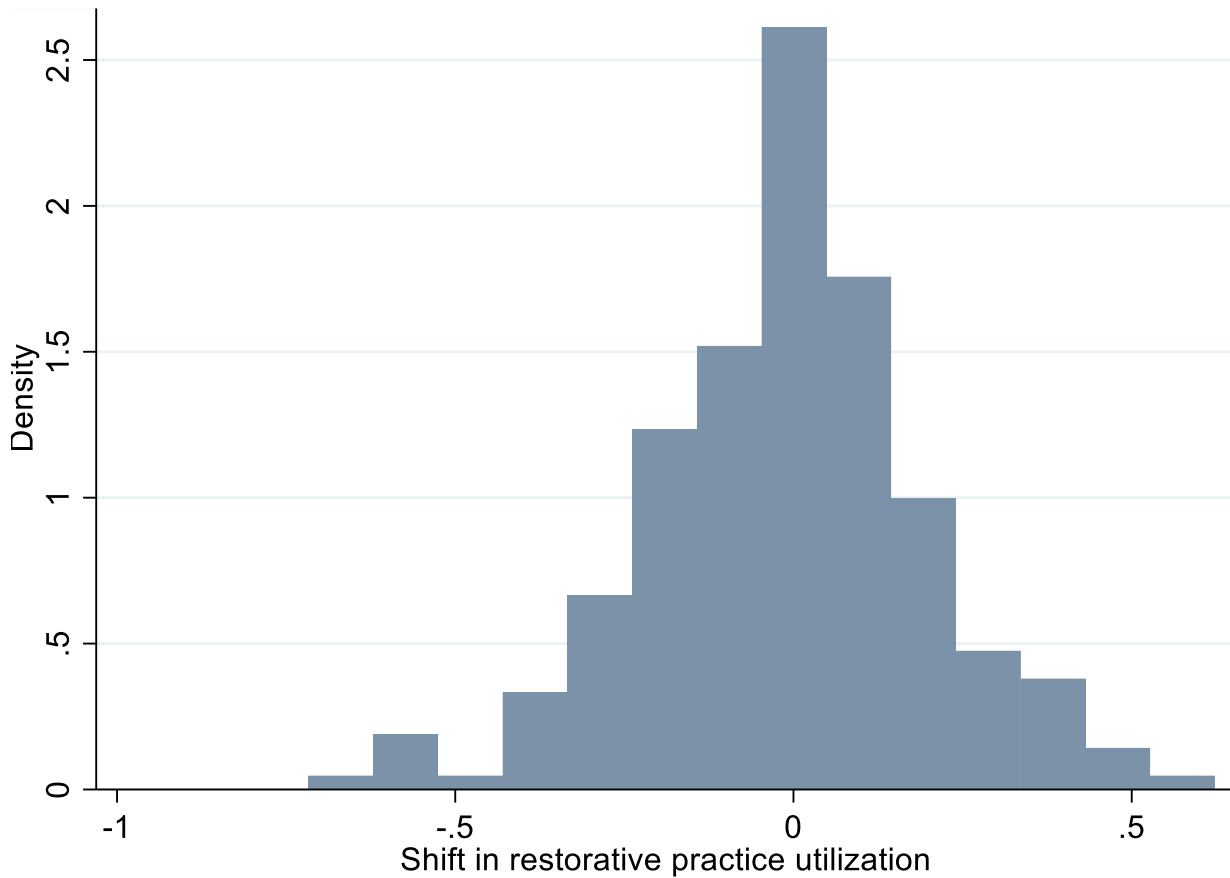
Note. For 220 schools that had adequate data (50 or more student surveys) in the first (2013–2014 through 2015–2016) and second (2016–2017 through 2018–2019) time waves to precisely identify shifts over time.

Figure 22

Distribution of School-Level Shifts in Depressive Symptom Rate
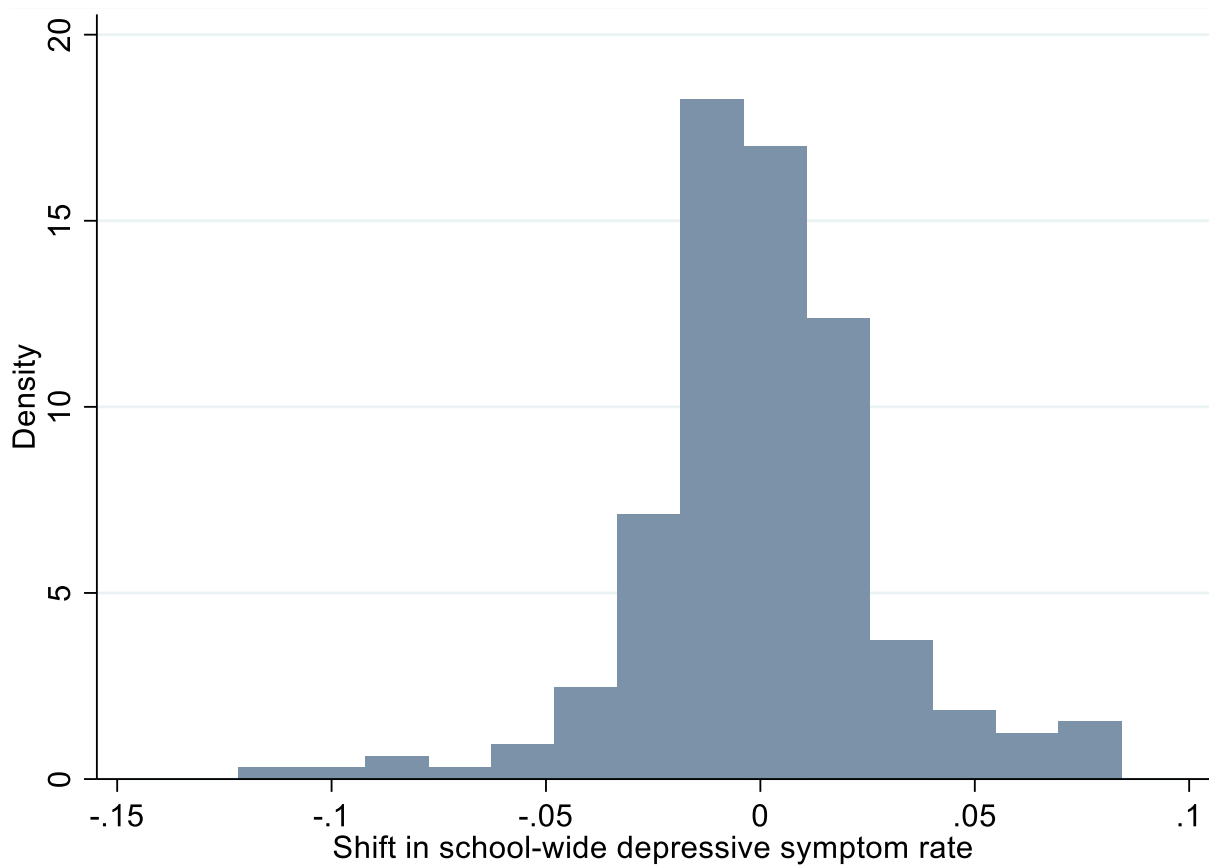
Note. For 220 schools that had adequate data (50 or more student surveys) in the first (2013–2014 through 2015–2016) and second (2016–2017 through 2018–2019) time waves to precisely identify shifts over time.

Earlier, in the section about student-level modeling, I indicated that schools evidenced considerable stability in their use of restorative practices over time. As depicted in Figure 23, 47% of schools saw their scores shift by less than one half of one standard deviation, and 78% of schools saw scores shift by less than a standard deviation. Only 3% of schools saw their scores shift by two or more standard deviations.

Figure 23

Percentage of Schools Seeing Varying Degrees of Change in Restorative Practice Utilization (in Standard Deviations)
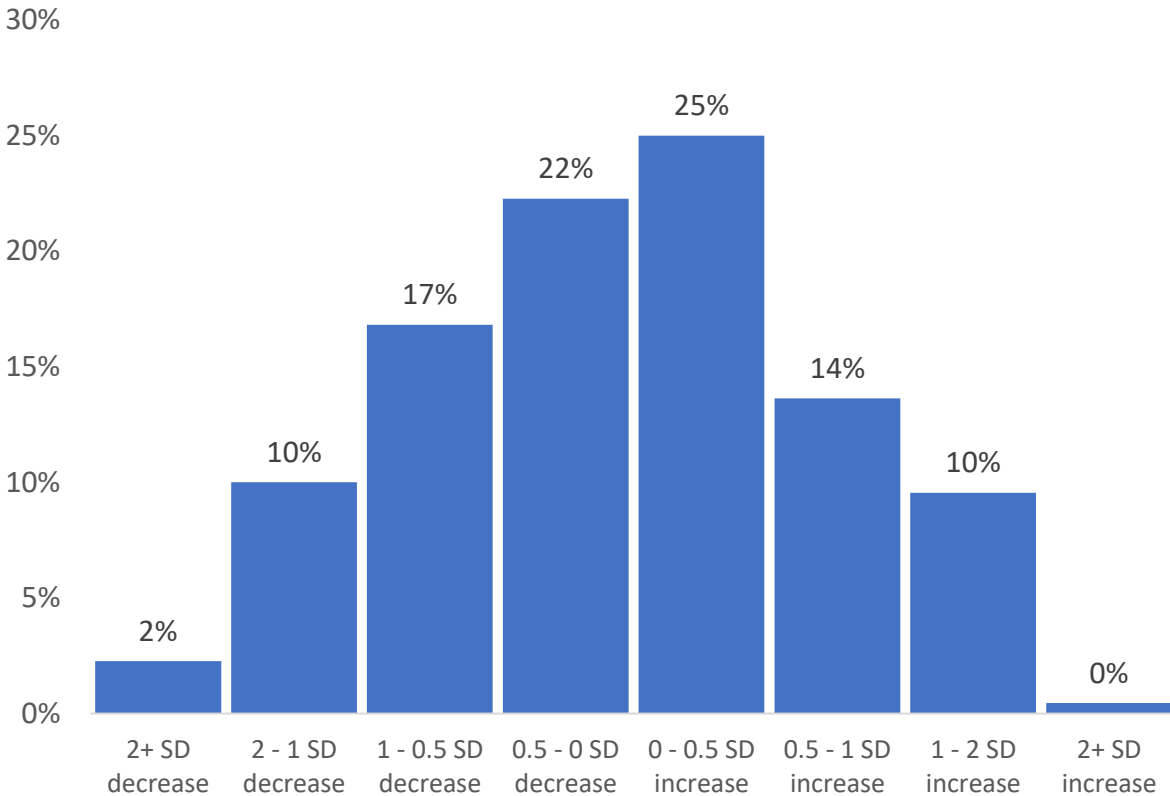
Note. Standard deviations were calculated based on the standard deviation of schools' restorative practice utilization scores in the early time period. Schools were included in this analysis if they had 50 student surveys in both the early and late time periods. These restrictions ensured precise measurement of the early, late, and delta restorative practice utilization scores. The restrictions also yielded a sample of 220 schools.

This is not to say that there is *no* temporal variation in schools' utilization of restorative practices. Indeed, there was not a single school that had a score of "0" on its level of temporal variation. However, that there is generally a small amount of temporal variation indicates that using "within-school" modeling strategies may yield particularly conservative estimates of the impact of restorative practice utilization on school-wide outcomes.

That said, as depicted in Figure 24, when we regress changes in school-wide outcomes on changes in school-wide restorative practice utilization, we find strong evidence that restorative practices drive positive shifts across all measures, and results hold even after we adjust for compositional changes in schools. For example, in adjusted models, a one standard deviation increase in restorative practice utilization was related to a 0.45 standard deviation decrease in misbehavior, a 0.46 standard deviation decrease in victimization, a 0.39 standard deviation decrease in depressive symptoms, and a 0.74 standard deviation increase in school climate.

Figure 24

Relationship Between School-Level Changes in Restorative Practice Utilization and School-Level Changes in Various Outcomes
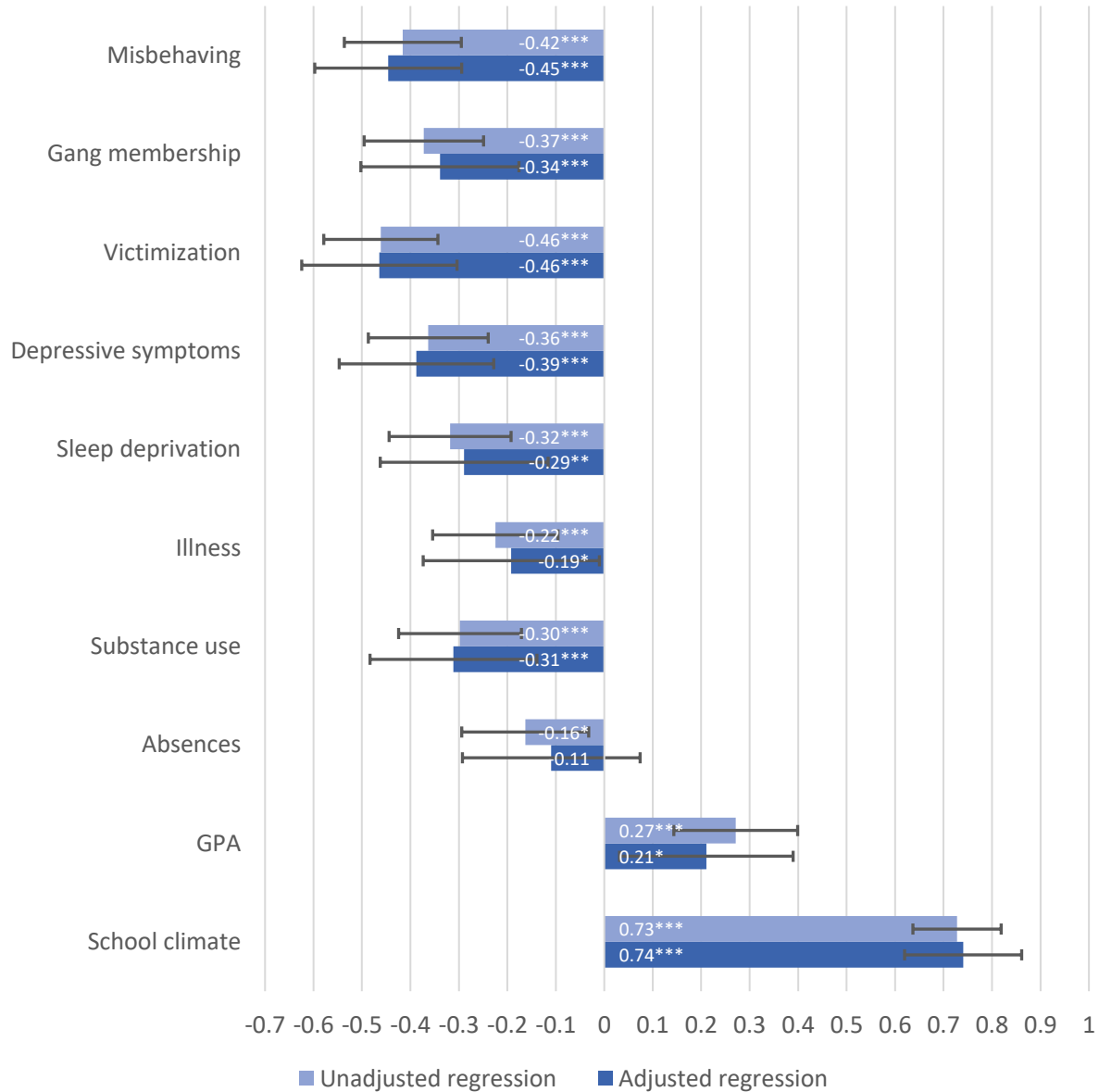
Note. Figure depicts standardized regression coefficients. So, for example, we see that an increase of 1 standard deviation in restorative practice utilization was related to a decrease of 0.42 standard deviations in school-wide misbehavior in unadjusted models, and it was related to a decrease of 0.42 standard deviations in adjusted models. Adjusted models include terms for composition shifts in student demographics (race, ethnicity, gender, free- and reduced price-lunch status, and parental education).
$^*p < .05.$ $^{**}p < .01.$ $^{***}p < .001.$

Figure 25 visually depicts the relationship between shifts in restorative practice utilization and school climate. As the figure illustrates, schools that increased their restorative practice utilization generally saw improvements in school climate. Equally importantly, schools that *reduced* their utilization of restorative practices generally saw school climate diminish.

Figure 25

Relationship Between Shifts in Restorative Practice Utilization and Shifts in School Climate
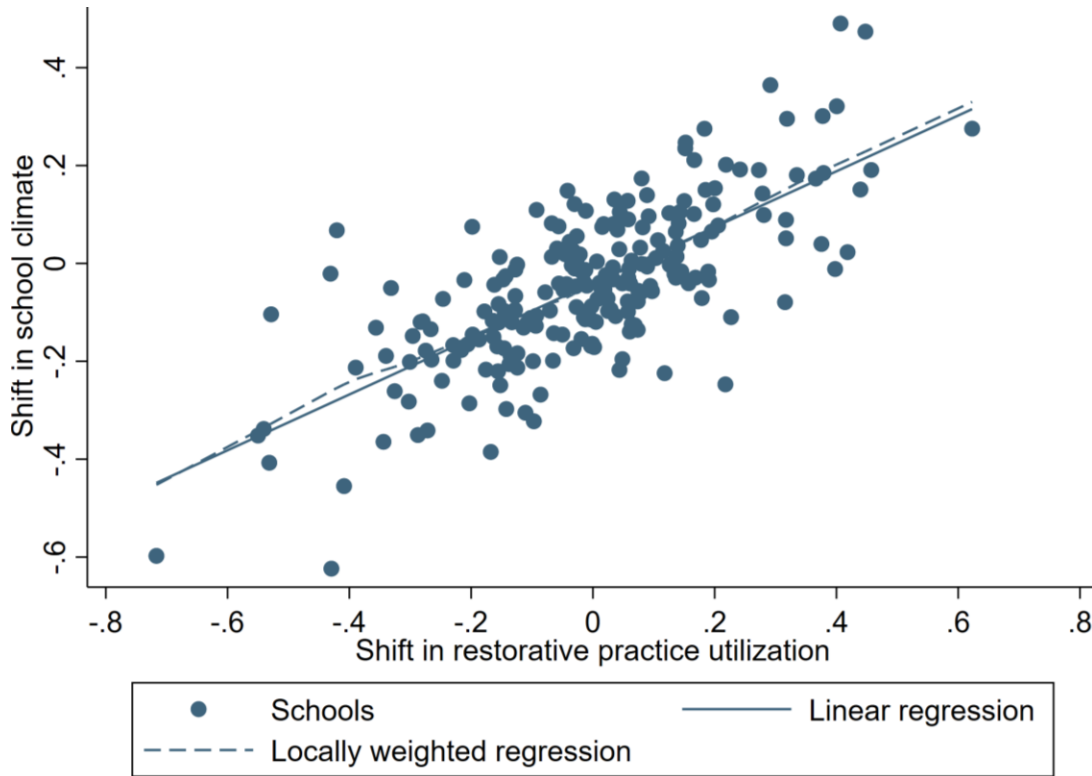

Note. Figure depicts data for 220 schools that had adequate data (50 or more student surveys) in the first (2013–2014 through 2015–2016) and second (2016–2017 through 2018–2019) time waves to precisely identify shifts over time.

Research Limitations

This research aimed to identify the relationship between exposure to restorative practices and student outcomes. To achieve this goal, we developed a consistent definition of restorative practice exposure, measured school-level restorative practice utilization over time, and merged our school practice data with longitudinal student outcome data. We then recruited models designed to overcome student and school-level selection effects. While the results presented here are encouraging, readers should bear in mind the following research limitations.

First, even if one grants that our models can glean causal effects, they are only able to glean the effect of student exposure to the specific restorative practices included in our scale

measure. Notably lacking from our scale is a direct measure of whether staff are engaging in various kinds of restorative circles (e.g., community building, harm repair, and reintroduction). Future work could seek to develop and field surveys designed to ascertain student exposure to more specifically defined restorative practices, and identify the impacts of exposure to these practices.

A second limitation relates to the timing of CHKS data collection. As noted, we leverage CHKS data to identify schools' levels of restorative practice utilization over time. However, CHKS data is collected relatively infrequently (biannually in most cases), and only for students in certain grades (mainly 7th, 9th, and 11th grades). As such, the data cannot be leveraged to understand, with temporal granularity, how or when schools shift their practices. As such, many sophisticated modeling approaches (such as event study designs) are impractical or inappropriate for this data. Future research could seek to identify data that can track school practices over time with sufficient granularity to empower models like event studies. Future research could also seek to ascertain whether the findings discussed above extend to other grade levels, such as elementary and high school grades.

Third, our models may omit important confounders related to student or school selection effects, and our model cannot rule out the possibility of reverse causality. An ideal model would exploit exogenous variation in student sorting or school practice adoption to evaluate whether students who randomly gain access to restorative practices subsequently see improved outcomes relative to students who randomly do *not* get restorative practice exposure. Our models do not exploit exogenous variation, but instead seek to adjust for all confounders. Notably, our multivariate regression and PSM models include a strong set of control variables (including lagged outcome variables, student characteristics, and school characteristics); and our within-student and within-school models adjust for all time-invariant confounders at either level. Still, it is entirely possible that the failure to include unmeasured, time-variant confounders has resulted in the detection of a spurious relationship between restorative practice exposure and outcomes. For example, if students who sort into restorative schools are experiencing faster developmental and cognitive gains than students who sort into non-restorative schools (for example, due to lower levels of exposure to systematic disadvantage), then our within-student models will be biased. Similarly, if schools that adopt restorative practices tend to experience other transitions other than shifts in their student body demographics (such as the adoption of other policies, or increases in school funding), then our within-school models will be biased. While we cannot rule out the possibility that the results presented here reflect spurious relationships to unmeasured time-variant student or school-level factors, the stability of these findings across within-student and within-school specifications provides reassurance that what we are detecting is signal, not noise. In addition, even if we are indeed detecting a causal relationship between restorative practice exposure and student outcomes, it is possible, given our models, that the relationship runs in reverse. While this is mathematically feasible, it does not seem likely, particularly in the case of academic achievement: why would students who see atypically large gains in standardized test performance be more likely to subsequently sort into schools that use more restorative practices? Nonetheless, to address these limitations, future work should seek to exploit exogenous variation in student-level restorative practice exposure, and school-level restorative practice adoption. As discussed above, one strategy for generating useful exogenous variation might be the implementation of a *teacher*-level randomized controlled trial.

A final limitation is that the results presented here focus on how student outcomes shift when students gain exposure to restorative practices, but do not provide guidance regarding *how*

students might gain exposure to these practices. Indeed, this work is premised on the notion that restorative programming often fails to shift restorative practices (e.g., Acosta et al., 2019). In the language of policy analysis and causal effects, these findings lack a manipulable treatment and therefore do not point to a specific policy solution.

However, this work does answer a first-order question—exposure to restorative practices is related to positive student outcomes. Having provided an answer, I hope this work will empower school leaders to sustain their investments in restorative practices so they can identify professional development that drives widespread utilization, and innovate solutions to implementation challenges. And I hope it will empower researchers to ask second-order and third-order questions that can empower educators to improve outcomes for students. For example, future research could evaluate whether other policies and practices (such as professional development related to cultural sensitivity, or increases in workforce diversity) can work synergistically with restorative practices to generate outsized benefits for students. And it could ascertain whether the effects of exposure to restorative practices depend on school context.

V. Policy Implications and Legal Strategies

The models above indicate that restorative practices can be recruited to drive a range of positive academic, disciplinary, and health impacts; and to bridge racial disparities. Accordingly, schools and school districts may seek means of both overcoming typical implementation challenges, and of accelerating and accentuating both the reach and impact of restorative practices. Below, I therefore discuss policy approaches that might enhance the potential of restorative practices to realize positive outcomes. These recommendations are designed for educational decision-makers at all institutional levels, including the Department of Education, state education agencies, local education agencies, and schools. Throughout this section, I use the word "staff" to describe all adults employed by schools who interface with children (including, for example, teachers, guidance counselors, and school resource officers).

Commit to Culture Change

Recently, when I interviewed a restorative practice trainer, I was told that "restorative practices are a lifestyle, not just a program or something you can learn from reading a binder." Many researchers and practitioners (e.g., Brown, 2017; Garnett et al., 2020; Thorsborne et al., 2019) argue similarly that to ensure these practices realize intended impacts, schools must commit to a cultural transformation, shifting community members' views about the sources of misbehavior, the effects of punishment, the potential of relationships to improve with effort, the feasibility of inclusive communities, and even the morality of restitution as an alternative to exclusion. This transformation may be difficult if schools maintain their adherence to punitive regimes while attempting to implement restorative practices. In one study, Sadler (2021) concluded that the cultural incongruence between a punitive discipline regime and a new restorative program led to deleterious outcomes for Black students in one school.

We thus recommend that educational institutions seeking to implement RP encourage schools and community members to make real shifts in their discipline frameworks (for example, abating the use of punitive discipline) and provide a context that enables staff to shift their teaching philosophies (relaxing reliance on punitive mechanisms to manage classrooms, and encouraging the use of relational approaches). This could be achieved by communicating the negative impacts of exclusionary discipline, modeling relational alternatives to managing common tricky classroom situations, demonstrating the potential of relationships to grow with effort, and providing ample support and time for staff to practice relational approaches before they face classroom conflict. Of course, teachers only have so much control over disciplinary decisions, and many schools, districts, and states rely on regulatory and statutory guidance that *requires* exclusionary discipline be recruited whenever students engage in certain conduct (i.e., they require a "zero tolerance" approach to bullying or vandalism). When applied overly broadly, this approach may be incongruous with a restorative paradigm. Thus, to empower schools to realize a restorative culture shift, leaders at all levels may want to reevaluate their disciplinary policies to ensure that exclusionary discipline is not a default when it need not be.

Another means of empowering culture change is to leverage "mindset interventions" designed to help shift teachers' attitudes about misbehavior and student-teacher relationships. As discussed above, at least two psychological studies (Okonofua et al., 2016, 2020) have evaluated teacher-facing interventions that seem able to not only encourage teachers to adopt what could be termed "restorative mindsets," but also empower teachers to respond equitably to students of different racial backgrounds when they misbehave. School leaders seeking to implement

restorative practices could leverage these psychological interventions to help empower teachers to adopt and promote a restorative culture.

Empower Perseverance

In another recent interview with a practitioner, I was told that the "work [of shifting to RP] is not for the timid. It takes time, and patience, for these practices to work." My review of RP implementation research demonstrated see that changes over time to outcomes (like academic performance) are often "u-shaped," meaning there are short-term declines followed by long-term gains (e.g., Sadler, 2021). This trajectory may indicate growing pains that must be weathered before positive impacts can be realized. Schools may be tempted to abandon RP during this early period of implementation if they fear that district funds will subside if immediate results (e.g., on discipline rates or discipline disparities) are not positive, or if they experience potentially short-lived declines in academic performance. Thus, institutions hoping to realize the positive impacts of RP should seek (or provide) funding that is structured to support multiple years of implementation. They should also consider providing clear guidance that funding is not tied to near-term results, and communicating to stakeholders that it is important to persevere through growing pains. This may require updating accountability systems which often require drastic action be taken if a school or district experiences what could be a temporary academic setback. New accountability systems could build in flexibility for schools and districts to maintain current school practices for sufficient durations to whether potential growing pains and reap downstream benefits.

Preempt Caregiver Concerns

Another major threat to RP program perseverance and effectiveness is caregiver concerns that when schools adopt RP, their children will experience negative outcomes (such as more bullying and classroom disruptions; or declines in academic performance). In an article for the International Institute of Restorative Practices, Phillips (2017) provides five tips for ensuring parent and family buy-in for restorative practices, which are: 1) host sessions to introduce families to RP; 2) provide ongoing information online; 3) invite parents to serve on RP leadership groups; 4) have students bring RP information packets home; and 5) recruit an RP consultant to facilitate communication with caregivers. In another guide on this topic, Community Organizing and Family Issues (2015) suggests that schools recruit parents to join "Parent Peace Centers," providing RP training and employing parents to conduct RP circles and to provide intensive tutoring and mentoring to students exhibiting disruptive behavior as an alternative to suspension. The key insight from both pieces is that leaders should proactively overcome parents' reservations by heading off misunderstandings; by communicating the value of restorative practices for achieving goals that are important to parents (e.g., a positive school climate, *less* student misbehavior, and social and emotional growth); and, if possible, by providing parents with opportunities to "feel" restorative practices in action by inviting them to participate in restorative activities.

Close the Program/Practice Gap

In Acosta et al.'s (2019) RCT, they notably did not find that students randomly assigned to teachers who received RP training saw certain improvements (such as less bullying).

However, they did find that students who got more exposure to RP did see these improvements. My models find, similarly, that exposure to restorative practices can drive marked improvements in students' academic, disciplinary, and health outcomes; and can bridge stubborn racial disparities. Collectively, research thus demonstrates a program/practice gap. While exposure to RPs can have huge benefits, RP programs that provide trainings to staff (e.g., teachers) may not always accrue to staff actually using (and students getting exposed to) restorative practices. So, what can be done to close the program/practice gap and ensure students get exposure to these practices?

One potential solution is to ensure staff actually want RP training when they receive it. Via sophisticated, randomized controlled trials, C. R. Cook et al. (2018) and Duong et al. (2019) find that RP training for a cohort of willing teachers not only improved student-teacher relationships, but also reduced student misbehavior and improved student attention in class. Unlike Acosta et al.'s (2019) and Augustine et al.'s (2018) RCTs (which randomized at the level of the school, and required that all educators in the treatment schools participate in multiple RP trainings), C. R. Cook et al.'s and Duong et al.'s research processes began with the recruitment of a group of teachers who had opted-in to RP training. The teams then randomized from a subset of entirely willing staff to determine which would actually receive the training. One explanation for their relatively superior results is that their research design ensured staff buy-in, and that staff buy-in is a precondition to program success. Indeed, Evans and Lester's (2013) review of implementation guidance surfaces the importance of securing staff buy-in before implementing RP, and specifically "recommended spending the necessary time for discussion and dialogue about school practices, as opposed to unilaterally deciding to implement" (p. 62). And research regarding other forms of trainings indicates the importance of voluntary participation (Gegenfurtner et al., 2016).

In this light, schools might take one of two approaches. The first is to provide RP training to staff who would like to volunteer to receive them. This can avoid the drawbacks of requiring unwilling staff to use RP. However, RP practitioners (Kidde, 2017) and researchers (González et al., 2019) have argued that a whole-school RP model (where all staff receive RP training) is more effective. Relatedly, the second option then is to prepare staff for a whole school model by proactively facilitating discussions about school practices before choosing RP (let alone implementing it) to ensure all staff feel they have chosen RP for their schools (and for themselves).

Another challenge schools face when trying to close the program/practice gap is that even when staff receive RP training and are willing to use it, they may not be able to sustain their use of restorative practices. When faced with challenging relational dynamics or classroom conflicts, they may abandon RP and revert to prior punitive practices. Research (e.g., Evans & Lester, 2013) shows that staff sometimes worry RP is "too soft" and can encourage students to misbehave. This preconception (while out of step with extant research) could lead staff to abandon RP when the going gets tough. To address these issues, educational institutions can take a two-pronged approach. First, they can help staff shift their preconceptions by presenting relatable case studies and examples showing declines in misbehavior following sustained RP implementation. Secondly, they can provide continuous professional development, coaching, and partner learning so teachers can weather the temptation to abandon RP, and can—slowly but surely—make RP their new modus operandi.

Educational institutions may also try leveraging insights from behavioral science and attempt to "nudge" the use of restorative practices (such as "community-building circles"). In

one of the more famous nudge experiments, Ashraf et al. (2014) found that "social signaling" was a powerful driver of prosocial behavior. HIV is a major public health issue in Zimbabwe, and condoms are considered a low-cost, effective strategy. However, encouraging condom purchases is an enduring challenge. The authors recruited Zambian hairdressers to sell condoms, and randomly assigned them to various conditions. In one, hairdressers were simply encouraged to sell condoms; in two others, they were given small or large financial incentives per condom sold; and in a final condition, hairdressers were given "a 'thermometer' display, showing condom sales and stamps on it, one star for each sale" (Ashraf et al., 2014, p. 3). The thermometer provided hairdressers with a mechanism for signaling to customers and peers that they were committed to doing their part to slow and stop the spread of HIV. The hairdressers in the thermometer condition sold more than twice as many condoms as those in any other condition. Schools could provide thermometer displays to be placed outside of teachers' doors where they could indicate the number of community building circles they have held.

Expand Access

Our models (as those of Payne & Welch, 2015) indicate that RP exposure is lower for Black students and low-income students (two groups that are particularly at risk of exposure to exclusionary discipline). To ensure RP realizes its intended impacts, schools should take steps to ensure students of all backgrounds (and particularly those most often subjected to harsh discipline) are not only exposed to RP, but experience RP in a manner that deepens their connection to the school. Kervick et al. (2019) argue that a critical step in achieving widespread and productive exposure to RP is to ensure teachers receive training in equity literacy, critical consciousness, bias awareness classrooms, and culturally responsive teaching. These practices can help teachers identify and overcome structural barriers that discourage certain groups from participating in RP activities, and communicate with students of varied backgrounds in ways that make them want to engage in RP activities.

A critical driver of the extent to which teachers use RP, and the quality of implementation, is the strength of their relationships with students (Brown, 2017). Evidence from psychology (e.g., Okonofua & Eberhardt, 2015; Okonofua et al., 2020) shows that, absent training, teachers may be more likely to label Black students as "troublemakers," and may be less likely to form positive relationships with Black students. To ensure teachers are able to leverage RP in their interactions with Black students, schools may want to help stem this "troublemaker labeling" process and empower teachers to improve relationships with Black students. On this topic, Okonofua et al. (2020) report on a randomized controlled trial finding that teachers who were encouraged to view student-teacher relationships as capable of improving over time, encouraged to view students as being capable of growing in their social and emotional skills, and provided with opportunities to hear students' perspectives showed smaller Black-White disparities in their disciplinary responses to misbehavior. Interventions akin to that implemented by Okonofua et al. (2020) may thus be a powerful tool for ensuring teachers can form positive relationships with, and subsequently leverage RP when interacting with, students of all backgrounds.

Discipline disparities also impact students with learning differences, and educational institutions may therefore seek to ensure teachers leverage RP when interacting with special education students. Kervick et al. (2019) provide the following guidance for schools hoping to overcome accessibility issues, and ensure RP reaches students with learning differences:

> [M]any common RP essential practices, such as sitting quietly in a circle, taking turns, and using perspective taking and affective statements, must be presented in optioned ways for all students to be able to participate in a restorative classroom. For example, circle facilitators could represent the circle prompts in multiple formats (projecting the circle prompts on a screen so all students have a visual prompt rather than delivering the question only in an auditory format). Circle facilitators could consider the size of the circle itself and consider flexible grouping to maximize student engagement and limit the amount of time needed to wait one's turn. Circle facilitators might also provide response options so that students with communication challenges can still respond to the prompts. For example, framing a question so that students can respond gesturally (e.g., thumb up, thumb down) ensures that all students can participate regardless of language ability. (p. 601)

A final strategy for ensuring equitable access to restorative practices is to aim accountability paradigms towards a goal of equitable access. Under the federal Every Student Succeeds Act (2015), in order for states to retain Title I funding, they must set "ambitious," "long term" goals for students and subgroups of students. To demonstrate progress towards these goals, states' accountability systems must annually measure at least one measure of school quality or student success, which can include measures of school climate. And states must use these systems to identify, and inform districts about, schools deemed in need of "comprehensive support" due to their low performance on state accountability systems. Once a school is identified as needing comprehensive support, the district and school must take steps to help the school improve. Taken together, accountability paradigms under ESSA allow states to identify and support struggling districts and schools, and provide clear goals and benchmarks for districts and schools to aim for. States have substantial leeway in determining their goals for student subgroups, and for setting their measures of school quality. School leaders seeking equitable access to restorative justice could therefore establish that one of their ambitious school climate goals is that student subgroups have equitable access to restorative practices, as measured by student surveys that allow for disaggregation of data regarding restorative practice exposure. If collected as part of accountability frameworks, data regarding differential access to restorative practices could help leaders identify districts and schools in need of support or encouragement to realize equity.

Leverage RP in Unprecedented Times

While this chapter is largely written without reference to the COVID-19 pandemic, here, I want to acknowledge the elephant in the room. At least in the 2021-22 school year, school has not felt "normal." With due care, RP can be recruited in these unprecedented times.

A defining characteristic of school in the age of COVID-19 is the use of online learning platforms and the shift to remote instruction. How can schools continue to leverage RP in a remote framework? Das et al. (2019) discuss the potential of "virtualized" RP as a tool for reducing cyberbullying and creating a more inclusive and engaging online learning for k-12 students. They conceptualize restorative coordinators creating "virtual peace rooms" when a conflict arises. Facilitators could virtually invite relevant students to the room to help them address any conflicts and repair relationships. They note that because conflict often surrounds students' use of social media, students should be able to add content from popular social media platforms to the peace rooms. And they argue that virtual RP would provide students with new

and exciting ways to become active participants in enhancing their school climate, such as participating in collective moderation and curation.

Another uncertainty in these times is how best to support students as they return to school after lengthy closures. Students may feel more anxiety than usual during school reopenings for at least three reasons. First, they may be worried about contracting the COVID-19 virus, or spreading it to medically vulnerable family or community members. Second, they may be anxious about interacting socially after a long hiatus. Third, they may feel growing unease due to recent social conflicts and tragedies, such as Black Lives Matter protests in response to the killing of unarmed Black individuals; a deadly riot at the US Capitol; severe disasters (such as floods and wildfires) that have been linked to global warming; culture wars about the very existence of the COVID-19 and the kinds of text books that should be in schools; and global conflicts such as the war in Ukraine. Community building circles are tailor-made to provide students with opportunities to share feelings about these big issues while learning (through teacher guidance) how to empathize with, and reassure, one another. Educational institutions that have already trained staff in RP may therefore want to strongly encourage teachers to use proactive RP approaches; and institutions that have not provided RP training to staff may, time permitting, want to provide training for staff before the next school year begins.

Guide School Resource Officers Through Training, Trust, and Transformation

According to the most recent data available (Diliberti et al., 2019), 58% of schools employ an on-site police officer, and 45% specifically employ a school resource officer (SRO). While schools ostensibly employ SROs to enhance student safety, research regarding the impacts of SROs has been discouraging. Of particular importance to this dissertation, Fisher and Hennessy (2016) find that SRO presence is associated with more exclusionary discipline; and Finn and Servoss (2014) find that schools that employ more security measures (such as employing SROs) also have larger Black-White discipline disparities. Relatedly, research indicates that Black students have more negative views of school police than their peers (Nakamoto et al., 2019). And qualitative research by Fisher et al. (2022) suggests that SROs may sometimes perceive Black, but not White, students as threats.

Given these challenges, can schools employ SROs while successfully implementing RP? Rosiak (2021) provides useful guidance on this front, arguing that SROs can indeed leverage RP, and should. The author indicates that for SROs to utilize RP successfully, educational institutions should ensure SRO buy-in by involving them in pre-implementation discussions and by helping SROs understand that a restorative regime actually requires more accountability for misbehavior than an exclusionary regime as only in the former, misbehaving students must take steps to "make things right." They also recommend providing universal RP training for SROs; and taking steps to build student trust in SROs' abilities to facilitate relational repair. On this latter point, they quote Keith Hickman, Executive Director of Collective Impact at the International Institute of Restorative Practices, as saying that the essential ingredient of good implementation of RP among SROs is "strong hiring practices that look at the officer's disposition, competencies, and skills" (Rosiak, 2021, p. 17).

To this list, I suggest the addition of one more critical point to ensure SROs can implement RP. Trainings should help SROs transform their view of their job. While, traditionally, SROs address student safety in large party by identifying and responding to "dangerous" students, in a restorative regime, SROs must shift their philosophical orientation and expand their work to include nurturing and repairing relationships to proactively enhance

student safety. If SROs are unable to make this cognitive shift, RP implementation may prove at least challenging and potentially damaging, not just for SROs, but also for the students and school communities they serve. A final consideration is that schools that employ SROs and hope to implement RP should also be mindful of the impact SRO activities can have on school culture. As noted previously, practitioners often argue that RP is most successful when a school has created a restorative culture characterized by trust and respect. When SROs respond to everyday incidents of misbehavior, it may sap students' sense of trust and make them feel disrespected. Thus, educational institutions implementing RP may want to limit SRO functions to proactive community building, and reactions to severe incidents of violence or threat.

Conclusion

Thousands of schools have taken on the brave work of implementing restorative programing. If my review of research and practitioner guidance has demonstrated anything, it is that while many schools face common challenges and roadblocks, each school's implementation journey is unique. Thus, the recommendations above are by no means a panacea. Instead, the suite of solutions any given school must identify is unique to the set of challenges that school faces. Nonetheless, my hope is that many of these recommendations will prove useful in *avoiding* common pitfalls, and in catalyzing iterative processes designed to identify and improve solutions.

Culture change is deliberative work. It can be daunting and can seem Sisyphean. But, as Nelson Mandela once said, "It always seems impossible until it is done." I hope that these recommendations, paired with the results of the analyses in the prior sections, will be like "wind in the sails" for administrators, schools, students, caregivers, and communities navigating their unique path to creating a truly restorative community.

References

Acosta, J., Chinman, M., Ebener, P., Malone, P. S., Phillips, A., & Wilks, A. (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*, 876–890. https://doi.org/10.1007/s10964-019-01013-2

Adams, J. M. (2014, Month DD). Superintendents surveyed about suspensions. *EdSource*. https://edsource.org/2014/superintendents-surveyed-about-suspensions/65874

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist*, *49*(1), 15–24. https://doi.org/10.1037/0003-066X.49.1.15

Aizer, A., & Doyle, J. J., Jr. (2015). Juvenile incarceration, human capital, and future crime. Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, *130*(2), 759–803. https://doi.org/10.1093/qje/qjv003

Alfred, R. (2013). *Stemming the school-to-prison pipeline: Applying restorative justice principles to school discipline practices*. Department of Education. https://safesupportivelearning.ed.gov/sites/default/files/sssta/20130321_SSDWebinar4RestorativeJusticePresentation.pdf

Alvarado, S. E. (2020). The complexities of race and place: Childhood neighborhood disadvantage and adult incarceration for Whites, Blacks, and Latinos. *Socius*. https://doi.org/10.1177/2378023120927154

American Psychological Association Zero Tolerance Task Force. (2008). Are zero tolerance policies effective in schools? An evidentiary review and recommendations. *American Psychologist*, *63*(9), 852–862. https://doi.org/10.1037/0003-066X.63.9.852

Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Anyon, Y., Gregory, A., Stone, S., Farrar, J., Jenson, J. M., McQueen, J., Downing, B., Greer, E., & Simmons, J. (2016). Restorative interventions and school discipline sanctions in a large urban school district. *American Educational Research Journal*, *53*(6), 1663–1697. https://doi.org/10.3102/0002831216675719

Armour, M. (2014). *Ed White Middle School restorative discipline evaluation: Implementation and impact, 2012/2013 sixth & seventh grade*. The Institute for Restorative Justice and Restorative Dialogue. http://sites.utexas.edu/irjrd/files/2016/01/Year-2-Final-EW-Report.pdf

Ashraf, N., Bandiera, O., & Jack, K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics 120*, 1-17. https://doi.org/10.1016/j.jpubeco.2014.06.014

Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., & Joseph, A. A. (2018). *Can restorative practices improve school climate and curb suspensions? An evaluation of the impact of restorative practices in a mid-sized urban school district*. RAND. https://www.rand.org/pubs/research_reports/RR2840.html

Bacher-Hicks, A., Billings, S., & Deming, D. (2019). *The school to prison pipeline: Long-run impacts of school suspensions on adult crime*. NBER. https://www.nber.org/system/files/working_papers/w26257/w26257.pdf

Bagley, W. C. (1914). *School discipline*. MacMillan.

Baker, M. (2009). *DPS Restorative Justice Project: Year three*. Denver Public Schools.

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

Baliga, S. (2021). Whose harm? The role of the state in restorative justice. *New Political Science, 43*(1), 35–45. https://doi.org/10.1080/07393148.2021.1880700

Barrett, N., McEachin, A., Mills, J. N., & Valant, J. (2021). Disparities and discrimination in student discipline by race and family income. *Journal of Human Resources*, *56*(3), 711–748. https://www.muse.jhu.edu/article/798142

Beckett, K., Nyrop, K., & Pfingst, L. (2006). Race, drugs, and policing: Understanding disparities in drug delivery arrests. *Criminology*, *44*(1), 105–137. https://psycnet.apa.org/doi/10.1111/j.1745-9125.2006.00044.x

Bennett, M. D. (2020). So much trouble on my mind: African American males coping with mental health issues and racism. *Urban Social Work*, *4*(2), 152–172. https://doi.org/10.1891/USW-D-19-00005

Beyers, J. M., Bates, J. E., Pettit, G. S., & Dodge, K. A. (2003). Neighborhood structure, parenting processes, and the development of youths' externalizing behaviors: a multilevel analysis. *American Journal of Community Psychology*, *31*(1–2), 35–53. https://doi.org/10.1023/a:1023018502759

Blood, P., & Thorsborne, M. (2005, Month DD–Month DD). *The challenges of culture change: Embedding restorative practice in schools* [Paper presentation]. Sixth International Conference on Conferencing, Circles and Other Restorative Practices: Building a Global Alliance for Restorative Practices and Family Empowerment, Sydney, Australia.

Brenner, P. S., & DeLamater, J. (2016). Lies, damned lies, and survey self-reports? Identity as a cause of measurement bias. *Social Psychology Quarterly*, *79*(4), 333–354. https://doi.org/10.1177/0190272516628298

Brown, M. (2017). Being heard: How a listening culture supports the implementation of schoolwide restorative practices. *Restorative Justice*, *5*(1), 53–69. https://doi.org/10.1080/20504721.2017.1294792.

Casella, R. (2003). Zero tolerance policy in schools: Rationale, consequences, and alternatives. *Teachers College Record*, *105*(5), 872–892. https://doi.org/10.1111/1467-9620.00271

Cavanagh, T., Vigil, P., & Garcia, E. (2014). A story legitimating the voices of Latino/Hispanic students and their parents: Creating a restorative justice response to wrongdoing and conflict in schools. *Equity & Excellence in Education*, *47*(4), 565–579. https://doi.org/10.1080/10665684.2014.958966

Cohen, J., McCabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, *111*(1), 180–213. https://doi.org/10.1177%2F016146810911100108

Community Organizing and Family Issues. (2015). *Parent-to-Parent Guide: Restorative justice in Chicago Public Schools*. http://www.cofionline.org/COFI/wp-content/uploads/2016/06/COFI-P2P-guide-update-2015.pdf

Cook, C. R., Coco, S., Zhang, Y., Fiat, A. E., Duong, M., Renshaw, T., Long, A. C., & Frank, S. (2018). Cultivating positive teacher–student relationships: Preliminary evaluation of the establish–maintain–restore (EMR) method. *School Psychology Review*, *47*(3), 226–243. https://doi.org/10.17105/SPR-2017-0025.V47-3

Cook, P. J., MacCoun, R., Muschkin, C., & Vigdor, J. (2008). The negative impacts of starting middle school in sixth grade. *Journal of Policy Analysis and Management*, *27*(1), 104–121. https://psycnet.apa.org/doi/10.1002/pam.20309

Daly, T., Keeling, D., Grainger, R., & Grundies, A. (2008). *Mutual benefits: New York City's shift to mutual consent in teacher hiring*. The New Teacher Project. https://tntp.org/assets/documents/MutualBenefits.pdf

Darling-Hammond, S. (2017). Designed to fail: Implicit bias in our nation's juvenile courts. *UC Davis Journal of Juvenile Law & Policy*, *21*(2), 169–193. https://sjlr.law.ucdavis.edu/archives/vol-21-no-2/JJLP-Vol21-Issue2-Darling-Hammond.pdf

Darling-Hammond, S., Fronius, T. (2022). Restorative practices in schools. In Espelage, D. (In Press). *Handbook of Classroom Management*. New York, NY: Routeledge.

Darling-Hammond, S., Fronius, T., Sutherland, H., Guckenberg, S., Petrosino, A., & Hurley, N. (2020). Effectiveness of restorative justice in US K–12 schools: A review of quantitative research. *Contemporary School Psychology*, *24*, 295–308. https://doi.org/10.1007/s40688-020-00290-0

Darling-Hammond, S., Lee, R. T., & Mendoza-Denton, R. (2021). Interracial contact at work: Does workplace diversity reduce bias? *Group Processes & Intergroup Relations*, *24*(7), 1114–1131. https://doi.org/10.1177/1368430220932636

Darling-Hammond, S., Trout, L., Fronius, T., & Cerna, R. (2021). *Can restorative practices bridge racial disparities in schools? Evidence from the California Healthy Kids Survey*. WestEd. https://www.wested.org/wp-content/uploads/2021/08/Restorative-Practices-Bridging-Racial-Disparity-Research-Brief-3.pdf

Das, A., Macbeth, J., & Elsaesser, C. (2019). Online school conflicts: expanding the scope of restorative practices with a virtual peace room. *Contemporary Justice Review, 22*(4), 351-370. https://doi.org/10.1080/10282580.2019.1672047

Davis, F. (2014). Discipline with dignity: Oakland classrooms try healing instead of punishment. *Reclaiming Children and Youth*, *23*(1), 38–41. https://eric.ed.gov/?id=EJ1038853

Davis, F. (2019). *The Little Book of Race and Restorative Justice: Black Lives, Healing, and US Social Transformation*. New York, NY: Good Books.

de Brey, C., Snyder, T. D., Zhang, A., & Dillow, S. A. (2019). *Digest of education statistics: 2019*. U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED611019.pdf

Del Toro, J., & Wang, M. (2020). School cultural socialization and academic performance: Examining ethnic-racial identity development as a mediator among African American adolescents. *Child Development*, *92*(4), 1458–1475. https://doi.org/10.1111/cdev.13467

Deming, D. J. (2011). Better schools, less crime? *The Quarterly Journal of Economics*, *126*(4), 2063–2115. https://doi.org/10.1093/qje/qjr036

Denver School-Based Restorative Practices Partnership. (2017). *School-wide restorative practices: Step by step.* https://www.skidmore.edu/campusrj/documents/Denver-2017-School-Wide-RP- Aarons, G Implementation-Guide.pdf

Department of Education. (2014). *Application for new grants under the School Climate Transformation Grant–Local Education Agency Grants Program: CFDA 84.184G.* https://www2.ed.gov/programs/schoolclimatelea/2019-184g.pdf

Department of Education & Department of Justice. (2014). *Joint "dear colleague" letter*. https://www2.ed.gov/about/offices/list/ocr/letters/colleague-201401-title-vi.html

Diliberti, M., Jackson, M., Correa, S., & Padgett, Z. (2019). *Crime, violence, discipline, and safety in U.S. public school: Findings from the School Survey on Crime and Safety: 2017–2018* (NCES Report No. 2019-061). National Center for Education Statistics. https://nces.ed.gov/pubs2019/2019061.pdf

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*(4), 315–319. https://doi.org/10.1111/1467-9280.00262

Downey, D., & Pribesh, S. (2004). When race matters: Teacher evaluations of students' classroom behavior. *Sociology of Education*, *77*(4), 267–282. https://psycnet.apa.org/doi/10.1177/003804070407700401

Dubow, E. F., Boxer, P., & Huesmann, L. R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill-Palmer Quarterly*, *55*(3), 224–249. https://doi.org/10.1353/mpq.0.0030

Duong, M. T., Pullmann, M. D., Buntain-Ricklefs, J., Lee, K., Benjamin, K. S., Nguyen, L., & Cook, C. R. (2019). Brief teacher training improves student behavior and student–teacher relationships in middle school. *School Psychology*, *34*(2), 212–221. https://doi.org/10.1037/spq0000296

Eden, M. (2020). *Two steps to restoring school safety*. American Enterprise Institute. https://www.aei.org/wp-content/uploads/2020/06/Two-Steps-to-Restoring-School-Safety.pdf?x91208

Evans, K., & Lester, J. (2013). Restorative justice in education: What we know so far. *Middle School Journal*, *44*(5), 57–63. https://doi.org/10.1080/00940771.2013.11461873

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Ewing, C. P. (2000). Sensible zero tolerance protects students. *Harvard Education Letter*, *16*(1). https://digitalcommons.law.buffalo.edu/journal_articles/504

Eyllon, M., Salhi, C., Griffith, J. L., & Lincoln, A. K. (2020). Exclusionary school discipline policies and mental health in a national sample of adolescents without histories of suspension or expulsion. *Youth & Society*, *54*(1), 1–20. https://doi.org/10.1177/0044118X20959591

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. Council of State Governments Justice Center. https://www.ojp.gov/ncjrs/virtual-library/abstracts/breaking-schools-rules-statewide-study-how-school-discipline

Fagan, J., Geller, A., Davies, G., & West, V. (2010). Street stops and broken windows revisited. In S. K. Rice & M. D. White (Eds.), *Race, ethnicity, and policing* (pp. 309–348). New York University Press.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research*, *73*(3), 321–368. https://doi.org/10.3102/00346543073003321

Finn, J. D., & Servoss, T. J. (2014). Misbehavior, suspensions, and security measures in high school: Racial/ethnic and gender differences. *Journal of Applied Research on Children*, *5*(2). https://digitalcommons.library.tmc.edu/childrenatrisk/vol5/iss2/11

Fisher, B., & Hennessy, E. (2016). School resource officers and exclusionary discipline in U.S. high schools: A systematic review and meta-analysis. *Adolescent Research Review*, *1*(3), 217–233. https://doi.org/10.1007/s40894-015-0006-8

Fisher, B., Higgins, E. M., Kupchik, A., Viano, S., Curran, F. C., Overstreet, S., Author, X. X., & Author, X. X. (2022). Protecting the flock or policing the sheep? Differences in school resource officers' perceptions of threats by school racial composition. *Social Problems*, *69*(2), 1–19. https://doi.org/10.1093/socpro/spaa062

Fowler, B., Rainbolt, S., & Mansfield, K. (2016, Month DD–Month DD). *Re-envisioning discipline in complex contexts: An appreciative inquiry of one district's implementation of restorative practices* [Paper presentation]. University Council for Educational Administration Annual Conference, Detroit, MI, United States.

Fronius, T., Darling-Hammond, S., Persson, H., Guckenburg, S., Hurley, N., & Petrosino, A. (2019). *Restorative justice in U.S. schools: An updated research review*. WestEd. https://www.wested.org/resources/restorative-justice-in-u-s-schools-an-updated-research-review/

Garnett, B., Moore, M., Kidde, J., Ballysingh, T. A., Kervick, C. T., Bedinger, L., Smith, L. C., & Sparks, H. (2020). Needs and readiness assessments for implementing school-wide restorative practices. *Improving Schools*, *23*(1), 21–32. https://doi.org/10.1177/1365480219836529

Gegenfurtner, A., Könings, K. D., Kosmajac, N., & Gebhardt, M. (2016). Voluntary or mandatory training participation as a moderator in the relationship between goal orientations and transfer of training. *International Journal of Training and Development, 20*(4), 290-301. https://doi.org/10.1111/ijtd.12089

Gerber, A. S., & Green, D. (2012). *Field experiments: Design, analysis, and interpretation.* W.W. Norton.

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions?* Yale University Child Study Center. https://medicine.yale.edu/childstudy/zigler/publications/Preschool%20Implicit%20Bias%20Policy%20Brief_final_9_26_276766_5379_v1.pdf

Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide.* Princeton University Press.

Golann, J., & Debs, M. (2019, Month DD). The harsh discipline of no-excuses charter schools: Is it worth the promise? *EducationWeek*. https://www.edweek.org/leadership/opinion-the-harsh-discipline-of-no-excuses-charter-schools-is-it-worth-the-promise/2019/06

Golann, J. W., Debs, M., & Weiss, A. L. (2019). "To be strict on your own": Black and Latinx parents evaluate discipline in urban choice schools. *American Educational Research Journal*, *56*(5), 1896–1929. https://doi.org/10.3102/0002831219831972

Goldys, P. (2016). Restorative practices: From candy and punishment to celebration and problem-solving circles. *Journal of Character Education*, *12*(1), 75–80. https://eric.ed.gov/?id=EJ1151544

González, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the school to prison pipeline. *Journal of Law and Education*, *41*(2), 281–335. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2658513

González, T. (2015). Socializing schools: Addressing racial disparities in discipline through restorative justice. In D. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 151–165). Teachers College Press.

González, T., Sattler, H., & Buth, A. J. (2019). New directions in whole-school restorative justice implementation. *Conflict Resolution, 36*(3), 207-220. https://doi.org/10.1002/crq.21236

Government Accountability Office. (2018). *Discipline disparities for Black students, boys, and students with disabilities*. https://www.gao.gov/assets/gao-18-258.pdf

Grant, A. A. , Mac Iver, D. J., & Mac Iver, M. A. (2022). The impact of restorative practices with Diplomas Now on school climate and teachers' turnover intentions: Evidence from a cluster multi-site randomized control trial. *Journal of Research on Educational Effectiveness*. https://doi.org/10.1080/19345747.2021.2018745

Gregory, A., Clawson, K., Davis, A., & Gerewitz, J. (2016). The promise of restorative practices to transform teacher-student relationships and achieve equity in school discipline. *Journal of Educational and Psychological Consultation*, *26*(4), 325–353. https://doi.org/10.1080/10474412.2014.929950

Gregory, A., & Evans, K. R. (2020). *The starts and stumbles of restorative justice in education: Where do we go from here?* National Education Policy Center. https://nepc.colorado.edu/publication/restorative-justice

Gregory, A., Huang, F. L., Anyon, Y., Greer, E., & Downing, B. (2018). An examination of restorative interventions and racial equity in out-of-school suspensions. *School Psychology Review*, *47*(2), 167–182. https://doi.org/10.17105/SPR-2017-0073.V47-2

Gregory, A., Huang, F. L., & Ward-Seidel, A. R. (2021). *Evaluation of the Whole School Restorative Practices Project: One-year implementation and impact on discipline incidents* [Technical report]. Rutgers University.

Gregory, A., Ward-Seidel, A. R., & Carter, K. V. (2021). Twelve indicators of restorative practices implementation: A framework for educational leaders. *Journal of Educational and Psychological Consultation*, *31*(2), 147–179. https://doi.org/10.1080/10474412.2020.1824788

Griffith, D., & Tyner, A. (2019). *Discipline reform through the eyes of teachers*. Thomas Fordham Institute. https://fordhaminstitute.org/sites/default/files/publication/pdfs/20190730-discipline-reform-through-eyes-teachers.pdf

Grissom, J. A., & Jones, A. (2020). Racial and ethnic diversity in the public sector workforce: Insights from public education. In A. Rutherford & J. J. Meier (Eds.), *Race and Public Administration* (pp. 16-38). Routledge.

Hashim, A., Strunk, K., & Dhaliwal, T. (2018). Justice for all? Suspension bans and restorative justice programs in the Los Angeles Unified School District. *Peabody Journal of Education*, *93*(2), 174–189. https://doi.org/10.1080/0161956X.2018.1435040

Hatzenbuehler, M. L., Keyes, K., Hamilton, A., Uddin, M., & Galea, S. (2015). The collateral damage of mass incarceration: Risk of psychiatric morbidity among nonincarcerated residents of high-incarceration neighborhoods. *American Journal of Public Health*, *105*(1), 138–143. https://doi.org/10.2105/AJPH.2014.302184

Hinze-Pifer, R., & Sartain, L. (2018). Rethinking universal suspension for severe student behavior. *Peabody Journal of Education*, *93*(2), 228–243. https://doi.org/10.1080/0161956X.2018.1435051

Hoffman, S. (2014). Zero benefit: Estimating the effect of zero tolerance discipline polices on racial disparities in school discipline. *Educational Policy*, *28*(1), 69–95. https://doi.org/10.1177/0895904812453999

Hoxby, C. (2002). The power of peers. *Education Next*, *2*(2). https://www.educationnext.org/the-power-of-peers/

Huang, F. L., & Cornell, D. G. (2017). Student attitudes and behaviors as explanations for the Black–White suspension gap. *Children and Youth Services Review*, *73*, 298–308. https://doi.org/10.1016/j.childyouth.2017.01.002

Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *The American Economic Review*, *102*(5), 2048–2082. https://doi.org/10.1257/aer.102.5.2048

Jain, S., Bassey, H., Brown, M., & Kalra, P. (2014). *Restorative justice in Oakland schools: implementation and impacts*. Oakland Unified School District. https://www.ousd.org/cms/lib/CA01001176/Centricity/Domain/134/OUSD-RJ%20Report%20revised%20Final.pdf

Johnson, J., Whitestone, E., Jackson, L., & Gatto, L. (1995). Justice is still not colorblind: Differential racial effects of exposure to inadmissible evidence. *Personality and Social Psychology Bulletin*, *21*(9), 893–898. https://doi.org/10.1177/0146167295219003

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics, 27*(2), 213-256. https://doi.org/10.1086/599334

Jordan, K. L. (2012). Juvenile transfer and recidivism: A propensity score matching approach. *Journal of Crime and Justice*, *35*(1), 53–67. https://doi.org/10.1080/0735648X.2011.632133

Kafka, J. (2011). *The history of "zero tolerance" in American public schooling*. Palgrave MacMillan.

Kalu, S. R., Menon, S. E., & Quinn, C. R. (2020). The relationship between externalizing behavior and school and familial attachments among girls from diverse backgrounds. *Children and Youth Services Review*, *116*, 105170. https://doi.org/10.1016/j.childyouth.2020.105170

Kang, J., Dasgupta, N., Yogeeswaran, K., & Blasi, G. (2009). Are ideal litigators White? Measuring the myth of colorblindness. *Journal of Empirical Legal Studies*, *7*(4), 886-915. https://www.researchgate.net/publication/228189941_Are_Ideal_Litigators_White_Measuring_the_Myth_of_Colorblindness

Karp, D., & Breslin, B. (2001). Restorative justice in school communities. *Youth and Society*, *33*(2), 249–272. https://doi.org/10.1177%2F0044118X01033002006

Kerstetter, K. (2016). A different kind of discipline: Social reproduction and the transmission of non-cognitive skills at an urban charter school. *Sociological Inquiry*, *86*(4), 512–539. https://doi.org/10.1111/soin.12128

Kervick, C. T., Moore, M., Ballysingh, T. A., Garnett, B. R., Smith, L. C. (2019). The emerging promise of restorative practices to reduce discipline disparities affecting youth with disabilities and youth of color: Addressing access and equity. *Harvard Educational Review, 89*(4), 588-610. https://doi.org/10.17763/1943-5045-89.4.588

Kidde, J. (2017). Whole-school restorative approach resource guide. *Vermont Agency of Education*. https://education.vermont.gov/sites/aoe/files/documents/edu-integrated-educational-frameworks-whole-school-restorative-approach-resource-guide_0_0.pdf

Kinsler, J. (2013). School discipline: A source or salve for the racial achievement gap? *International Economic Review*, *54*(1), 355–383. https://doi.org/10.1111/j.1468-2354.2012.00736.x

Kirk, D. S. (2008). The neighborhood context of racial and ethnic disparities in arrest. *Demography*, *45*(1), 55–77. https://dx.doi.org/10.1353%2Fdem.2008.0011

Kupchik, A., & Ward, G. (2014). Race, poverty, and exclusionary school security: An empirical analysis of U.S. elementary, middle, and high schools. *Youth Violence and Juvenile Justice*, *12*(4), 332–354. https://doi.org/10.1177/1541204013503890

Lacoe, J., & Steinberg, M. P. (2019). Do suspensions affect student outcomes? *Educational Evaluation and Policy Analysis, 41*(1), 34–62. https://doi.org/10.3102/0162373718794897

Lansing School District. (2008). *Lansing School District restorative justice annual report*. http://www.lansingschools.net/downloads/restorative_justice_files/lansing_school_district_restorative_justice_annual_report_07-08.pdf

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, *61*(5), 534–547. https://doi.org/10.1177/1470785318805305

Legewie, J., & Fagan, J. (2019). Aggressive policing and the educational performance of minority youth. *American Sociological Review*, *84*(2), 220–247. https://doi.org/10.1177/0003122419826020

Levine, C. S., Markus, H. R., Austin, M. K., Chen, E., & Miller, G. E. (2019). Students of color show health advantages when they attend schools that emphasize the value of diversity. *Proceedings of the National Academy of Sciences*, *116*(13), 6013–6018. https://doi.org/10.1073/pnas.1812068116

Lewis, S. (2009). *Improving school climate: Findings from schools implementing restorative practices*. International Institute for Restorative Practices. https://www.iirp.edu/pdf/IIRP-Improving-School-Climate-2009.pdf

LiCalsi, C., Osher, D., & Bailey, P. (2021). *The empirical examination of the effects of suspension and suspension severity on behavioral and academic outcomes*. American Institutes for Research. https://www.air.org/sites/default/files/2021-08/NYC-Suspension-Effects-Behavioral-Academic-Outcomes-August-2021.pdf

Losen, D., & Skiba, R. (2010). *Suspended education: Urban middle schools in crisis*. Civil Rights Project. https://escholarship.org/uc/item/8fh0s5dv

Loveless, T. (2013). *How well are American students learning?* Brookings. https://www.brookings.edu/wp-content/uploads/2016/06/2013-brown-center-report-web-3.pdf

Lustick, H. (2020). Culturally responsive restorative discipline. *Educational Studies*, *56*(5), 555–583. https://doi.org/10.1080/00131946.2020.1837830

Macallair, D. E. (2015). *After the doors were locked: A history of youth corrections in California and the origins of twenty-first-century reform*. Rowman & Littlefield.

Matthews, S. K., & Agnew, R. (2008). Extending deterrence theory: Do delinquent peers condition the relationship between perceptions of getting caught and offending? *Journal of Research in Crime and Delinquency*, *45*(2), 91–118. https://doi.org/10.1177/0022427807313702

McChesney, K., & Aldridge, J. (2018). The relationships between school climate and adolescent mental health and wellbeing: A systematic literature review. *International Journal of Educational Research*, *88*, 121–145. https://www.doi.org/10.1016/j.ijer.2018.01.012

McCold, P. (2002). *Evaluation of a restorative milieu: CSF Buxmont School/day treatment programs 1999–2001, evaluation outcome technical report*. International Institute of Restorative Practices. http://www.iirp.edu/pdf/erm.pdf

McCold, P. (2008). Evaluation of a restorative milieu: Restorative practices in context. In M. H. Ventura (Ed.), *Restorative Justice: From Theory to Practice* (pp. 99-137). Emerald Group Publishing Limited. https://doi.org/10.1016/S1521-6136(08)00405-3

McMorris, B. J., Beckman, K. J., Shea, G., Baumgartner, J., & Eggert, R. C. (2013). *Applying restorative justice practices to Minneapolis Public Schools students recommended for possible expulsion*. University of Minnesota. https://www.legalrightscenter.org/uploads/2/5/7/3/25735760/lrc_exec_summ-final.pdf

Meier, K. J., & Stewart Jr., J. (2003). The impact of representative bureaucracies: Educational systems and public policies. In J. Dolan & D. H. Rosenbloom (Eds.), *Representative Bureaucracy: Classic Readings and Continuing Controversies* (pp. 125-133). M.E. Sharpe.

Michels, S., Checler, A., Agish, M., & Ferriss, S. (Directors). (2016). *How zero tolerance blurred the lines between schools and criminal justice* [Film]. Retro Report. https://www.retroreport.org/video/unraveling-zero-tolerance/

Mirsky, L., & Wachtel, T. (2007). "The worst school I've ever been to": Empirical evaluations of a restorative school and treatment milieu. *Reclaiming Children and Youth*, *16*(2), 13–16. https://eric.ed.gov/?id=EJ771340

Moore, W. L., & Cooper, H. (1984). Correlations between teacher and student background and teacher perceptions of discipline problems and disciplinary techniques. *Psychology in Schools*, *21*(3), 386–392. https://doi.org/10.1002/1520-6807(198407)21:3<386::AID-PITS2310210319>3.0.CO;2-P

Nakamoto, J., Cerna, R., & Stern, A. (2019). *High school students' perceptions of police vary by student race and ethnicity: Findings from an analysis of the California Healthy Kids Survey, 2017/18*. WestEd. https://www.wested.org/wp-content/uploads/2019/05/resource-high-school-students-perceptions-of-police.pdf

National Assessment of Educational Progress. (2019). *Name of the page*. Retrieved March 1, 2022, from https://www.nationsreportcard.gov/dashboards/achievement_gaps.aspx

Nittler, K., & Gerber, N. (2018). *Teacher transfers: Finding the right fit*. National Council on Teacher Quality. https://www.nctq.org/blog/Teacher-transfers:-finding-the-right-fit

Norris, A. (2009, Month DD–Month DD). *Gender and race effects of a restorative justice intervention on school success* [Paper presentation]. American Society of Criminology Annual Conference, Philadelphia, PA, United States.

Okonofua, J., & Eberhardt, J. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. https://doi.org/10.1177/0956797615570365

Okonofua, J., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences*, *113*(19), 5221–5226. https://doi.org/10.1073/pnas.1523698113

Okonofua, J. A., Perez, A. D., & Darling-Hammond, S. (2020). When policy and psychology meet: Mitigating the consequences of bias in schools. *Science Advances*, *6*(42), eaba9479. https://doi.org/10.1126/sciadv.aba9479

Rutkowski, D., Rutkowski, L, Bélanger, J., Knoll, S., Weatherby, K., & Prusinski, E. (2013). Teaching and Learning International Survey, TALIS 2013: Conceptual Framework. *Organisation for Economic Co-operation and Development.* https://www.oecd.org/education/school/TALIS%20Conceptual%20Framework_FINAL.pdf

Owens, J., & McLanahan, S. S. (2020). Unpacking the drivers of racial disparities in school suspension and expulsion. *Social Forces*, *98*(4), 1548–1577. https://doi.org/10.1093/sf/soz095

Papageorge, N. W., Gershenson, S., & Kang, K. M. (2020). Teacher expectations matter. *The Review of Economics and Statistics*, *102*(2), 234–251. https://doi.org/10.1162/rest_a_00838

Payne, A. A., & Welch, K. (2015). Restorative justice in schools: The influence of race on restorative discipline. *Youth and Society*, *47*(4), 539–564. https://doi.org/10.1177%2F0044118X12473125

Payne, A. A., & Welch, K. (2018). The effects of school conditions on the use of restorative justice in schools. *Youth Violence and Juvenile Justice*, *16*(2), 224–240. https://doi.org/10.1177/0044118X12473125

Pearman, F. A., Curran, F. C., Fisher, B., & Gardella, J. (2019). Are achievement gaps related to discipline gaps? Evidence from national data. *AERA Open*, *5*(4), 1–18. https://doi.org/10.1177/2332858419875440

Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, *3*(3), 314–338. https://doi.org/10.1111/j.1751-9004.2009.00183.x

Perry, B. L., & Morris, E. W. (2014). Suspending progress: Collateral consequences of exclusionary punishment in public schools. *American Sociological Review*, *79*(6), 1067–1087. https://doi.org/10.1177/0003122414556308

Pesta, R. (2021). School punishment, deterrence, and race: A partial test of defiance theory. *Crime & Delinquency*, *68*(3), 463–494. https://doi.org/10.1177/00111287211005396

Phillips, R. (2017). *Five keys for gaining parent and family buy-in for restorative practices*. International Institute of Restorative Practices. https://www.iirp.edu/news/five-keys-for-gaining-parent-and-family-buy-in-for-restorative-practices

Pollack, A., Eden, M., & Pollack, H. (2019). *Why Meadow died: The people and policies that created the Parkland shooter and endanger America's students*. Post Hill Press.

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, *89*(4), 499–535. https://doi.org/10.3102/0034654319853545

Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, *116*(17), 8255–8260. https://doi.org/10.1073/pnas.1808307116

Riestenberg, N. (2003). *Restorative schools grants final report, January 2002–June 2003: A summary of the grantees' evaluation*. Minnesota Department of Education. http://crisisresponse.promoteprevent.org/webfm_send/1200

Roch, C. H., Pitts, D. W., & Navarro, I. (2010). Representative bureaucracy and policy tools: Ethnicity, student discipline, and representation in public schools. *Administration & Society*, *42*(1), 38–65. https://doi.org/10.1177/0095399709349695

Rocha, R. R., & Hawes, D. P. (2009). Racial diversity, representative bureaucracy, and equity in multiracial school districts. *Social Science Quarterly*, *90*(2), 326–344. http://www.jstor.org/stable/42940590

Rosiak, J. (2021, Summer). How school-based law enforcement can engage in restorative practices. *Journal of School Safety*, p. 14-18. https://www.mydigitalpublication.com/publication/?m=9648&i=709622&p=14&ver=html5

Sadler, K. (2021). *No-excuses in restorative justice clothing: The effects of adopting restorative justice in a no-excuse setting* [Doctoral dissertation, University of North Carolina at Chapel Hill]. Name of the Archive.

Saucier, D. A., Miller, C. T., & Doucet, N. (2005). Differences in helping Whites and Blacks: A meta-analysis. *Personality And Social Psychology Review*, *9*(1), 2–16. https://doi.org/10.1207/s15327957pspr0901_1

School of Teacher Education. (2020). *Teacher credential program handbook*. San Diego State University. https://education.sdsu.edu/documents/ste-credential-handbook-2020-2021.pdf

Shem-Tov, T., Raphael, S., & Skog, A. (2021). *Can restorative justice conferencing reduce recidivism? Evidence from the make-it-right program* (NBER Working Paper No. 29150). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w29150/w29150.pdf

Shi, Y., & Zhu, M. (2022). Equal time for equal crime? Racial bias in school discipline. *Economics of Education Review*, *88*, 102256. https://doi.org/10.1016/j.econedurev.2022.102256

Shin, Y., & Raudenbush, S. W. (2011). The causal effect of class size on academic achievement: Multivariate instrumental variable estimators with data missing at random. *Journal of Educational and Behavioral Statistics*, *36*(2), 154–185. https://doi.org/10.3102/1076998610388632

Shollenberger, T. (2015). Racial disparities in school suspension and subsequent outcomes: Evidence from the National Longitudinal Survey of Youth. In D. Losen (Ed.), *Close the school discipline gap: Equitable remedies for excessive exclusion* (pp. 31–43). Teachers College Press.

Skiba, R. J., Horner, R. H., Chung, C., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*(1), 85–107. https://doi.org/10.1080/02796015.2011.12087730

Skiba, R., Medriatta, K., & Rausch, M. K. (2016). *Inequality in school discipline*. Palgrave MacMillan.

Smith, R. J., Levinson, J. D., & Robinson, Z. (2015). Implicit White favoritism in the criminal justice system. *Alabama Law Review*, *66*(4), 871–923. https://www.semanticscholar.org/paper/Implicit-White-Favoritism-in-the-Criminal-Justice-Levinson-Smith/1490d75c1cef56b5bbc4ee5d125286b6608840b4

Smolkowski, K., Girvan, E. J., Mcintosh, K., Nese, R. N. T., & Horner, R. H. (2016). Vulnerable decision points for disproportionate office discipline referrals: Comparisons of discipline for African American and White elementary school students. *Behavioral Disorders*, *41*(4), 178–195. https://doi.org/10.17988/bedi-41-04-178-195.1

Stinchcomb, J. B., Bazemore, G., & Riestenberg, N. (2006). Beyond zero tolerance. *Youth Violence & Juvenile Justice*, *4*(2), 123–147. https://doi.org/10.1177/1541204006286287

Sullivan, A. L., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children*, *79*(4), 475–494. https://doi.org/10.1177/001440291307900406

Sumner, D., Silverman, C., & Frampton, M. (2010). *School-based restorative justice as an alternative to zero-tolerance policies: Lessons from West Oakland*. University of California, Berkeley, School of Law. https://www.law.berkeley.edu/files/thcsj/10-2010_School-based_Restorative_Justice_As_an_Alternative_to_Zero-Tolerance_Policies.pdf

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864

Terrill, S. (2018, Month DD–Month DD). *Discipline that restores: An examination of restorative justice in the school setting* [Paper presentation]. MidAmerica Nazarene University Colloquium, Olathe, KS, United States.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 2013. https://doi.org/10.3102/0034654313483907

Thorsborne, M., Riestenberg, N., & McCluskey, G. (2019). *Getting more out of restorative practice in schools: Practical approaches to improve school wellbeing and strengthen community engagement*. Jessica Kingsley Publishers.

Todic, J., Cubbin, C., Armour, M., Rountree, M., & González, T. (2020). Reframing school-based restorative justice as a structural population health intervention. *Health and Place*, *62*, 102289. https://doi.org/10.1016/j.healthplace.2020.102289

Tyler, T. (2006). Restorative justice and procedural justice: Dealing with rule breaking. *Journal of Social Issues*, *62*(2), 307–326. https://doi.org/10.1111/j.1540-4560.2006.00452.x

von Stumm, S., Smith-Woolley, E., Ayorech, Z., McMillan, A., Rimfeld, K., Dale, P. S., & Plomin, R. (2020). Predicting educational achievement from genomic measures and socioeconomic status. *Developmental Science*, *23*(3), Article e12925. https://doi.org/10.1111/desc.12925

Walker, S. C., & Herting, J. R. (2020). The impact of pretrial juvenile detention on 12-month recidivism: A matched comparison study. *Crime & Delinquency*, *66*(13–14), 1865–1887. https://doi.org/10.1177/0011128720926115

Wang, M. T., & Degol, J. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Way, S. M. (2011). School discipline and disruptive student behavior: The moderating effects of student perceptions. *The Sociological Quarterly*, *52*(3), 346–375. https://www.jstor.org/stable/23027541

Williams, J. A., III, Johnson, J. N., Dangerfield-Persky, F., & Mayakis, C. G. (2020). Does employing more novice teachers predict higher suspensions for Black students? A hierarchical multiple regression analysis. *The Journal of Negro Education*, *89*(4), 448–458. https://www.muse.jhu.edu/article/794949

Williams, J. A., III, & Wiggans, G. (2016). Models of success, teacher quality and student disciplinary infraction: A critical analysis of Chicago's Urban Preparatory Academies

and Harlem Children's Zone. *Journal of Educational Issues*, *2*(2), 73–89. http://dx.doi.org/10.5296/jei.v2i2.9788

Wright, A., Gottfried, M. A., & Le, V.-N. (2017). A kindergarten teacher like me: The role of student–teacher race. *American Educational Research Journal*, *54*(Suppl. 1), 78S–101S. https://doi.org/10.3102/0002831216635733

Xu, Y. (2022). *Where are charter schools located?* National Alliance of Public Charter Schools. https://data.publiccharters.org/digest/charter-school-data-digest/where-are-charter-schools-located/

Yao, L., & Robert, S. A. (2008). The contributions of race, individual socioeconomic status, and neighborhood socioeconomic context on the self-rated health trajectories and mortality of older adults. *Research on Aging*, *30*(2), 251–273. https://doi.org/10.1177/0164027507311155

Yusem, D., Curtis, D., Johnson, K., McClung, B., Davis, F., Kumar, S., Mayo, T., & Hysten, F. (2016). *Oakland Unified School District restorative justice implementation guide: A whole school approach*. Publisher Name. http://rjoyoakland.org/wp-content/uploads/OUSDRJOYImplementation-Guide.pdf

Zehr, H. (2002). *The little book of restorative justice*. Good Books.