Evaluation of the Whole School Restorative Practices Project: One-Year Implementation and Impact on Discipline Incidents

Anne Gregory, Ph.D., Rutgers University
Francis Huang, Ph.D., University of Missouri
Allison Rae Ward-Seidel, Ed.M., Rutgers University

Corresponding author: Anne Gregory, annegreg@gsapp.rutgers.edu

The evaluation was supported by funding from the United States Federal Department of Education Investing in Innovation (i3 Grant), through Grant Award #U411C160035 to the Morningside Center for Teaching Social Responsibility. This report is part of a larger evaluation of a Whole School Restorative Practice and Racial Equity Project.

Suggested Citation: 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. New Brunswick, NJ: Rutgers University

Contents

Exhibits

Abstract

There is a need for rigorous experimental research on restorative practices in schools as a means to reduce exclusionary discipline and the disproportionate sanctioning of Black students, male students, and students with disabilities. The current study addressed this need in its evaluation of Morningside Center for Teaching Social Responsibility's Whole School Restorative Practices (RP) Project. The study was conducted in a large Northeastern city using a cluster randomized control experimental design in 18 elementary, middle, and high schools. The RP Project was a 2-year initiative which was interrupted in its second year of implementation given the global health pandemic and the shift to remote schooling. After two years, it was anticipated that the nine RP project schools, relative to nine comparison schools, would have lower discipline incidents and narrowed discipline disparities across subgroups. Given the pandemic disruption, the effects of one year of implementation were examined despite that change was anticipated after two years. The 5,878 student sample was drawn from discipline incident record data provided by the public school district. After only one year of implementation, the study found that overall, students in the RP Project schools were less likely to receive a discipline incident record (11.1%) compared to students in the comparison schools (18.2%). In the single year of the two-year project, no differential effects of the intervention were noted based on student race, gender, or disability status. Given the project anticipated change after two years, the reduction in discipline incidents was striking. Future research will need to examine whether reducing incidents in year one of implementation sets the stage for narrowed discipline disparities in year two.

Introduction

For over a decade, educators, researchers, and advocates have been documenting the detrimental effects of exclusionary school discipline (e.g., removal from classrooms, office discipline referrals, suspension, expulsion; e.g., Fabelo et al., 2011; Losen et al., 2015). A meta-analysis of school suspension studies showed significant associations between exclusionary discipline and reduced academic achievement and heightened risk of dropping out (Noltemeyer et al., 2015). The sheer number of lost instructional days due to suspension is striking: Losen and Martinez (2020) found that over 11.3 million days of instruction were lost due to out of school suspension in the U.S. in the 2015-16 school year. Moreover, beyond instructional loss, receipt of suspension increases students' risk for dropping out of school (Balfanz, Byrnes, & Fox, 2013; Fabelo et al., 2011) and later arrest (Mittleman, 2018). Most negative discipline encounters originate in the classroom (McIntosh, Frank, & Spaulding 2010). Discipline incidents and teacher-issued office discipline referrals (ODRs) often result in students being sent out of the classroom thereby losing instructional time.

The urgency to prevent disciplinary events from occurring in the first place has been underscored by persistent disparities in school discipline (Huang, 2020; Skiba et al. 2014). Black students are suspended at double the rate (8%) of their White peers (3.8%), and students with disabilities are suspended at double the rate (8.6%) of their peers without disabilities (4.1%; U.S. Department of Education Office for Civil Rights, 2018; Harper et al., 2019). A concern is that exclusionary discipline can initiate a downward spiral whereby one discipline incident or suspension leads to another (Anyon et al., 2016). In fact, research shows that discipline sanctions in the past is one of the strongest predictors of receipt of discipline sanctions in the future (e.g., Anyon et al., 2016; Blake et al., 2020; Mittleman, 2018). Some students can enter a "school-to-prison pipeline." For example, analysis of nationwide data from nearly 3,000 15-year olds found that adolescents who were suspended during childhood were more than twice as likely to be arrested than non-suspended peers who had the same observable risk for a suspension (Mittleman, 2018).

States, districts, and schools have been aiming to prevent discipline incidents and reduce the use of exclusionary discipline (Losen & Martinez, 2020) and many are striving to eradicate discipline disparities (Ritter, 2018). Restorative practices (RP) and restorative justice initiatives have been adopted across the nation (Fronius et al., 2019). The U.S. Department of Education (2014) formally recommended RP as means to address racial disparities in school discipline. Since then, states such as Colorado have adopted statewide legislation encouraging the use of RP (Silva et al., 2019).

RP has roots in indigenous traditions of community building and repairing harm based on accountability and restoring relationships (Zehr, 2014). RP is often cited as originating in indigenous cultures, including among the Maori in New Zealand (Tauri, 2018; Zehr, 2014). The theoretical foundation of RP in education is that accountability is based on collective, fair process and relationships rather than top-down compulsory control (Evans & Vaandering, 2016; Thorsborne & Blood, 2013). Many RP practitioners and scholars note that RP is a paradigm shift away from a punitive approach and toward a supportive approach to student behavior (e.g., Berkowitz, 2012; Evans & Vaandering, 2016).

More recently there has been increased attention to RP with an explicit equity focus (Davis, 2019; Lustick, 2019; Manassah et al., 2018; Valandra, 2020; Winn, 2018). Scholars note that RP inherently draws on equitable principles through the circle process, which can elicit student voice and disrupt typical school hierarchies by equally valuing multiple perspectives (Evans & Vaandering, 2016; Zehr, 2014). Intentional, recurring time for students and staff to participate in circles may foster empathetic perspective taking (Sandwick et al., 2019) and allow opportunities for meaningful relationship building. Practicing perspective-taking, empathy, and individuating (i.e., focusing on an individual's personal, relatable characteristics rather than social group or race) may counter racial bias that contributes to the overrepresentation of students of color in discipline (Ispa-Landa, 2018). RP may also offer a format for developing a critical consciousness about systemic racism and oppression. Winn (2018) describes restorative circles as a process that can help teachers develop the language and mindset to initiate discussions about race, privilege, and bias. Moreover, RP may offer opportunities for Social Emotional Learning (SEL) including recognizing emotions, managing anger, active listening, empathy, and bias awareness which could, in turn, help prevent conflicts and reduce bias in disciplinary moments (Ispa-Landa, 2018; Jagers et al., 2019).

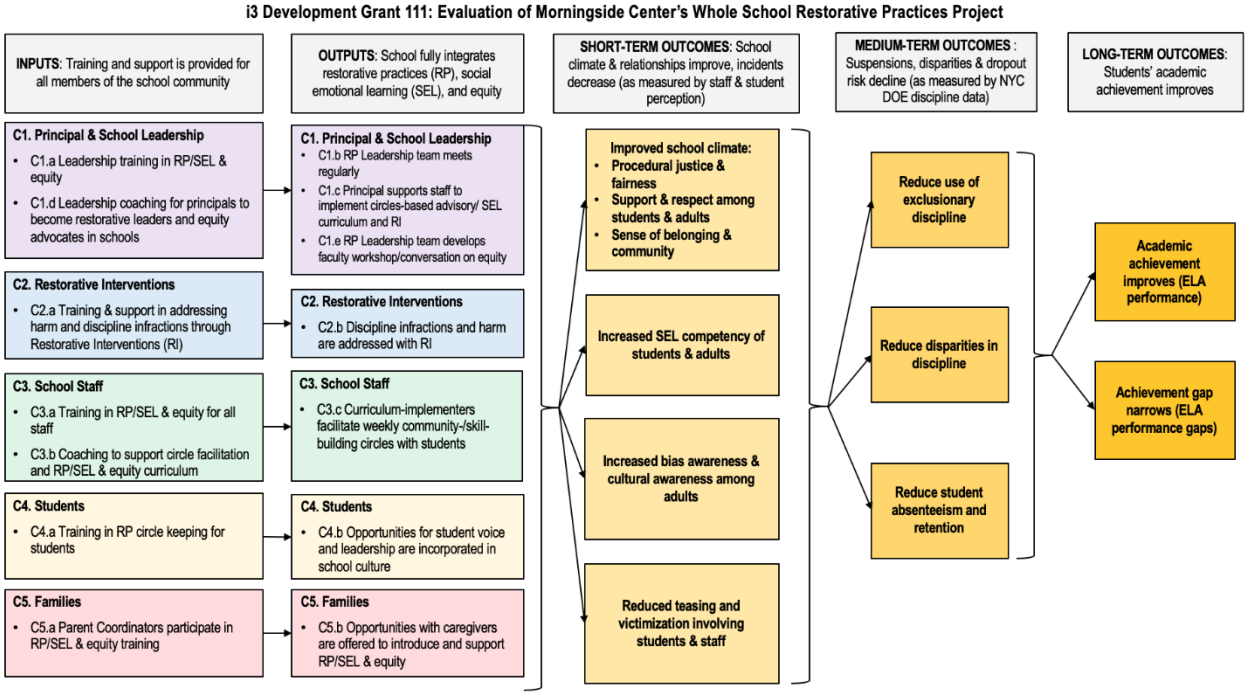
Whereas over 60 single case studies of school-based RP have shown positive outcomes associated with RP implementation (Zakszeski & Rutherford, 2021), the few U.S. randomized control trials (RCTs) of RP have had mixed results (Acosta et al., 2019; Augustine et al., 2018). The dearth of RCTs have led scholars to note that RP implementation in schools has outpaced the evidentiary base (Song & Swearer, 2016; Zakszeski & Rutherford, 2021). Moreover, RP has been conceptualized as a comprehensive intervention yet some schools neglect to explicitly integrate Social Emotional Learning (SEL) and an equity focus (Gregory & Evans, 2020). Thus, the current evaluation addresses an urgent need for an RCT that examines a robust approach to RP that integrates SEL and equity. The evaluation examines the first year implementation and impact of a schoolwide initiative in reducing exclusionary discipline and its disproportionate use with minoritized groups.

Whole School Restorative Practices Project

This study presents results from the first year of a larger evaluation of a Whole School RP Project. The project builds on a three-decade collaboration between Morningside Center for Teaching Social Responsibility and a large district located in the United States. Through a federally-funded i3 Development Grant under Absolute Priority 3 (Improving School Climate, Behavioral Supports, and Correctional Education), Morningside Center developed an innovative model of RP with a focus on racial equity and SEL. Outside evaluators from Rutgers University and the University of Missouri were subcontracted to conduct the evaluation.

The RP Project had a three-strand approach integrating RP with racial equity conversations and social emotional development for adults and students (Manassah et al., 2018). Morningside Center set forth that strengthening relationships, developing social and emotional skills, including cultural awareness, and using restorative interventions instead of punitive discipline measures could create a positive climate for learning and address the root causes of racial disparities in discipline (Exhibit 1 Logic Model).

Exhibit 1.
Logic Model for the Evaluation of Morningside Center's Whole School RP Project



The current report provides evaluation findings from Year 1 of the two-year RCT with a focus on discipline outcomes. The current study examines RP implementation and impact after one year (2018-2019) of the two-year initiative, which was interrupted given the shift to remote schooling due to the global health pandemic (2019-2020).

We address the following research questions (RQ).

RQ1.To what degree was the Whole School RP Project implemented with fidelity?

RQ2. In a single year of implementation, did the Whole School RP Project reduce discipline incidents (i.e., office discipline referrals)?[1]

RQ3. In a single year of implementation, did the Whole School RP Project reduce disparities in discipline incidents (i.e., office discipline referrals) for Black students, male students, and students in special education relative to their peers?

[1] The term "Incident" is used throughout the report to align with the District's terminology. In some states and districts, discipline incidents are referred to as Office Discipline Referrals or ODRs.

Research Design

To estimate the impact of the RP initiative on student discipline (i.e., involvement in one or more discipline incidents), we conducted a cluster randomized control trial (RCT) where schools were randomly assigned to be in either the treatment/intervention or the control/business-as-usual condition. Sample selection and statistical techniques to answer RQ2 and RQ3 were guided by a design document preregistered by the project evaluation team in June 2018 with Abt Associates, the i3 evaluation technical assistance provider. As required by the funder, preregistration was done prior to randomizing schools. The statistical analytic techniques to answer RQ2 and RQ3 were pre-established to increase the credibility of findings. The current study reports the intent to treat (ITT) impact estimates of being assigned to receive the RP initiative for the first year of this two-year study. The Institutional Review Boards of Rutgers University, University of Missouri, and the District's Department of Education approved the study protocol. All demographic and disciplinary data were provided by the District's Department of Education in a de-identified manner.

Random Assignment Procedure

Eighteen schools (i.e., six elementary schools, six middle schools, four high schools, and two combined schools with both middle and high school students) were recruited to participate in one large district. The district serves around 47% Black/African American and 42% Latinx students who reside in families and neighborhoods that are challenged by enduring health, wealth, and employment disparities relative to neighborhoods in the surrounding areas – disparities that have been attributed to histories of racism and discrimination (often referred to as structural racism).

The evaluators and Morningside Center staff met with all 18 principals to discuss the project and the randomization procedures. All 18 principals agreed to participate. As the RP program was a whole-school intervention, randomization was done at the group level with all members in the school assigned to the same condition. Schools were blocked by grade configuration (i.e., elementary, middle, combined, and high) and percentage of enrolled Black students (for high schools and elementary schools only, which had varying levels of Black student enrollment) before being randomized by the outside evaluation team. R syntax was used to randomize schools within each block to treatment or control conditions. We recorded the randomization process on video which was made available to schools and the Morningside Center to demonstrate the integrity of the procedure. Within each block, each school had an equal probability of assignment to each condition.

There were equal numbers of schools assigned to treatment and control conditions across blocks. Based on the randomization, nine schools being offered the treatment or intervention and nine schools were assigned to the "business-as-usual" or control condition, where schools continued their own professional development and programming. There were no post-random assignment exclusions from the sample. Baseline equivalence analyses demonstrated that intervention and control schools had no statistically significant differences in terms percent of Black students enrolled, percent eligible for free-reduced priced meals, and enrollment size.

Business-as-usual schools were provided $5,000 over a two-year period and could avail of the intervention services (i.e., training, coaching, curriculum) after study completion at a discounted rate. Business-as-usual schools could continue their professional development in the areas of school discipline. Yet, they did not have access to an outside-funded coordinator (1-3 days a week for nine months) who integrated and supported individual level and systems change via a comprehensive, schoolwide RP initiative with SEL and equity foci.

Intervention Condition

Morningside Center in their Whole School RP Project integrates RP with SEL and racial equity efforts to foster systemic change (policy and practice shifts) and individual change (educator and student intrapersonal and interpersonal change). Broadly, the RP Project is comprised of five components (see Exhibit 2), which include professional development and coaching with school leaders and staff so that they have a greater capacity to foster equitable, relationship-focused, and restorative school environments. Morningside Center approaches SEL and equity as a set of attitudes, practices, and policies that are fully integrated into the life and culture of the school, rather than as add-on "programs." They define SEL as the process by which individuals develop capacity to 1) understand and manage feelings; 2) relate well to others; 3) deal well with conflict and other life challenges; 4) make good decisions; and 5) take responsibility for improving communities (Roderick, 2018).

The nine intervention schools were matched with a Morningside Center Staff Developer (herein called "RP Coordinator" for comparability across other U.S. RP initiatives) and an RP/equity principal coach. The eight RP Coordinators were experienced Morningside Center trainers (M = 11.5 years), with a diverse set of professional backgrounds (e.g., classroom teacher, community organizer) and racial/ethnic backgrounds (four identify as Black and four identify as White). RP Coordinators worked in the intervention schools for 1-3 full days per week for nine months in the 2018–2019 school year. With the aim of building capacity through training programs and experiential learning, RP Coordinators sought to provide opportunities for community-building, critical self-reflection on racial equity practices, restoring harm, and SEL for students, staff, and families. In addition, the RP Coordinators met as a team two days per month to jointly advance implementation. The five components of the intervention are as follows:

Principal and School Leadership Component. The principal and leadership component aimed to increase the capacity of leaders for whole school systems change (see Exhibit 2). At each school, the RP Coordinator and principal established a racially and culturally diverse RP Leadership Team representative of various school roles (e.g., support staff, teachers, administrators). Importantly, the RP leadership team members themselves participated in RP and equity activities (e.g., "Brave Conversations about Race," cultural artifact share, "Rethinking Discipline" policy review). The intervention was designed to build capacity in the RP leadership team through experiencing restorative circles and conversations about race in order to turnkey and lead equity conversations to the whole school staff during monthly professional development sessions.

In addition, principals worked individually with one of three RP principal coaches during regular 90-minute sessions (Roderick, 2018). RP principal coaches were former school principals, who

had worked with Morningside Center when leading schools themselves. Individual principal coaching focused on developing principals' own social emotional skills, racial equity mindset, and restorative leadership capacity alongside creating schoolwide systemic change where adults participate in a culture of ongoing learning to identify and address inequity. Coaching sessions addressed a range of systems-change activities, including those oriented toward racial equity (e.g., alignment of vision/values) and self-critical reflection for continuous improvement (e.g., 360-degree leadership evaluation).

Restorative Intervention Component. RP Coordinators trained key staff (e.g., Deans, Guidance Counselors, Assistant Principals, support staff) to facilitate restorative interventions and develop related policies and procedures. Morningside Center's 10-hour restorative intervention professional development includes restorative approaches to deescalating conflict and resolving conflict when harm occurs. Restorative interventions are problem-solving processes during which disputants (and those harmed) identify how their actions have affected others and, when appropriate, develop action or support plans for repairing harm, rebuilding community, and fostering social emotional skills.

RP Schoolwide Staff Development Component. In this component in Year 1, RP Coordinators held a 3-day training for all school staff including teachers, paraprofessionals, administrators, and support staff (e.g., office, nutrition, and custodial staff). The training programs were designed to leverage the community-building RP circle format to develop social emotional competencies of adults (e.g., developing cultural awareness through sharing during circles) as a means to set the foundation for fostering a more equitable and inclusive approach to student behavior. Moreover, the training directly engaged participants in critical self-reflections about race, racism, and oppression. The training also included an orientation to circle keeping with students and how to use the Morningside Center K-12 SEL curricula, which is delivered during weekly student circles (Brown et al., 2010; Jones et al., 2010; 2011). A sample of curricula topics include: Listening, Anger Triggers, Problem Solving, Stress Tamer, Assertiveness, Academic goal-setting, Standing up to Institutionalized Oppression, Sharing Identities, The Challenge of Being an Ally, and Standing up to Stereotypes.

After the multi-day training, implementation supports included coaching and follow-up sessions with staff (e.g., implicit bias awareness, culturally relevant book studies, and an examination of equitable classroom practices). RP Coordinators coached a wide range of staff (e.g., school administrators, paraprofessionals) related to systems change (policy and practice shifts) and individual change (e.g., culturally relevant relationship building skills). They also supported the implementation of the K-12 SEL curricula delivered through circles. The coaching model includes an RP-Observe framework that guides circle implementation and feedback for elements of safety, belonging, voice, learning, and engagement (Gregory et al., 2017). RP Coordinators coached circle keepers or advisors who were implementing the curriculum through a cycle of observing circles, co-facilitating, and debriefing to plan next steps. Topics included ways to improve RP facilitation, SEL curriculum delivery, and equitable classroom practices (e.g., greeting students by name, classroom arrangement, equitable response strategies).

Exhibit 2.
The Whole School RP Project

Components	Description of Components
RP Leadership Team Component	(a) Two-day principal cohort training (e.g., RP experiential learning, cultural awareness activities, and identification of racialized determinants in school discipline) (b) Coaching support for the RP Leadership Team planning schoolwide equity professional development and engaging in a Rethinking Discipline process (e.g., examining the purpose of discipline and align school mission, vision, and policies with restorative and cultural values) (c) Principal supports schoolwide implementation of weekly circles-based advisory program based on social emotional learning curriculum and restorative intervention (d) Monthly principal coaching sessions to develop restorative and equitable leadership skills (e.g., 360-degree assessment of SEL skill, and setting personal and professional goals) (e) Delivery of equity activities in regular schoolwide professional development (e.g., culturally responsive book study, racial autobiography, implicit bias training, equitable classroom practice assessment)
Restorative Intervention Component	(a) Leadership training and support in addressing harm and discipline infractions through restorative interventions (e.g., community problem solving, peer mediation, restorative conversation, restorative conference, re-entry circle) (b) Discipline infractions and harm addressed restoratively (i.e. a problem-solving process including what happened, what people were thinking/feeling, who was impacted AND actions for repairing the harm, skill-building, or community building. Student behavior addressed in a restorative manner (e.g., restorative questions or restorative conferences based on a strengths-based, skill-building, and harm-restoring approach)
Schoolwide Staff Development Component	(a) Multi-day training in RP, SEL, and equity offered to all school staff (e.g., restorative mindset, understanding race, racism, and oppression, and SEL curricula) (b) Coaching with staff who implement the SEL curricula to improve RP circle facilitation skills and equitable classroom habits (e.g., relational approaches to conflict, social justice lesson plan development, increasing equitable practices) (c) Weekly community- and skill-building circles with all students guided by SEL curriculum and addressing social justice (e.g., cultural identity, oppression, stereotypes, prejudice, discrimination, social injustice in current events)
Student Leadership and Voice Component	(a) Restorative training opportunities for students in Year 2 (e.g., circle facilitation, peer mediation) (b) Opportunities for student leadership and voice in Year 2 (e.g., student leadership teams, student surveys, peer mentors)
Family RP Opportunities Component	(a) Training and collaboration with the Parent Coordinator (a district-employed, family liaison in every school) (b) Schoolwide family opportunities to experience RP (e.g., adult SEL exercises, mindfulness, supporting children's emotion management)

Student Leadership and Voice Component. This component included weekly community- and SEL skill-building circles with students in small-group advisory settings or a designated class. Additional intervention components for student leadership and voice were designed for Year 2 of the Whole School RP Project (e.g., training students as circle facilitators, student leadership teams, student surveys, or peer mentors).

Family RP Opportunities Component. In this component, Morningside Center aimed to develop the capacity of the full time, district-employed, family liaison in each school (see Table 1). Families were invited to participate in RP training and, with RP Coordinators, co-plan and co-facilitate family RP opportunities. Events were designed to be accessible for diverse families (e.g., offered childcare, transportation, and multiple languages).

School Sample

All 18 schools completed Year 1 of the evaluation. Evaluation activities included a principal interview and student and staff surveys. At the school level, the average school size was 366 students per school and the percent of low SES students was 91% enrolled. On average the percent of Black student enrollment was 55%. There was no school-level attrition, meaning no schools dropped out of the evaluation.

Student Sample

The original sample consisted of 6,507 students enrolled from 1st to the 12th grade. Of those students, 421 were considered "late joiners" as they had enrolled six weeks after the start of class and were excluded (n = 6,086) as was prespecified in accordance with the What Works Clearinghouse (WWC) guidelines. Of the remaining enrolled 6,086 students, 208 students were considered attritors who completely left the NYC school system for various reasons (e.g., transferred out of state, withdrew from school after repeated absences, was deceased). The overall attrition rate was 3.4% where 4.1% left from the control group and 2.7% left from the treatment group. Based on WWC guidelines, this represents a tolerable threat of bias under both optimistic and cautious conditions.

In the final analytic sample (n = 5,878), 253 students had transferred to another school within the NYC school system but were retained in the analytic sample as a result of the intent to treat (ITT) analysis prespecified which preserved the original randomized design. The analytic sample was composed of 2,919 students in the control group and 2,959 students in the intervention group (see Exhibit 3).

Of the sample (male = 50%), the majority were Black (54%), followed by Latinx (38%), some other race/ethnicity (7%), and White (2%) students. Twenty-four percent had an identified disability and 91% were classified as coming from low socioeconomic status (SES) backgrounds.

Exhibit 3.
Student Attrition by Treatment Status

	Control	Intervention	Overall
Assigned n	3,044	3,042	6,086
Attrition	125 (4.1%)	83 (2.7%)	208 (3.4%)
Analytic n	2,919	2,959	5,878

Student Demographics
Students in the intervention and comparison schools were equivalent across demographic characteristics (See Exhibit 4).

Exhibit 4.
Descriptive Statistics Comparing Students and Schools Assigned to the Intervention and Control Conditions (*n* = 5,878)

	Control (*n* = 2,919)		Intervention (*n* = 2,959)		
Student level					
Race	*n* (%)	*M (SD)*	*n* (%)	*M (SD)*	*p*[1]
Black	1,650 (56.5)		1,510 (51.0)		.317
Hispanic	970 (33.2)		1,242 (42.0)		
Other	216 (7.4)		168 (5.7)		
White	83 (2.8)		39 (1.3)		
Male	1,484 (50.8)		1,464 (49.5)		.689
With a disability	719 (24.6)		692 (23.4)		.439
Poverty	2,668 (91.4)		2,689 (90.9)		.832
Prior Disc. Incident	424 (14.5)		361 (12.2)		.439
Missing prior disc.	182 (6.2)		136 (4.6)		.153
School level (n = 18)					
% Black enrolled		0.57 (0.15)		0.53 (0.19)	.593
School type					1.00
Middle	3 (33.3)		3 (33.3)		
Combined	1 (11.1)		1 (11.1)		
Elementary	3 (33.3)		3 (33.3)		
High	2 (22.2)		2 (22.2)		
Enrollment size		364.33 (123.13)		367.67 (120.13)	.954
% Poverty		0.90 (0.06)		0.91 (0.06)	.860

Notes. [1]All differences between intervention and control groups are not statistically significant (all *p*s > .05). For categorical student-level variables, *p* values are from Rao-Scott (1981) χ^2 tests which account for the clustered nature of the data. Continuous variable differences based on *t*-tests. Prior Disc Incident = Discipline Incident (i.e., ODR) in the year prior to the RP Project.

Measures of Year 1 Implementation Fidelity

Although fidelity is conceptualized as a multi-dimensional construct (i.e., adherence, exposure, quality; Schulte, Eason, & Parker, 2009), feasibility issues restricted measurement of implementation to "dosage" of each program component. At the beginning of the project, thresholds of adequate dosage were identified (See Appendix A). Twelve measures of implementation fidelity/dosage across the five components described above were measured in Year 1 of the project. Implementation data were collected through monthly, structured fidelity checks along with one annual 60-minute interview with each RP Coordinator. In addition, 60-minute annual interviews with a school administrator from each school were also conducted.

Answering RQ1, the fidelity data showed that, overall, all nine schools met the pre-established threshold for adequate fidelity of project implementation in Year 1 (See Appendix A). Although meeting overall school-level thresholds, there was variability in implementation at each school and across project components (See Exhibit 2 for listed activities within each component).

Fidelity of RP Leadership Team Component (9/9 schools met threshold)

(a) The nine principals attended Morningside Center's two-day leadership training.
(b) The nine school RP leadership teams met regularly throughout the year (M = 8; range = 4 – 13).
(c) The nine principals supported RP implementation by creating time in the schedule for schoolwide circles-based advisory and/or restorative interventions (M = 2 out of 3 support criteria met; range = 1 – 3).
(d) The nine principals met with their coaches for individual sessions addressing restorative and equitable leadership (M = 11 sessions; range = 7 - 14).
(e) RP leadership teams in 8 of the 9 schools held whole school racial equity professional development sessions during a designated staff meeting time. The number of schoolwide staff equity sessions ranged in those schools (M = 5; range = 2 - 8).

Fidelity of Restorative Interventions Component (9/9 schools met threshold)

(a) In Year 1, three out of nine schools trained designated staff in tier 2 restorative interventions
(b) The nine schools addressed more than 50% of discipline infractions in a restorative manner (M = 69%; range = 53% - 88%).

Fidelity of Schoolwide Staff Development Component (7/9 schools met threshold)

(a) In Year 1, all 9 schools implemented the 15-hour, intensive RP training which integrated SEL and equity. RP Coordinators trained, on average, 54% of all staff in each school (range = 24% - 90%; e.g., instructional staff, support staff, paraprofessionals, office assistants, nutrition staff, custodial staff, and administrators).
(b) On average, RP Coordinators trained 95% of the staff who were RP circle keepers implementing the SEL curriculum (range = 77% - 100%) and coached an average of 82% of those trained staff (range = 43% - 100%). The number of coaching sessions per RP circle keeper varied (M = 5 coaching sessions, range = 1 - 24). RP Coordinator coaching extended beyond circle keeping and SEL curriculum. They coached staff on a range of

initiative efforts (e.g., relational approaches to conflict, social justice lesson plan development, increasing equitable practices; *M* = 3 coaching sessions, range = 1 - 15).

(c) Eight out of nine interventions schools scheduled weekly circles for all students during small-group advisory periods (middle and high schools) or designated classes (elementary schools). The remaining school scheduled circles for one grade level with plans to expand to the entire school in the following year. On average, 86% of designated curriculum implementers facilitated weekly circles with students (range = 39% - 100%).

Fidelity of Student Leadership and Voice Component: (not applicable in Year 1)

(a) Restorative training opportunities for students was not measured in Year 1 (i.e. student circle facilitator training and peer mediator training designed for Year 2 implementation)

(b) Opportunities for student leadership and voice were not measured in Year 1 (i.e. student leadership teams, student surveys, peer mentors were designed for Year 2 implementation).

Fidelity of Family RP Opportunities Component: (4/9 schools met threshold)

(a) Four of the nine schools trained the Parent Coordinator in RP, SEL, and equity (Parent Coordinator is a designated staff position in every school in the participating district).

(b) Six of the nine schools offered schoolwide opportunities with family members to introduce and experience restorative practices (*M* = 1 family RP opportunity, range = 1 - 3). Overall, the family RP opportunities component had the lowest implementation relative to the pre-established fidelity thresholds in Year 1. Seven out of nine intervention schools met adequate fidelity in at least one measure of the family RP opportunities component, yet two schools were unable to implement Parent Coordinator training or opportunities for families to participate in RP in Year 1).

Intervention Contrast

Following Year 1 of the project implementation, we conducted structured interviews using a checklist of intervention components with administrators in the control schools to identify the degree to which controls school initiatives differed from the Whole School RP Project in intervention schools. Based on administrator interviews and accompanying checklist, three control schools were engaging in schoolwide RP, SEL, and equity initiatives. Four of the nine control schools were implementing schoolwide RP (i.e., schoolwide community-building circles, some staff trained in RP, and a documented restorative approach to discipline). Of the remaining control schools, four had schoolwide initiatives related to SEL (i.e., regularly scheduled explicit instruction following an SEL curriculum). Only one remaining control school did not report a schoolwide initiative related to RP, SEL, or equity.

Thus, the "business-as-usual" condition reflected reforms that are common throughout the nation. The large urban city where the study took place had existing initiatives in place to reduce exclusionary discipline and address the racial discipline gap. For example, the district had reformed the discipline code to include restorative interventions and begun implicit bias training district-wide. As a result, the nine Whole School RP Project schools were compared to the "business-as-usual" schools, of which eight out of nine had some discipline reform initiatives in place.

Exhibit 5. *Comparing Initiatives in i3 Control Versus Intervention Schools in Year 1*

Administrator- reported schoolwide initiatives*	Control Schools	Intervention schools
RP + SEL + equity	3	9*
RP + SEL	0	0
SEL + equity	3	0
RP + equity	1	0
RP	0	0
SEL	1	0
Equity	0	0
No RP, SEL, equity	1	0
Total:	9 schools	9 schools

Note. * The nine schools in the intervention reached adequate threshold for fidelity of implementation (See Appendix A)

Student- and School-level Measures

The student-level outcomes of interest were if a student had been recorded in one or more discipline incidents in the 2018-19 school year. Prior year's Incidents (SY 2017-18) were also included as baseline measures. Throughout this report, the term Incident includes all student conduct occurrences recorded in the school and district database whether or not a disciplinary response or consequence was imposed. Incidents could result in exclusionary discipline (e.g., suspension) or other forms of intervention or support (e.g., parent outreach, referral to counseling services, conflict resolution). In some states and districts, Incidents are referred to as Office Discipline Referrals or ODRs. Incidents or ODRs are a meaningful dependent variable given evidence of their concurrent and predictive validity; students' receipt of one or more referrals is associated with negative teacher behavioral ratings (Pas, Bradshaw, & Mitchell, 2011) and, years later, with being off track for graduation (Tobin & Sugai, 1999).

All 18 schools were within the same school district which share a discipline code of conduct. This helped facilitate similar practices for issuing Incident records. In other words, the shared code facilitated some consistency in "data collection" – increasing the evaluation's credibility in its use of Incident records as a key dependent variable in the evaluation. The Whole School RP Project has multiple components, including a focus on strengthening relationships and belonging. Thus, the project was not singularly focused on staff reducing the frequency of discipline incidents (or those which get recorded), which ensures the dependent measure (e.g., Incidents) was not "overly aligned" with the intervention.

Categorical student-level covariates included: poverty status, disability status, race, and gender. At the school level, the school type (i.e., elementary, middle, high, combined), enrollment size, the percent of economically disadvantaged students enrolled, and the percent of Black students enrolled were included. A comparison of demographic and baseline measures between treatment and control participants/schools is shown in Exhibit 4.

Baseline Equivalency

Although there was low attrition, we show comparisons between the intervention and control groups. There were no statistically significant differences between the intervention and control groups at both the student and the school level (See Exhibit 4). Based on the preregistration and WWC protocols, baseline equivalence[2] would be established by comparing standardized mean differences using Cox's *d* for prior Incident (*d* = -0.12), gender (*d* = -0.03), and poverty status (*d* = -0.04).

Minimal differences resulting from sampling variation are shown in Exhibit 4 with all *p*s > .10. Student level measures account for clustering by using cluster adjusted standard errors with continuous variables and a Rao-Scott χ^2 test of independence (which accounts for clustering) for the categorical variable showing race/ethnicity breakdowns. School-level characteristic comparisons are made using standard *t*-tests.

Analytic Strategy

For the analysis of the likelihood of receiving an Incident, multilevel linear probability models (LPMs) were prespecified as the analytic approach to evaluate the ITT impact of being assigned to receive the RP intervention or not. LPMs are valid for experimental studies (Deke, 2014; Huang, 2019, 2021) though we had prespecified robustness checks using logistic regression models as well. Logistic regression models were run using a CR2 standard error adjustment (Bell & McCaffrey, 2002) and using the Mancl and DeRouen (2001) correction using a generalized estimating equations approach (Liang & Zeger, 1986). Of note is that both standard error adjustments are used to account for the limited number of clusters which traditional cluster robust standard errors do not account for (Huang & Li, 2021; Huang, 2021).

Intent to Treat (ITT) Estimates

RQ2 asked whether assignment to the RP intervention condition resulted in lower Incident rates for students. As the intervention status was assigned at the school level, to account for the nesting of students within schools, two-level multilevel models were used. Multilevel LPMs were conducted using restricted maximum likelihood estimation with the lme function in the nlme package (Pinheiro et al., 2014) in R 4.1 (R Core Team, 2020).

As there was minimal missing data (5.4%) and only coming from the baseline measure (e.g., Incident recorded in the prior year), the Institute of Education Sciences (IES, WWC, 2020, p. 40) recommends the use of the dummy variable indicator method. Although other popular methods for handling missing data include using multiple imputation or full information maximum likelihood, using the dummy variable method for missing data is straightforward, produces unbiased treatment effects due to the randomization, and is specifically acceptable only when a RCT is used and only if the baseline measure is missing (Puma et al., 2009, p. 34) which is exactly the condition in the current study. Specifically, missing data was due to students not having a prior year's discipline data given they were not enrolled in the school district that year.

---

[2] See https://ies.ed.gov/ncee/wwc/Docs/OnlineTraining/wwc_training_m3.pdf

Note, in order to evaluate the effect of the intervention, only the randomized intervention assignment variable is necessary and results in unbiased estimates, however, the inclusion of covariates (e.g., baseline measures) improves model power to detect effects (Murnane & Willett, 2011).

The student-level covariates included the dummy-coded baseline measure (e.g., prior year Incident), race (Hispanic, White, Other with Black as the reference group), gender, poverty status, and disability status. An additional dummy coded variable was included if a student was missing the baseline measure (1 = missing, 0 = not missing). At the school-level, intervention status (1 = treatment, 0 = control; the variable of interest), enrollment size (in hundreds of students), the percent of students living in poverty, and the percent of Black students enrolled were included. The continuous school-level covariates were mean centered. School type, which was a blocking variable, was included as a series of dummy codes with middle schools as the reference group. The overall combined impact evaluation formula can be expressed as:

$$Y_{ij} = \delta \text{Treat}_j + \alpha \text{Elem}_j + \tau \text{Comb}_j + \gamma \text{High}_j + \zeta_{2\text{-}4}\, \text{SchDemo}_j + \beta_{5-10}\, \text{Demo}_{ij} + \beta_{11} \text{Prior}_{ij} + \beta_{12} \text{Miss}_{ij} + u_{0j} + u_{1j}\, \text{Prior}_{ij} + e_{ij}$$

where Y_{ij} represents the outcome (Incident involvement = 1) of student i in school j and δ represents the coefficient of interest (i.e., the ITT effect). Potential heteroscedasticity was accounted for by allowing the slopes for the baseline measure to randomly vary to investigate whether the effect of prior sanctioning varied across the schools. The residual variance for e_{ij} was estimated based on the type of school attended to account for violations of homoscedasticity as well. A likelihood ratio test (LRT; LaHuis & Ferguson, 2009) was performed to evaluate if the inclusion of a random slope and allowing residual variances to be estimated based on the type of school attended resulted in an improvement in model fit. If the LRT was not statistically significant ($p > .05$), the simpler, more parsimonious model was supported.

Several moderators were specified which assessed if there was an interaction between intervention status with student gender, disability status, and race (RQ3). An LRT was used to test (using maximum likelihood as fixed effects were of interest) for an improvement in model fit of the interaction model compared with the main effects model.

Results

Intervention Effects on Discipline Incidents

Descriptively, during the school year prior to the initiative (SY 2017-18), the Incident rate was 12.2% in the intervention schools and 14.5% in the control schools. During the initiative year (SY 2018-2019), the Incident rate was 11.1% in the intervention schools and 18.2% in the control schools (see Exhibit 6). The difference in Incident rates between conditions in the initiative year was approximately 7 percentage points in favor of the intervention schools. Exhibit 6 shows that Incident rates in intervention schools were lower compared to control schools.

Exhibit 6.

Disciplinary Incident Rates in Initiative Year 1 by Intervention Status and School Type (*n* = 5,878 students in 18 schools)

	Incident rates				n	
	Control	Intervention	Diff	*d*	Control	Intervention
Elementary	10.2%	7.7%	-24%	-0.18	637	839
Middle	17.4%	10.0%	-43%	-0.39	1,124	964
High	27.4%	17.3%	-37%	-0.36	614	665
Combined	18.8%	10.6%	-44%	-0.40	544	491
Overall	18.2%	11.1%	-39%	-0.35	2,919	2,959

Notes. Diff = difference in rates comparing intervention with control schools. *d* = Cox's *d* (standardized mean difference). Combined = combined schools with middle and high school students

An unconditional null model results indicated that the intraclass correlation coefficient (ICC) (based on the LPM) for the Incident outcome was .07. LRTs indicated that both the random slope, $\chi^2(2) = 33.5$, $p < .001$, and allowing residual variances to be estimated separately by school type, $\chi^2(3) = 232.2$, $p < .001$, resulted in improved model fit and were both warranted.

Exhibit 7.
Multilevel Linear Probability Models for Discipline Incidents
(*n* = 5,878 students from 18 schools)

	Main Effects	Interaction
(Intercept)	0.116*** (0.025)	0.118*** (0.025)
Student level		
Prior Incident	0.250*** (0.027)	0.250*** (0.027)
Race[1]		
Hispanic	-0.046*** (0.009)	-0.051*** (0.013)
Other	-0.066*** (0.018)	-0.080*** (0.024)
White	-0.038 (0.029)	-0.045 (0.035)
With a disability	0.038*** (0.010)	0.038** (0.014)
Male	0.026** (0.008)	0.027* (0.012)
Poverty	0.034* (0.015)	0.034* (0.015)
Missing prior Incident	0.075*** (0.018)	0.075*** (0.018)
School level		
Intervention	**-0.045* (0.018)**	**-0.051* (0.022)**
% Black enrollment	-0.002 (0.001)	-0.002 (0.001)
Enrollment size (in 100s)	-0.030* (0.010)	-0.030* (0.010)
% Poverty	0.002 (0.003)	0.002 (0.003)
School type[2]		
Combined	0.039 (0.039)	0.037 (0.039)
Elementary	-0.067* (0.025)	-0.067* (0.025)
High	0.026 (0.035)	0.026 (0.035)
Interactions		
Intervention x Other		0.030 (0.036)
Intervention x Hispanic		0.010 (0.018)
Intervention x White		0.020 (0.063)
Intervention x Male		-0.002 (0.017)
Intervention x Disability		0.001 (0.020)

Notes [1]Black is the reference group. [2]Middle school is the reference group. Prior Incident = Student had prior year discipline incident. Intervention effects in **bold**. ***$p < .001$; **$p < .01$; *$p < .05$. Standard errors within parentheses.

Results of the model showing the main effects of the intervention (RQ2) are shown in Exhibit 7. Findings indicated that the effect of the RP initiative was a reduction of Incidents by 4.5 percentage points ($p = .03$) in the intervention schools compared to the control schools, controlling for all other variables in the model. Although the intervention effect is directly interpretable as a 4.5 percentage point reduction in Incidents, we computed Cox's index as a measure of effect size as specified by the WWC. Cox's standardized mean difference index was $d_{Cox} = -0.30$ (covariate adjusted), a small to moderate effect size. The next model tested interactions of intervention status with gender, race, and disability status. All interaction effects were not statistically significant and an LRT indicated no statistically significant improvement over the main effects model, $\chi^2(5) = 0.83$, $p = .98$. This indicates a lack of moderation effects.

Robustness Checks

Additional robustness checks (see Exhibit 8) were performed for the main effects of the intervention for Incidents. All the results controlled for other student- and school-level variables included in the main model. Results all indicated that the prevalence rates for Incidents were lower for the intervention condition. Population-averaged logit models are shown using logits and the more commonly understood odds ratios (*OR*s). Both models also indicate a decrease in Incidents associated with the intervention.

Exhibit 8.

Robustness Checks for Main Treatment Effect using Logistic Regression Models

	Treatment Group		Comparison Group		Estimated Impact (Robustness checks)				
Outcome	N Students	Rate	N Students	Rate	T-C difference	Standard error	p	OR	Model
Incidents	2,959	11.1	2,919	18.2	-0.684	0.206	.013	0.50	GLM binomial with CR2
					-0.684	0.289	.017	0.50	GEE approach with ML

Notes. Only treatment contrast shown. All school and student-level controls are included. CR2 = Bell and McCaffrey's (2002) bias reduced estimator. GLM = generalized linear model. GEE = generalized estimating equations. ML = Mancl and DeRouen (2001) standard error adjustment.

Summary

Overall, models indicate a small to moderate ITT main effect of the RP Project on discipline incidents. In the comparison schools, 18.2% of students received a discipline incident record compared to 11.1% of students in the RP Project schools. No differential effects of the intervention were noted based on gender, race, or disability status.

This study presents analysis after one year of program implementation. Scholars recognize that RP initiatives may take upwards of three years to fully implement training, policy, and practice shifts that lead to improvements (e.g., Gregory & Evans, 2020). This suggests that future evaluation of the Whole School RP Project after two to three years of implementation is needed to further understand its impact on reducing discipline disparities for Black students, male students, and students in special education. However, it is noteworthy that after only one year of the Whole School RP Project schools had lower rates of discipline incidents.

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*(5), 876–890. https://doi.org/10.1007/s10964-019-01013-2

Anyon, Y., Gregory, A., Stone, S. I., 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. A*merican Education Research Journal, 53*, 1663-1697. https://doi.org/10.3102/0002831216675719

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 Corporation.

Balfanz, R., Byrnes, V., & Fox, J. (2013). *Sent Home and Put Off-Track: The Antecedents, Disproportionalities, and Consequences of Being Suspended in the Ninth Grade.*

Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*, 169–182.

Berkowitz, K. (2012). Whole School Restorative Practice Implementation Guide. San Francisco Unified School District. Retrieved from https://www.healthiersf.org/RestorativePractices/Resources/documents/SFUSD%20Whole%20School%20Implementation%20Guide%20final.pdf.

Blake, J. J.; Smith, D. M., Unni, A., Marchbanks III, M. P., Wood, S., & Eason, J. M. (2020). Behind the eight ball: The effects of race and number of infractions on the severity of exclusionary discipline sanctions issued in secondary school. *Journal of Emotional and Behavioral Disorders, 28*(3), 131-143. https://doi.org/10.1177/106342662093769

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving Classroom Quality: Teacher Influences and Experimental Impacts of the 4Rs Program. *Journal of Educational Psychology, 102*(1), 153–167. https://doi.org/10.1037/a0018160

Davis, F. (2019). *The little book of race and restorative justice*. New York, NY: Good Books.

Deke, J. (2014). *Using the linear probability model to estimate impacts on binary outcomes in randomized controlled trials*. Department of Health and Human Services. https://www.hhs.gov/ash/oah/oah-initiatives/assets/lpm-tabrief.pdf

Evans, K. R., & Vaandering, D. (2016). *The little book of restorative justice in education: Fostering responsibility, healing, and hope in schools*. Good Books.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P. I., & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement.* Coucil of State Governments Justice Center. https://knowledgecenter.csg.org/kc/system/files/Breaking_School_Rules.pdf

Fronius, T., Darling-Hammond, S., Persson, H., Guckenburg, S., Hurley, N., & Petrosino, A. (2019). *Restorative justice in U.S. schools: An updated research review.* https://www.wested.org/wp-content/uploads/2019/04/resource-restorative-justice-in-u-s-schools-an-updated-research-review.pdf

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

Gregory, A., & Roberts, G. (2017). Teacher Beliefs and the Overrepresentation of Black Students in Classroom Discipline. *Theory into Practice, 56*(3), 187–194. https://doi.org/10.1080/00405841.2017.1336035

Harper, K., Ryberg, R., & Temkin, D. (2019). Black students and students with disabilities remain more likely to receive out-of-school suspensions, despite overall declines. *Child Trends*. https://www.childtrends.org/publications/black-students-disabilities-out-of-school-suspensions

Huang, F. L. (2019). Alternatives to logistic regression models in experimental studies. *The Journal of Experimental Education*, 1–16. https://doi.org/10.1080/00220973.2019.1699769

Huang, F. L. (2020). Prior problem behaviors do not account for the racial suspension gap. *Educational Researcher, 49*(7), 493-502. doi:10.3102/0013189X20932474

Huang, F. L. (2021). Analyzing cross-sectionally clustered data using generalized estimating equations. *Journal of Educational and Behavioral Statistics*, 10769986211017480.

Huang, F. & Li, X. (2021). Using cluster robust standard errors when analyzing group randomized trials with few clusters. B*ehavior Research Methods*.

Ispa-Landa, S. (2018). Persistently harsh punishments amid efforts to reform: Using tools from social psychology to counteract racial bias in school disciplinary decisions. *Educational Researcher, 47*(6), 384–390. https://doi.org/10.3102/0013189X18779578

Institute of Education Sciences. (2020). *Education research grant program*. https://ies.ed.gov/funding/pdf/2021_84305A.pdf

Jagers, R. J., Rivas-Drake, D., & Williams, B. (2019). Transformative social and emotional learning (SEL): Toward SEL in service of educational equity and excellence. *Educational Psychologist, 54*(3), 162–184. https://doi.org/10.1080/00461520.2019.1623032

Jones, S. M., Brown, J. L., Hoglund, W. L. G., & Aber, J. L. (2010). A School-Randomized Clinical Trial of an Integrated Social-Emotional Learning and Literacy Intervention: Impacts After 1 School Year. Journal of Consulting and Clinical Psychology, 78(6), 829–842. https://doi.org/10.1037/a0021383

Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention: An Experiment in Translational Developmental Research. *Child Development, 82*(2), 533–554. https://doi.org/10.1111/j.1467-8624.2010.01560.x

LaHuis, D. M., & Ferguson, M. W. (2009). The accuracy of significance tests for slope variance components in multilevel random coefficient models. *Organizational Research Methods*, *12*(3), 418–435.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22.

Losen, D. J., Hodson, C., Keith, M. A., II, Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* Los Angeles: University of California, The Civil Rights Project.

Losen, D. J., & Martinez, P. (2020). *Lost opportunities: How disparate school discipline continues to drive differences in the opportunity to learn.* Palo Alto, CA/Los Angeles, CA: Learning Policy Institute; Center for Civil Rights Remedies at the Civil Rights Project, UCLA.

Lustick, H. (2019). Going restorative, staying tough: Urban principals' perceptions of restorative practices in collocated small schools. Education and Urban Society. doi:10.1177/0013124520974335

Manassah, T., Roderick, T., & Gregory, A. (2018). A promising path toward equity: Restorative circles develop relationships, build communities, and bridge differences. *Learning Forward, 39*(4)*,* pp. 36-40. https://learningforward.org/journal/august-2018-vol-39-no-4/a-promising-path-toward-equity/

Mancl, L. A., & DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics, 57*(1), 126–134.

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing research-based trajectories of office discipline referrals for individual students. *School Psychology Review*, *39*, 380–394.

Mittleman, J. (2018). A downward spiral? Childhood suspension and the path to juvenile arrest. *Sociology of Education, 91*, 183-2014. https://doi.org/10.1177/0038040718784603

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*(2), 224-240. https://doi.org/10.17105/spr-14- 0008.1

Pas, E. T., Bradshaw, C. P., & Mitchell, M. M. (2011). Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools*, *48*, 541–555.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2014). *nlme: Linear and Nonlinear Mixed Effects Models.* http://CRAN.R-project.org/package=nlme

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). What to do when data are missing in group randomized controlled trials. NCEE 2009-0049. *National Center for Education Evaluation and Regional Assistance*.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rao, J. N. K., & Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, *76*(374), 221–230. https://doi.org/10.2307/2287815

Ritter, G. W. (2018). Reviewing the Progress of School Discipline Reform, *Peabody Journal of Education, 93*(2), 133-138, DOI: 10.1080/0161956X.2018.1435034

Roderick, T. (2018). *Whole school racial equity (RE) program personal and professional leadership coaching for principals: A manual for coaches*. Morningside Center for Teaching Social Responsibility internal report: Unpublished.

Sandwick, T., Hahn, J. W., & Hassoun Ayoub, L. (2019). Fostering community, sharing power: Lessons for building restorative justice school cultures. *Education Policy Analysis Archives, 27*(145). https://doi.org/10.14507/epaa.27.4296

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. School Psychology Review, 38(4), 460–475.

Silva, S. M., Han, T. M., Samimi, C., Golieb, K., McCurdy, J., & Forte, A. (2019). *State of the state: Restorative Justice in Colorado*. Colorado Restorative Justice Coordinating Council. https://www.rjcolorado.org/_literature_181293/STATE_OF_THE_STATE_Restorative_Justice_in_Colorado_%E2%80%93_June_2019

Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014) More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education, 47*(4), 546-564. DOI: 10.1080/10665684.2014.958965

Song, S. Y. & Swearer, S. M. (2016) The Cart Before the Horse: The Challenge and Promise of Restorative Justice Consultation in Schools. *Journal of Educational and Psychological Consultation,* 26(4), 313-324. DOI: 10.1080/10474412.2016.1246972

Tauri, J. M. (2018). Restorative Justice as a colonial project in the disempowerment of Indigenous peoples. In T. Gavrielides (Ed.), *Routledge International Handbook of Restorative Justice*. Routledge.

Thorsborne, M., & Blood, P. (2013). *Implementing restorative practices in schools: A practical guide to transforming school communities.* London, UK: Jessica Kingsley Publishers.

Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, *7*, 40–53.

United States Department of Education. (2014). *Guiding principles: A resource guide for improving school climate and discipline.* Washington, DC: Authors.

United States Department of Education, Office for Civil Rights. (2018). *2015-2016 civil rights data collection: School climate and safet*y. U.S. Department of Education Office for Civil Rights. https://www2.ed.gov/about/offices/list/ocr/docs/school-climate-and-safety.pdf

Valandra Waŋbli Wapȟáha Hokšíla, E. C. (Ed.) (2020). *Colorizing restorative justice: Voicing our realities.* Living Justice Press.

Winn, M. T. (2018). *Justice on both sides. Transforming education through restorative justice*. Harvard Education Press.

Zakszeski, B. & Rutherford, L. (2021). Mind the gap: A systematic review of research on restorative practices in schools. *School Psychology Review*. Published online: https://doi.org/10.1080/2372966X.2020.1852056

Zehr, H. (2014). *The Little Book of Restorative Justice*. New York, New York: Good Books.

Appendix A

The National Evaluation of i3 Technical Assistance Team

i3 Data Collection Templates for Reporting Fidelity of Implementation Findings

August 10, 2018,

Revised February 2019

The National Evaluation of i3 Technical Assistance Team

i3 Data Collection Templates for Reporting Fidelity of Implementation Findings

1. Logic Model and Fidelity Measure [All Grant Types]

Provide a copy of (a) the final intervention logic model and (b) the fidelity measure that Was used to generate the results in Table 2.1 below.

1. If there have been ***no*** changes to the versions of the logic model or fidelity measure that are "registered" with the AR team, there are two options:
    a. Referring reviewers to the relevant files in the i3 Sharepoint folder for evaluation documents
    b. Attaching/appending the logic model and fidelity measure to this template.
2. If there have been changes to the registered versions of the logic model or fidelity measure, a copy of the revised model or measure should be attached/appended to this template.

Please choose one option for submission of the logic model and one option for submission of the fidelity measure to communicate which version the AR team should use in its review.

SUBMISSION OF LOGIC MODEL

Check *one* box below:

Logic model to be reviewed by the AR team is the version that is *currently-registered* with the AR team. There have been no changes to the registered version of the logic model.

☐ See file name on Sharepoint, or

☐ Logic model is attached/appended to this template.

Logic model to be reviewed by the AR team is a revised version. There have been changes to the currently registered version of the model.

✓ New logic model is appended/attached to this template.

SUBMISSION OF FIDELITY MATRIX

Check *one* box below:

Fidelity matrix to be reviewed by the AR team is the version that is *currently-registered* with the AR team. There have been no changes to the registered version of the fidelity matrix.

☐ See file name on Sharepoint, or

☐ Fidelity matrix is attached/appended to this template.

Fidelity matrix to be reviewed by the AR team is a revised version. There have been changes to the currently registered version of the matrix.

✓ New fidelity matrix is appended/attached to this template.

2. Fidelity of Implementation of Intervention(s) by Year [All Grant Types]: Table 2.1

Findings from Evaluator Study of Implementation: IMPLEMENTATION YEAR 1

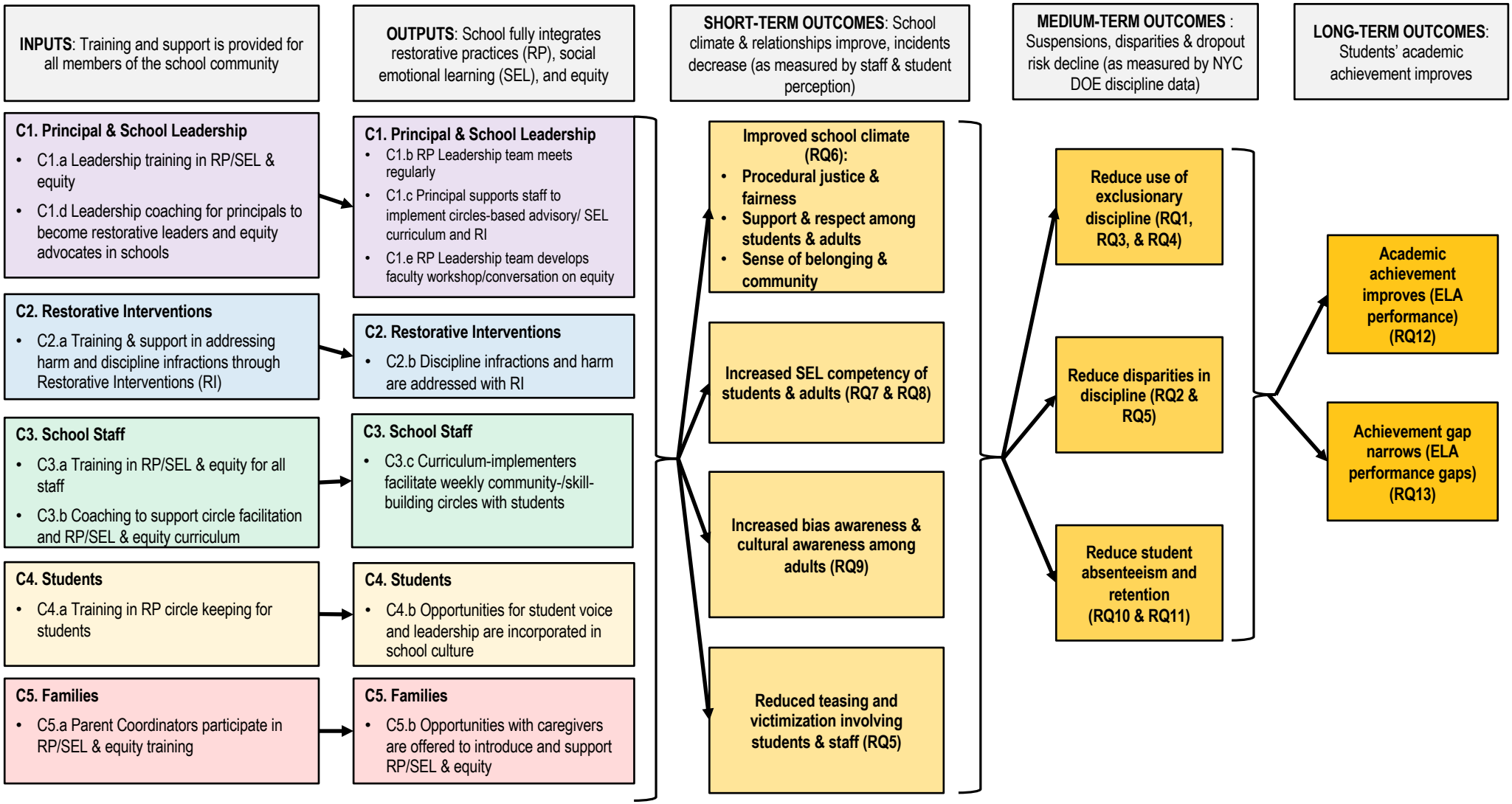
Enter calendar year: September 2018 – June 2019

Intervention Component	Implementation measure (total number of measurable indicators representing each component	Number of Units in Which Fidelity of Component Was Measured (# of schools, districts, etc.)	Number of Units in Which the Intervention Was Implemented (# of schools, districts, etc.)	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
C1. Principal and School Leadership	5	9 schools	9 schools	Adequate fidelity = principal trained in RP/Equity; 5 or more leadership team meetings; 2/3 principal support criteria; 5 or more principal coaching sessions; 1 or more whole school equity PD	7/9 (75%) of schools with adequate implementation	9/9 schools with adequate implementation	Yes
C2. Restorative Intervention Team	2	9 schools	9 schools	Adequate fidelity = staff trained in RI; 50% of discipline handled restoratively	7/9 (75%) of schools with adequate implementation	9/9 schools with adequate implementation	Yes
C3. School Staff	3	9 schools	9 schools	Adequate fidelity = 50% or more of staff trained in RP/equity; 50% or more of curriculum implementers participate in coaching; 50% or more of curriculum implementers hold weekly circles	7/9 (75%) of schools with adequate implementation	7/9 schools with adequate implementation	Yes
C4. Students	2	N/A in year 1	N/A in year 1	N/A in year 1	N/A	N/A	N/A
C5. Families	2	9 schools	9 schools	Adequate fidelity = parent coordinator trained in RP/equity & 1 or more RP opportunities for families	7/9 (75%) of schools with adequate implementation	4/9 schools with adequate implementation	No
Mediators[a] [if choosing to report findings for fidelity of Intervention OPTIONAL]							

[a] Mediators refer to measurement of short-term outcomes, e.g., changes in student/teacher/parent behaviors or attitudes that are assumed to have a direct connection to long-term outcomes. If these mediators have been measured in both the treatment and comparison groups **and** are being reported as part of an exploratory contrast, they should not be entered here. If these mediators have been measured in the treatment group only or in both the treatment and the comparison group **and** are not being reported as part of an exploratory contrast, the findings for just the treatment group would be reported here. If

Appendix A

i3 Development Grant 111: Evaluation of Morningside Center's Whole School Restorative Practices Project



Appendix B

Fidelity of Implementation of the Whole School RP Project:

***RP-Assess* Quantity, Quality, Equity**

Component 1: Principal and School Leadership – Support for principals and school leaders in becoming restorative leaders and fostering schoolwide RP/SEL

Component 2: Restorative Intervention Team- Coaching and fostering school leadership for schoolwide RP/SEL

Components 3: School Staff- differentiated training in RP/SEL and equity schoolwide practices for curriculum implementers, key discipline staff, and all support staff; curriculum implementer coaching in facilitating community- and skill-building circles

Component 4: Students- Opportunities for student leadership and voice

Component 5: Families- Opportunities for family member participation in circles and Restorative Practices

NEi3 Fidelity Matrix for QUANTITY and THRESHOLD for QUANTITY

Indicators	Definition	Unit of Implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score & threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Key Component 1: Support for Principals/School Leaders in becoming restorative leaders and fostering schoolwide RP/SEL and equity									
C1.a	Principal will complete leadership training in RP/SEL and equity with Morningside Center (activity 1.2.1)	Principal (school-level)	Morningside Center staff developer	Morningside Center staff developer submit rosters with training completion checklist (Oct & May)	0 (low) = principal completes 0-4 hours of training 1 (moderate) = principal completes 5-9 hours of training 2 (high) = principal completes 10 hours of training			All Treatment Schools (n = 9	Year 1 (Summer/ Fall 2018)
C1. b	Principal organizes RP Leadership Team to engage in the Rethinking Discipline Process and Courageous Conversations to develop a schoolwide plan for RP/SEL and equity (activity 1.2.2)	RP Leadership Team (school-level)	Staff Developer 15-minute monthly fidelity survey	Morningside Center Staff Developer will record data and report in a monthly 15-minute fidelity survey check-in through Qualtrics or phone	0 (low) = 0-4 RP leadership team meetings in a full school year 1 (moderate) = 5-7 RP leadership team meetings in a full school year 2 (high) = 8 or more RP leadership team meetings in a full school year (approximately once a month).			All Treatment Schools (n = 9)	Year 1 (Spring 2019) & Year 2 (Spring 2020)

C1.c	Principal supports staff in implementing circles-based advisory program/SEL curriculum in classrooms and restorative interventions (activity 1.2.3)	Principals (school-level)	Staff Developer Interview Criteria = 3 yes/no questions about time, RI personnel, staff/student ratio in circles; 2 yes/no questions in elementary level	1-hour phone interview with staff developer twice a year (Oct & May); data from 2 timepoints will be combined for one fidelity score per year	0 (low) = 0-1 out of 3 criteria are met (Y/N) 0 (low) Elementary = 0 out of 2 criteria are met 1 (moderate) = 2 out of 3 criteria are met (Y/N) 1 (moderate) Elementary = 1 out of 2 criteria are met 2 (high) = 3 out of 3 criteria are met (Y/N) 2 (high) Elementary = 2 out of 2 criteria are met			All Treatment Schools (n = 9)	Year 1 (Spring 2019) & Year 2 (Spring 2020)
C1. d	Principal participates in one-on-one coaching in RP/SEL and equity leadership with a retired-principal coach.	Principal (school-level)	Morningside Center report	Morningside Center Program Staff will report number of principal coaching sessions with the retired-principal coach to evaluation team once a year in June	0 (low) = principal participated in 0-4 RP/SEL/equity coaching sessions 1 (moderate) = principal participated in 5-9 RP/SEL/equity coaching sessions 2 (high) = principal participated in 10 or more RP/SEL/equity coaching sessions			All Treatment Schools (n = 9)	Year 1 (Spring 2019) & Year 2 (Spring 2020)
C1.e	RP Leadership Team and staff developer lead staff conferences/workshops/PD/conversations addressing issues of equity (SEL, implicit bias, culturally relevant education, empathetic mindset, countering stereotype threat, disproportionality, etc.; activity 2.2.2)	RP Leadership Team (school-level)			0 (low) = 0 staff workshops/PD/ conferences on equity occurred 1 (moderate) = 1 staff workshops/PD/ conferences on equity occurred 2 (high) = 2 or more staff workshops/PD/ conferences on equity occurred			All Treatment Schools (n = 9)	Year 1 (Spring 2019) & Year 2 (Spring 2020)
All indi					School level implementation scores in C1 range from 0 – 10 in year 1.	School level threshold: adequate implementation =	Adequate implementation at program level	All schools in which intervention	Years 1 (Spring 2019) &

cato rs					School level implementation scores in C1 range from 0 – 8 in year 2. Overall Principals and School Leadership Implementation <u>in Year 1</u> High: 10 out of 10 Moderate: 5-9 out of 10 Low: 0-4 out of 10 <u>Year 2</u> High: 8 out of 8 Moderate: 4-7 out of 8 Low: 0-3 out of 8	moderate-high (5-10 in year 1 and 4-8 in year 2)	= 75% of schools (7/9) with adequate implementation in years 1 and 2	is being implemente d (n = 9 schools)	Year 2 (Spring 2020)

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score & threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Component 2: Training and coaching in Restorative Interventions (RI) for schoolwide RP/SEL leadership									
C2.a	RI Team members (key school discipline staff e.g., Principals, APs, deans, and guidance counselors) participate in 10-hour / 2-day Tier II: Restorative Intervention training	Restorative Intervention Team (school-level)	Morningside Center staff developer	Morningside Center staff developer submit rosters with training completion checklist (Oct & May)	0 (low) = 0 staff trained in 10-hour/2-day Tier II: RI 1 (moderate) = less than 1:150 ratio of staff trained in 10-hour/2-day Tier II: RI (less than 1 RI trained staff per 150 students) 2 (high) = 1:150 ratio of staff trained in Tier II: RI (1 RI trained staff per 150 students)			All Treatment Schools (n = 9)	Year 1 (Summer/Fall 2018)
C2.b	RI Team members facilitate Restorative Intervention*s to address "more serious" harm** and discipline infractions	Restorative Intervention Team (school-level)	Staff Developer 15-minute monthly fidelity survey & Principal interview	Morningside Center Staff Developer will record data and report in a monthly 15-minute fidelity survey check-in through Qualtrics or phone & 45-minute phone interview with school principal in November of each program year	0 (low) = 0%-49% of "more serious"** discipline infractions are being handled restoratively* 1 (moderate) = 50-74% of "more serious"** discipline infractions are being handled restoratively* 2 (high) = 75% or more of "more serious"** discipline infractions are being handled restoratively*			All Treatment Schools (n = 9)	Years 1 (Spring 2019) & Year 2 (Spring 2020)
All indicators					School level implementation scores in C2 range from 0 – 4 Overall RI Team Implementation at the school level: High: 4 out of 4 Moderate: 2-3 out of 4 Low: 0-1 out of 4	School level threshold: adequate implementation = 2-4 (moderate-high) in years 1 and 2	Adequate implementation at program level = 75% of schools (7/9) with adequate (high) implementation in years 1 and 2	All schools in which intervention is being implemented (n = 9 schools)	Years 1 (Spring 2019) & Year 2 (Spring 2020)

* Restorative Interventions: a response to a behavior incident / discipline infraction including a problem-solving process (what happened, what people were thinking/feeling, and who was impacted) AND actions for repairing the harm, skill-building, or community building (e.g., community problem solving, peer mediation, restorative conversation, restorative conference, restorative circle, re-entry circle)

** "more serious" harm: any incident or behavior infraction in which an additional adult (i.e. administrator, dean, security agent, discipline team member) is needed to respond to the situation, OR any incident or behavior infraction when a student is asked to leave the class

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score & threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Component 3: Staff differentiated training in RP/SEL and equity									
C3.a	Staff participate in differentiated levels of training in curriculum implementation, circle keeping, and RP/SEL and equity	School Level (% of staff trained in any of the Morningside Center training models)	Morningside Center staff developer	Morningside Center staff developer submit rosters with training completion checklist (Oct & May)	0 (low) = 0%-29% of staff complete training 1 (moderate) = 50%-79% of staff complete training 2 (high) = 80%-100% of staff complete training			All Treatment Schools (n = 9)	Year 1: Summer/Fall 2018 & Year 2: Summer/Fall 2019
C3.b	Curriculum-implementer staff participate in coaching with a Morningside Center Staff Developer to support carrying out the RP/SEL & equity curriculum and continuing to develop SEL skills and facility with RP and equity (activity 1.4.5)	School Level (% of curriculum implementers receiving coaching)	Staff Developer 15-minute monthly fidelity survey	Morningside Center Staff Developer will record data and report in a monthly 15-minute fidelity survey check-in through Qualtrics or phone	0 (low) = 0%-49% of curriculum-implementers participate in at least 2 push-in coaching sessions 1 (moderate) = 50%-79% of curriculum-implementers participate in at least 2 push-in coaching sessions 2 (high) = 80% or more of curriculum-implementers participate in at least 2 push-in coaching sessions			All Treatment Schools (n = 9)	Year 1 (2018-2019) & Year 2 (2019-2020)
C3.c	Curriculum-implementer staff facilitate consistently (weekly) scheduled RP/SEL circles with	School Level (% of curriculum implementers holding weekly	Staff Developer Interview & Principal interview	1-hour phone interview with staff developer twice a year (Oct & May); data from 2 timepoints	0 (low) = 0%-49% of curriculum-implementers hold consistent, weekly circles with students			All Treatment Schools (n = 9)	Year 1 (2018-2019) & Year 2 (2019-2020)

	students (activity 1.4.2)	community-/skill-building circles)		will be combined for one fidelity score per year & 45-minute phone interview with Principal in November of years 1 and 2	1 (moderate) = 50%-79% of curriculum-implementers hold consistent, weekly circles with students 2 (high) = 80% or more of curriculum-implementers hold consistent, weekly circles with students				
All indicators					School level implementation scores in C3&4 range from 0 – 6 Overall School Staff Implementation at the school level: High: 6 out of 6 Moderate: 4-5 out of 6 Low: 0-3 out of 6	School level threshold: adequate implementation = 4-6 (moderate-high) in years 1 and 2	Adequate implementation at program level = 75% of schools (7/9) with adequate implementation in years 1 and 2.	All schools in which intervention is being implemented (n = 9 schools)	Years 1 (Spring 2019) & Year 2 (Spring 2020)

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score & threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Component 4: Staff developer support for student participation in circles, facilitation of circles, and leadership									
C4.a	Staff Developers train selected students as circle keepers (activity 2.3.1)	School-level	Staff Developer Interview	1 hour phone interview with staff developer twice a year (Oct & May); data from 2 timepoints will be combined for one fidelity score per year	0 (low) = 0 students trained to be circle keepers 1 (moderate) = 1-12 students trained to be circle keepers 2 (high) = ≥13 students trained to be circle keepers			All treatment schools (n = 9)	Year 2 (Spring 2020)
C4.b	Opportunities for student leadership are systematically added to school and authentic student voice is incorporated in school culture as a result of the project	School-level	Staff Developer Interview	1-hour phone interview with staff developer twice a year (Oct & May); data from 2 timepoints will be combined for one fidelity score per year	0 (low) = no additional opportunities for student leadership or incorporating student voice result in the project 1 (moderate) = at least 1 additional opportunity for student leadership and incorporating student voice 2 (high) = multiple additional opportunities for student leadership and incorporating student voice			All treatment schools (n =9)	Year 2 (Spring 2020)
All indicators	NA	NA	NA	NA	School level implementation scores in C4 range from 0 – 4 Overall implementation in year 2 High: 4 out of 4 Moderate: 2-3 out of 4 Low: 0-1 out of 4	School level threshold: adequate implementation = 2-4 (moderate-high) in year 2	Adequate implementation at program level = 75% of schools (7/9) with adequate implementation in year 2	All schools in which intervention is being implemented (n = 9 schools)	Year 2 (Spring 2020)

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to program level (score & threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Component 5: Opportunities for Family/Caregiver participation in circles									
C5.a	Parent Coordinators (designated staff in each NYC school) participate in RP/SEL and equity training	School level	Morningside Center staff developer	Morningside Center staff developer submit rosters with training completion checklist (Oct & May)	0 (low) = Parent Coordinator is not trained 1 (moderate) = Parent Coordinator completes partial RP/SEL & equity training 2 (high) = Parent Coordinator completes RP/SEL & equity training			All treatment schools (n = 9)	Year 1 (Spring 2019) & Year 2 (Spring 2020)
C5.b	Opportunities with families are offered to introduce and support RP/SEL and equity	School level	Staff Developer 15-minute monthly fidelity survey	Morningside Center Staff Developer will record data and report in a monthly 15-minute fidelity survey check-in through Qualtrics or phone	0 (low) = 0-2 RP opportunities held for families 1 (moderate) = 1-3 RP opportunities held for families 2 (high) = 4 or more RP opportunities held for families			All treatment schools (n = 9)	Year 1 (Spring 2019) & Year 2 (Spring 2020)
All indicators					School level implementation scores in C5 range from 0 – 4 Overall Caregiver Implementation High: 4 out of 4 Moderate: 2-3 out of 4 Low: 0-1 out of 4	School level threshold: adequate implementation = 2-4 (moderate-high) in years 1 and 2	Adequate implementation at program level = 75% of schools (7/9) with adequate implementation in years 1 and 2	All schools in which intervention is being implemented (n = 9 schools)	Years 1 (Spring 2019) & Year 2 (Spring 2020)