



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Reducing Disproportionality in School Discipline among Black Male High School Students: A Randomized Evaluation of a Comprehensive, Whole-School Intervention

Author(s): Anna Yaros, Cheryl Roberts, James Trudeau, Alan Barnosky, Jason Williams, Chloe McIntosh, Alex Rabre, Olga Khavjou, Kayla Keys

Document Number: 309969

Date Received: January 2025

Award Number: 2017-CK-BX-0015

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

April 2023

Reducing Disproportionality in School Discipline among Black Male High School Students: A Randomized Evaluation of a Comprehensive, Whole-School Intervention

Final Report

Prepared for

National Institute of Justice
U.S. Department of Justice
Attention: Cathy Girouard
810 7th Street, NW
Washington, DC 20531

Prepared by

**Anna Yaros, Cheryl Roberts, James Trudeau, Alan Barnosky,
Jason Williams, Chloe McIntosh, Alex Rabre, Olga Khavjou,
Kayla Keys**

RTI International
3040 East Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0216140

This project was supported by Award No. 2017-CK-BX-0015, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.



This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Contents

Executive Summary	7
1. Introduction	13
1.1 The Impact of Discipline Disproportionality on Black Male Students	13
1.2 Interventions to Address Disproportionality in Discipline	14
1.3 Advances in Analysis of Discipline Disproportionality	15
2. Method	17
2.1 Three-Pronged Intervention.....	17
2.1.4 Coaching of the Interventions	19
2.2 Study Goals, Design, and Methods.....	19
2.2.1 Study Goals.....	19
2.2.2 Study Design	20
2.2.3 Study Methods.....	21
2.3 Data Sources.....	22
2.3.1 Process Evaluation.....	22
2.3.2 Outcome Evaluation.....	24
2.3.3 Cost and Benefit Data	26
2.4 Analytic Plan	27
2.4.1 Process Evaluation.....	27
2.4.2 Outcome Evaluation.....	29
2.5 Cost Evaluation.....	30
2.5.1 Analytic Perspective and Timeframe	30
2.5.2 Cost Estimation.....	30
2.5.3 Benefit Estimation	31
3. Results	33
3.1 Implementation	33
3.1.1 Training Delivery	33
3.1.2 Fidelity of Implementation	36
3.1.3 Implementation Challenges.....	42
3.1.4 Implementation Facilitators.....	45
3.1.5 Other Future Considerations	49
3.2 Leadership and Teacher Perspectives on Early Results	50
3.3 Outcome Evaluation	55

3.3.1	Administrative Data Disproportionality Metrics	55
3.3.2	Hypothesis Tests	61
3.3.3	Staff Survey Outcomes	69
3.3.4	Student Survey Outcomes	74
3.3.5	Student Focus Group Outcomes.....	78
3.4	Cost Evaluation.....	81
3.4.1	Economic Analysis Results	81
3.4.2	Cost Analysis	81
3.4.3	Net Benefit Analysis.....	82
4.	Discussion	85
4.1	Implementation	85
4.1.1	Study Challenges and Limitations	85
4.1.2	Implementation Levels.....	87
4.1.3	Implementation Fidelity.....	88
4.1.4	Challenges and Facilitators	90
4.2	Outcomes	92
4.2.1	Administrative Data	93
4.2.2	Staff Survey Results	96
4.2.3	Student Survey Results	97
4.2.4	Qualitative Outcomes from Student Focus Groups	97
4.3	Cost Evaluation.....	98
4.4	Interpretation of the Study	99
4.5	Implications	100
	References	101

Figures

Number		Page
Figure 3.1.	ODR Risk Ratios, by Group and School Year	56
Figure 3.2.	ODR RDR per 100 Students, by Group and School Year.....	57
Figure 3.3.	ODR Disciplinary Rate, by Group and School Year	58
Figure 3.4.	OSS Risk Ratio, by Group and School Year	59
Figure 3.5.	OSS RDR-100, by Group and School Year	60
Figure 3.6.	OSS Disciplinary Rate, by Group and School Year.....	61
Figure 3.7.	ODR Outcomes, by Year, Treatment Condition, and Race	62
Figure 3.8.	OSS Outcomes, by Year, Treatment Condition, and Race.....	63
Figure 3.9.	Number of Absences, by Year, Treatment Condition, and Race.....	64
Figure 3.10.	Changes in OSS Binary Outcome as a Function of Year 2 TFI Scores	66
Figure 3.11.	Changes in ODR Count Outcome as a Function of Year 1 SET Score	67
Figure 3.12.	Average Staff Rating of School Climate, by Survey Wave and Treatment Condition	70
Figure 3.13.	Average Staff Rating of Authoritative School Climate, by Survey Wave and Treatment Condition	71
Figure 3.14.	Average Staff Rating of Perceived Safety, by Survey Wave and Treatment Condition	72
Figure 3.15.	Average Staff Rating of Culturally Responsive Self-Efficacy, by Survey Wave and Treatment Condition.....	72
Figure 3.16.	Student Surveys' CRT Scale Scores by Wave, Treatment Condition, and Race 75	
Figure 3.17.	Additional Student Survey Scale Scores by Wave, Treatment Condition, and Race	76
Figure 3.18.	Student Surveys' CRT Scale Score by RP Implementation Level, Wave, and Race.....	78

Tables

Number	Page
Table 2.1. Three-Pronged Intervention Approach	18
Table 2.2. Evaluation Questions	21
Table 2.3. Staff Survey Constructs, Measures and Reporting Time Frames.....	25
Table 2.4. Measures and Data Sources for Resource Quantities and Unit Costs.....	27
Table 2.5. Estimates and Sources of Lifetime Economic Outcomes of Dropping Out 32	32
Table 3.1. Intervention Training Start Dates and Grades Targeted by School	33
Table 3.2. PBIS Tier 1 SET Scores, Treatment and Control Schools, Year 1	37
Table 3.3. PBIS TFI Score in Treatment Schools, Year 2.....	38
Table 3.4. Percentage of Ninth Grade Teachers in Treatment Schools Reporting Facilitating RP Circles in the Classroom.....	39
Table 3.5. Percentage of 9th Grade Teachers Reporting Feeling Confident Facilitating RP Circles in their Classroom, Treatment Schools	41
Table 3.6. Level of CRI Training and Early Implementation, Treatment Schools, Spring 2020 (Pre-COVID-19).....	41
Table 3.7. Significance Tests of Associations between Outcomes and SET/TFI Implementation Levels.....	66
Table 3.8. Significance Tests of Associations between Outcomes and RP Implementation Levels.....	68
Table 3.9. Significance Tests of Associations between Staff Survey Outcomes and SET/TFI Implementation Levels.....	73
Table 3.10. Significance Tests of Associations between Staff Survey Outcomes and RP Implementation Levels	74
Table 3.11. Significance Tests of Associations between Student Survey Outcomes and RP Implementation Levels	77
Table 3.12. Significance Tests of Associations between Student Survey Outcomes and PBIS Implementation Levels	77
Table 3.13. Total Activity Costs by Treatment Condition	82
Table 3.14. Mean Costs per School and per Student.....	82
Table 3.15. Monetized Benefits per Student	84
Table 3.16. Per Student Benefit, Cost, and Net Benefit	84

List of Abbreviations and Acronyms

ASC	Authoritative School Climate Survey
CMS	Charlotte-Mecklenburg Schools
COVID-19	coronavirus disease 2019
CRI	culturally responsive instruction
ISS	in-school suspension
MTSS	multi-tiered systems of support
NIJ	National Institute of Justice
ODR	office disciplinary referral
OSS	out-of-school suspension
PBIS	positive behavioral intervention supports
PLC	professional learning community
RDR	raw differential representation
RDR-100	raw differential representation per 100 students
RP	restorative practices
SEL	social-emotional learning
SET	School-wide Evaluation Tool
TFI	Tiered Fidelity Inventory

Executive Summary

The project team endeavored to study how a combination of three interventions could prevent the racially disproportionate levels of exclusionary discipline observed in high schools in a large southeastern U.S. public school district. Like most districts in the United States, Charlotte Mecklenburg Schools' Black male high school students were more likely to experience exclusionary discipline, which means being suspended or expelled from school, than their non-Black male or female student peers. This disparity is apparent nationally from 2015–2016 data available from the U.S. Commission on Civil Rights showing that, Black students are subject to more suspensions and expulsions (2019). Discipline disparities are associated with negative outcomes for Black male students, including lower levels of school connectedness, school engagement, and academic achievement (Skiba et al., 2018). These challenges can also relate to increased dropout and arrests (Monahan et al., 2014).

Literature points to several evidence-based interventions that can be implemented throughout an entire school and that show promise in some cases in reducing exclusionary discipline across students and reducing the disparate enforcement on Black students.

Positive Behavior Interventions and Supports (PBIS) has long been used to reduce office discipline referrals in high schools (Flannery et al., 2014; Freeman et al., 2016), but to date has not been sufficient by itself to reduce disproportionality among Black students compared to non-Black students. **Restorative Practices (RP)** have shown similar results in the reduction of exclusionary discipline (see Lodi et al., 2021, for a review); results from experimentally controlled studies are mixed as to whether or not discipline disparities for Black students were reduced (e.g., Augustine et al., 2018; Acosta et al., 2019). **Culturally responsive instruction (CRI)** directly targets the cultural competency of teachers with the goal of improving the relevance of classroom material and increasing the engagement of students of diverse cultures in schools.

Our study is a randomized controlled trial of the bundle of these three interventions in high schools in a diverse Southeastern US school district that had previously documented discipline disparities in most of its schools. About half (54%) of students in the district were identified as economically disadvantaged. The three-pronged intervention approach was designed to implement PBIS as an overall framework within which multi-tiered intervention teams provided support for evidence-based practices. PBIS also included data-driven decision-making and classroom management skills training. RP involved training in and implementation of proactive community-building and circles to support relationship-building among students and staff in classrooms. Lastly, CRI involved training and practice in lesson design for Black male students and a brief training on implicit bias. In addition to the three distinct interventions, trained intervention coaches provided support for implementation in each individual school that was randomized to receive the intervention. Our study had five goals to understand and report on the implementation and impact of the intervention.

- **Goal 1: Intervention implementation.** With support from the study grant, CMS will implement three interventions in eight randomly selected high schools (labelled treatment schools), providing professional development, coaching, and data-driven feedback for staff in PBIS, RP, and CRI.
- **Goal 2: Process evaluation.** RTI International will conduct an in-depth process evaluation of the extent to which and the fidelity with which each component is implemented in treatment schools in comparison with control schools.
- **Goal 3: Outcome evaluation.** RTI will evaluate the effectiveness of the three-intervention model in reducing disparities in school discipline for Black male students, including office discipline referrals (ODRs), out-of-school suspensions (OSSs), and expulsions. We will evaluate the effectiveness of the interventions in creating a more positive and safe school climate for every student and determine how implementation affects outcomes.
- **Goal 4: Cost-benefit analysis.** RTI will complete a cost-benefit analysis to estimate the net benefits to society minus the cost of the three-intervention model in schools based on academic, economic, and criminal justice outcomes for students.
- **Goal 5: Dissemination.** CMS and RTI will disseminate study findings and develop recommendations for research, policy, and practice to reduce discipline especially among Black male students.

Method. A total of 16 high schools were randomized to treatment or control conditions using propensity score matching. Propensity scores were created using a range of variables representing the characteristics of each school. Schools were matched in pairs based on the similarity of their propensity scores. In each pair, one school was randomly assigned to the treatment condition and the other to the control condition. One school assigned to the treatment condition was unwilling to engage in the intervention and was re-assigned as a control school, resulting in seven treatment schools and nine control schools.

The study began during the 2018-2019 school year. School disruptions caused by the coronavirus disease 2019 (COVID-19) included school closures in March 2020 (Year 2) followed by remote learning until February 2021 (Year 3). This meant that implementation of PBIS was interrupted by COVID-19 and that full implementation of RP and CRI was impeded.

The process evaluation relied on qualitative and quantitative data from implementation and fidelity measures (e.g., training data, observations, staff survey) as well as focus groups. The outcome evaluation relied on a staff survey, student survey, and administrative data about office discipline referrals, attendance, out-of-school suspensions, and in-school suspensions. For the 2019-2020 school year, we used administrative data for the period until the school building closures in March 2020. In Year 3, administrative data varied

significantly from previous years because instruction was conducted for part of the year using remote learning, for part of the year using in-person learning, and a hybrid of the two. Because of this, we did not use administrative data as outcomes from the 2020-2021 school year (Year 3). The cost benefit analysis used costs of implementation of the three interventions along with benefit estimations based on the literature about the long-term economic consequences of dropping out coupled with attendance records from administrative data.

Results: Process Evaluation. Results of the process evaluation describe the staggered implementation of the three interventions and the severe impact of school disruption due to the COVID-19 pandemic. PBIS was implemented near the beginning of Year 1 (2018-2019) in all seven treatment schools, except for one school, in which staff did not receive PBIS training until the end of Year 1. RP was implemented with less regularity, with one school having staff trained a year prior to the grant beginning, two schools not receiving any training, one school receiving training in Year 1, and three schools receiving training in Year 2 (2019-2020). Six of seven treatment schools trained at least some staff in CRI in August 2019 (Year 2). Throughout the first year, each school had a coach to assist them in training and implementation. Coaching support decreased in Year 3.

Fidelity observations found that treatment schools, in general, showed more fidelity toward PBIS during Year 1 than control schools, as expected. Only one of seven schools reached full implementation fidelity for PBIS in Year 1 as measured by observations. In Year 2, fidelity was measured by self-report and four schools achieved full PBIS implementation fidelity. In Year 1, only three of seven treatment schools were engaging in RP circles in their 9th grade classrooms. By Years 2 and 3, five of seven schools were implementing RP to some degree, though lower percentages of teachers were implementing circles in Year 3 (post-COVID) than in Year 2. Implementation of CRI was rated as low, medium, and high. One of seven schools was not engaging in CRI by Spring 2020 (Year 2), three schools were rated as low implementers, two schools were rated as medium implementers, and one school was rated as having high implementation. With the interruption by the pandemic, none of the schools were able to fully implement the three-intervention model. Therefore, our outcome evaluation could not test the effects of the conceptual model as designed but instead assessed early effects of partial implementation of PBIS in most schools and initial limited implementation of RP or CRI in some schools.

Qualitative data provided important contextual information about the challenges and facilitators of implementation. Several early challenges identified as themes related to unclear expectations, delayed start of coaching, need to engage more teachers in leadership roles in some schools, and dissatisfaction with initial RP training. Challenges that persisted included other competing district initiatives, the complexity of implementing three interventions, lack of teacher buy-in, demands on staff time, need for more skill-building among staff, staff turnover, and lack of implementation plans for RP. We continued to

assess challenges after the onset of COVID-19, which included problems with student engagement in remote learning, overload of staff, and student behavior problems. Facilitators for implementation of the three interventions include coaching, administration and teacher champions, teacher buy-in, and other intervention-specific structures of support.

Results: Outcome Evaluation. Disproportionality of discipline was measured using administrative data that included ODRs and OSSs. The first disproportionality metric we used was the risk index (the proportion of students in a group that received one or more ODRs), which was higher for Black male students at baseline than their non-Black and non-male peers. Risk ratios showed the risk index of Black male students divided by the risk index of non-Black male students. Risk ratios for ODRs did not show a significant improvement over time in the treatment schools compared to the control schools. Similar analyses were completed for OSSs. Risk ratios indicated that the risk of at least one OSS for Black males in treatment schools, compared to non-Black males, decreased from Year 1 to Year 2, but still remained greater than risk ratios in control schools.

Analyses were conducted to test the hypothesis that treatment schools would show an improvement in the disproportionate amount of discipline that Black males experienced compared to non-Black males. Comparisons of year-to-year changes in ODRs, OSSs, and absences among Black male students and non-Black male students in treatment schools and control schools offered mixed or limited support for the hypothesis. In treatment schools, ODRs sharply increased from baseline to Year 1 for both Black and non-Black male students, though increases were less for non-Black male students. These increases were not seen in control schools where the variable indicating receiving one or more decreased from Year 0 to Year 1 for Black male students and was relatively flat for non-Black male students. The increases from Year 0 to Year 1 in the ODR binary seen in treatment schools did not persist during Year 2.

The delays in implementation and challenges discussed above weakened any differences we would have expected to see between the treatment and control groups. Therefore, we conducted additional analyses that took advantage of the variation in levels of implementation among treatment schools. One statistically significant finding, in Year 1, showed that Black male students in schools with lower PBIS implementation were more likely to receive an OSS than students in schools with average PBIS implementation, and even more likely than students in schools with high PBIS implementation. For non-Black male students, the likelihood of receiving an OSS was not related to PBIS implementation. For analyses using the percentage of teachers who reported leading RP circles as the implementation measure, there were no statistically significant interactions of implementation level, year, and race.

Our staff survey showed improvements in staff ratings of school climate over each year for measures related to overall climate and authoritative school climate (meaning climate characterized by equal support and structure for students). No significant effects were found for measures of staff perceived safety or their self-ratings of their culturally responsive self-efficacy.

The most striking finding from analysis of the student survey data was that at Wave 3, administered in spring of the 2020–2021 school year, after in-person instruction resumed and following implementation of aspects of the intervention in the preceding 2 school years. Black male students in treatment group schools showed a large increase in their reports of their teachers' level of CRI. This finding suggests benefits of the targeted intervention.

In spring of Year 3, focus groups with Black male students in three treatment schools provided an opportunity to explore their daily perceptions of equity in their schools. Findings were mixed. In two schools, students reported that staff generally applied discipline equally, regardless of students' race and gender; however, participants from one school (which did not have improvements in PBIS fidelity scores), reported that discipline was harsher for Black students. For RP, students reported no current restorative circles, but were able to identify being engaged in restorative circles in at least one of their classrooms prior to the pandemic. They said that engaging in restorative practice circles improved their relationships with their peers and allowed sharing perspectives between different types of peers. Students were also asked whether they could relate to class lessons and the extent to which they reflected their culture and background. Overall, students reported being able to relate to some lessons, but not most of them.

Results: Cost-benefit Analysis. We conducted a cost-benefit analysis by calculating the cost per student of the project at treatment schools (about \$62) in comparison to the cost per student at control schools (about \$2). To calculate the benefit for treatment and control schools, we used absences as a proxy for dropout, which can predict financial outcomes including earnings, health, property crime, and arrests. The difference of the two benefit estimates indicates that the treatment group results in a benefit loss of \$268 per student compared to control (though not statistically significant), and the difference of the two cost estimates indicates the treatment intervention costs \$60 per student. The net benefit is the difference of these differences, and results in a total loss of \$328 per student for the treatment group. This loss may be due to a lack of effect of the intervention on absences.

Discussion. The implementation of a three-pronged approach to reducing discipline disproportionality for Black male high school students was met with a number of challenges, including the COVID-19 pandemic. The complexity of three interventions integrated together also caused implementation problems. None of the schools were able to fully implement the three-intervention model in the time period before the pandemic, but all implemented some elements of the intervention. We think this was the primary reason that

we did not detect significant improvements in discipline disparities for Black male students in the treatment condition. Despite the limitation, we had several significant results that pointed to the benefit of our treatment. First, absences appeared to go down for Black male students in our treatment schools during Year 1 of treatment, though this effect was not observed in Year 2. Another important finding was improvements in two measures of school climate (as rated on staff surveys) within treatment schools. This suggests that a schoolwide intervention, such as PBIS, may have impacts directly on the perceptions of teachers. Those perceptions may take more time than the study had to flow down to actions or student behaviors.

This study has important implementation for whole school interventions. We learned that they are quite challenging, especially in the high school setting. Moreover, we learned that integrating multiple interventions in high schools can take much longer than expected and requires more training and follow-up than were possible during this study. Overall, we experienced successes in using intervention coaching to provide implementation support. The result was an improvement in some metrics for staff and students across our study. More research on the interventions of PBIS, RP, and CRI at the high school level, both in isolation and in combination, is needed to understand how they affect students and disproportionate discipline.

1. Introduction

Black male students in schools across the United States are suspended, expelled, and referred for disciplinary action at a higher rate than their peers. These disparities exist regardless of the type of disciplinary action, level of school poverty, or type of public school attended (Nowicki, 2018). Exclusionary discipline—defined as suspensions and expulsions—takes youth out of school. Black students across the United States missed five times as many days of school than White students due to out-of-school suspensions (OSSs) in the 2015–2016 school year (U.S. Commission on Civil Rights, 2019).

Research suggests that racial disparities in discipline are not explained by actual differences in youth behavior (Huang & Cornell, 2017; Skiba et al., 2014) because Black male students do not commit more disciplinable offenses than their peers in school (Commission on Civil Rights, 2019). Contributing factors for this race gap appear to include lack of teachers' skills in classroom management (Vavrus & Cole, 2002), lack of cultural competence (Skiba et al., 2018), and implicit bias related to racial stereotypes (Staats, 2014). Implicit biases have a substantial effect on subjective decision-making, and therefore can influence how teachers and school personnel interpret and respond to subjective discipline infractions, which typically have the highest levels of disproportionality for Black students (Gregory & Weinstein, 2008).

1.1 The Impact of Discipline Disproportionality on Black Male Students

Racial disparities in discipline at the school level are associated with Black students reporting lower levels of school belonging and equity and more adjustment problems (Skiba et al., 2018). Exclusionary discipline also results in lower school engagement and academic achievement among Black male students (Commission on Civil Rights, 2019). Harsh discipline policies and practices also negatively affect the overall school climate, making schools less safe and supportive (Commission on Civil Rights). Data suggest that zero-tolerance policies have not made schools safer, despite that being their goal (Skiba, 2013).

Zero-tolerance policies, which rely on exclusionary discipline, have sharply increased the number of OSSs, expulsions, and referrals to the juvenile justice system for behaviors that were previously handled in school (American Psychological Association Zero Tolerance Task Force, 2008). The process, often referred to as the "school-to-prison pipeline," disproportionately affects minority students (Fabelo et al., 2011). Students who are suspended or expelled have much greater risk of dropping out, truancy, and arrest (Monahan et al., 2014). Given that the youth most likely to be arrested or dropout are in the high school age range, our study's focus on high schools is particularly pertinent.

1.2 Interventions to Address Disproportionality in Discipline

To address the complex problem of racial disparities in discipline requires a multifaceted approach. The three approaches in our study are briefly discussed here: positive behavioral intervention supports (PBIS), restorative practices (RP), and culturally responsive instruction (CRI).

1.2.1 PBIS

PBIS is a schoolwide, data-driven “systems approach to establishing the social culture and behavior supports” that foster children’s social and academic success (Horner et al., 2007). PBIS teaches and reinforces consistent, positive behavioral expectations throughout the school system and provides three levels, or tiers, of behavioral support, depending on student need. PBIS has been most widely implemented in elementary and middle schools, where rigorous evaluations, including multiple randomized trials, have shown subsequent reductions in behavioral problems and ODRs, increases in perceived school safety, increases in academic achievement, and improved school organizational effectiveness (for elementary school, see Bradshaw et al. 2012, and Horner et al., 2009; for middle school, see Caldarella et al., 2011).

PBIS is now being implemented in thousands of U.S. high schools. Two recent quasi-experimental studies of PBIS in high schools have shown reductions in ODRs (Flannery et al., 2014; Freeman et al., 2016), but no randomized controlled studies have tested PBIS in high schools. Although PBIS has been very effective in reducing ODRs, PBIS has not been sufficient by itself to reduce disproportionality among Black students (Skiba et al., 2014; Vincent et al., 2011; Vincent & Tobin, 2011). The literature suggests that multiple strategies are needed to address this disproportionality problem. The Discipline Disparities study presented in this report used PBIS in high schools in combination with two research-informed practices: RP and CRI.

1.2.2 RP

RP has emerged as a promising approach to reducing disproportionality as well as total suspensions (Commission on Civil Rights, 2019; Lodi et al., 2021). RP evolved from restorative justice, an evidence-based practice in the criminal justice system used to repair harm to victims by offenders, but it expands upon this to include both prevention and response to discipline problems in schools. RP focuses on building healthy relationships and repairing harm, instead of relying on punishment and exclusion (International Institute for Restorative Practices, 2017). Indeed, a recent review of the literature showed that RP can be challenging to implement in schools but can show positive impacts in the areas of student development of social and emotional skills, school connectedness, positive school climate, and reductions in overall school discipline and conflict among students (Lodi et al., 2021).

Across the United States, RP has been implemented in many high schools with promising results, including declines in OSS and some reductions in disproportionality. In one study in Denver, CO, district suspension rates declined by nearly 47%, with the largest reductions among Black students (González, 2014). Gregory and colleagues reported a study of RP in 29 high school classrooms that showed fewer office discipline referrals (2016). Another study in 14 middle schools in Maine employing a cluster-randomized design did not show significant differences in discipline between intervention and control schools, but authors suggest that may have been due to implementation challenges (Acosta et al., 2019). A randomized trial in Pennsylvania from elementary through high school did find significant differences in discipline between intervention and control schools for elementary schools, non-statistically significant improvement for middle schools, and no improvements for high schools during the study period (Augustine et al., 2018). More research is needed to understand the effects of RP in reducing discipline disparities, especially in high schools. Some school systems are integrating RP into a PBIS framework (e.g., in Illinois and Minnesota; see Beckman, McMorris, & Gower, 2012), but this approach has yet to be rigorously evaluated.

1.2.3 CRI

In our increasingly diverse society, American educators need to be able to effectively engage and teach students whose cultural background differs from their own. To promote education equity and student empowerment, CRI draws on strengths of students' cultures to make learning relevant (Gay, 2020; Ladson-Billings, 1994). Gay (2000) defined CRI as "using the cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse students to make learning encounters more relevant to and effective for them" (p. 29). Building educators' knowledge, skills, and self-efficacy in CRI, in contexts where the backgrounds of students and teachers differ, requires training and professional development. CRI strategies include training to increase teachers' cultural competence and adapt lessons to be more culturally relevant for nondominant cultural groups. Research related to school disparities in achievement and discipline suggests that CRI may be an important part of the solution to reduce disparities in achievement and discipline (Tucker et al., 2005; Vincent et al., 2011). However, a review in 2018 suggested that rigorous quantitative research related to culturally responsive practices was too sparse to draw any conclusions (Bottiani et al., 2018). More research is needed to understand the impact of CRI, especially in combination with RP and in the framework of PBIS.

1.3 Advances in Analysis of Discipline Disproportionality

Disproportionality in discipline has typically been measured using risk ratios in the past, but new measures of disproportionality have emerged (Curran, 2020). In 2017, RTI submitted the proposal that led to this project: A Randomized Evaluation of a Comprehensive, Whole-School Intervention, also known as Discipline Disparities. As proposed, our primary outcome

evaluation focuses on the extent to which implementation of the intervention resulted in a smaller difference between Black male students and other populations in the rates of school discipline. Since that time, substantial advances have been made in understanding and improving metrics used to study disproportionality.

A seminal article was entitled (in reference to the allegory of men who are blind forming different impressions of an elephant): “Tail, Tusk, and Trunk: What Different Metrics Reveal About Racial Disproportionality in School Discipline” (Girvan et al., 2019). Girvan et al. strongly recommended that “researchers and policymakers should be deliberate in their specific aims in measuring discipline disproportionality and select a combination of metrics that provides information most responsive to their goals” (2019, p. 1). We took that advice to heart and present findings based on several metrics that provide complementary and comprehensive information on different aspects of disproportionalities in discipline.

Girvan et al. (2019) warned of limitations of, and challenges posed by, commonly used metrics:

Risk ratios are, for example, known to be sensitive to small changes, particularly when the number of individuals in one or both of the groups being compared is small, and to provide no information about differences or changes in the absolute rates or numbers of individuals actually impacted: Schools with very high or very low rates of discipline incidents may nevertheless have identical risk ratios. (p. 41)

A subsequent article by Curran (2020) built upon Girvan et al. (2019). This article described strengths and weaknesses of various disproportionality metrics and demonstrated “how the choice of metric of the Black-White discipline gap can drastically change conclusions about whether the gap is closing or widening” (Curran, 2020, p. 382).

We drew heavily upon both these articles in updating and strengthening our analytic approach. Therefore, this report presents descriptive information about various disproportionality metrics. We base our hypothesis-testing analyses not on summary metrics such as risk ratios, but instead on the student-level source data used to compute the disproportionality metrics. These source data are the administrative data on each student’s race and gender and the number of office disciplinary referrals (ODRs) and OSSs received each school year. We describe the disproportionality metrics we used and our outcome analysis methods in detail in **Section 2**.

2. Method

Our study is a randomized controlled trial of a bundle of three interventions to support school staff in reducing discipline disparities. PBIS, RP, and CRI were combined in randomly selected high schools in a large school district in the southeastern United States. Charlotte-Mecklenburg Schools (CMS), the 18th largest school system in the country, continually seeks new ways to improve the education and lives of its students, across all demographics. Like many districts nationwide, CMS has identified Black male youth as disproportionately likely to receive discipline referrals and suspension compared to other student populations. CMS is a semi-urban, culturally diverse school district located in the Piedmont region of North Carolina. It is ideal for testing this combination of interventions because of its large size and high-need student population. CMS has 164 schools, 144,000 students, and over 18,000 staff members. More than half (54%) of all CMS students are economically disadvantaged, and the district is racially and ethnically diverse, with the student body being 42% African American, 32% Caucasian, 18% Latino/a, and 8% another ethnicity.

2.1 Three-Pronged Intervention

To build on the existing literature to promote equity and reduce disparities in discipline and other areas to improve discipline disparities in high schools, we combined three types of schoolwide interventions—PBIS, RP, and CRI—to create a culturally responsive behavioral intervention at the school level.

2.1.1 *PBIS*

The overarching framework for our three-pronged intervention is the strongly evidence-based, schoolwide PBIS. PBIS was implemented according to the training and fidelity criteria outlined by the U.S. Office of Special Education Programs' National Technical Assistance Center on PBIS. PBIS employs three tiers to address the needs of all students (Tier 1), students exhibiting some problem behaviors (Tier 2), and students exhibiting significant problem behaviors (Tier 3). Across tiers, the primary elements of PBIS are the systems it creates within each school building, evidence-based practices that schools adopt for addressing behavior and discipline, and data-driven decision-making that allows continuous improvement in the system.

2.1.2 *RP*

RP is an intervention that impacts the quality of the interpersonal interactions between students and staff and provides alternatives to traditional discipline practices. For this initiative, CMS emphasized the proactive community-building circles. RP was supported by CMS trainers certified by the International Institute for Restorative Practices who have implemented RP in some CMS middle schools.

2.1.3 CRI

CRI is a process of working with educators to assist in their ability to engage students of different cultures, races and ethnicities with lesson plans, relevant instructional content, and learning tools. CRI is designed to reduce the effects of implicit biases in instruction by increasing the relevance of learning materials by aligning it to reflect the cultures of the students being taught. **Table 2.1** describes the elements of CRI taught to school staff, including intentional lesson design that incorporates culturally relevant examples to relate to Black male students. Staff also received an introduction to implicit biases to increase their awareness of this issue and their own potential biases.

The three interventions, as they were intended to be implemented, are described in Table 2.1.

Table 2.1. Three-Pronged Intervention Approach

Intervention Component	Level of Intervention	Interventionists	Elements of Intervention
PBIS	Infrastructure for school discipline	Certified CMS PBIS Trainers CMS Coaches Office of Special Education Programs' PBIS Technical Assistance Center	<ul style="list-style-type: none"> ▪ Systems <ul style="list-style-type: none"> – Three-tiered staff teaming: implementation teams – Classroom matrix: schoolwide behavior expectations – Coaching and technical assistance ▪ Evidence-based practices <ul style="list-style-type: none"> – Classroom management skills professional development – RP (see below) – CRI (see below) ▪ Data-driven decision-making <ul style="list-style-type: none"> – Educator's Handbook: software that tracks all ODRs and fidelity assessments – Classroom observations and performance feedback
RP	Discipline methods that address student-staff and student-student relationships	Certified CMS RP Trainers CMS coaches	<ul style="list-style-type: none"> ▪ Informal RP: Affective statements and questions ▪ Small impromptu circles ▪ Proactive circles ▪ Responsive circles and restorative conferences ▪ Coaching

(continued)

Table 2.1. Three-Pronged Intervention Approach (continued)

Intervention Component	Level of Intervention	Interventionists	Elements of Intervention
CRI	Individual students and staff bias	CMS coaches Consultation from: Dr. Chance Lewis (University of North Carolina at Charlotte)	<ul style="list-style-type: none"> ▪ Implicit bias training ▪ Teacher professional development <ul style="list-style-type: none"> – Parent–teacher communication – Intentional lesson design strategies for Black students

Note: Elements of the intervention approach were implemented to various levels, due in part to school disruption from the coronavirus disease 2019 (COVID-19) pandemic, as described in the process evaluation results section.

2.1.4 Coaching of the Interventions

The intervention model also included coaching. To help treatment schools implement these three interventions, in Year 1, CMS hired eight full-time coaches to work with each of the eight treatment schools. The coaches were intended to:

- provide guidance and technical assistance to school administrators and each school’s PBIS leadership team,
- collect and analyze schoolwide data for PBIS,
- provide formal group training (e.g., RP trainings) to schools and district staff,
- assessing teachers’ needs and progress in implementing the interventions; and
- provide coaching to individual teachers.

2.2 Study Goals, Design, and Methods

2.2.1 Study Goals

Our study was designed to understand the impact of three interventions on students, specifically Black male students¹, who were identified as the most at-risk group for exclusionary discipline in our participating school district, as is common among many U.S. school districts. The goals of the study are as follows:

- **Goal 1: Intervention implementation.** With support from the study grant, CMS will implement three interventions in eight randomly selected high schools (labelled treatment schools), providing professional development, coaching, and data-driven feedback for staff in PBIS, RP, and CRI.
- **Goal 2: Process evaluation.** RTI International will conduct an in-depth process evaluation of the extent to which and the fidelity with which each component is implemented in treatment schools change in comparison with control schools.
- **Goal 3: Outcome evaluation.** RTI will evaluate the effectiveness of the three-intervention model in reducing disparities in school discipline for Black male

¹ In the CMS administrative data, students’ parents reported their race and gender identities.

students, including ODRs, OSSs, and expulsions. We will evaluate the effectiveness of the interventions in creating a more positive and safe school climate for every student and determine how implementation variability affects outcomes.

- **Goal 4: Cost-benefit analysis.** RTI will complete a cost-benefit analysis to estimate the net benefits to society minus the cost of the three-intervention model in schools based on academic, economic, and criminal justice outcomes for students.
- **Goal 5: Dissemination.** CMS and RTI will disseminate study findings and develop recommendations for research, policy, and practice to reduce discipline especially among Black male students.

2.2.2 Study Design

In developing the research design, RTI worked closely with CMS to build on its current infrastructure and capacity to implement this combination of programs in treatment group high schools. We utilized a randomized controlled trial with school-level randomization of 16 of CMS' high schools into two groups, using propensity score matching². This resulted in one school being selected for the treatment group which was not willing to engage in the intervention. This school became a control school, thereby resulting in seven schools in the treatment and nine schools in the control condition.

The treatment group received the three components, including RP and CRI within the framework of schoolwide PBIS. School disruptions caused by the coronavirus disease 2019 (COVID-19) included school closures in March 2020 (Year 2) followed by remote learning until February 2021 (Year 3), with a hybrid of in-person and remote learning after schools re-opened. This meant that implementation of PBIS was interrupted by COVID-19 and that full implementation of RP and CRI was impeded. This is discussed further in **Section 3: Results.**

Exclusion criteria

Prior to randomization, nine CMS high schools were excluded from the study. Two schools that were already implementing PBIS were excluded from the pool of high schools. Seven small, nontraditional high schools were also excluded.

Control schools

The nine schools randomly selected as control schools continued to receive "treatment as usual" conditions for discipline and behavioral supports in CMS high schools. This includes adhering to the CMS Code of Conduct, which outlines unacceptable behavior. Other behavior programs are selected by each principal and implemented to varying extents, depending

² Propensity score matching was completed by creating propensity scores of being selected for the treatment condition based on a range of variables: enrollment numbers of students of different ethnicities, percentage of students receiving free and reduced lunch, whether or not the schools had completed previous RP training, out of school suspensions per 100 students, the risk index and the risk ratio for Black male students. Based on propensity scores, schools were matched with the school with the closest propensity score to it. One school from each pair was randomly assigned to the treatment condition while the other was assigned the control condition.

upon the school. To understand activities in the “treatment as usual” condition schools that may have overlapped with the interventions in treatment schools (PBIS, RP, and CRI), we obtained information prior to COVID-19 from the district about other interventions occurring in control schools.

2.2.3 Study Methods

This project builds upon inputs, including the broader Charlotte community context, extant CMS student services (i.e., CMS-certified PBIS and RP trainers), and CMS’ internal discipline structure at each high school. CMS leveraged these inputs to implement three interventions, including a schoolwide PBIS framework within which RP and CRI were conducted with the support of coaches. The process evaluation will measure outputs of these activities, such as the number of school staff who receive training and the fidelity and frequency with which the interventions were implemented. Finally, outcomes to be measured over time include changes in overall school levels and disproportionality in school discipline and attendance of Black male students, as reflected in administrative data provided by CMS. Student and staff surveys will measure outcomes of school safety and climate and self-efficacy, as well as staff implicit bias. The evaluation will answer the questions summarized in **Table 2.2**.

Table 2.2. Evaluation Questions

Area	Topic	Evaluation Questions	Hypotheses
Process Eval.	PBIS	To what extent is the PBIS framework established by schools? What are the barriers and supports to adopting tiered teams and implementing PBIS to fidelity in different high schools?	PBIS is expected to be fully implemented in some intervention schools more quickly than others with supports such as strong intervention coaches overcoming barriers such as the complexity of the high school settings.
	RP	Do school staff implement elements of RP with fidelity? How are RP implemented as an alternative to suspension, expulsion, or school-based arrest? What are the barriers and supports to implementing RP in high schools?	Elements of RP such as RP and proactive circles will likely need support from coaching to reach fidelity. Schools are expected to develop standardized alternatives to suspension and expulsion. Barriers may include a lack of administrator support or difficulty with staff and student buy-in.
	CRI	Does CRI reduce staff implicit biases? To what extent do school staff engage in culturally responsive professional development? What factors serve as barriers and supports for implementing CRI?	CRI, and implicit bias training in particular, is expected to reduce implicit bias in some staff. Barriers could include stigma against identifying ones’ own biases. Training will likely be associated with changes in teachers’ self-efficacy in cultural competence, and, ultimately, disproportionality outcomes for Black male students.

(continued)

Table 2.2. Evaluation Questions (continued)

Area	Topic	Evaluation Questions	Hypotheses
Outcome Eval.	Disproportionality outcomes	To what extent does implementation of each of the three interventions result in a difference between Black male students and other populations in the rates of school discipline? Does the intervention relate to lower disproportionality in rates of dropout? Does the intervention reduce disparities in reportable offenses?	Intervention schools will display less disproportionality between Black male students and other student populations in school discipline. Reduced disproportionality will be.
	School-level outcomes	Do overall school levels of discipline improve over time? To what extent does school safety and school climate improve over time in response to the intervention? To what extent does student and staff self-efficacy increase with the implementation of the intervention?	Schools receiving the interventions are expected to show overall lower levels of discipline problems, suspension, which are expected to be associated with improvements in school safety, school climate, and self-efficacy.
Cost-Benefit Eval.	Cost-benefit analysis	What are the benefits to society of the combination of these three interventions in relation to the cost?	The benefit to society of three intervention outcomes is anticipated to be greater than the cost of the intervention.

2.3 Data Sources

In this section we describe data used for each component of the evaluation: 1) for the process evaluation, the data on trainings, fidelity measures, and focus groups; 2) for the outcome evaluation, surveys of school staff and students as well as administrative data; and 3) for the cost study, estimated costs for staff time, materials, and space.

2.3.1 Process Evaluation

Training data (treatment schools)

Approximately twice a month, the school district project coordinator sent a Microsoft Excel spreadsheet with staff training data to RTI's project economist, who compiled the training data for analysis. The spreadsheet included data on type of training, total number of hours, and number and positions of staff trained. For the process evaluation, we used the training data to assess which schools received training for each intervention component and the start of training at each school. The training data were also used for the cost-benefit evaluation.

Fidelity data

PBIS schoolwide evaluation tool (treatment and control schools). PBIS, as a well-developed, evidenced-based framework, has extensive standardized resources for implementation and measurement. RTI used the School-wide Evaluation Tool (SET), which

provides a standard validated tool to measure the Tier 1 (universal) schoolwide PBIS implementation fidelity. The SET evaluates 28 research questions across seven feature areas: (1) expectations defined, (2) behavioral expectations taught, (3) acknowledgement procedures, (4) correction procedures, (5) monitoring and evaluation, (6) management, and (7) district-level support (Sugai et al., 2001; Todd et al, 2012). The SET provides a summary score of overall schoolwide implementation, as well as a score indicating the level of implementation of each of the seven feature areas. Schools scoring at least 80% overall and 80% in the area of teaching behavioral expectations have reached the high-fidelity threshold for implementing schoolwide positive behavioral support at a universal level.

To assess the level of PBIS implementation at treatment schools and the difference from control schools, two trained RTI evaluators, without knowledge of the schools' condition, administered this tool to treatment and control schools in Fall 2018 before implementation began (i.e., baseline) and then toward the end of Year 1, in Spring 2019. Data collection included interviews of the school principal, school staff, and students; document review; and observations in hallways and classrooms. The original evaluation plan called for three annual SET site visits; however, the COVID-19 pandemic prevented the RTI team from visiting the schools as planned in 2020 and 2021.

Tiered Fidelity Inventory (TFI; treatment schools). On a quarterly basis from Fall 2018 through Spring 2020, the district's school coaches for this initiative conducted the PBIS TFI with members of each treatment school's PBIS leadership team. The TFI provided a valid and reliable measure of the extent to which schools have the core features of PBIS in place (Algozzine et al., 2014). In contrast with the SET, the TFI was based on self-report of the PBIS leadership team, rather than including the interviews and walk-through observations. In addition, the TFI assessed the extent of implementation of all three tiers of PBIS: Tier 1, universal; Tier 2, targeted; and Tier 3, intensive. This initiative focused on implementing the Tier 1 level of PBIS, which provided the system of schoolwide behavioral supports for all students. The TFI included subscales to rate the elements within each tier, as well as an overall score for each tier. Schools scoring at least 70% overall for Tier 1 for three quarters were considered to have met the threshold for high fidelity for Tier 1.

School staff survey implementation data (treatment schools). Discussed in more detail in **Section 2.3.2 Outcome Evaluation** below, a staff survey included questions for treatment school staff related to fidelity of implementation of RP, such as whether teachers conducted RP community-building circles in their classroom and the frequency of the circles. Focus groups and meetings (treatment schools)

Focus groups with PBIS leadership team. During the study implementation period, RTI conducted 90-minute virtual focus groups each spring with representatives from the treatment schools PBIS leadership teams. This included a total of one focus group in April 2019, one in June 2020, and two in April–May 2021. The primary purpose of the focus

groups was to learn about implementation progress, barriers, and facilitators to implementation, and any perceived effects.

Focus groups with teachers. RTI also conducted four 90-minute virtual focus groups with ninth grade teachers, including two in May–June 2020 and two in November 2020. Staff focus groups aimed to learn about teachers’ perspectives on progress implementing these interventions during the 2019–2020 school year before the COVID-19 pandemic; the extent to which teachers were able to continue implementing any of the interventions during school closures; and barriers and facilitators to implementation.

Focus groups with students. In April 2021, RTI conducted three 45-minute virtual focus groups with 10th grade Black male students from three treatment schools. Schools were selected for their demographic diversity and greater range of implementation experiences. The focus groups intended to learn about the perspectives of Black male students on the discipline structure at school, school safety, their relationships with adults at school, and the interventions, including RP. The focus groups took place during school hours.

Meetings with district project coordinators and review of project documents. In addition to the formal data sources discussed above, RTI’s evaluation was informed by regular meetings with the district project coordinators, as well as review of documents they developed related to the initiative.

2.3.2 Outcome Evaluation

School staff survey

We surveyed staff at four points in time across 3 school years. The staff survey combines subscales from various sources to measure school climate and school safety:

- To gauge **personal safety**, three sets of questions asked participants how often they felt unsafe before, during, or after school in various locations on school property. Respondents answered using a three-point scale: “never,” “occasionally,” or “most or all of the time.”
- **School climate** was assessed using four subscales on the Inventory of School Climate for staff (Brand et al., 2008) and an additional four subscales from the Authoritative School Climate Survey (ASC; Cornell, 2017). The subscales from Inventory of School Climate measured disruptiveness (five items), peer sensitivity (five items), support for cultural pluralism (five items), and safety problems (four items). Scales from the ASC measured the prevalence of teasing and bullying (five items), teacher respect for students (four items), student engagement in school (six items), and school disciplinary structure (nine items).
- **Staff cultural responsiveness** was measured by the Culturally Responsive Teaching Self-Efficacy (CRTSE) Scale (Siwatu, 2007). The scale asks about staff confidence in their ability to execute specific teaching practices and tasks that are associated with teachers who have adopted CRI. In the original scale, participants were asked to rate 40 items related to teaching practices by indicating the degree of confidence from 0 (no confidence at all) to 100 (completely confident). Before our

data collection, we revised the CRTSE scale to reduce respondent burden. In our study, the response options were reduced from a 0–100 scale to a 1–10 scale. In both scales, higher values indicate increased confidence. Additionally, this study used 34 of the original 40 items. The six items removed (nos. 18, 22, 23, 29, 30, and 31) had the lowest factor loadings from a previous factor analysis of the full scale and pertained to English language learners, and as such were not relevant to this intervention. Due to skewness in the data, following data collection, the response options were recoded from the 10-point ordinal scale to a three-categorical variable. The recodes were made such that responses 1–7 were recoded into a value of “1”; responses 8 and 9 were recoded into a value of “2”; and response 10 was recoded into a value of “3.”

The school district provided staff lists with name, email addresses, and school information at the participating schools to the evaluation team. All staff at each participating school received an email invitation from the investigator that described the purpose of the study and provided a statement of consent, instructions, and a link, unique to each staff member, to the web-based survey for completion within the data collection period (approximately 4–6 weeks). Using a unique link allowed linking individual surveys over time. The full survey took an estimated 20–30 minutes to complete.

Staff survey data were collected and managed using REDCap (Research Electronic Data Capture) tools hosted at RTI (Harris et al., 2019; Harris et al., 2009). REDCap is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for data integration and interoperability with external sources.

Table 2.3 provides information about the measures used in the staff surveys.

Table 2.3 Staff Survey Constructs, Measures and Reporting Time Frames

Construct	Summary of Measures	Reporting Time Frame
Personal safety	How safe does respondent feel before, during, and after school in each of three locations?	Past 30 days
Inventory of School Climate	Four subscales: disruptiveness (5 items), peer sensitivity (5 items), support for cultural pluralism (5 items), and safety problems (4 items).	None specified
Authoritative School Climate Survey	Four subscales: prevalence of teasing and bullying (5 items), teacher respect for students (4 items), student engagement in school (6 items), and school disciplinary structure (9 items).	None specified
Staff cultural responsiveness	Confirmatory factor analysis indicated one subscale.	None specified

Administrative data

CMS provided the evaluation team with a wealth of administrative data that included information on each student's ODRs, tardiness and attendance, in-school suspensions (ISSs), and OSSs for the academic year before the intervention started (2017–2018) and each of the three academic years of the intervention implementation (Year 1: 2018–2019, Year 2: 2019–2020, and Year 3: 2020–2021). For the purposes of this project, we concentrated on ODRs and OSSs as measures of discipline, focusing on discipline disparities.

CMS provided metrics for ODRs as counts of the number of ODRs per student per year. For OSS, metrics were number of days per year that a student had been suspended, rather than the number of referrals resulting in suspension. In addition to count variables, we coded ODR and OSS variables as binary to indicate whether a student had any ODRs or OSSs each year.

We also examined absences, both excused and unexcused, as measured for each student in our treatment and control schools. The total number of absences for the year was used in models for outcomes.

In March 2020, the COVID-19 pandemic caused CMS to close schools and switch learning to a remote learning model. Students remained in a remote learning model until returning to in-person classes, mixed with remote learning, in February 2021. They continued in a hybrid in-person and remote learning mode for most of the rest of the spring 2021 term. Following the onset of COVID-19, and into the 2020–2021 school year, the frequency and measure of ODRs, OSSs, and absences were significantly changed because of the format of classes. As a result, we did not use administrative data from the 2020–2021 school year. To compensate for missing 2019–2020 data in the final quarter of the academic school year, we adjusted outcomes based on methods detailed in **Section 2.5.3: Benefit Estimation**.

2.3.3 Cost and Benefit Data

We estimated the cost of the interventions in treatment schools and similar costs in control schools and conducted a cost-effectiveness analysis in treatment schools compared with control schools.

Our cost data structure derives from the principles of activity-based costing (Drummond et al., 2015). Estimating economic effects entails four sequential steps:

1. Identify the activities needed to implement the program.
2. Identify the resources used to execute each activity.
3. Determine the quantity of each resource used.
4. Assess the unit cost for each resource.

Table 2.4 describes the general categories for each of the activities and resources, along with the quantity measure, unit cost measure, and data source for each. To best measure real-world implementation costs, we did not include costs that are research-only elements of the study, such as staff time for completing grant paperwork.

Table 2.4. Measures and Data Sources for Resource Quantities and Unit Costs

Activity	Resource	Measure		Data source
		Quantity	Unit Cost	
Trainings	Staff time	Hours	Wage rates	Training guide, study records, BLS
	Materials	Cost of materials	Cost per material	Training guide
	Space	Square footage	Cost per sq ft	Hollands et al., 2015
Tier 1 team	Staff time	Hours	Wage rates	Tier 1 team guide; BLS
SET	Staff time	Hours	Wage rates	Study records; BLS
Coaching	Staff time	Full-time equivalent estimate	Wage rates	Study records
RP circles	Staff time	Hours	Wage rates	Administrative data; BLS

Note: BLS = Bureau of Labor Statistics.

We identified program activities and resources through semi-structured telephone interviews with the CMS grant coordinator. We used existing sources whenever possible to collect the quantity measure and unit cost data, including study records, administrative data, the Bureau of Labor Statistics data, and the CostOut Database (Hollands et al., 2015). When needed, we developed data collection guides to obtain additional information that the CMS grant coordinator completed and returned (i.e., the training guide and the Tier 1 team guide). We used CMS administrative data to apportion district-wide costs to schools and to generate cost-per-student estimates.

We derived long-term economic benefits by combining absence records from CMS administrative data with estimates from the literature. We describe the sources and methods for the benefit calculations within **Section 2.5.1** below.

2.4 Analytic Plan

2.4.1 Process Evaluation

Measurement of implementation fidelity

The study uses a variety of data sources to measure fidelity of implementation of the three interventions. The fidelity measures and their scope vary by intervention. PBIS has the most standardized and comprehensive measures for assessing implementation fidelity. Those measures primarily focus on adherence: whether the core features of PBIS are in place

schoolwide. For RP, we used administrative data related to teacher exposure to training and staff survey data to estimate teachers' use of RP in the classroom (i.e., dosage. For CRI, we use data related to teacher exposure to training and school-level structures of support to implement CRI.

Although this section focuses on the quantitative measures of implementation fidelity, qualitative data from multiple sources also helped to provide a fuller understanding of implementation progress.

- **PBIS.** To measure PBIS implementation during Year 1, the study uses the PBIS SET summary scores, collected at baseline in fall 2018 and for follow-up in spring 2019. Because the pandemic prevented RTI staff from returning to the schools to collect SET data in 2020 or 2021, we used the TFI Tier 1 summary scores to measure level of PBIS implementation in Year 2. The TFI data are available for treatment schools for fall 2019 and spring 2020 before the COVID-19 pandemic. The outcomes analysis uses the spring 2019 SET score and the spring 2020 TFI score to examine relations between levels of PBIS implementation and study outcomes.
- **RP.** For RP, the core practice in this initiative is conducting regular proactive community-building RP circles in the classrooms. We collected data on the prevalence and frequency of this practice through the annual staff survey, which asked teachers if they conduct the proactive RP circles in their classroom, and, if so, the average monthly frequency. To increase the accuracy and consistency of teachers' responses, the survey question (below) included a definition of RP circles:

Do you facilitate proactive restorative practices (RP) circles? [Proactive RP circles are structured group discussions, normally held in a seated or standing circle to give students and educators an opportunity to learn about one another and build mutual trust. A facilitator or circle keeper guides but does not control the discussion. There is typically a sequential go-around with a talking piece. An RP circle is NOT an impromptu conversation or traditional class discussion.]

- The staff survey also asked staff about average duration in minutes of their RP circles to learn whether teachers were dedicating substantial time to this. In addition, the developer of RP-Observe (Gregory et al., 2017) trained project coaches to use this validated tool to observe and assess multiple dimensions of RP circle quality. The coaches began conducting these assessments with school staff, but the COVID-19 pandemic interrupted this work before sufficient data could be collected for the evaluation.

Although two treatment schools tried to implement RP schoolwide, three focused on ninth grade. To investigate the relation between level of RP implementation and study outcomes, the study uses the annual measure of the percentage of ninth grade teachers implementing RP circles in their classrooms, conditional on their school having received training in RP.

- **CRI.** To estimate the level of implementation of the CRI intervention among treatment schools, we considered three primary factors: if schools (1) trained a significant number of teachers; (2) completed a CRI implementation plan; and (3) added any CRI lesson plans for Black male students to their school's repository of CRI lesson plans. We assigned one point for each of these dimensions, creating a fidelity score with a range of 0–3, in which 0 = none, 1 = low, 2 = medium, and 3 = high. The CRI fidelity data are used to summarize scope of implementation for

the process evaluation. Because CRI was the last component of this initiative and the pandemic began early in CRI implementation, the analysis of study outcomes does not include CRI fidelity data, instead focusing on PBIS and RP.

Analysis of qualitative data

All focus groups were audio-recorded and professionally transcribed. To analyze the data, the research team used a rapid qualitative analysis technique. This included summarizing transcripts by topic using structured templates and matrix displays that summarize findings by topic across groups (see Hamilton & Finley, 2019). The lead analyst collaborated closely with the analysis team members in planning and executing the analysis, including reviewing transcripts, templates, and data displays and summaries.

2.4.2 Outcome Evaluation

Administrative data outcomes

We used administrative data to examine the following five outcomes per school year:

1. Whether each student had any ODRs
2. For students with any ODRs, the number the student had
3. Whether each student had any OSSs
4. For students with any OSS in a school year, the number the student had
5. The number of absences (excused and unexcused) each student had

We looked at ODR and OSS counts *conditional on a student having any* because most students did not have any. Therefore, analysis of counts for all students would have been driven by the large number of students with zero and thus mirrored the binary analysis to a large extent. Making these analyses conditional on a student having one or more ODR, or one or more OSS, as appropriate, allowed us to more cleanly examine trends in the counts of ODRs and OSSs for students who received any. We did not apply this approach to absences because most students had one or more absences. Examining absence as a binary outcome did not strengthen the analyses.

Target and reference group

Our analysis of administrative data outcomes focuses on comparisons of Black male students with all non-Black male students. This approach provides the most appropriate reference group for Black male students—non-Black male students, only—which sharpens the focus on disproportionality due to someone being Black and avoids the ambiguity of how much of any effect is due to being Black *and* male.

Analytic approach

To make full use of student-level administrative data, we ran two-level models, students nested within schools. This applies to the administrative data (ODRs, OSSs, and absences) and student survey data.

For staff survey data, we nested respondents within schools but did not do any subgroup analysis.

Change over time

In analyzing changes in outcomes over time, we nested time within school only. We did not analyze individual students longitudinally. For staff survey data, we used survey wave as the unit of time (with fall 2018 as baseline).

2.5 Cost Evaluation

2.5.1 Analytic Perspective and Timeframe

The analytic perspective for the economic analysis was societal, meaning that the analysis included costs and benefits accruing to society overall. The costs of the intervention—including staff time, space, and materials—accrue to the district but also accrue to the broader society, because taxpayers ultimately fund the program. Long-term economic benefits of the intervention (such as changes in earnings or crime through intervention activities effecting graduation rates) similarly accrue to society. The societal perspective allows for a net benefit analysis that includes all such costs and benefits.

We intended to collect and analyze three school years of cost and benefit data: 2018–2019, 2019–2020, and 2020–2021. However, the COVID-19 pandemic interfered with implementation and data collection beginning in March 2020 and continued into the 2020–2021 school year. Because of the interruption, our cost analysis includes only the first 2 years of the intervention. Similarly, to analyze the impact of the intervention on the benefits, we used data from the pre-intervention period (2017–2018) and just the first 2 years of the intervention.

2.5.2 Cost Estimation

We collected costs at the school level. The intervention was a series of trainings and meetings for school staff, as well as RP circles. Therefore, the costs accrue to staff, and to evaluate student-level outcomes, we aggregated staff costs to the school level. To calculate costs at the school level, we first estimated the cost of each program activity by combining the unit cost and quantity of each resource and then summed across activities. We then divided total school-level costs by student enrollment in each year in the corresponding school to calculate cost per student.

In several instances we needed to make generalizing assumptions when estimating costs.

- The staff survey collected data on frequency of RP circles in the 2019–2020 school year only. We retroactively applied the estimates for that year to the 2018–2019 school year.
- Staff time for delivering the SET was provided in aggregate, not by school. We apportioned SET costs to schools by the number of SETs each school received. Most

schools received two SETs, and therefore there was little variation in the SET cost across schools.

- Coaches reported their time in aggregate, not by school. We apportioned coaching time equally across all schools that received coaching.
- The members of Tier 1 teams fluctuated often. Based on guidance from CMS, we assumed a team consisted of 10 people: one administrator, one support staff, one behavior management technician, and seven teachers. In the 2018–2019 school year, each school provided estimates on how often they met and how long the meetings lasted; those estimates were applied to both study years.

2.5.3 Benefit Estimation

We calculated long-term economic consequences of dropping out using attendance records from administrative data. Our approach was: 1) first, impute absences for quarter 4 of the 2019–2020 school year; 2) generate model-adjusted marginal predictions of the count of absences per student by treatment condition (i.e., treatment vs. control group) and school year; 3) use an estimate from the literature to calculate the probability of dropping out; and 4) use estimates from the literature to calculate the net present value of long-term economic consequences associated with dropping out.

We outline additional details on each step of our approach below.

1. **Imputation of absences.** Quarter 4 of the 2019–2020 school year had missing absence records for all students because of the COVID-19 pandemic. Not accounting for this lack would have resulted in underestimating the total number of absences in the year, and therefore a biased estimate of future benefits. We used last observation carry-forward analysis to impute absences for quarter 4 of the 2019–2020 school year.
2. **Accounting for student and school level effects.** We used a mixed-effects negative binomial regression with absences as the dependent variable and random effects being the interaction of treatment condition and year; the interaction of race and gender; the count of incidents in the school at baseline; the percentage of students who received free or reduced lunch at baseline; and OSSs per 100 students at baseline. The model also included school-level fixed effects, propensity score weights derived from the outcomes study, and school-level clustered standard errors. We then generated marginal predictions showing the model-adjusted count of absences per student by treatment condition and year. The purpose of the step is to generate model-based estimates of the average number of per-student absences for each year and each treatment condition.
3. **Calculation of probability of dropping out.** Kirksey (2019) reported that the percentage of school missed is associated with a 0.08 increased likelihood of dropping out ($p < 0.001$). We multiplied this estimate by the adjusted absences described in the previous step to calculate the probability of dropping out for each treatment condition and year.
4. **Calculation of the net present value of economic consequences.** We used additional sources to provide the lifetime net present value cost of dropping out for a variety of relevant economic outcomes: earnings, health, property crimes, and arrests. **Table 2.5** shows these sources and estimates. We multiplied the probability of dropping out (described above) by these estimates, yielding an anticipated cost of dropping out by treatment condition and year. We then calculated the change in the net benefit from

the baseline (2017–2018) to the latest year of usable data (2019–2020), representing the monetary benefit of reduced dropout for each treatment condition. All price estimates were converted to 2021 U.S. dollars.

Table 2.5. Estimates and Sources of Lifetime Economic Outcomes of Dropping Out

Outcome	Estimate	Source
Decreased earnings	\$361,844 (adjusted from 2007\$)	Cohen & Piquero, 2009
Increased health cost	\$236,491 (adjusted from 2007\$)	Muennig, 2007
Increase in property crimes	0.638 crimes	Bjerk, 2012
Cost of property crime	\$6,023 (adjusted from 2008\$) ^a	McCollister et al., 2010
Increase in arrests	0.831 arrests	Bjerk, 2012
Cost of arrest	\$1,560 (adjusted from 1987\$) ^b	Cohen et al, 1994

^a We used the estimate for vandalism, the lowest priced crime, as a conservative assumption

^b Includes the cost of arrest and booking

Net benefit analysis

To perform the net benefit analysis, we used the average of total costs per student and change in total benefits per student from baseline (2017–2018) to 2019–2020 by treatment arm. The net benefit is the difference between those two benefits: the change in the monetary benefit of the intervention minus the cost.

3. Results

In this section we present results, organized around the major components of the study: implementation, outcomes, and costs.

3.1 Implementation

3.1.1 Training Delivery

This section provides a summary of the training provided at each treatment school for each of the three interventions: PBIS, RP, and CRI. To kick off this initiative, the district’s project director made a 1.5-hour presentation to school staff at treatment schools (five of seven participated), providing an overview of the initiative and contextual information on implicit bias and how it may affect Black male students at school. **Table 3.1** presents by treatment school the training start dates for each intervention, which we consider the intervention start date, and the grades targeted; interventions either targeted the whole school or ninth grade. Note that to protect school identification, we assigned numbers 1–16 to treatment and control schools; this table only includes treatment schools.

Table 3.1. Intervention Training Start Dates and Grades Targeted by School

School	Schoolwide PBIS (Fall 2018–Spring 2019)		RP Circles (Spring 2019–Fall 2019)		CRI (Fall 2019)	
	Grades Targeted	Module 1 Date	Grades Targeted	Circles Training Start Date ^a	Grades Targeted	First Training Date
School 2	Whole school	August 2018	9th grade	November 2019	9th grade	August 2019
School 5	Whole school	August 2018	9th grade	January 2018; September 2019 (booster) ^b	9th grade	August 2019
School 6	Whole school	July 2018	9th grade	July 2018	9th grade	August 2019
School 8	Whole school	November 2018	Not trained	N/A	9th grade	August 2019
School 10	Whole school	October 2018	Whole school	October 2019	9th grade	August 2019
School 11	Whole school	April 2019	Not trained	N/A	Not trained ^d	N/A
School 16	Whole school	July 2018	Whole school	August 2017 ^c	9th grade	August 2019

^a When there were multiple RP circle training dates for some schools, we counted the main training date (i.e., when most of the staff were trained) as the staff training start date.

^b The original training for school 5 occurred before this National Institute of Justice (NIJ) grant.

^c The original training for school 16 occurred before this NIJ grant.

^d Although school 11 sent school representatives to the training, they did not train any significant number of school staff.

First, the district launched PBIS, which was intended to provide the foundational behavioral framework for each school. Training in RP and CRI began over the next 12 months. Whereas all schools received the PBIS training, three schools received little to no training in RP (although one had received training a year prior to this initiative) and one school did not receive any training in CRI. However, even among the schools that participated in CRI training, the level of staff participation in training varied greatly by school. The following provides more detail by intervention, which is important to understanding the activities, scope, and variation in the receipt of training. Training and coaching comprise the core of the intervention.

PBIS

PBIS targeted each school at a schoolwide level. CMS launched PBIS first, mostly in fall 2018, beginning with the 12-hour Module 1 core training for Tier 1. This module focused on training the PBIS school leadership teams, which typically included representation from school administration, counseling, teaching, and other school leadership. Their purpose was to develop multi-tiered systems of support (MTSSs) in the school, which included PBIS, RP (specifically, how to facilitate conversations), and CRI. In the district, prior to this study, the behavior side of MTSS lacked structure, and CMS intended PBIS to provide that structure for the high schools participating in this study.

Following the Module 1 training, all schools except one participated in follow-up PBIS leadership team training in the summer of 2019 (12 hours). Most schools participated in the follow-up Tier 1 leadership team training in June 2020 (3 hours) and supplemental team training for Tier 2 in September 2020 (3 hours). In addition to the schoolwide PBIS training, three schools participated in additional classroom-level training, typically for ninth grade teachers, using the trainings “Routines and Procedures,” “Behavior Specific Praise,” or both.

RP

For RP, three schools focused on the ninth grade and two targeted the whole school. Due to scheduling challenges, implementation of RP training was more spread out over time than the PBIS training. In late fall 2018, several treatment schools started the Introduction to RP training (6 hours), which provided the conceptual framework for RP. We count the intervention as starting after schools received the RP circles training, because the RP circle training taught school staff how to facilitate proactive community-building circles in the classroom. One treatment school completed both the Introduction to RP and the circles training (6 hours combined) in the summer before the grant started; however, most treatment schools (three of seven) trained staff in RP circles in fall 2019. Two schools (school numbers 5 and 16) had received RP circle training prior to this initiative. One of these schools completed a 1-hour booster training in fall 2019.

CRI

CRI training focused exclusively on the ninth grade. In August and October 2019, CMS provided training sessions in CRI related to the development of lesson plans that are culturally responsive to Black male students. An expert from University of North Carolina at Charlotte led the training. All schools sent at least one person to each of the two trainings. Some schools only sent their school trainers to the August CRI training (6 hours), so they could in turn train school staff, whereas other schools sent varying numbers of teachers to the training. The October training (3 hours at a professional learning community [PLC] meeting) was for the PBIS school leadership teams and addressed development of implementation plans.

Additional professional development and coaching

School leadership teams received technical assistance from their school coaches and support in three PLC meetings per year (1.25 hours each). The PLC meetings focused primarily on PBIS, except for the one in October 2019, which focused on CRI.

Although coaches started working with the schools during Year 1, due to hiring delays and the need to train the coaches, the coaching did not operate in full capacity in many schools until Year 2. CMS trained the coaches in the three interventions (PBIS, RP, and CRI), implicit bias, Educator's Handbook (the student data system used by CMS), data drill down, root cause analysis, adaptive leadership (Heifetz et al., 2009), and the RP-Observe tool (Gregory et al., 2017) to conduct observations to assess the quality of proactive RP circles in the classroom. (Observations were conducted in spring of Year 1 and fall of Year 2 but stopped after that because of COVID-19.)

The coaches primarily worked with each school's administration and PBIS leadership team to build their capacity and help guide their implementation. Coaches worked with administrators to build training schedules, identify team members, connect with teachers, and help with implementation. Coaches attended the monthly PBIS leadership team meetings and coached team members to help them plan agendas, facilitate discussions, plan implementation, and solve problems. To a lesser extent, coaches also worked with the school professional learning communities to provide technical assistance and coaching on implementation. The coaches also worked with schools to analyze and use PBIS schoolwide data to identify and understand patterns of disproportionality and use these data to inform school efforts. Through the grant, in Year 1, CMS for the first time developed a data dashboard to provide key disciplinary and academic indicators by race, ethnicity, and gender to share with each of the treatment schools. The dashboard comprised the following data:

- Risk index and risk ratio (CMS coaches and project staff used these and other data in training with schools to understand the problem and help schools start to diagnose the reasons for disproportionality)

- ODRs
- OSSs
- Attendance
- D and F grade reports (for Math 1 and English 1 only)
- Retentions

The coaches also provide some of the RP training and conducted classroom observations for RP and PBIS. Observations were conducted in spring of Year 1 and fall of Year 2 but stopped after that because of COVID-19. To a limited extent, coaches also provided some direct coaching to teachers for PBIS and RP.

Over the course of the initiative, CMS also decided to create stipend positions within each school to support the PBIS leadership team. In Year 2, one coach left their position; for that school, CMS provided part-time coaching support and created stipend positions for two staff members at that school for the roles of PBIS team facilitator and data analyst. Having these in-school positions increased the initiative and productivity of the PBIS leadership team at that school; therefore, CMS expanded this model to all the other treatment schools in Year 3 when coaching support decreased. In Year 3, CMS reduced the number of coaches to two full-time and one part-time coach for the eight schools, decreasing the ratio from one coach per school, to one coach for two schools. Through the two stipend positions at each school, CMS aimed to increase the capacity, responsibility, and accountability of school staff.

3.1.2 Fidelity of Implementation

This section presents available data on implementation fidelity for each intervention. Fidelity measures and their scope vary by intervention.

PBIS

As discussed in the Methods section, for PBIS prior to COVID, RTI gathered fidelity data for treatment and control schools using the validated SET. Trained RTI staff, blinded to treatment condition, visited the study schools at baseline in fall 2018 and spring 2019 toward the end of Year 1 of implementation. Data collection included interviews of the school principal, school staff, and students; document review; and observations in hallways and classrooms. RTI evaluators followed the standard SET protocol and scoring rubric that measures multiple dimensions of PBIS implementation and identifies areas of strength and areas for improvement.

After the COVID-19 pandemic started, RTI staff could no longer visit the schools; therefore, the evaluation only has one year of SET data. On a quarterly basis, CMS coaches administered the TFI among treatment schools, and we include TFI data for Year 2 of the evaluation before COVID-19 lockdowns. The TFI also provides a valid and reliable measure

of the extent to which school personnel apply the core features of schoolwide PBIS. The TFI has the same areas of focus as the SET; however, the National Institute of Justice (NIJ)-grant-funded coaches conducted the TFI in collaboration with the PBIS leadership team and relied on self-reporting of the PBIS leadership team. Whereas the SET requires an overall score of 80% to achieve fidelity, the TFI requires a score of 70% over repeated quarters.

Table 3.2 presents SET data for Year 1 of implementation, including the fall 2018 baseline and spring 2019 follow-up SET scores for both treatment and control schools. At baseline, all treatment and control schools have SET scores well below the overall fidelity score of 80%, with a mean of 45% for treatment schools and 35% for control schools. By spring 2019, the mean SET score among treatment schools increased to 57%, whereas the mean among control schools stayed flat at 32%. Among treatment schools, seven of eight schools showed an increase in their SET scores of at least 7 percentage points, with one school increasing 25 percentage points to achieve an overall SET score of 84%. By contrast, two control schools had declining SET scores, and none came close to the overall fidelity score of 80%.

Table 3.2. PBIS Tier 1 SET Scores, Treatment and Control Schools, Year 1

	Fall 2018 (Baseline)	Spring 2019
Treatment Schools		
School 2	48%	59%
School 5	63%	64%
School 6	58%	65%
School 8	39%	51%
School 10	25%	33%
School 11	25%	44%
School 16	59%	84% ^a
MEAN	45%	57%

(continued)

Table 3.2. PBIS Tier 1 SET Scores, Treatment and Control Schools, Year 1 (continued)

	Fall 2018 (Baseline)	Spring 2019
Control Schools		
School 3	27%	49%
School 4	N/A	30%
School 7	34%	38%
School 9	40%	23%
School 13	N/A	N/A
School 14	N/A	19%
School 15	37%	34%
MEAN	35%	32%

^a Indicates score passed 80% threshold for fidelity.

Note: The SET was administered by RTI evaluation staff, blinded to treatment status. RTI staff were not able to get access to administer the SET in some of the control schools.

For Year 2 of implementation, **Table 3.3** provides data for the CMS-administered TFI, including fall 2019 and spring 2020 before the COVID-19 pandemic. As shown in Table 3.3, the mean TFI score reached 67% in fall 2019 and 70% in spring 2020. Two schools reached overall fidelity scores of at least 70%, and by spring 2020, four schools reached overall fidelity scores. During this period, two schools reached and sustained fidelity over both quarters. TFI criteria for implementation fidelity require an overall score of 70% for three consecutive quarters. Note that one school, number 16, was the only school to achieve fidelity in Year 1, but completely regressed in Year 2 and never recovered fidelity; this school lost its principal after Year 1 and had an extended leadership void.

Table 3.3. PBIS TFI Score in Treatment Schools, Year 2

School	October 2019	February 2020
School 2	83% ^a	80% ^a
School 5	57%	60%
School 6	80% ^a	97% ^a
School 8	67%	57%
School 10	53%	70% ^a
School 11	67%	73% ^a
School 16	60%	53%
MEAN	67%	70%

^a Indicates score passed 70% fidelity threshold.

Note: The TFI was administered by CMS project staff.

One important component of PBIS is establishing and reinforcing a small number of positively stated behavioral expectations for the school and classroom. According to the staff survey administered by RTI, in spring 2019, 78% of treatment school staff reported teaching students a small number of positively stated expectations for routines in their classroom; this increased modestly to 82% in 2020 and 84% in 2021 (a marginally significant change between 2019 and 2021, $p = .05$). Although we do not have pre-implementation data, this nonetheless shows an increase of 6 percentage points during the implementation period. In control schools, in spring 2019, a higher percentage (86%) of staff reported teaching a small number of positive stated classroom expectations, but this stayed relatively flat in the two subsequent years at 87%.

To summarize PBIS fidelity, from the baseline SET to the final TFI before the pandemic, five out of seven schools had increased fidelity scores, and four out of seven schools reached or exceeded the fidelity threshold score.

RP

As discussed above, although some schools implemented RP with the whole school, more schools focused RP on the ninth grade. For RP, we measured fidelity as the percentage of 9th grade teachers reporting that they facilitate RP circles in their classroom, conditional on the school receiving any training in RP circles. This measure approximates the prevalence of this core RP practice in classrooms. If a school has a high proportion of teachers facilitating circles, that suggests that, on average, students in that school have a higher “dose” of RP than students in a school with a low proportion of teachers facilitating circles. **Table 3.4** presents these data from the annual staff survey for the years 2019, 2020, and 2021.

Table 3.4. Percentage of Ninth Grade Teachers in Treatment Schools Reporting Facilitating RP Circles in the Classroom

School	Spring 2019	Spring 2020	Spring 2021
School 2	N/A	42%	14%
School 5	53%	53%	59%
School 6	83%	63%	45%
School 8	N/A	N/A	N/A
School 10	N/A	70%	37%
School 11	N/A	N/A	N/A
School 16	60%	67%	70%

Note: Only includes schools that received any RP circle training. Also note that for some schools, the sample size of ninth grade teachers was small, especially in spring 2021 (range 7-81).
Data source: RTI-administered staff survey.

As shown in Table 3.4, the percentage of teachers who reported facilitating any RP circles varied across schools and years. Two schools were not able to train any significant numbers

of staff, whether before or during the initiative. Among the schools that implemented RP, most ninth grade teachers reported facilitating any RP circles in the classroom in three schools in 2019 and in four schools in 2020. The COVID-19 pandemic interrupted the momentum and reported circle facilitation declined in three schools in 2021. Nonetheless, in the two schools that had had prior experience with RP (numbers 5 and 16), the proportion of teachers facilitating circles increased modestly in 2021.

The frequency of circles also matters. According to the 2020 and 2021 staff surveys, among all teachers who reported facilitating RP circles in their classroom, 70% reported facilitating circles at least twice per month and about one-third did so at least four times per month. Teachers reported that, on average, circle time tended to last about 20 minutes.

The 2021 staff survey also asked teachers whether they had facilitated any circles virtually when schools were closed during the pandemic. The 2021 staff survey asked the following question about RP circle facilitation in-person and virtually:

Do you facilitate proactive restorative practices (RP) circles? [Proactive RP circles are structured group discussions, normally held in a seated or standing circle to give students and educators an opportunity to learn about one another and build mutual trust. A facilitator or circle keeper guides but does not control the discussion. There is typically a sequential go-around with a talking piece. An RP circle is NOT an impromptu conversation or traditional class discussion.]

Nearly half (48%) of ninth grade teachers in treatment schools reported facilitating any virtual circles.

The annual staff survey also asked treatment school staff about their confidence in facilitating RP circles in their classroom. **Table 3.5** presents the percentage of ninth grade teachers who reported that they strongly or somewhat agree that they feel confident facilitating RP circles. The percentages feeling confident generally increased between 2019 and 2020, but some schools showed declines in 2021 after school reopened from the COVID-19-related closures.

RP aims to build stronger relationships between students and teachers, and one of the ways it accomplishes this is through the RP circles in which students express their thoughts and feelings. In the annual staff surveys, one question asks school staff how often they ask students to express their feelings. Among treatment school staff, the percentage answering often or always increased modestly from 80% in 2019 to 85% in 2021, but this did not reach statistical significance ($p = 0.12$). Among control schools, the percentage stayed flat: 87% in 2019 and 86% in 2021. We do not know why in spring of Year 1 (2019) control school staff reported greater frequency of asking students to express their feelings than did treatment school staff. In general, many teachers likely overestimate this for reasons including social desirability and difficulty in accurately assessing how much they encourage

students to express their feelings. A change in score is more important than the initial score.

Table 3.5. Percentage of 9th Grade Teachers Reporting Feeling Confident Facilitating RP Circles in their Classroom, Treatment Schools

School	Spring 2019	Spring 2020	Spring 2021
School 2	N/A	85%	38%
School 5	81%	87%	90%
School 6	75%	78%	74%
School 8	N/A	N/A	N/A
School 10	N/A	90%	77%
School 11	N/A	N/A	N/A
School 16	80%	95%	100%

Note: Conditional on the school receiving any RP circle training. Percentage of teachers who strongly agree or somewhat agree that they feel confident (versus somewhat disagree or strongly disagree). Note that for some schools the sample size of ninth grade teachers was small, especially in 2021

Data source: RTI administered staff survey.

CRI

For fidelity of CRI, we estimated level of implementation by considering whether (1) a school trained a significant number of teachers in CRI (beyond sending a representative to the training), (2) schools completed an implementation plan, and (3) schools added any CRI lesson plans to their school’s repository of CRI lesson plans. We assigned one point for each of those dimensions, creating a fidelity score with a range of 0–3: 0 = none, 1 = low, 2 = medium, and 3 = high. **Table 3.6** below presents the score for each school. Only one school achieved a high level of fidelity, including creating CRI lessons for their school’s repository. Two other schools achieved partial implementation. These schools generally started implementing CRI around January 2020 and implementation was curtailed by the COVID-19 pandemic and school closures just more than 2 months later. Thus, schools were not able to achieve full CRI implementation.

Table 3.6. Level of CRI Training and Early Implementation, Treatment Schools, Spring 2020 (Pre-COVID-19)

School	Score (0–3) ^a	School	Score (0–3) ^a
School 2	2	School 10	3
School 5	2	School 11	0
School 6	1	School 16	1
School 8	1		

^a Scoring: 0 = none, 1 = low, 2 = med, 3 = high.

In spring 2021, the staff survey asked teachers if, in the current school year, they developed or implemented any lesson plans that specifically focus on cultural responsiveness to Black male students. Among treatment schools, the proportion of teachers in each school responding “yes” ranged from 19% to 50%, including 46% for the school with the highest level of CRI implementation. However, in schools where teachers were not trained or few were trained in CRI, a substantial proportion of teachers still reported implementing CRI lesson plans for Black male students in their classrooms. It may have been challenging for teachers to accurately assess their cultural responsiveness to specific groups of students. Some teachers may feel that their teaching is culturally responsive for Black male students even if they did not participate in this specific training. We do not have data on the quality of their lesson plans.

3.1.3 Implementation Challenges

Early challenges

Several PBIS leadership team participants shared that they had a rocky start early in implementation for several related reasons, including the following:

- **Lack of clear expectations and roles.** In Year 1, several PBIS team leaders felt confused about their and the coaches’ roles and how the PBIS team was meant to support teachers and staff in implementation. In addition, they did not adequately understand future plans for the grant and implementation milestones.
- **Delayed start of the coaches working with schools.** The district had to hire and train the coaches and then they had to build relationships with the schools. PBIS leadership team members from multiple schools felt that they needed the coaches more present in the schools at the beginning. They wanted more help at the outset to plan how to implement these initiatives, especially in a high school context, which is larger and more complex than that of lower grades, in which PBIS has been more commonly implemented.

According to CMS project staff, although the coaches had prior knowledge of PBIS, they initially lacked the school-level application skills and had to build the expertise in PBIS and the other interventions to coach. In addition, coaches without experience at the high-school level had a disadvantage because high school teachers have different attitudes and ways of operating than teachers in lower grades.

- **Lack of teacher leadership in the PBIS team in some schools.** Several respondents in the PBIS leadership focus groups noted that having an administration-heavy PBIS implementation team at the beginning did not help with teacher buy-in or engagement, and they had to recruit more teachers. Therefore, they had to get teachers invested and show teachers how this initiative would benefit their students and the school.
- **Some dissatisfaction with the initial RP training.** Some teachers felt that they needed a more hands-on training with role playing to feel more confident and less vulnerable facilitating RP circles.

Teachers were slower to buy in because of these hurdles.

Ongoing challenges

Other major challenges to implementation during the initiative included the following:

- **Competing district initiatives.** In both the 2019 and 2020 focus groups, some participants highlighted that PBIS and RP competed with the rollout of other district initiatives, which overwhelmed staff in 2019 and challenged scheduling of both PBIS meetings in 2020 and the trainings with the treatment schools overall.
- **Complexity of this initiative, with three components.** Mastering three interventions requires a greater level of effort than just one. Some school PBIS leadership team members suggested that the three components of the initiative competed with each other, adding stress to teachers and staff and impeding the rollout of the initiative:

One administrator said,

I'll tell you that has been during the 2 years we've been working on PBIS that has been especially when you add you know the restorative circles and the culturally relevant instruction piece. You just got to be careful of how many layers you pile on your teachers because...you know, like everybody, you only have so much energy. And so you got to be careful how much energy you're asking to be invested in one, two, or three things. So that's been an obstacle that I've been constantly trying to kind of dance.

One PBIS leadership team member noted that, as a result, their school had to deprioritize CRI, to below PBIS and RP circles.

- **Need for staff buy-in and effective communication to increase buy-in.** At the high school level, teachers focus on academics; high schools have traditionally not focused on social-emotional learning (SEL) and teaching or reinforcing student behavior. In focus groups with school PBIS leadership team members, teacher buy-in was a commonly reported challenge. Some teachers did not see the need for a system of positive behavioral supports for high school students and some felt they just need to exert strict control of the classroom. RP emphasizes building relationships with students and encouraging student voice. PBIS and RP required changing staff mindsets and priorities as well as developing new skills. Cultivating buy-in and changing mindsets takes time and ongoing support.

In addition, in the staff focus groups, multiple teachers wanted help to understanding the relationship of these different interventions, including the nuances of the different programs and their purposes, and how they were supposed to fit together to effect equity.

- **Time demands and perceived staff burden.** In general, competing demands presented a challenge to adopting these new practices. For example, some teachers were concerned that RP circles took away from instructional time, and some had challenges with managing the length of time of the circles.
- **Knowledge and skills needed for the interventions.** Although schools made progress in training and implementation, many teachers still needed to build more skills to be able to confidently apply PBIS, RP, and CRI to their day-to-day practices. For RP, many teachers lacked skills in facilitating proactive circles and needed more practice, guidance, and feedback. Some teachers also struggled with discomfort

giving some control to students in the circles. For CRI, many teachers needed more knowledge and skills of how to develop culturally responsive lesson plans.

- **Staff turnover and lack of a process to onboard new staff.** PBIS leadership team members from multiple schools pointed out that new teachers to the school were not adequately briefed and trained on the initiative.
- **Lack of financial resources for PBIS incentives.** Throughout the initiative, PBIS leadership team members from multiple schools reported the lack of resources for PBIS incentives for students as a barrier. One school gained support for this from their alumni, another from a separate grant.
- **Lack of implementation plans for RP.** PBIS has established guidelines and measures for implementation, whereas RP and CRI do not. For CRI, the last intervention to be launched, CMS asked each treatment school to develop an implementation plan. For RP, the schools did not develop implementation plans, which would have been helpful to provide more structure.
- **School leadership changes and challenges.** Two schools' principals changed within the first year and a half of implementation, causing implementation to lag during the periods of instability in these schools. One of those schools, which was the only school to achieve PBIS fidelity in Year 1, had multiple leadership changes and was without a principal for an extended period; they did not recover PBIS fidelity during the initiative. In the other school, the principal left after a shooting in which a student was killed. Finally, in a third school, the principal was not invested and did not assign an administrator to the initiative, which impeded progress.

COVID-19 pandemic and barriers during COVID-19 school closures

The pandemic began in Year 2, before achievement of widespread fidelity of implementation. Schools physically closed for 11 months, preventing implementation of most of the interventions during this time and halting momentum of the initiative. Distinct challenges implementation challenges associated with the COVID-19 pandemic included the following:

- **Staff overload and stress.** The steep learning curve for virtual learning overwhelmed staff. School leadership was concerned about increasing their burden.
- **Lack of student engagement in remote learning.** In the focus groups with PBIS leadership team representatives and teachers, many staff cited challenges engaging students virtually, especially because students had the option to keep their cameras off. Some teachers noted the difficulty of building relationships with students online.
- **Needing more staff training and resources for virtual implementation.** Although the district developed some tools for online adaptations (e.g., virtual PBIS behavior matrix), multiple staff members in focus groups reported that they did not know how to translate these interventions into a virtual classroom. Specifically mentioned was how to create an online class culture to effectively implement RP circles.
- **Behavioral challenges.** Teachers also dealt with different behavioral challenges with students in the virtual classroom, with the greatest challenges being student absences and lack of engagement.

3.1.4 Implementation Facilitators

Many supports and resources helped to facilitate implementation of this initiative. The following summarizes key supports:

Coaching

The coaches, who worked closely with the school leadership teams and PLCs, were a central component of this initiative. In focus groups, school PBIS team leadership members emphasized the major role of the coaches in guiding their schools' implementation.

School administrative leadership and teacher champions

The initiative needed administrator support and consistent involvement to succeed, including regular administrator participation in the PBIS leadership team. An administrator had to make it a priority for school leadership and staff.

PBIS leadership team members also emphasized the need for teacher leader champions and active teacher participation in the PBIS leadership teams. To address the challenges of implementation in Year 2, representatives from the PBIS leadership team in two schools described how having low barriers to entry to joining the leadership team helped them. They did not force anyone to join the team, and they attempted to integrate the PBIS team into the school leadership team. Leaders in one school strategically transitioned part of their school leadership team to the PBIS team, because they found that using school leadership team meetings to discuss activities related to both addressed the challenges around scheduling a time when everyone is available to meet.

In the focus groups, PBIS team leaders widely cited the need to have some sort of champion who could carry the initiative and keep it as a top priority in the school. Two schools' PBIS teams were active in making sure activities were implemented in classrooms and spreading the word about them in PLCs and other committees members were parts of. At another school, the administration followed up with teachers to make sure that RP was occurring and that all teachers had the PBIS matrix. A PBIS team member from a fourth school had a dedicated group of teachers who worked well with their coach; seeing their example had pushed school staff to continue the work. In the school with the greatest level of implementation of CRI, the school's leadership had selected teachers who had interest in CRI to serve as champions.

Structures of support

In addition to integrating the PBIS leadership team with the school leadership team, other important support structures included clear roles, regular communication and meeting times, PLCs dedicating consistent time to this initiative, and data to inform the initiative. Leaders from three PBIS leadership teams mentioned the helpfulness of PBIS implementation "pacing" guides to provide milestones and timelines to keep them on track.

As part of this initiative, the district also collected and analyzed schoolwide data for PBIS (e.g., TFI, Educator’s Handbook) and developed a data dashboard on key indicators by race and gender. The coaches worked with the schools to analyze and understand their school’s data to develop a shared understanding of problems associated with discipline disparities and strategies to address it.

Teacher buy-in

In Year 1 of implementation, some PBIS leadership team members described how they had been educating teachers on the intervention to gain buy-in. For example, this included reviewing and sharing data on disparities and outcomes among Black male students, as well as demonstrating the benefits of the intervention through modeling and peer learning. Staff and students also had opportunities to offer feedback on implementation, which helped the team pause, reassess, and restart at different points throughout implementation.

Across schools, staff also varied in their receptiveness to different interventions. For example, in one school there was significant staff pushback against the CRI component, and the school’s leadership had to pull back on implementation. In another school, however, staff were especially enthusiastic about CRI, with teacher champions leading implementation.

A couple of PBIS leadership team focus group participants noted the importance of transparency and scaffolded support to gaining teachers’ trust and instilling confidence in them. One recommended:

Say, “Hey I’ll be here with you.” Or “I’ll check in with you. We can do an I do, we do, you do, if it makes you feel more comfortable.” But it’s necessary because those teachers are interacting with those kids on a regular basis and if they’re not building that community in their classroom, it’s going to come out in one way or another. It’s going to show. So you might as well give them a tool, some guided practice, maybe even a lesson topic or whatever else to help them feel more comfortable.

One PBIS leadership team member felt that teachers may be slower to buy into CRI if they perceived it as mandatory. Another said that leaders should not present it “top down,” but rather have a champion whom teachers already trust to encourage them. He shared the importance of choosing language carefully when presenting the initiative to teachers, and, when possible, trying to align that language with language teachers already know. He also proposed:

Then also for circles...model it. I mean, you want to explain it, but I think you would just start a meeting with a circle of some kind or model it first before you explain, “Well, this is the circle, it wasn’t that bad. You guys got engaged, you did this and that...and this is kind of like what we want our students to do at least on a weekly basis, somewhere on campus.”

One PBIS leadership team focus group participant said that the monthly meetings for this initiative may have helped their teachers to do a “mind shift” and get on board because they got to hear student success stories and network with other teachers with enthusiasm for the initiative:

I think the networking helps gaining ideas from other schools, like the rapport idea, so the networking definitely helps, and the PDs [professional development] helped just everybody on board, because it was a mind shift for some of our staff members. I mean to where, “Oh, you’re not going to suspend them?” “No, I’m not, you’re going to come in my office and we’re going to have a conversation about it. And we’re going to, because if I do, they’re still coming back to your classroom.” And then to go to the point of where then people would want that, they would say, “Ms. [Name], can we do a restorative conference with this?” Then I was like, “Oh yes, absolutely.” Before that wasn’t something that was even suggested.

Structures of support for specific interventions

Schools also benefited from putting in place supporting structures to facilitate implementation of the specific interventions.

RP supporting structures. In focus groups with PBIS leadership team members and teachers, some schools created supports to facilitate implementation of RP, including:

- **Modeling and teacher champions.** Modeling the proactive circles and teaching a smaller group of enthusiastic teachers to become experienced “champions” instilled confidence in the other teachers to implement RP.
- **Instituting a set day and time for weekly RP circles.** According to focus group participants, some schools established a schoolwide day and time to do restorative circles. Staff agreed that this helped implementation fidelity.

Personally, from what I’ve heard that you guys are implementing in other schools. ... You had the alternate Monday schedule so that there was time for it. I definitely think having just time of the day, especially with us being those core teachers, math and English, we have that EOC [End-of-Course Test], and we already do kind of rush through the curriculum a little bit. Somehow structuring it and then also...having someone kind of guide you on what to talk about. I think other schools are implementing really cool things that I feel like people at my school didn’t hear about.

- Some other benefits to having a scheduled time included the administration and PBIS team being able to visit teachers in their classrooms and provide feedback during the RP block, students sharing which teachers were not implementing, and cohesion with other schools also implementing.
- **Providing teachers with guidance and structure for RP circle topics.** Some schools created a repository of RP circle topics. In addition, schools can tailor topics to the school as needed. According to staff, it helped to have someone in charge of creating or coordinating these resources for the teachers. This also applied to CRI

lesson plans. In one school, teacher champions worked with their PLCs to start a repository of CRI lesson plans.

Considerations for training.

- **Providing follow-up training.** Participants also noted the importance of having follow-up training and coaching to help teachers feel more confident and comfortable to facilitate RP circles. One PBIS team leader suggested refresher circle training between semesters in the same school year, when many teachers change classes and need to restart the circles again.
- School staff greatly appreciated coaches who were willing to go to the schools personally to do the trainings and especially benefited when the techniques were modeled for them. For future professional development, some teachers would like to see more collaboration across the district, hoping to see effective practices being implemented by other teachers in other schools.
- **Timing of training.** Due to scheduling challenges, for most schools, the RP training occurred well into the school year. Multiple teachers emphasized the importance of timing of the rollout. When possible, providing training before or at the beginning of the school year facilitates teachers incorporating the new practice into their plans and routines for the school year. One teacher emphasized that introducing new initiatives or practices properly to students at the beginning of the year helps the students to get invested. At the same time, teachers acknowledge how busy the beginning of the year is for administrators and challenges of fitting in training into the schedule:

I don't think we had our restorative circles training until...maybe the end of September. So, we had already established all of our classroom routines and everything and then we also had a lot of our pacing calendar already set out. So, we had to figure out which lessons we could go back and try and work time into to make that work and still not lose any instructional time.

You know how important those first few days for teachers and students are, getting that routine down and I know that there's already a lot going on at the beginning of the year. But I think introducing kids to the normality of a circle at the beginning of the year will kind of help throughout.

It would've been nice if somehow it could've worked out where we could've had the training before school started or even that first week and then we could've sat with our PLC and worked out this week, this lesson would be a good place to set time aside to do this.

PBIS equity lens.

- **Celebrating all students.** In focus groups, multiple staff members emphasized the importance of keeping an equity lens and celebrating all kinds of achievements, not just honor roll and attendance, considering what is realistic for different students.

I think it's important to... celebrate all kinds of different things. Some kids just can't get to school on time, so the only thing that you reward is being present on time, you're not going to hit those kids. Maybe I can't get to school on time because I have to help my siblings get ready for their school or whatever, but I'm really good at getting my homework done. I think it's just really important. We did something

through community in schools as well where every week was a different challenge. One week was no missing work, one week was perfect attendance, and I felt just having at some point a goal that's achievable for everyone. It's not realistic for someone who's always had Ds to say you only get a reward if you get an 85 or higher on a test. I think it's just definitely having different levels of achievable things to make sure everyone gets celebrated at some point.

Increased district attention to SEL as a structure of support during COVID-19

The pandemic drew attention to SEL and mental health. During this time, the district created a required a weekly SEL period. This created additional space for teachers to hold RP circles during this time.

Other structures of support including providing online tools and other resources for teachers, such as the virtual PBIS behavior matrix and online RP circle topics. In addition, the pandemic led the district to create more online independent training for the interventions.

3.1.5 Other Future Considerations

In addition to the challenges and facilitators discussed above, the process evaluation identified other elements to be considered in implementing this type of complex, multi-component initiative.

- **Time for training coaches.** For this initiative, coaches needed to develop competence in multiple interventions. Therefore, planning more lead time (e.g., 3–6 months) for training coaches would help the launch of the initiative go more smoothly. Coaches also need skills in staff professional development and increasing staff buy-in.
- **Sequencing of interventions and time for building fidelity and system change.** A longer time horizon would have been beneficial for implementing these three interventions to achieve fidelity and create system change. The initiative design called for PBIS to provide the schoolwide behavioral framework on which the other interventions would build. Implementation could have been improved by allowing more time to achieve fidelity with PBIS before other interventions were added. The initiative operated within the time period of the grant and was interrupted by the COVID-19 pandemic; however, this was an ambitious undertaking for high schools in a 3-year period.
- **Facilitating integration of interventions.** Integrating these interventions and within existing structures and initiatives helps to facilitate adoption. For example, the district supports MTSS and PBIS should be integrated with MTSS. One school encouraged continuation of RP circles during COVID-19 shutdowns because it coincided with a suicide prevention initiative. Staff also need to understand how these interventions fit together and ways to support synergy. The district's new weekly SEL period provided an opportunity to further reinforce these interventions. As one PBIS leadership team member suggested, support from the state or district could either help with sustainability or competing district or state mandates could draw away attention from the initiative.

- **Promoting integration of behavioral supports with academic supports.** One of the PBIS leadership team members had this recommendation to further advance the goals of the initiative in the future:

Having behavioral standards infused into the curriculum. This would be a set of standards for behavior similar to academic standards. High school teachers are so tied to standards, they live and die by their curriculum standards, well, then I would say, I know we have ASCA [American School Counselor Association] standards for counseling, but maybe we need a set of behavioral standards that could be infused into the curriculum, so that teachers understand that academia and the restorative practices, for example, they're not separate from each other, all of these things should be happening simultaneously.

3.2 Leadership and Teacher Perspectives on Early Results

In focus groups, school PBIS leadership team members and teachers discussed how the initiative has contributed to building more positive relationships, improving school climate, and increasing equity in their schools. The following provides examples of perceived benefits of these interventions so far.

3.2.1 RP

Building relationships

Although most PBIS leadership team participants initially reported that a number of teachers had reservations about implementing RP, as implementation progressed, school representatives had more positive feedback to share. In focus groups, school representatives discussed how the RP circles allowed teachers and students to build a community of trust where students could open up and share more about themselves personally and discuss topics that mattered to them. One teacher said that, as a result, when COVID-19 hit and staff were dealing with the challenge of disengagement, more students were responsive to their efforts to connect virtually.

Last year when we went out for COVID, because of the relationships we built, kids actually were more out to turn their videos on and talk and engage with us, especially when we had to start sending out messages going, "Hey, I need you to log in, because I have to check and make sure you're okay." They were willing to do that, and I had a lot more of that when we went out for COVID from March to June. But when we started [virtual] school in August [with new students], not so much, and it echoes to what [other focus group participant] said about how we couldn't make the kids put their videos on. So because that relationship hadn't been built, the kids weren't going to turn their cameras on.

Although they acknowledged that some teachers can be afraid to discuss race-related topics with students, multiple PBIS team leaders indicated how important topics that bring in race and culture are to their students. Race-related topics that some staff mentioned discussing

during proactive circles included the Civil War, slavery, and other topics not previously covered in as much depth. One teacher said that in her class, their proactive circles hit on topics that teachers may not have known that their students had an interest in, even fun and silly topics to create a rapport:

Since I make mine (proactive circles) completely random, I've learned that some of my students like pineapple pizza and others turn their nose up at it. Then they make fun of me, because I like broccoli on my pizza. And we've had recurring discussions about what we put on pizza, but that wouldn't have happened without having that component of it and learning the quirks, and [their] knowing me. It's a human kind of relationship and we[re] all in the same storm, different boat kind of mentality.

One teacher described how the RP circles became the students' favorite part of class:

I can tell you in my classroom, one of the things that my students said at the end of the year that they liked the most were the restorative circles that we did, and that was the highlight. So I don't know how I feel about that, because I don't know, you don't like my teaching, but you like the circle a lot, I don't know. But I took it as a win, the fact that we did it and they were like, "Wait, it's Monday, and we're not having our circle. What's going on?" And so that, at least in my small little microcosm, they did enjoy that and there was some good dialogue and less of an issue with behavior from there.

The restorative circles...really did well in the freshman academy. ... I know that was a good part of building rapport for some teachers. Even though a lot of teachers were uncomfortable, maybe not have done it as often, but it did open up some things in the school.

Resolving conflicts and behavioral challenges more effectively

In addition to appreciating the benefits of the proactive community-building circles in class, PBIS team leaders and teachers described the benefits of using restorative circles to resolve conflicts instead of using exclusionary discipline.

And I think the restorative practices, especially the conversations, we didn't really talk about it here, but we got specific questions and things we could ask when a situation was escalated, and we were trying to have a restorative conversation with a student. I think that's certainly beneficial, and I think that African American male students appreciate the opportunity to have their voice heard.

Our ninth-grade academy assistant principal used restorative practices several times this year with students that I referred to him.... I found it to be extremely effective in terms of healing relationships between students that had experienced conflict.

School PBIS leadership team members also discussed the data they had been tracking on students disciplined with ISS. Before COVID-19, two schools had been tracking which students were most referred for ISS and holding restorative circles with them. One school leader said that they also started tracking which teachers sent students to ISS and who sent the most. Then they practiced restorative circles with those teachers to make sure they were implementing them in their classroom to help build rapport and relationships. Team leaders from both schools reported that restorative circles decreased ISSs and OSSs.

One PBIS team leader believed RP would play a huge role following the return to in-person lessons for students who have dealt with trauma and learning loss as a result of trauma:

One of our big issues that we're having is sending students out of the classroom for very minor issues. So we're hoping that those restorative practices, the teachers really learning those restorative practices and being able to build those relationships with those students, can give them a better understanding of what that student might be going through, and help them process certain situations before having a student removed from the classroom for an entire block and missing an entire block of learning, or writing referral on a student to where the student is getting in trouble for a situation that could have been easily talked through.

In discussing the benefits of this initiative, one school leader summarized from his experience the importance of getting to know his Black male students:

I remember... that I talked to one of the struggling Black male students.... I was just talking to him about his goals and dreams and he was like, "Nobody's ever even thought about what I wanted to do after I left high school, nobody even asked that question." He said that you asking me that question tells me that you care.... I had kids come back in 2 or 3 years and say, "You know what? I don't think I ever would have gone to college or the military or Central Piedmont, or even try to develop a trade, if it wouldn't have been for you just asking me, 'What do you want to do?'" [They thought,] "I am good at that, so I'm going to try and do that, just because I see that you see it in me. If you see it in me, I know it could be possibly done."

3.2.2 PBIS

Consistent behavioral expectations and positive reinforcement

In two focus groups with teachers in 2020, teachers reflected on early benefits of PBIS. Some teachers discussed the clarity and consistency of the PBIS behavior matrix:

At my school I think that ...what has really been successful has been the PBIS and the behavior matrix, the consistency that every situation...that each student is going to be looked at through the same lens for the same behavior.... Even though we talk about all this flexibility and all this

understanding where the student's coming from and that's happening at the classroom or the teacher level, but when they get to admin, it's following a specific criterion. They know that if it was a White girl or a Black male they're being channeled through the same behavior discipline. They're getting the same consequences.

Another stated,

And I have seen a difference in how the students respond to the behavior that's done at the high school under this PBIS model versus when they were in middle school and they felt like they were being targeted or things weren't being handled the same.

Teachers also discussed how PBIS creates a positive environment for students and provides the positive reinforcement that they need. One teacher noted how praising students for trying really benefits their morale. Another talked about how PBIS helps to provides "a base level of joy" in the school:

It does provide a better aspect of making sure there's a base level of joy to your school. I don't think a lot of PBIS stuff creates the entire culture and the enjoyment that kids should have at school, but at least creates the entry-level expectation of making sure that kids can smile while they're at school and laugh and be rewarded for being the awesome humans that they are.

3.2.3 CRI

A PBIS team leader discussed how their school had focused on adapting lessons so students could see themselves in them:

That's been a big focus this year as well, so that also helped change the mindset of things that teachers [were] representing in their classroom or the images they were using. Could any child in [name] High School walk into a classroom and see themselves in that room? So those are some things that did change.

Leadership team perspectives on changes in staff attitudes, knowledge, and behavior

Early in implementation, many PBIS team leaders said that their teachers were slow to take up the initiative. They said that many teachers felt that it added more to their already full plate of responsibilities, felt that this would eventually go away like other initiatives, did not fully understand the objectives of the initiative, or felt it was unrealistic and asking too much of them.

Teachers think they have to be strict because "If I don't think certain group will take over in *my* classroom. You guys aren't in here when I'm dealing with this, and you want me to participate in something that might not work for me in my classroom at this time."

The word “positive” throws teachers and they think that’s all that’s involved. Teachers say, “We can’t be positive all the time.” We try to [explain to] them that’s not all [that] PBIS means. Slowly teachers are acknowledging benefits of what we’re doing.

As implementation progressed, PBIS leadership team participants said they saw favorable shifts in teachers’ attitudes toward the initiative, even though there was still a long way to go. Some participants discussed how they had been working with their teachers to encourage them to resolve disciplinary issues using RP. They tried to use consistent language to describe what they have learned as “tools” and a “toolbox.” One of the PBIS team leaders was known among his teachers as the least likely administrator to suspend, and said they redirected teachers back to their RP tools with prompts like:

“What did the conversation look like with the kid? How did you respond to this? What did the parents say?” If I can get that consistently across the board, I do see it benefiting not just the students, but also the teachers. I think [they’ll] feel a lot more empowered and know that they do have the power to make a shift with these students where they’re good for sure.

One PBIS team leader felt that the initiative was about giving teachers the tools to help them connect with students in ways they had not been able to before:

What I’ve learned is that it’s not so much about the kids sometimes getting those tools. I mean it is, they do learn language and ways to express themselves. But a lot of times it’s a way for people at the school to figure out what’s going on with the kids to better connect with them, because we had issues happening with students that were beyond what we would have ever imagined they would be dealing with, and definitely beyond what we might have dealt with at their age. And so having a consistent time or space or the language for them to be able to talk about these things and for people to get that insight [was important].

Again, everyone doesn’t have that skillset, so every teacher, every admin doesn’t have that skillset to naturally have kids feel compelled or comfortable talking to them. So this provides a space for us to be able to gain more insight into what’s going on with our students, what’s our priority for them? What they’re dealing with, so we can know how to redirect them, know what they do need to stay out of trouble or whatever. So using all these things together combined, it helps better equip us, I feel like, more than just to equip the students. Like [participant] was saying, we have to be very intentional and very honest and transparent about everything that’s happening with our Black male students. And, honestly, Black female students as well.

3.3 Outcome Evaluation

This section presents two main categories of outcome evaluation results. First, we present the various disproportionality metrics described in Section 3.3.1, which are based on administrative data. Second, we present results of statistical analyses to test (1) differences between treatment and control group schools, and (2) the extent to which changes in outcomes within treatment schools are a function of implementation levels of PBIS and, separately, RP. We present results of these statistical analyses for the administrative data outcomes, staff survey data, and student survey data.

3.3.1 Administrative Data Disproportionality Metrics

In this section, we discuss highlights of the administrative data metrics, first for ODRs and then for OSSs. This section discusses both male students, and female students, although the study and the analyses presented in the Hypothesis Tests section focused on male students. This section focuses on the across-school averages for the treatment group and the control group. To derive these averages, we calculated each score once per school year at each school, then averaged scores across the schools. Student population did not weigh scores more or less heavily; each school contributed equally to the average. The data on which the metrics are based were subject to the limitations related to the COVID-19 pandemic school closures, as discussed in Section 2.3.2.

ODRs

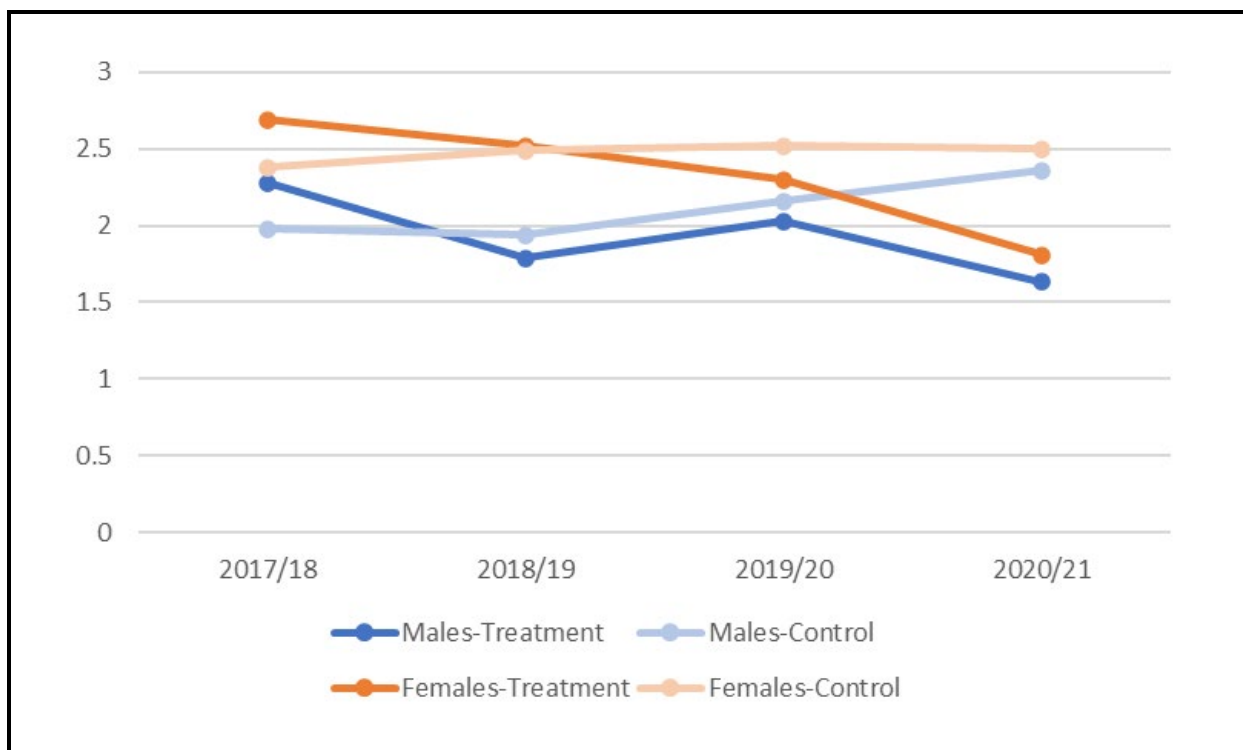
Key metrics for ODR are presented in **Exhibit ODR-Metrics** in **Appendix A**. As expected based on research in the field, in the pre-intervention, 2017–2018 school year, scores had substantial disproportionalities between Black and non-Black male students, with school staff being more likely to give Black male students one or more ODRs. The **risk index** (the proportion of students in a group who received one or more ODRs) was 0.34 for Black male students in both the treatment and the control group schools, whereas non-Black male students had risk indices of 0.16 at treatment schools and 0.18 at control schools. We used these risk index scores to generate **risk ratio** scores, calculated as Black male students' risk index divided by non-Black male students' risk index. The risk ratios for Black male students were 2.28 in treatment schools and 1.98 in control schools. That means, in the 2017–2018 school year, Black male students in treatment schools were, on average, 2.28 times as likely to receive at least one ODR as non-Black male students; in control schools, the risk ratio was less, 1.98.

Similar risk disproportionalities in ODR were seen among female students. Black female students' risk index scores (0.27 in treatment schools, 0.24 in control schools) were somewhat lower than those of Black male students, but higher than both non-Black male students (0.16 in treatment schools, 0.18 in control schools) and non-Black female students' risk index scores (the lowest of the four groups' scores, 0.12 in treatment schools

and 0.11 in control schools). The risk ratio for female students overall was 2.69 in treatment schools and 2.38 in control schools, higher than the respective risk ratios for male students.

Figure 3.1 shows ODR risk ratios of students in each school year, disaggregated by gender. (ODR risk ratios in the 2020–2021 school year were artificially low due to in-person school closures and are not displayed.) No clear and consistent group differences appear in the trendlines. The lower 2020–2021 ODR risk ratios for students in treatment schools appears promising, but we must bear in mind the cautions about interpreting changes in risk ratios discussed earlier (e.g., see Curran, 2020). Figure 3.1 is meant to convey descriptive information; our analysis of change over time in disproportionalities is the focus of **Section 3.2.2**.

Figure 3.1. ODR Risk Ratios, by Group and School Year

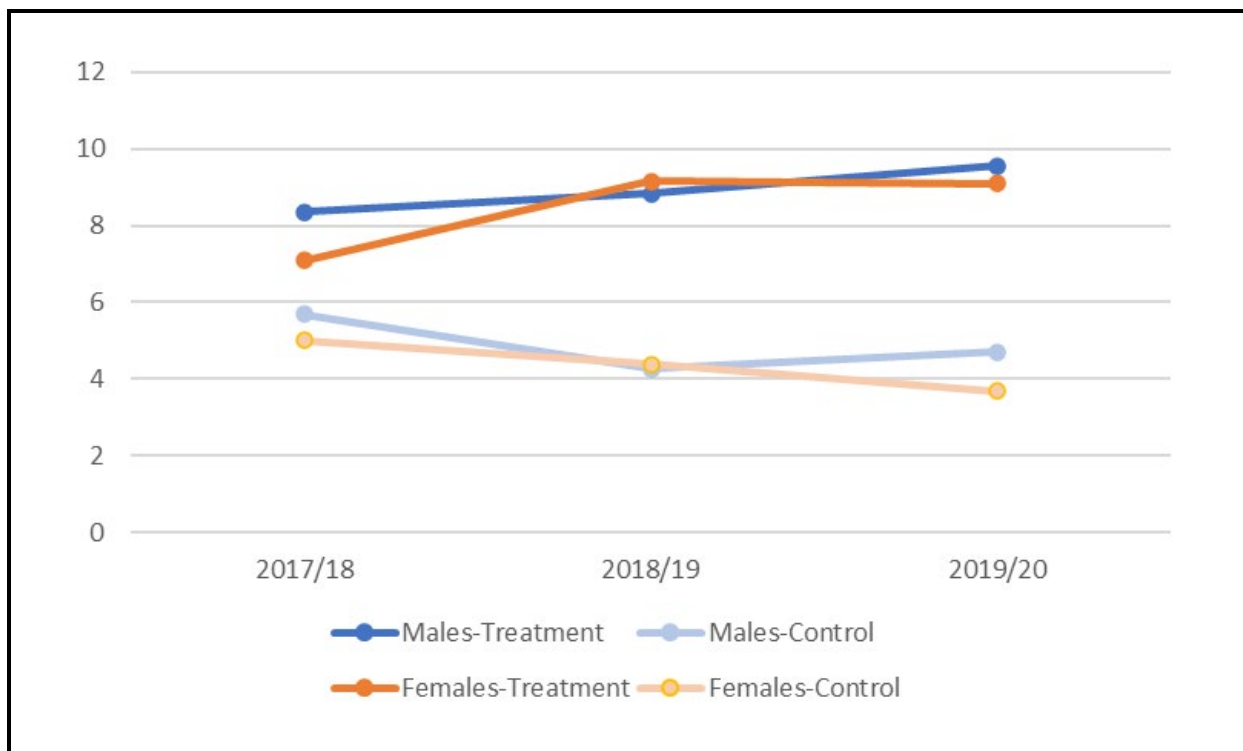


ODR disproportionality is also evident in the scores for **raw differential representation (RDR)**, which Girvan et al. (2019) described as, “the estimated number of students in a target group who did experience discipline but who would not have if students in that group were subject to discipline at the same rate as students in the reference group.” In the 2017–2018 school year, the RDR for male students was 87.7 in treatment schools and 67.8 in control schools; for female students the respective RDRs were 69.4 and 56.2. RDR conveys a sense of how many Black students per school, on average, were impacted by risk

disproportionalities, but each school’s RDR is partly a function of its enrollment size. **RDR per 100 students (RDR-100)** adjusts for enrollment size and is comparable across schools of different sizes. The 2017–2018 RDR-100 for male students was 8.36 in treatment schools and 5.68 in control schools; for female students the respective RDR-100s were 7.10 and 5.00.

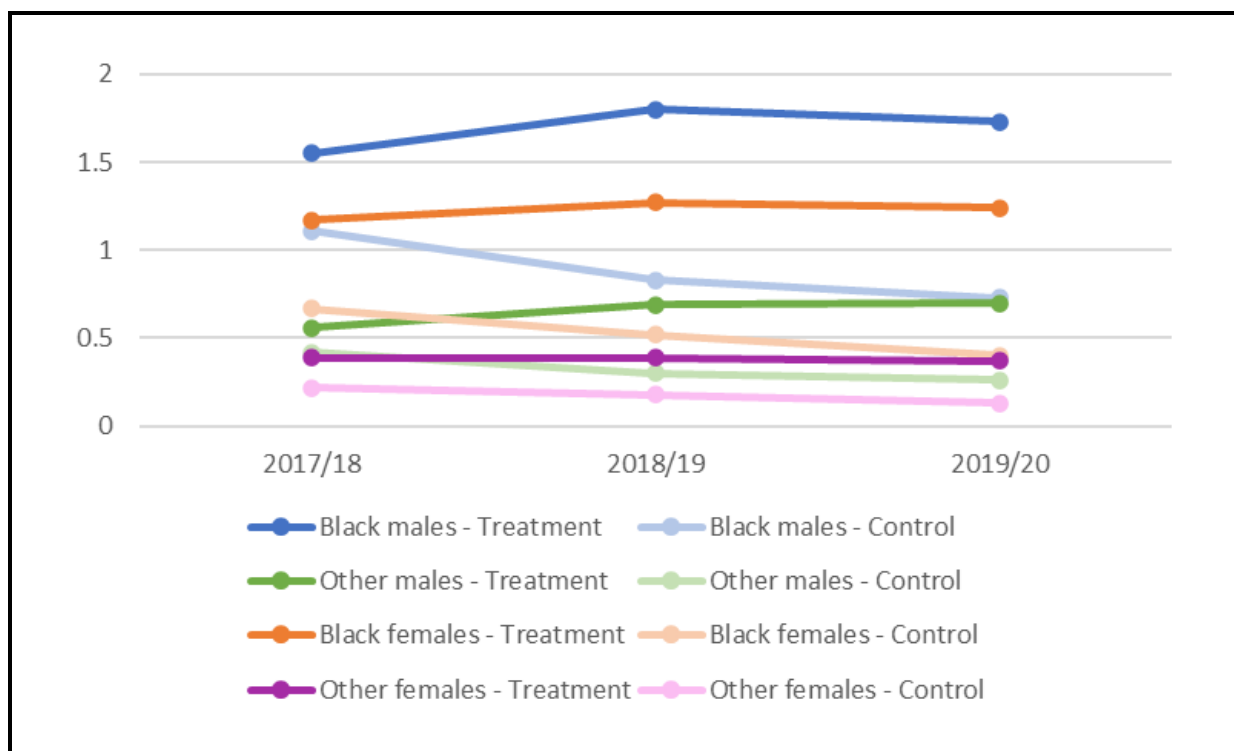
Figure 3.2 displays ODR RDR-100s by gender in each school year. Each school year, ODR RDR-100 was higher in treatment than control schools for all genders. Again, this Figure is meant to convey descriptive information and not test for change over time; the latter tests are discussed in **Section 3.3.2: Hypothesis Tests**.

Figure 3.2. ODR RDR per 100 Students, by Group and School Year



The final ODR metric is **discipline rate**, scaled to the number of students (divided by 100) and the number of school days. In the 2017–2018 school year, the rate in treatment schools was highest for Black male students (1.55), then Black female students (1.17), then non-Black male students (0.56), and lowest for non-Black female students (0.39). A similar pattern was seen in control schools, but rates were lower for each group: Black male students, 1.11; Black female students, 0.67; non-Black male students, 0.42; and non-Black female students, 0.22. In the 2018–2019 and 2019–2020 school years, treatment group Black male students and, to a lesser extent, Black female students were disciplined at higher rates than other students (see **Figure 3.3**).

Figure 3.3. ODR Disciplinary Rate, by Group and School Year



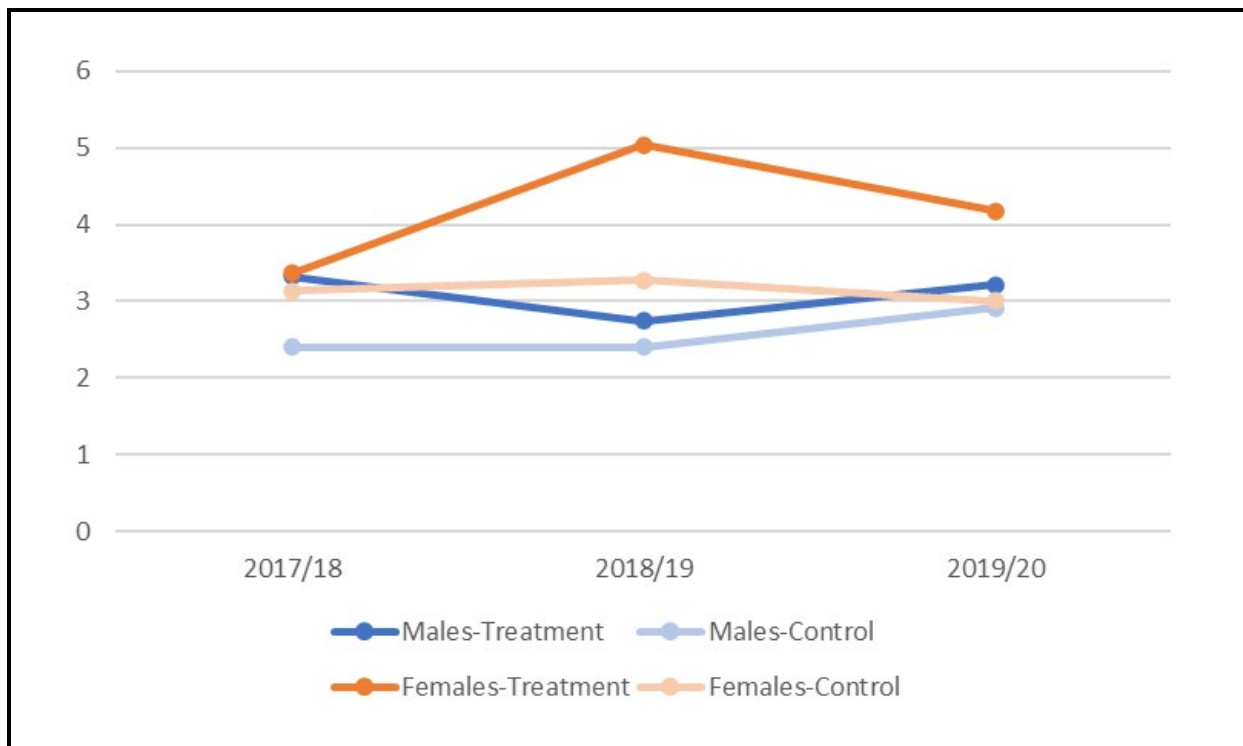
OSSs

The same key metrics for OSS are shown in Appendix A. As with ODR, in the pre-intervention (2017–18) school year, there were substantial disproportionalities between Black and non-Black male students in their likelihood of receiving one or more OSSs. The OSS **risk index** was 0.18 for Black male students in treatment schools and 0.17 for Black male students in the control group schools, compared with non-Black male students, whose risk index was 0.08 in treatment schools and 0.07 in control schools. These risk index scores translate into **risk ratio** scores of 3.33 in treatment schools and 2.40 in control schools. That is, in the 2017–2018 school year, Black male students in treatment schools were, on average, 3.33 times as likely to receive an OSS as other male students; in control schools, the risk ratio was slightly less but still substantial, 2.40.

Similar risk disproportionalities in OSS were seen among female students. Black female students’ risk index scores (0.14 in treatment schools, 0.11 in control schools) were somewhat lower than Black male scores, but higher than both non-black male students’ scores (0.08 in treatment schools, 0.07 in control schools) and non-Black female students’ risk index scores (the lowest of the four groups’ scores, 0.05 in treatment schools and 0.04 in control schools). The risk ratio for female students was 3.37 in treatment schools and 3.13 in control schools, both higher than the risk ratios for male students.

Figure 3.4 shows OSS risk ratios in each school year, disaggregated by gender. (OSS risk ratios in the 2020–2021 school year were artificially low due to in-person school closures and are not displayed.) The most noteworthy finding is a large increase for treatment group female students in the 2018–2019 school year (5.04, up from 3.37 the preceding year). In the 2018–2019 school year, Black female students in treatment schools were, on average, five times as likely to receive at least one OSS as non-Black female students in the same schools. This risk ratio decreased in the 2019–2020 school year but was still very high at 4.18.

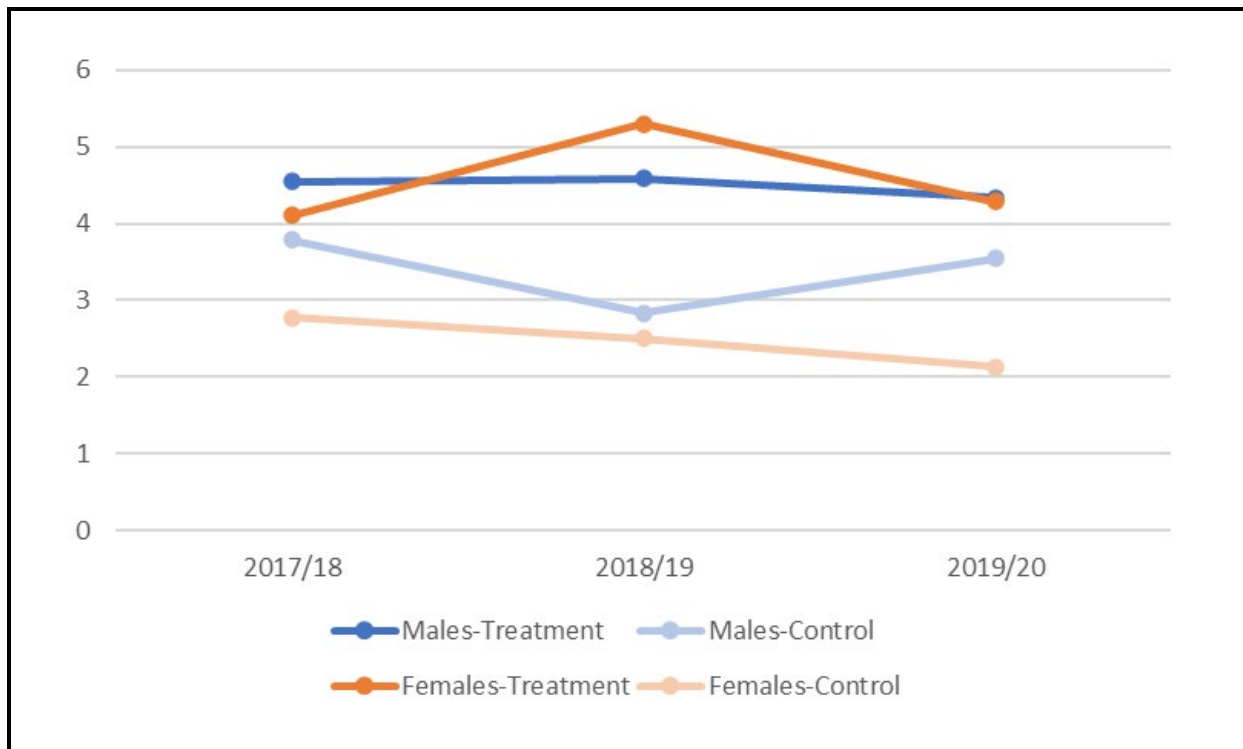
Figure 3.4. OSS Risk Ratio, by Group and School Year



OSS disproportionality is also evident in the scores for **RDR**. In the 2017–2018 school year, the RDR for male students was 49.2 in treatment schools and 44.6 in control schools; for female students, the RDRs were 39.6 in treatment schools and 30.4 in control schools. The 2017–2018 OSS RDR-100 for male students was 4.55 in treatment schools and 3.79 in control schools; for female students, the RDR-100s were 4.11 in treatment schools and 2.77 in control schools.

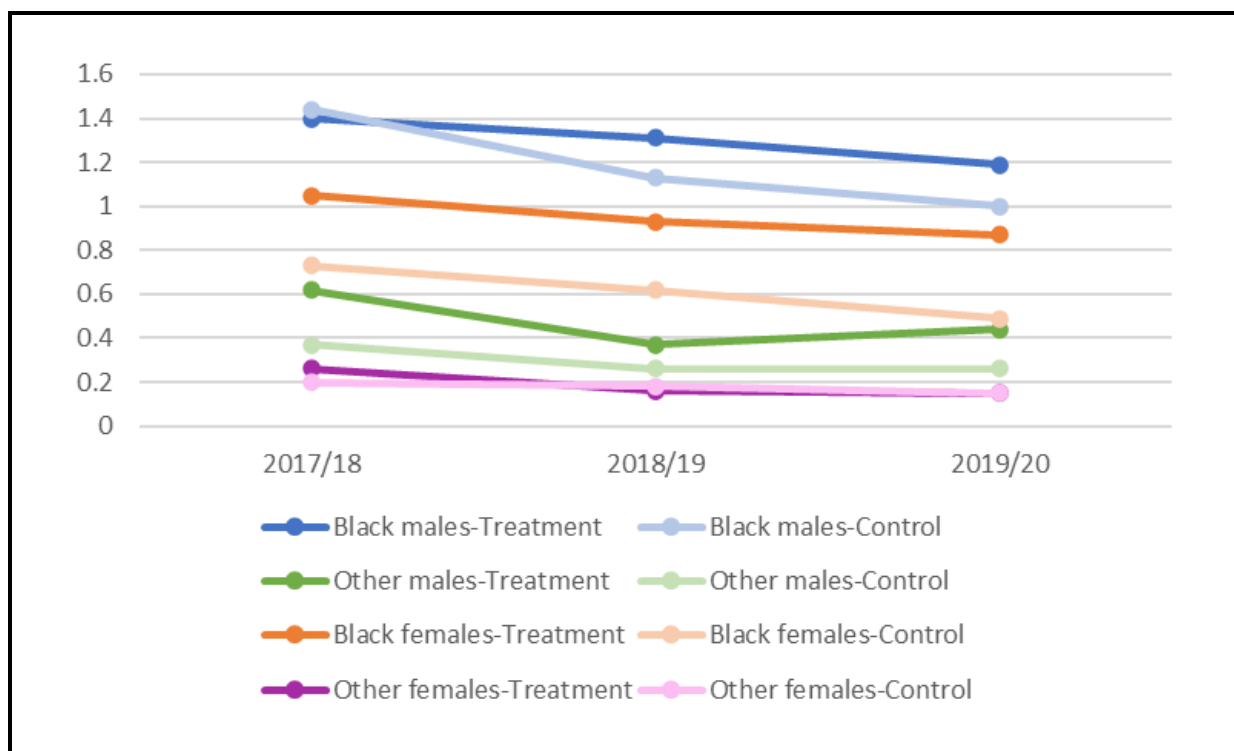
Figure 3.5 shows OSS RDR-100 for male students and female students in each school year. (Like ODR, OSS RDR scores in the 2020–2021 school year were artificially low due to in-person school closures and are not displayed.) Each school year, OSS RDR-100 was higher in treatment than control schools for all students. Again, this Figure is meant to convey descriptive information and not test for change over time.

Figure 3.5. OSS RDR-100, by Group and School Year



The final OSS metric is **discipline rate**. As was the case for ODR, in the 2017–2018 school year, the OSS rate in treatment schools was highest for Black male students (1.40), then Black female students (1.05), then non-Black male students (0.62), and lowest for non-Black female students (0.26). A similar pattern was seen in control schools: Black male students, 1.44; Black female students, 0.73; non-Black male students, 0.37; and non-Black female students, 0.26. In the 2018–2019 and 2019–2020 school years, treatment group Black male students and, to a lesser extent, non-Black male students were disciplined at higher rates than other students (see **Figure 3.6**).

Figure 3.6. OSS Disciplinary Rate, by Group and School Year



3.3.2 Hypothesis Tests

This section presents findings related to outcomes, first for the administrative data on ODRs, OSSs, and absences, and then for the staff survey data. For each type of data, we present results of statistical analyses to test (1) differences between treatment and control group schools, and (2) the extent to which changes in outcomes within treatment schools are a function of implementation levels of PBIS and, separately, RP.

Administrative data results

We based our hypothesis-testing outcome analyses on the source data used to compute the disproportionality metrics, namely administrative data on each student’s race and gender, the number of ODRs and OSSs each student received each school year, and the number of days the student was absent (whether excused or unexcused). For each outcome, we applied two analytic frameworks.

1. Based on the school-randomized design, we compared changes in outcomes for schools in the treatment condition with schools in the control condition.
2. Within the treatment condition, we used each school’s PBIS and RP implementation levels as independent variables (separately for PBIS and RP) to explain differential changes in outcomes.

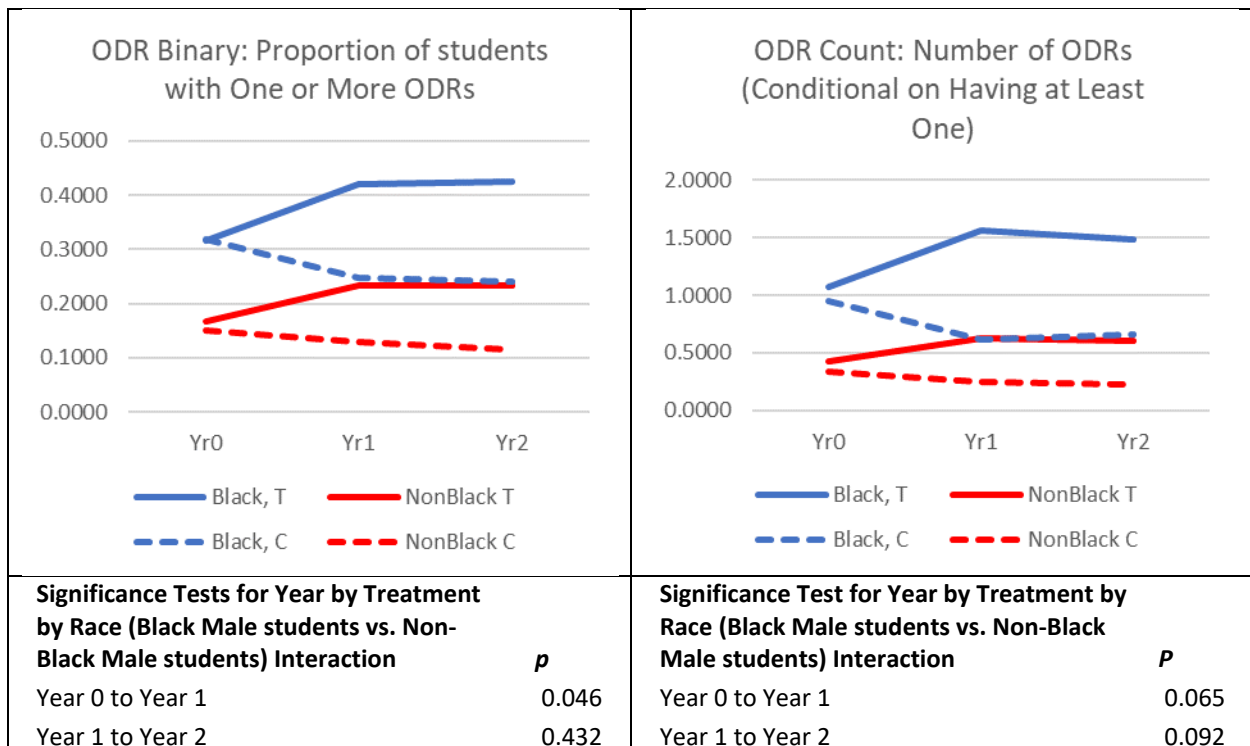
To help interpret the following results, we provide this reminder of the timing of intervention trainings in treatment group schools.

- **Year 0** is the 2017–2018 school year, prior to any intervention training as part of the study (although staff in two treatment schools received training in RP in 2017 and 2018).
- **Year 1** is the 2018–2019 school year; four treatment schools received PBIS training in July or August 2018, one school in October 2018, one school in November 2018, and one school not till April 2019.
- **Year 2** is the 2019–2020 school year; six treatment schools received RP training August 2019, whereas the school that was late in receiving PBIS training did not receive RP training.

Comparing outcomes for treatment and control group schools

Figure 3.7 shows the ODR binary and count trendlines for Black male students and non-Black male students in treatment and control schools, along with the significance tests for three-way interactions between (1) year by (2) treatment by (3) race, first comparing Year 0 with Year 1 and then comparing Year 1 with Year 2.

Figure 3.7. ODR Outcomes, by Year, Treatment Condition, and Race



Note: T = treatment group, C = control group.

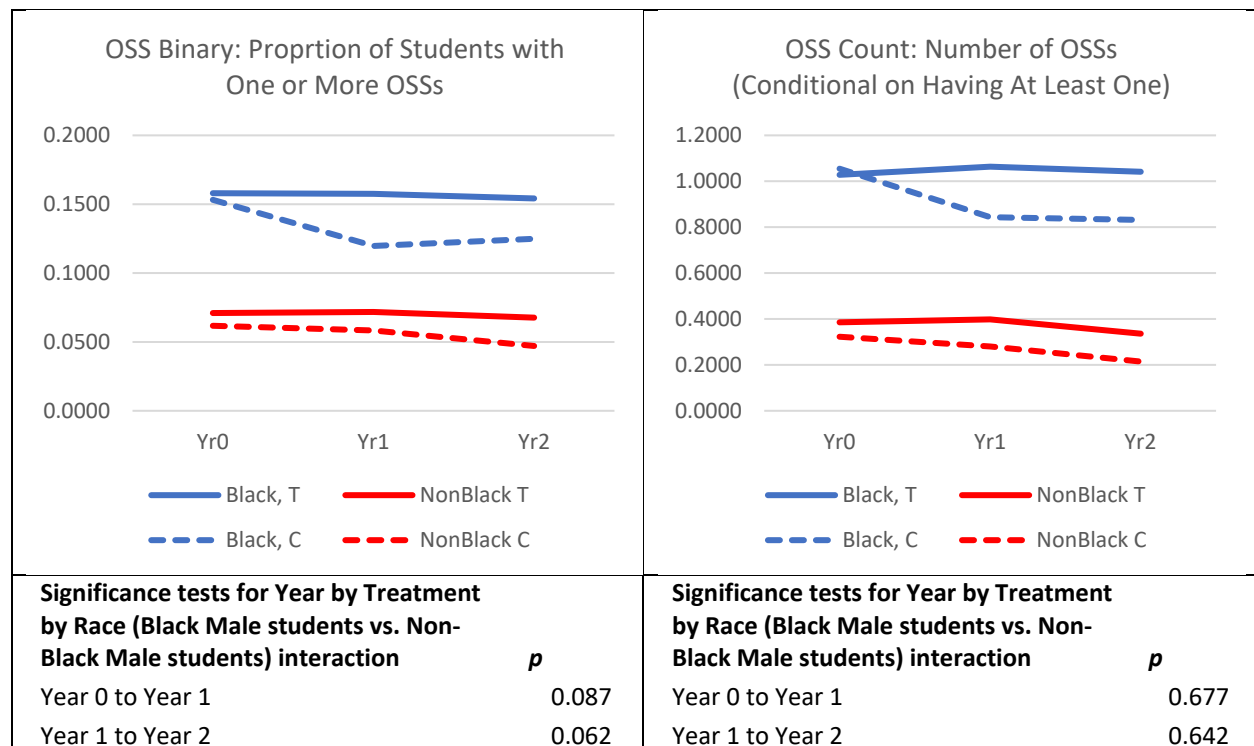
For the ODR binary outcome, the three-way interaction was significant in comparing Year 0 with Year 1, but not significant in comparing (in a separate test) Year 1 with Year 2. In treatment schools there were dramatic increases from Year 0 to Year 1 for both Black male

students (0.32 to 0.42) and non-Black male students (0.17 to 0.23), though at lower levels for non-Black male students. These increases were not seen in control schools. Conversely, ODR binary decreased from Year 0 to Year 1 for Black male students (0.32 to 0.25) and was relatively flat for non-Black male students (0.15 to 0.13). The Year 0 to Year 1 increases in ODR binary seen in treatment schools did not persist during Year 2. It is tempting to ascribe this leveling off to a lagged effect of PBIS training in Year 1, but that interpretation begs the question of why the treatment schools experienced the initial sharp increases for Black and non-Black male students during the school year in which PBIS training took place.

The trendlines for the ODR count outcome resembled the ODR binary trendlines, but the three-way interaction was not quite statistically significant for either interval: from Year 0 to Year 1, $p = .065$, and from Year 1 to Year 2, $p = .092$.

Figure 3.8 shows the OSS binary and count trend lines for Black male students and non-Black male students in treatment and control schools, along with the significance test for three-way interaction (year by treatment by race).

Figure 3.8. OSS Outcomes, by Year, Treatment Condition, and Race



Note: T = treatment group, C = control group.

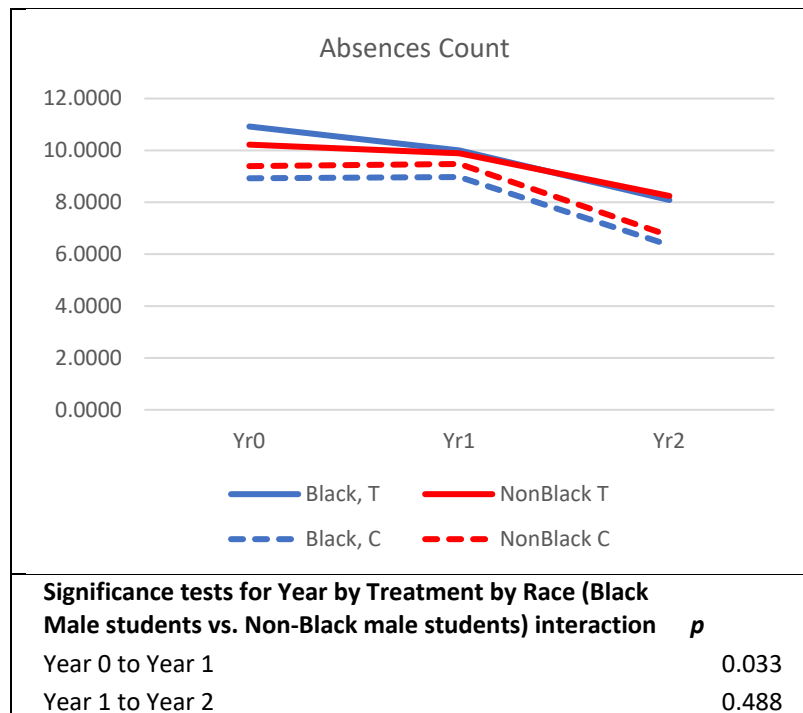
For the OSS binary outcome, the three-way interaction fell slightly short of statistical significance in comparing Year 0 with Year 1 ($p = 0.087$) and in comparing (in a separate test) Year 1 with Year 2 ($p = 0.062$). In treatment schools, the OSS binary trend was fairly

flat for both Black male students and non-Black male students. In control schools, the OSS binary trend for Black male students decreased somewhat from Year 0 to Year 1, which was not seen for non-Black male students.

The trendlines for the OSS count outcome resembled the OSS binary trendlines, but the three-way interaction terms were far from statistically significant for either interval: from Year 0 to Year 1, $p = .677$ and from Year 1 to Year 2, $p = .642$. The OSS count analysis was conditioned on a student having one or more OSS events; values below 1 were possible because the data recorded OSS *days* and some OSSs were for less than a full school day.

Figure 3.9 shows the absences count trend lines for Black male students and non-Black male students in treatment and control schools, along with the significance test for three-way interaction (year by treatment by race).

Figure 3.9. Number of Absences, by Year, Treatment Condition, and Race



Note: T = treatment group, C = control group.

For the absences outcome, the three-way interaction is significant ($p = 0.03$). This effect is driven by a substantial decrease from Year 0 to Year 1 for treatment group Black male students (10.92 to 10.00), a much steeper decrease than that for treatment group non-Black male students (10.22 to 9.89). Stated differently, in treatment group schools, Year 0 absences for Black male students were substantially higher than for non-Black male students (10.92 vs. 10.22), but in Year 1 the difference in absences was greatly reduced (10.00 vs. 9.89). In control group schools, absences did not change substantially from

Year 0 to Year 1 for either Black male students (8.93 to 8.97) or non-Black male students (9.39 to 9.48). These results suggest that PBIS training and implementation in Year 1 helped to reduce absences, for Black male students in particular but also to a lesser degree for non-Black male students.

When comparing Year 1 with Year 2, the three-way interaction term was not significant. Absences decreased for all four groups (treatment group Black male students and non-Black male students, control group Black male students and non-Black male students), which is most likely due to in-person school being suspended in March 2020 because of the COVID-19 pandemic. Of importance here is that the *relative* decrease in absences from Year 1 to Year 2 did not differ among the four groups, suggesting that the greater reduction in absences from Year 0 to Year 1 for treatment group male students compared with their control group counterparts was not sustained in Year 2.

Using PBIS and RP implementation levels to explain outcomes

We explored whether outcomes were associated with differing levels of implementation of PBIS and, separately, RP among schools. In these analyses, we used data on two aspects of implementation described previously: (1) the SET and the TFI measure of PBIS implementation, and (2) the percentage of teachers who reported leading RP circles. One set of models examined changes from the 2017–2018 school year (Year 0) to the 2018–2019 school year (Year 1), and a second set examined changes from the 2018–2019 school year (Year 1) to the 2019–2020 school year (Year 2). Each model examined the three-way interaction of implementation level, school year (using Year 0 as the reference), and race (comparing Black and non-Black male students).

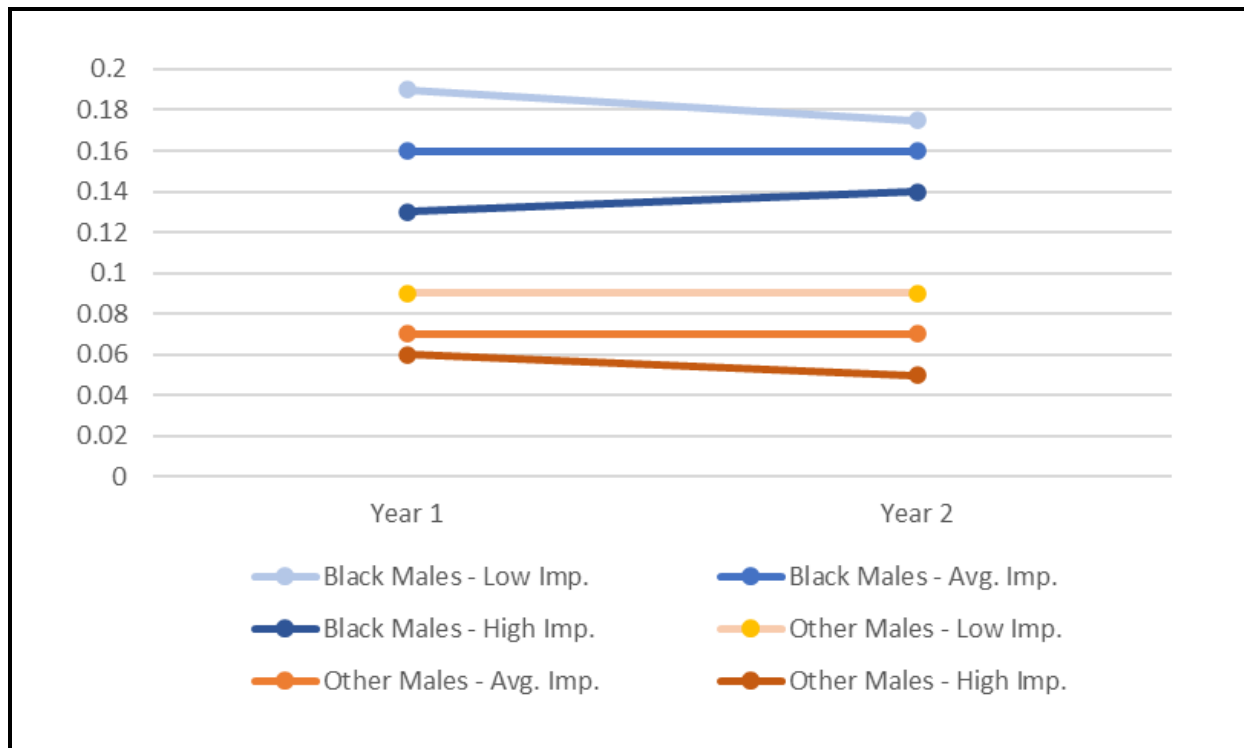
Associations between outcomes and SET/TFI implementation levels. For models using the SET/TFI score as the implementation measure, just one statistically significant interaction ($p = .046$) involved Year 1 to Year 2 change in the OSS binary outcome and one that fell slightly short of significance ($p = .052$) involving Year 0–Year 1 change in the ODR count outcome (see **Table 3.7**). These two statistically significant results are shown in bold font and highlighted cells in the table.

To explicate the interaction involving the OSS binary outcome, we plotted the model-estimated means for Black male students and non-Black male students at Year 1 and Year 2, for three illustrative levels of implementation: the average TFI score across schools in Year 2, one standard deviation below the average (which we call low implementation), and one standard deviation above the average (which we call high implementation). See **Figure 3.10**.

Table 3.7. Significance Tests of Associations between Outcomes and SET/TFI Implementation Levels

	Year 0 (2017/18) to Year 1 (2018/19)					Year 1 (2018/19) to Year 2 (2019/20)				
	Estimate	Standard Error	DF	t Value	Pr > t	Estimate	Standard Error	DF	t Value	Pr > t
ODR Binary	-0.3847	0.2615	51,197	-1.47	0.1412	-0.2272	0.4614	28,766	0.49	0.6225
OSS Binary	-0.2722	0.3154	51,197	-0.86	0.3882	-1.2624	0.6336	28,766	1.99	0.0463
ODR Count	-0.6038	0.3107	51,197	-1.94	0.0520	-0.6574	0.4996	28,766	1.32	0.1882
OSS Count	-0.3113	0.6180	51,197	-0.50	0.6145	-0.3502	1.1391	28,766	0.31	0.7585
Absences Count	-0.0116	0.1214	51,197	-0.10	0.9242	0.2202	0.2306	28,766	-0.96	0.3396

Figure 3.10. Changes in OSS Binary Outcome as a Function of Year 2 TFI Scores



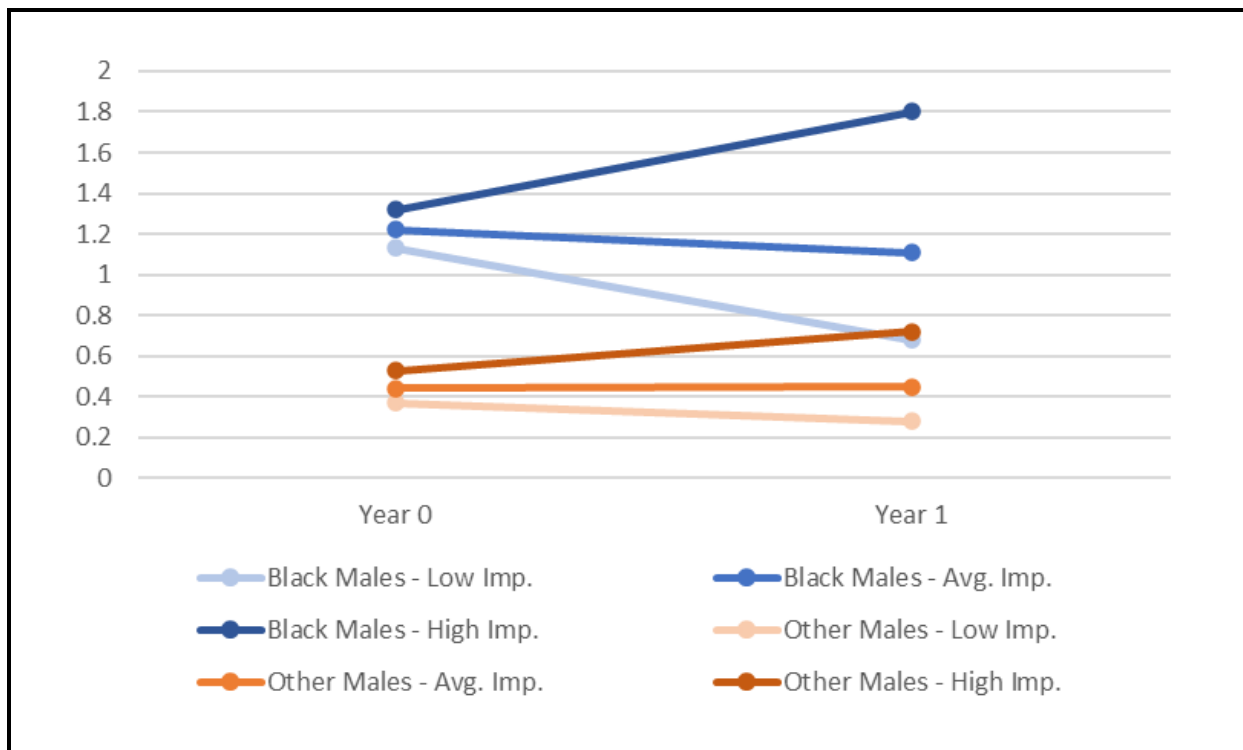
Note: Imp. = implementation.

As previously mentioned and seen above, school staff was more likely to give OSSs to Black male students (blue lines in Figure 3.10) than to non-Black male students (orange lines in Figure 3.10). Our model-based estimates suggest that in Year 1, Black male students in schools with *low* SET and TFI scores in Year 2 (i.e., lower implementation of PBIS) were more likely to receive at least one OSS than students with *average* SET scores, and Black male students in schools with average SET scores were more likely to receive at least one

OSS than those in schools with *high* SET and TFI scores. In Year 2, the differences in the OSS binary outcome for Black male students in schools with differing levels of SET scores remained but were somewhat reduced. Non-Black male students were similarly more likely to receive an OSS if they attended a school with low SET scores and less likely to if they attended a school with high SET scores.

The near-significant effect involved the change in ODR count from Year 0 to Year 1. As reported previously, ODR counts were routinely higher for Black male students than non-Black male students. Surprisingly, Black male students in schools with higher SET scores in Year 1 had higher ODR counts than those in schools with average SET scores; the latter had higher ODR counts than Black male students in schools with lower SET scores (see **Figure 3.11**). A similar pattern was seen for non-Black male students, though with smaller differences among schools with low, average, and high SET scores.

Figure 3.11. Changes in ODR Count Outcome as a Function of Year 1 SET Score



Note: Imp. = implementation.

Associations between outcomes and RP implementation levels. For models using the percentage of teachers who reported leading RP circles as the implementation measure, there were no statistically significant interactions of implementation level, year, and race (see **Table 3.8**).

Table 3.8. Significance Tests of Associations between Outcomes and RP Implementation Levels

	Year 0 (2017/18) to Year 1 (2018/19)					Year 1 (2018/19) to Year 2 (2019/20)				
	Estimate	Standard Error	DF	t Value	Pr > t	Estimate	Standard Error	DF	t Value	Pr > t
ODR Binary	0.2549	0.2021	27401	1.26	0.2073	0.2650	0.2080	28766	1.27	0.2026
OSS Binary	0.06069	0.2747	27401	0.22	0.8251	0.0326	0.3207	28766	0.10	0.9191
ODR Count	-0.2448	0.2211	27401	-1.11	0.2682	0.1879	0.2306	28766	0.81	0.4152
OSS Count	-0.4560	0.4479	27401	-1.02	0.3086	-0.5336	0.5235	28766	-1.02	0.3080
Absences Count	-0.0956	0.0949	27401	-1.01	0.3139	-0.0519	0.1072	28766	-0.48	0.6278

3.3.3 Staff Survey Outcomes

We examined the staff survey to answer questions about how school climate, staff-perceived safety, and staff-perceived self-efficacy for CRI were impacted by the intervention.

1. Based on the school-randomized design, we compared changes in staff survey outcomes for schools in the treatment condition with schools in the control condition.
2. Within the treatment condition, we used each school's PBIS and RP implementation levels as independent variables to explain differential changes in staff survey outcomes.

To help interpret the following results, we provide this reminder of the timing of intervention trainings in treatment group schools.

- **Wave 1** is the fall of the 2018–2019 school year, prior to any intervention training as part of the study (although three treatment schools previously received trained in RP in 2017 and 2018). This conceptually serves as baseline, as Year 0 does in the administrative data.
- **Wave 2** is the spring of the 2018–2019 school year; four treatment schools received PBIS training in July or August 2018, one school in October 2018, one school in November, and one school not till April 2019.
- **Wave 3** is the spring of the 2019–2020 school year before COVID-19; six treatment schools received RP training August 2019 and one school (the one that was late in receiving PBIS training) did not receive RP training.
- **Wave 4** is the spring of the 2020–2021 school year, after schools returned to in-person learning following remote learning. As might be expected, staff responses are notably different for this timepoint as the COVID-19 pandemic continued and schools were just beginning to manage in-person classes.

Comparing outcomes for treatment and control group schools

Figure 3.12 shows the staff-reported **school climate** trendlines for all staff respondents in treatment and control schools, along with the significance tests for wave by treatment interaction, comparing Wave 1 to Wave 2, Wave 2 to Wave 3, Wave 3 to Wave 4, and (to encompass the entire study period) Wave 1 to Wave 4.

When comparing Wave 1 to 2 for the school climate outcome, which is based on subscales from the Inventory of School Climate, the two-way interaction between wave and treatment condition was not significant ($p = 0.25$). Average ratings of climate on this scale decreased for both treatment and control groups, which is commonly due to seasonality: specifically, ratings of school climate typically decrease over the course of the school year, from higher in the fall of a school year, to lower in the spring.

Comparisons of Wave 2 to Wave 3 and of Wave 3 to Wave 4—that is, comparisons of data from one spring to the next—were not statistically significant. However, when comparing Wave 1 to Wave 4, a significant interaction effect was observed between wave and

treatment condition ($p = 0.01$). Least squares means comparisons suggest that from Wave 1 to Wave 4, treatment group school climate improved more than that of control schools.

Figure 3.12. Average Staff Rating of School Climate, by Survey Wave and Treatment Condition

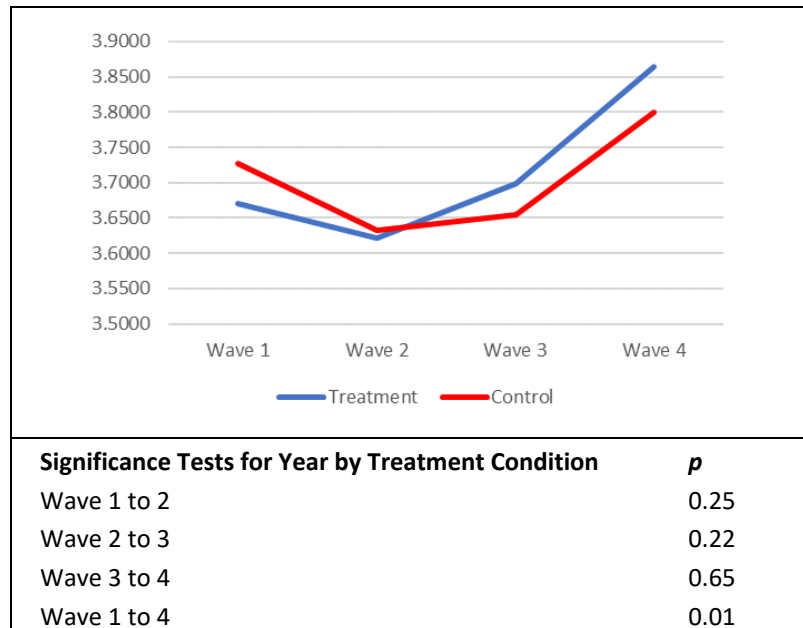


Figure 3.13 shows the staff-reported **authoritative school climate** trendlines for all staff respondents in treatment and control schools, along with the significance tests per wave by treatment interaction, again comparing Wave 1 to Wave 2, Wave 2 to Wave 3, Wave 3 to Wave 4, and Wave 1 to Wave 4.

For the authoritative school climate outcome, which is based on subscales from the Authoritative School Climate survey, the two-way interaction between wave and treatment condition was not significant between Wave 1 and Wave 2 ($p = 0.21$). Average ratings of authoritative school climate on this scale decreased for both treatment and control groups, which is commonly due to seasonality: specifically, ratings of school climate typically decrease over the course of the school year, from higher in the fall of a school year, to lower in the spring.

The wave by treatment condition interaction effects were also nonsignificant when comparing Wave 2 to Wave 3 ($p = 0.70$) and Wave 3 to Wave 4 ($p = 0.20$). By Wave 4, in comparison to Wave 1, results indicate a significant interaction between wave and treatment condition ($p = 0.04$). In treatment schools, staff ratings of authoritative school climate increased from Wave 1 to Wave 4, whereas in control schools Waves 1 and 4 were approximately the same.

Figure 3.13. Average Staff Rating of Authoritative School Climate, by Survey Wave and Treatment Condition

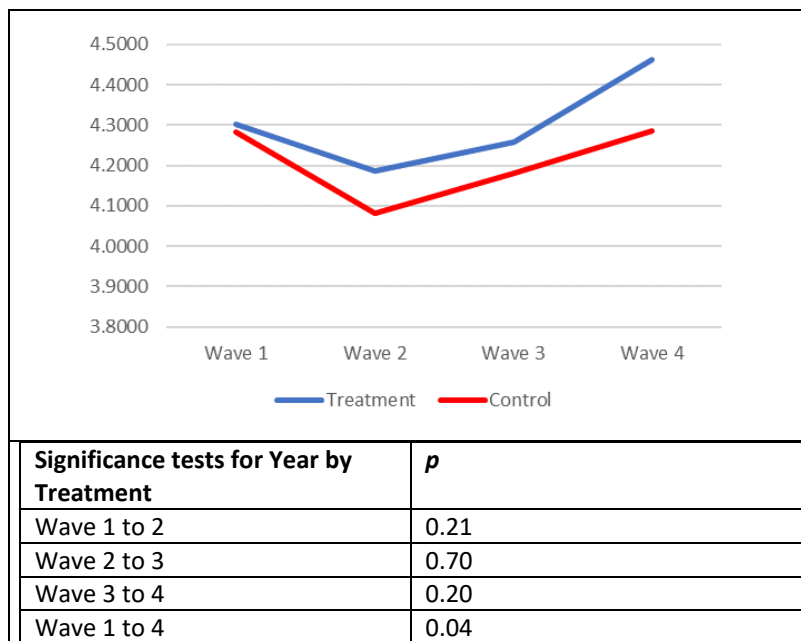


Figure 3.14 shows the staff-reported **perceived safety** trendlines for all staff respondents in treatment and control schools, along with the significance tests for Wave by treatment interaction, again comparing Wave 1 to Wave 2, Wave 2 to Wave 3, Wave 3 to Wave 4, and Wave 1 to Wave 4. Respondents answered using a three-point scale: “never,” “occasionally,” or “most or all of the time.”

For the perceived safety outcome, the two-way interaction between wave and treatment condition was not significant for each comparison between Wave 1 and Wave 2 ($p = 0.95$), Wave 2 and Wave 3 ($p = 0.10$), or Wave 3 and Wave 4 ($p = 0.73$). The final Wave 1 to Wave 4 interaction test approached significance at $p = 0.06$. In this case, there was a trend for the treatment group respondents to increase more than the control group respondents in perceived safety.

Figure 3.15 shows the staff-reported **culturally responsive self-efficacy** trendlines for all staff respondents in treatment and control schools, along with the significance tests for wave by treatment interaction, again comparing Wave 1 to Wave 2, Wave 2 to Wave 3, Wave 3 to Wave 4, and Wave 1 to Wave 4. Respondents used a scale of 1–10 to indicate low to high confidence. However, due to skewness in the data, following data collection, the response options were recoded: responses 1–7 were recoded into a value of “1”; responses 8 and 9 were recoded into a value of “2”; and response 10 was recoded into a value of “3.”

Figure 3.14. Average Staff Rating of Perceived Safety, by Survey Wave and Treatment Condition

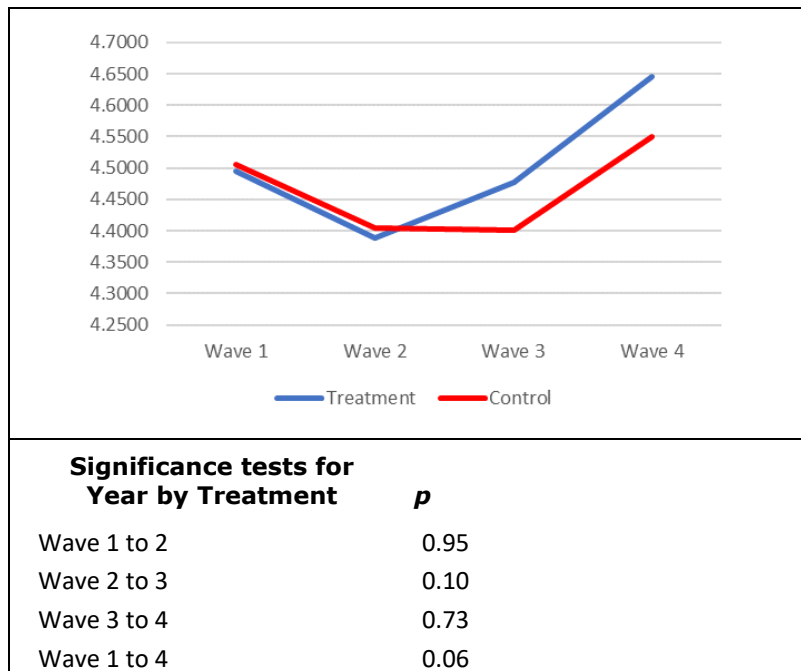
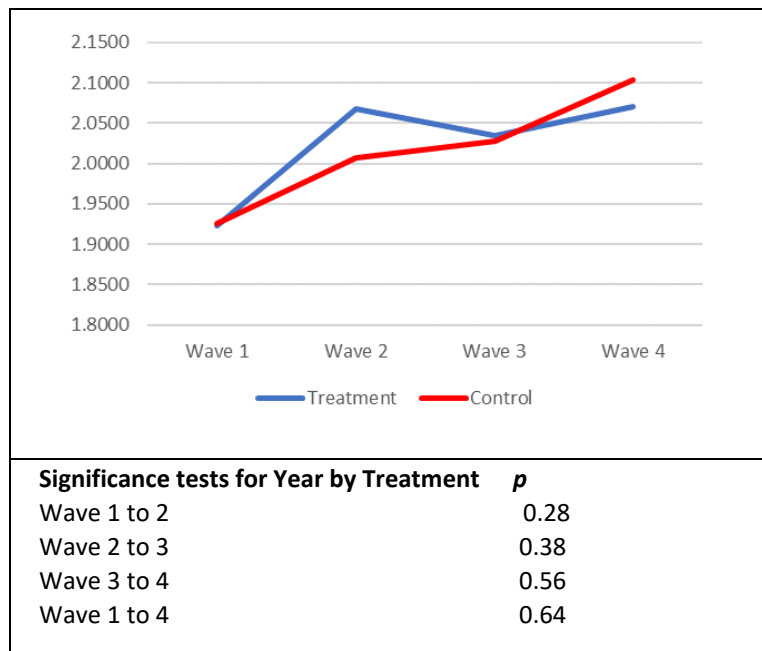


Figure 3.15. Average Staff Rating of Culturally Responsive Self-Efficacy, by Survey Wave and Treatment Condition



For the culturally responsive self-efficacy scale, the two-way interaction between wave and treatment condition was not significant for each timepoint comparison ($p = 0.28, 0.38, 0.56, \text{ and } 0.64$, respectively). Both treatment and control group schools tended to increase over time.

Using PBIS and RP Implementation Levels to Explain Staff Survey Outcomes

We explored whether staff survey outcomes were associated with differing levels of implementation of PBIS and, separately, RP among schools. In these analyses, we used data on two aspects of implementation described previously: (1) the SET/TFI measure of PBIS implementation, and (2) the percentage of teachers who reported leading RP circles. One set of models examined changes from the Wave 1 (fall 2018) to Wave 2 (spring 2019), and a second set examined changes from Wave 2 (spring 2019) to Wave 3 (spring 2020). Each model examined the three-way interaction of implementation level, wave, and race (comparing Black male students with non-Black male students).

Associations between outcomes and SET/TFI implementation levels

For models using the SET/TFI score as the predictor variable, there were no statistically significant interactions for any of the four staff survey outcomes (school climate, authoritative school climate, perceived safety, and culturally responsive self-efficacy) in either the comparisons between Wave 1 and Wave 2 or the comparisons between Wave 2 and Wave 3 (see **Table 3.9**).

Table 3.9. Significance Tests of Associations between Staff Survey Outcomes and SET/TFI Implementation Levels

Outcome	Wave 1 to Wave 2					Wave 2 to Wave 3				
	Estimate	Std. Error	DF	t Value	Pr > t	Estimate	Std. Error	DF	t Value	Pr > t
School Climate	0.0763	0.1240	1669	0.62	0.5386	-0.3032	0.1981	847	-1.53	0.1263
Authoritative School Climate	0.0524	0.1996	1626	0.26	0.7930	-0.3985	0.3163	826	-1.26	0.2081
Perceived Safety	0.0639	0.1555	1637	0.41	0.6812	-0.0875	0.2505	828	-0.35	0.7270
Culturally Responsive Self-Efficacy	0.1393	0.1735	1505	0.80	0.4220	-0.0025	0.2815	758	-0.01	0.9931

Associations between outcomes and RP implementation levels

For models using the percentage of teachers who reported leading RP circles as the implementation measure, there were no statistically significant interactions for any of the four staff survey outcomes (school climate, authoritative school climate, perceived safety, and culturally responsive self-efficacy) in either the comparisons between Wave 1 and Wave 2 or the comparisons between Wave 2 and Wave 3 (see **Table 3.10**).

Table 3.10. Significance Tests of Associations between Staff Survey Outcomes and RP Implementation Levels

Outcome	Wave 1 to Wave 2					Wave 2 to Wave 3				
	Estimate	Std. Error	DF	t Value	Pr > t	Estimate	Std. Error	DF	t Value	Pr > t
School Climate	0.03389	0.0783	944	0.43	0.6653	0.0124	0.1086	847	0.11	0.9093
Author. School Climate	-0.06068	0.1221	920	-0.50	0.6193	0.0999	0.1727	826	0.58	0.5634
Perceived Safety	0.0033	0.1007	926	0.03	0.9739	0.1340	0.1366	828	0.98	0.3269
Cult. Resp. Self-Efficacy	-0.0736	0.1099	861	-0.67	0.5032	0.1535	0.1526	758	1.01	0.3146

Note: Pr = probability.

3.3.4 Student Survey Outcomes

We examined the student survey data to answer questions about how scales related to culturally responsive teaching (CRT), academic efficacy (AE), equity in behavior expectations (EQ), and RP were impacted by the intervention. We applied the same analytic framework as for the staff survey:

1. Based on the school-randomized design, we compared changes in student survey outcomes between treatment and control schools. We tested the three-way interaction of condition (treatment vs. control), respondent race (Black male vs. non-Black male), and timeframe (Wave 1 vs. Wave 2; Wave 2 vs. Wave 3; or Wave 1 vs. Wave 3).
2. Within the treatment condition, we used each school’s PBIS and RP implementation levels as independent variables (separately) to explain differential changes in student survey outcomes. We tested the three-way interaction of implementation level (low, average, or high), respondent race (Black male vs. non-Black male), and timeframe (Wave 1 vs. Wave 2; Wave 2 vs. Wave 3; or Wave 1 vs. Wave 3).

To help interpret the following results, we provide this reminder of the timing of the student survey waves relative to intervention trainings in treatment group schools.

- **Wave 1** was administered in the spring of the 2018–2019 school year; four treatment schools received PBIS training in July or August 2018, one school in October 2018, one school in November, and one school not till April 2019.
- **Wave 2** was administered in the spring of the 2019–2020 school year before the COVID-19 pandemic; six treatment schools received RP training in August 2019 and one school (the one that was late in receiving PBIS training) did not receive RP training.
- **Wave 3** was administered in the spring of the 2020–2021 school year, after schools returned from remote learning to in-person learning. As might be expected, student responses are notably different for this timepoint as the COVID-19 pandemic continued and schools were just beginning to resume in-person classes.

Note that Wave 1 of the staff survey was in fall of the 2018–2019 school year and Wave 2 was in the spring of that school year, and the staff survey had four waves. In contrast, the

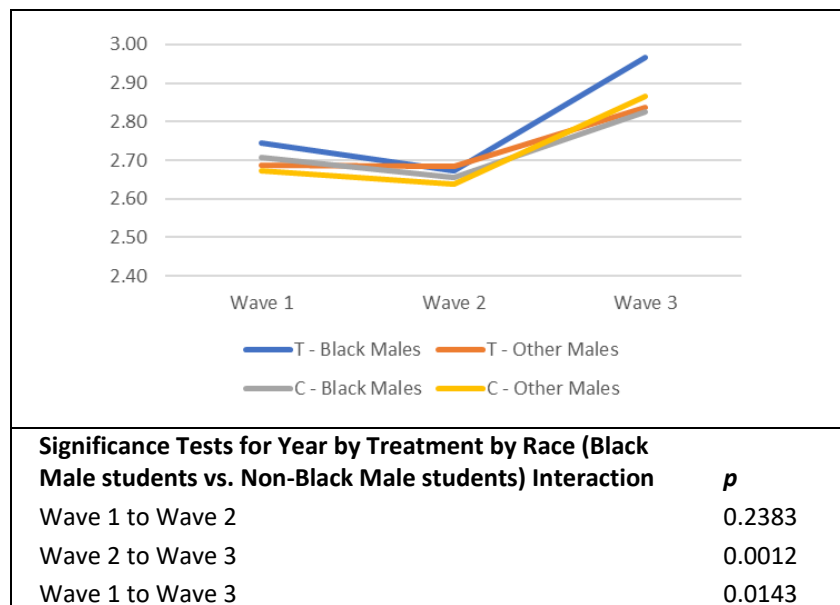
student survey was administered only in the spring of each year and analyses included only three waves of data.

CMS administered the student surveys, as it has for years preceding this study with a similar form. The Wave 1 instrument included the standard items used by CMS. Items for the PBIS, RP, and CRI scales were added to the Wave 2 and Wave 3 instruments for this study.

Comparing outcomes for treatment and control group schools

For each student survey outcome, we tested the three-way interaction of condition (treatment vs. control), respondent race (Black male vs. non-Black male), and timeframe (Wave 1 vs. Wave 2; Wave 2 vs. Wave 3; or Wave 1 vs. Wave 3). The interaction term was significant only for the **CRT** outcome (see **Figure 3.16**). This scale measured level of agreement with statements that the respondent’s teachers provided instructional materials and examples that reflect the respondent’s cultural background, were interested in their culture, and used real-life examples to help explain things.

Figure 3.16. Student Surveys’ CRT Scale Scores by Wave, Treatment Condition, and Race



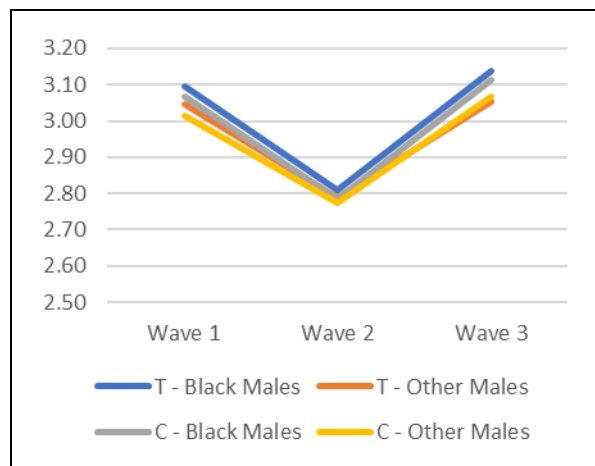
Note: T = treatment group, C = control group.

At Wave 1, the four groups responded similarly regarding the degree to which their teachers exhibited CRI. At Wave 2, all four groups' ratings decreased somewhat and to a similar extent; the interaction term was nonsignificant. At Wave 3, all four groups' ratings increased markedly, especially those of Black male students in treatment condition schools; the interaction term was statistically significant in comparing Wave 3 with both Wave 1 and with Wave 2. This exceptional increase among Black male respondents in treatment condition schools indicates targeted benefits of the intervention.

Similar effects were not observed for the other student survey outcomes: all other three-way interaction tests were not statistically significant. The trendlines and significance tests are shown in **Figure 3.17**.

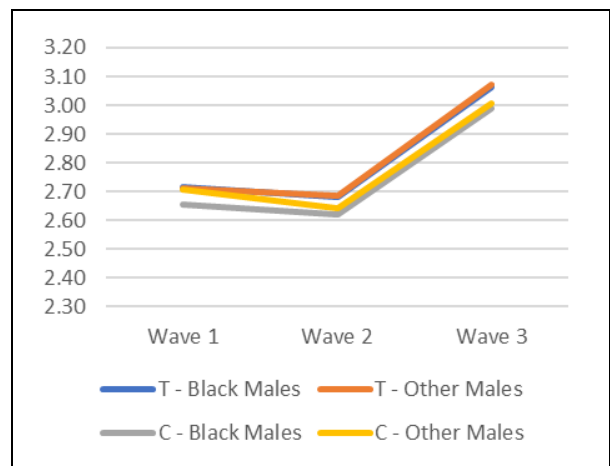
Figure 3.17. Additional Student Survey Scale Scores by Wave, Treatment Condition, and Race

Academic Efficacy



Significance Tests for Year by Treatment by Race (Black Male students vs. Non-Black male students) Interaction	
Wave 1 to Wave 2	
Wave 2 to Wave 3	
Wave 1 to Wave 3	

Equity in Behavior Expectations



Significance Tests for Year by Treatment by Race (Black Male students vs. Non-Black male students) Interaction		<i>p</i>
Wave 1 to Wave 2		0.535
Wave 2 to Wave 3		0.704
Wave 1 to Wave 3		0.354

(continued)

Using PBIS and RP implementation levels to explain student survey outcomes

Similar to staff survey analyses, we explored whether student survey outcomes were associated with differing levels of implementation of PBIS and, separately, RP among schools. In these analyses, we used data on two aspects of implementation described previously: (1) the SET/TFI measure of PBIS implementation, and (2) the percentage of teachers who reported leading RP circles. One set of models examined changes from Wave 1 (spring 2019) to Wave 2 (spring 2020), and a second set examined changes from Wave 2 (spring 2020) to Wave 3 (spring 2021). Each model examined the three-way interaction of implementation level, wave, and race (comparing Black male students with non-Black male students).

None of the interaction terms were statistically significant, neither using RP implementation (see **Table 3.11**) or PBIS implementation (see **Table 3.12**) as the predictor variable, nor for either timeframe comparison (e.g., Wave 1 vs. Wave 2).

Table 3.11. Significance Tests of Associations between Student Survey Outcomes and RP Implementation Levels

Outcome	Wave 1 to Wave 2					Wave 2 to Wave 3				
	Estimate	Standard Error	DF	t Value	Pr > t	Estimate	Standard Error	DF	t Value	Pr > t
CRT Scale	0.0121	0.1060	5378	0.11	0.9090	-0.0415	0.1319	3329	-0.31	0.7534
AE scale	-0.0336	0.1168	5107	-0.29	0.7736	-0.1700	0.1429	3067	-1.19	0.2342
EQ scale	-0.0235	0.1633	5122	-0.14	0.8856	0.0913	0.1823	3077	0.50	0.6163
RP scale	0.0327	0.1193	5367	0.27	0.7843	-0.0190	0.1413	3288	-0.13	0.8931

Note: Pr = probability.

Table 3.12. Significance Tests of Associations between Student Survey Outcomes and PBIS Implementation Levels

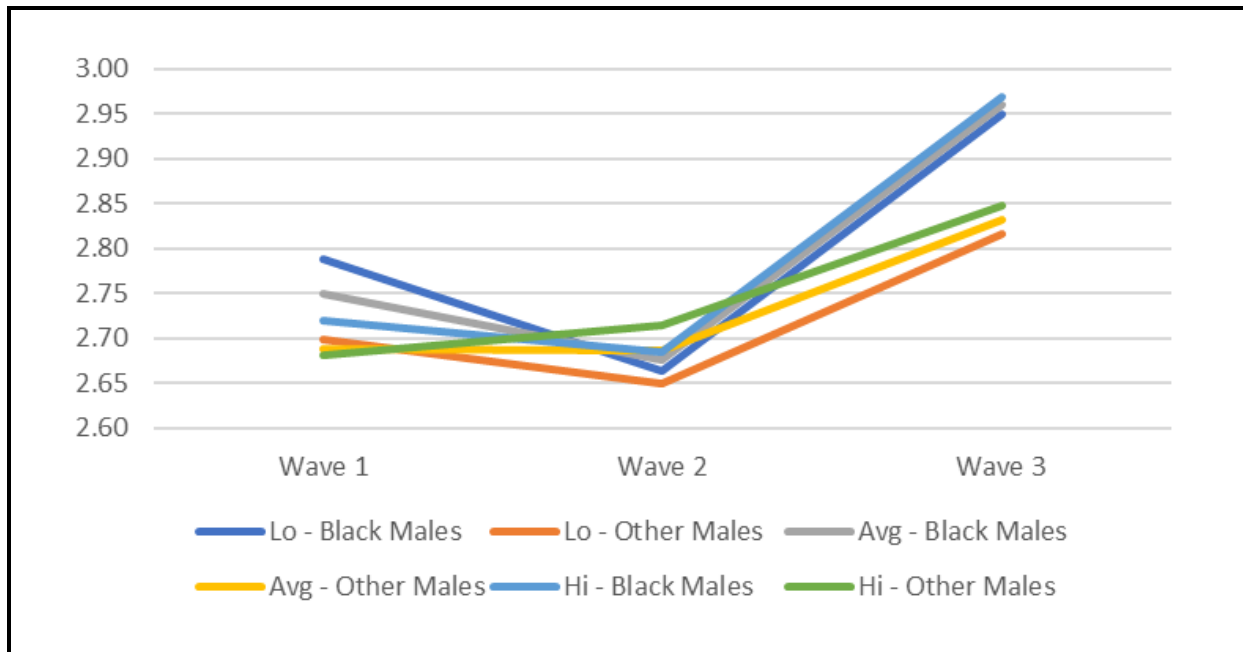
Outcome	Wave 1 to Wave 2				
	Estimate	Standard Error	DF	t Value	Pr > t
CRT Scale	-0.3902	0.2662	5378	-1.47	0.1427
AE scale	-0.3980	0.2957	5107	-1.35	0.1783
EQ scale	-0.1896	0.4145	5122	-0.46	0.6474
RP_scale	-0.1198	0.3042	5367	-0.39	0.6938

Note: Pr = probability.

We provide **Figure 3.18** as an example illustrating the lack of a relationship between the RP predictor variable and the CRT outcome. The graph shows that Black male students reported higher levels of the outcome at Wave 3 than did non-Black male students (echoing

the effect shown above in Figure 3.16), but the level of RP in the students' schools did not matter: students in schools with high, average, and low levels of RP implementation reported similar levels of the culturally responsive teacher outcome at Wave 3.

Figure 3.18. Student Surveys' CRT Scale Score by RP Implementation Level, Wave, and Race



3.3.5 Student Focus Group Outcomes

We conducted three virtual focus groups with Black male students from three treatment schools to qualitatively understand the impact of the interventions. Several important themes emerged from questions related to their experiences in their schools with some of the interventions. Please note, these focus groups were conducted in spring 2021 after most of the students had returned to in-person learning following the beginning of the COVID-19 pandemic.

PBIS

Behavior rules and expectations. Participants agreed their schools' behavioral rules and expectations were clear; rules are clearly posted within a class's syllabus and then teachers review and explain the rules. Some participants also reported administrators or other school staff being in the hallways to tell students what not to do and provide warnings if students committed an infraction. One participant noted that although school staff do a good job letting students know what to do and what not to do, some students just ignore this staff and the rules they provide. One student explained, "If you go in the classroom, it's [the code of conduct] near the door, but yeah, we don't really...after the first couple days, we don't even really look back at it. But, everybody basically knows it."

Fairness of punishment. In two schools, students generally reported that although punishment might be excessive, it was fairly distributed among all students. One student stated, “I feel Black kids, we’re treated equally, but it gets to a point where the teachers don’t have to write up as many kids because they handle the situations in the classroom.” In the third school, participants generally agreed that punishments were not given out fairly to all students. They provided examples of times when all students received punishments, but Black male students received a harsher version. One participant noted that Black female students tend to receive the same level of punishment as Black male students. One student stated:

I don’t think [...] that they’re [punishments are] given off fairly because I’ve seen a few times where somebody getting kicked out of class [...] and they were a Black student, but they [the teacher] might see somebody else do the same thing that’s not a Black student and they’ll... I’m not going to say they’re [the other student] [...] scot-free, but they’ll get less of a punishment.

Emotional safety

Students expressed differing levels of comfort with sharing their emotions and confiding in a teacher if they needed support at school. A couple of participants reported not feeling comfortable openly confiding their feelings to teachers. One participant said this lack of openness to sharing their feelings was a result of not having built a personal connection; others noted they were not the “type of person” to express how they are feeling to someone they are not close with. Participants agreed that teachers and the school as a whole “put themselves out there” and do make an active effort to allow students to share. When asked how comfortable and safe the general school environment made them feel as Black male students, one student noted his comfort in sharing the same race as his teachers. He said, “Pretty comfortable. It’s a lot of the same, our same race, so pretty much comfortable... [In addition] the teachers, they try a close, friendly relationship.”

Physical safety

When discussing physical safety, students usually felt safe in the school and did not feel that bullying and fighting were big problems in their school. One stated, “I’m only a sophomore and I didn’t even have a full freshman year, so I can’t even say that anything’s really happened, that I’ve witnessed, but I could say that I do feel pretty safe.” One student acknowledged that just because he does not feel unsafe does not mean that other students might not experience victimization: “I’ve never felt like I was in danger or threatened or bullied or anything by anybody like that, but I know for some people, they’d probably be feeling like that.”

RP

Focus group participants were asked if their teachers used RP circles in their classrooms before the COVID-19-related school building closures. In one focus group, the students

from that school had experienced RP circles in at least one class before the pandemic, but not during remote learning. One student stated, "They had a thing. I forget what it was called but in the last year, I think it felt like rhetorical something Mondays, but I know [Teacher], she had us [hold] a stuffed bear, and we all took turns talking about topics." In focus groups with students from the other two schools, only one student in each group recalled participating in RP circles in class.

For those students who recalled restorative circles, they were asked to reflect on if the circles helped them in any way or improved the classroom or school environment. Participants reported enjoying participating and hearing their classmates talk openly about their experiences. One said that the circles improved the classroom environment. They all noted that it had a positive effect on building connections in school. One participant reported the use of restorative circles created a connection between them and other students whom they may not have spoken with before. This connection helped improve their understanding of classmates and provided them the space to open up. This participant explained, "I feel more obligated to talk to you [another student] because it only improved my understanding for you and how you thought and how you think. So I feel like you can see everybody's personality a little bit more with the restorative circles." This participant felt the time spent on restorative circles was too short.

CRI

In the focus groups, perspectives of Black male students varied on the extent to which they could relate to class lessons and whether lessons reflected their culture and background. Many shared that it was difficult to relate to class lessons. For one school that had made more progress implementing CRI, two students in that focus group said that found some class lessons relevant to their lives to a certain extent; however, a third student said they had not connected to any assignments since starting high school. He said, "This wasn't in high school, so the last time I connected to a[n] actual assignment, was probably *To Kill a Mockingbird* in eighth grade. [...] I connected with that one the most."

A couple of participants from one school shared that they felt more of a connection to their African American history course. They felt there was less relevancy to an English or math course, and even with some U.S. history courses. One student said,

I know for U.S. history you really only get to learn about a certain amount of it, but they don't really focus on the African American history a lot like that. And I feel like that's not just a problem here, I think that's a problem globally, obviously.

However, a student from another school reported connecting with their weekly current events writing assignment in their U.S. history class, noting that sometimes those assignments focused on issues related to African Americans.

Implicit bias

Participants noted teachers needed to do their best to get rid of preconceived biases they may have for Black students. Teachers need to make active efforts to lift students and separate them from labels the staff may have incorrectly created. One participant explained:

I think it's just, try and get rid of the categor[ies] they have of us because you might look at the news and be like, a 13-year-old Black male got shot because he was selling drugs or something like that. So, a teacher or somebody looking at just every Black student and he's oh, he sells drugs, he's this and he's that.

3.4 Cost Evaluation

3.4.1 Economic Analysis Results

All values are reported in 2021 dollars and represent opportunity costs. Because staff attended trainings and meetings, and conducted RP circles, that time was not available to perform other necessary functions (e.g., teaching, grading). This represents an opportunity cost and shows the true societal cost of performing the intervention. These opportunity costs do not appear in budgets or expense reports. The budgetary costs to a school district would be significantly lower, as they exclude the opportunity costs of staff time and only include direct costs like paying trainers and purchasing materials. Opportunity costs are necessary for performing economic analysis and understanding the true costs of the intervention.

3.4.2 Cost Analysis

Table 3.13 presents the total costs of delivering the intervention by activity for each treatment condition. Total cost for the implementation of the program over 2 school years (2018–2019 and 2019–2020) is \$1,731,380, with nearly all of those costs (94%) incurred by the treatment condition. The most expensive element of the intervention was PBIS coaching, costing \$1,208,234, or 70% of total costs. This category was labor intensive, with each treatment school (and one school that switched conditions after the Year 1) receiving a dedicated PBIS coach. The next most expensive activity were trainings, with CRI training being the most expensive at \$178,883. The total combined cost of the trainings was \$395,181, or 23% of total costs. Trainings are expensive because of the time it requires for staff to attend when they otherwise could be performing other necessary job functions.

Most schools received the SET, which is why the control condition shows positive costs for the PBIS SET. Also, the one school that switched conditions in the 2018–2019 school year from treatment to control received PBIS coaching and CRI training, which is why the control condition shows positive costs for those two categories. We present total costs instead of annual because the costs of the intervention are cumulative over time.

Table 3.13. Total Activity Costs by Treatment Condition

Activity	Treatment (7 schools)	Control (10 schools)	Total
PBIS			
Training	\$85,046	\$0	\$85,046
Tier 1 team	\$66,216	\$0	\$66,216
SET	\$11,340	\$7,560	\$18,900
Coaching	\$1,127,685	\$80,549	\$1,208,234
CRI			
Training	\$166,682	\$12,201	\$178,883
RP			
Training	\$113,173	\$0	\$113,173
Circles	\$42,849	\$0	\$42,849
Other			
Training	\$18,080	\$0	\$18,080
TOTAL	\$1,631,070	\$100,310	\$1,731,380

Notes: Most schools received the PBIS SET, which is why the control schools show positive for the SET. One school switched from treatment to control in 2018–2019, which is why the control condition also shows positive costs for PBIS coaching and CRI training.

3.4.3 Net Benefit Analysis

Table 3.14 presents mean costs for each treatment condition, per school and per student. Means for the treatment schools are much higher than that of the control schools. The sum of the annual means is presented in the last row. The treatment schools incurred an average \$233,010 cost per school (\$61.62 per student) over both study years, while control schools incurred an average of \$11,129 (\$1.96 per student) over both study years. This difference is to be expected, as the intervention was delivered in treatment schools only, except for the minimal costs of the SET in 2018–2019, and for one school which received some trainings and coaching in 2018–2019 before switching to the control condition.

Table 3.14. Mean Costs per School and per Student

School Year	Mean Cost (SD)			
	Treatment (N = 7)		Control (N = 10)	
	School	Student	School	Student
2018–2019	\$110,489	\$29.49	\$10,978	\$1.92
	(\$19,326)	(\$7.92)	(\$29,186)	(\$4.98)
2019–2020	\$122,521	\$32.13	\$151	\$0.04
	(\$21,512)	(\$9.52)	(\$454)	(\$0.11)
TOTAL	\$233,010	\$61.62	\$11,129	\$1.96

Notes: Cost per student calculated by first estimating the cost per student per school, then taking the mean across schools. Total costs were estimated by summing mean costs across years.

Monetary benefit estimates were calculated based on the probability of dropping out, described at length in the Methods section and summarized here. Because this was a prospective study, we did not observe dropping out and therefore predicted them based on absence records. We used a mixed-effects negative binomial model to generate marginal predictions that provided average per-student absences by year and treatment condition, after controlling for model parameters. We then used estimates from Kirksey (2019) to transform these average per-student absences into probability of dropping out for each year and treatment condition.

Output from the mixed-effects negative binomial model (not presented) showed no statistically significant difference in changes in absences between treatment and control, either from baseline to 2018–2019 ($p = 0.703$) or from baseline to 2019–2020 ($p = 0.282$). In other words, the intervention had no statistically significant effect on reducing absences. By extension, when applying the estimates from Kirksey (2019), the intervention had no statistically significant effect on improving dropout. Therefore, the finding for the net benefit analysis is that over 2 years the intervention cost about \$59.66 per student to implement (\$61.62 for treatment minus \$1.96 for control, from Table 3.14), with no statistically significant monetary benefit in reducing dropout.

We continued to use the marginal predictions to perform the net benefit analysis for completeness, with the important caveat that the benefit estimates we present are not statistically significant. **Table 3.15** shows the monetized benefits per student by year for the treatment and control schools. Earnings is most affected by dropout, followed by health, property crime, and arrests. Total benefits improved for the treatment schools from 2017–2018 to 2018–2019 (from -\$2,715 to -\$2,557) but worsened by 2019–2020 (to -\$2,873) for a benefit drop of \$158 per student in 2019–2020 compared with baseline. The control group had a benefit improvement of \$110 per student between 2017–2018 and 2019–2020. Although this result is opposite to our hypothesis that the intervention would improve school attendance and thus generate long-term economic gains, as noted above, these estimates are not statistically significant.

Table 3.16 presents the net benefit analysis, using the rounded per student cost estimates of \$62 for treatment and \$2 for control from Table 3.14, and the per student benefit estimates of -\$158 for treatment and \$110 for control from Table 3.15. The difference of the two benefit estimates indicates that the treatment group results in a benefit loss of \$268 per student compared to control (though not statistically significant), and the difference of the two cost estimates indicates the treatment intervention costs \$60 per student. The net benefit is the difference of these differences, and results in a total loss of \$328 per student for the treatment group. We completed this full net benefit analysis for completeness, but because the benefit comparison between treatment and control was not statistically significant, we conclude the program costs \$60 per student, but with no measurable positive or negative monetary benefits.

Table 3.15. Monetized Benefits per Student

School Year	Earnings	Health	Prop. Crime	Arrests	Total	Change (Benefit) ^a
Treatment schools						
2017-18	-\$1,628	-\$1,064	-\$17	-\$6	-\$2,715	
2018-19	-\$1,533	-\$1,002	-\$16	-\$5	-\$2,557	
2019-20	-\$1,722	-\$1,126	-\$18	-\$6	-\$2,873	-\$158
Control schools						
2017-18	-\$1,548	-\$1,012	-\$16	-\$6	-\$2,582	
2018-19	-\$1,494	-\$977	-\$16	-\$5	-\$2,492	
2019-20	-\$1,482	-\$969	-\$16	-\$5	-\$2,472	\$110

^a Not statistically significant

Table 3.16. Per Student Benefit, Cost, and Net Benefit

	Benefit	Cost	Net Benefit
Treatment schools	-\$158	\$62	
Control schools	\$110	\$2	
<i>Difference</i>	<i>-\$268</i>	<i>\$60</i>	<i>-\$328</i>

4. Discussion

This ambitious study examined the implementation, outcomes, and costs and net benefits of three interventions in high schools. We discuss implementation, fidelity, and measurement of fidelity at length, reflecting the importance of understanding how the interventions were delivered in the treatment schools. This aspect of the study supports NIJ's interest in improving implementation science. To measure outcomes, we examined both actions as reflected in administrative data as well as staff and student experiences and perceptions as reflected in survey responses. We estimated intervention costs using rigorous economic approaches and sought to understand the net benefits that accrued as a result of those costs, focusing on absences and dropping out as the outcomes of interest to be able to draw from well-grounded studies of subsequent lifetime costs.

Unfortunately, the advent of the COVID-19 pandemic during Year 2 of implementation caused considerable disruption to the training of school staff and implementation of the intervention in treatment schools.

Before COVID-19, schools in the treatment condition varied in how much training and implementation they had, supported by both qualitative and quantitative data. Indeed, studies indicate that the interventions we engaged in can take longer than the 1.75 years we had before COVID-19 to be implemented with fidelity (Flannery et al., 2009; Beckman et al., 2012) and to be integrated together. We also embarked on an ambitious coaching support process that was designed to get schools engaged in interventions. The complexities of this intervention model exceeded those in other studies of interventions focused on reducing discipline disparities.

4.1 Implementation

4.1.1 *Study Challenges and Limitations*

Our study encountered a number of external challenges that limited the full implementation of the interventions and affected the results and their interpretation. The largest of these was the COVID-19 pandemic resulting in remote learning from March 2020 through February 2021. The effect of this disruption in school is partially described in sections 2 and 3. In addition to those effects, it should be noted that COVID-19 also substantially changed the climate of schools in our study, both when students were engaged in remote learning and when they returned to the school building during the 2020–2021 school year.

As is common in studies in school settings, there were changes in schools that could have impacted our results. First, in several of our treatment schools the principal and other leadership changed across the years, which appeared to affect implementation. In at least one case, the school principal did not assign administrators to the PBIS team, which made

it difficult to get buy-in and momentum with the intervention. Some respondents to the staff focus groups also noted that teacher turnover appeared be a challenge because new teachers were not readily trained in the intervention when they joined the school.

Tragically, a treatment school in CMS experienced a school shooting resulting in a student death during Year 1. During that same year, several other CMS high schools had incidents of students bringing guns to school. This may have shifted some of the attention of high schools during Year 1 to more immediately impactful school safety interventions, such as increased law enforcement presence and increased screening of students. On the other hand, these incidents had a clear implication of emphasizing the importance of school safety and managing students' peer relationships.

Another challenge across the study was schools changing treatment condition after going through our random assignment procedure. As described in the Methods section, we randomly assigned each school in matched pairs to treatment or control condition. Matched pairs were based on propensity scores including enrollment of Black male students, percentage of students using free and reduced-price lunches, and discipline levels throughout the school. In two instances, the school in the pair assigned to the treatment group chose not to participate in the intervention and became part of the control condition. These two schools that declined participation were affluent, majority White student schools. Although we used propensity scores to weight our outcomes, there may have been differences between the treatment and control schools for which we were not able to control. We did control for the percentage of the school enrollment that was Black male students in the propensity scores, but we did not explore the broader makeup of a school more specifically. It is reasonable that results may have differed for Black male students when they were in schools that were majority Black, majority White, or majority Latinx students.

Our project was spurred in part by a larger desire by CMS, like many other districts in the country, to address inequities in its schools. The cultural context of Charlotte, NC, and the nation is important to understanding how people responded to our three-pronged intervention. In Charlotte, as in many places around the country, a local officer-involved shooting of a Black man in 2016 caused a general sense of unrest, especially among communities of color. People protested in Charlotte following the shooting and adjudication of the officer involved in 2016. In summer 2020, protests continued as part of a larger national context of calls for racial justice. This context may have improved staff buy-in and engagement in an equity-focused intervention like ours. But also, the time was undoubtedly challenging for Black male students across schools, who expressed emotional turmoil during their student focus groups. This societal context also increased the probability that other CMS schools, including those in our control group, were engaging in informal efforts to improve equity in discipline in their schools.

Lastly, like many studies of interventions in schools, our study was influenced by other interventions occurring in our high schools. High schools in CMS implement programs related to academics, student behavior, and student and staff well-being at the behest of their principals. We are aware that some other programs that were not directly targeting discipline disparities could have influenced our results, though hopefully these were dispersed evenly across project groups.

4.1.2 Implementation Levels

This initiative combined PBIS, RP, and CRI. PBIS, which began first, was intended to provide a data-driven, schoolwide behavioral framework for each school to teach and support positive behavior. For RP, which aimed to strengthen relationships, schools most often targeted the ninth grade, but two schools targeted the whole school. CRI aimed to make academic instruction more meaningful to the lives of Black male students and focused exclusively on the ninth grade in the schools that received training. The following summarizes the delivery and timeline of training:

- **PBIS.** In Year 1, starting in summer 2018, the district launched PBIS. Four schools completed training in the summer of 2018, two began in the middle of fall 2018, and one in spring 2019. Leadership teams from all schools participated in the PBIS training.
- **RP.** In Year 1 and Year 2, the district added the RP component. Generally, the introduction to RP started in late fall of Year 1, with the RP circle training occurring in fall of Year 2. Two schools had received RP training in the year before this initiative. Two other schools received little to no RP training primarily due to scheduling challenges.
- **CRI.** During summer and fall of Year 2, the district added the CRI training component. One school did not receive any training in CRI. Among the schools that participated in CRI training, the level of staff exposure to training varied greatly by school. Because of the pandemic, only three schools trained a significant number of their ninth grade teachers during the study period. Among the three interventions, CRI had the least diffusion of training before the COVID-19 pandemic.

In summary, although the proportion of targeted staff trained varied by school, the project delivered core training to all seven treatment schools for PBIS, five schools for RP, and three schools for CRI. The training was not fully delivered as intended for RP and CRI.

In addition to training, the initiative included coaching, with a coach dedicated to each school in the first 2 years of the intervention and coaches shared between schools in Year 3. After some early delays getting coaches fully operational in some schools, the leadership team members from the treatment schools expressed high satisfaction with the coaching support.

4.1.3 Implementation Fidelity

Measurement and challenges

The measures of fidelity for each of the three interventions have strengths and limitations. To measure the fidelity of PBIS implementation, we used a combination of the SET (Sugai, et al. 2001; Todd et al. 2012) and TFI (Algozzine, 2014), both validated and standardized tools. Because of the onset of the COVID-19 pandemic, RTI staff only visited the schools in Year 1; thus, we have SET data only for spring 2019 in Year 1. For Year 2, we instead used the spring 2020 TFI data collected just before the pandemic.

Whereas the SET was administered independently by two RTI research staff blinded to treatment and control schools, the TFI was conducted by the school coaches in collaboration with the school leadership teams. The TFI, which is administered quarterly, relies on self-report of leadership teams and does not include the observations and interviews of staff and students conducted for the SET. Thus, the TFI may not be as accurate as the SET. In a comparison of the spring 2019 SET and TFI scores, the scores were about the same for three schools, but the TFI score was lower for one school and higher for three. Although we would have preferred to use the SET to measure fidelity of PBIS implementation throughout the study, as planned, using TFI data for Years 2 and 3 was the best possible solution to the disruptions caused by the pandemic. Nevertheless, this necessary change in measurement should be kept in mind in interpreting findings related to PBIS implementation and, indirectly, effects of PBIS on outcomes.

To estimate RP fidelity, we used staff survey data to estimate the prevalence of teachers' use of RP in the classroom (conditional on their school having had more than minimal exposure to RP training). An important limitation of the staff survey data is that they rely on self-report, which is subject to social desirability and reporting error. Although the survey question carefully defined the criteria for RP circles to help staff distinguish RP circles from other types of class discussions, it appeared that some teachers overreported their use of RP circles. Thus, the total percentage of teachers conducting RP circles in their classrooms is likely somewhat overestimated. However, given the training data, the relative prevalence across schools seemed plausible.

Finally, for CRI, we do not have reliable data about teachers' actual use of culturally responsive lesson plans focused on Black male students. Instead, the fidelity estimate used data on level of teacher exposure to the CRI training and having school-level structures in place to support CRI implementation. Of the implementation measures for the three interventions, the measure of CRI is the least robust. Given this and the low implementation of CRI, we have placed less emphasis on it in this report and did not use the measure to help account for differences in outcomes.

Results

Although the treatment schools made considerable progress in implementing the interventions, most schools did not consistently achieve high fidelity in implementing any of the interventions in the approximately 18-month period of implementation before the COVID-19 pandemic and resulting closures of in-person schooling. Almost all the treatment schools showed meaningful momentum in at least one of the interventions. Moreover, before the pandemic, two schools were able to achieve significant momentum in all three interventions, and one school achieved fidelity in PBIS across two quarters and had a majority of its ninth grade teachers reporting that they facilitated at least one RP circle with their students.

PBIS. Most treatment schools made gradual progress in their implementation fidelity as reflected by the mean SET and TFI scores. In Year 1 of implementation, the mean overall SET score increased modestly from fall to spring, from 45% to 57%. In Year 1, one school reached an overall SET fidelity score over 80%, but with a change of leadership in Year 2, this school did not sustain its gains. In Year 2, the mean overall TFI score of treatment schools reached 67% in October and 70% in February (range 53–97), with four of the seven treatment schools reaching the 70% fidelity threshold of the TFI in that last quarter before the COVID-19 pandemic. For implementation fidelity, the TFI requires an overall score of 70% over three consecutive quarters. Two treatment schools reached the TFI fidelity threshold for two consecutive quarters in 2020 before the pandemic. **Given the challenges of implementing PBIS in high schools, many of the district’s treatment schools achieved substantial progress.**

RP. According to the spring 2020 staff survey, in three of seven treatment schools, a majority of ninth grade teachers (range of 53%–70%) reported facilitating any RP circles in their classroom. Among all teachers who reported facilitating RP circles, about one-third did so approximately weekly, which means a minority of the teachers frequently facilitated circles. Therefore, during the study period, **the RP circle dose was relatively light.**

CRI. The CRI component was implemented last and had the least time for diffusion before the COVID-19 pandemic. Only one school achieved a high-fidelity score in spring 2020. Before the pandemic, that school had started rolling out its CRI implementation plan with its ninth grade teachers, begun developing a repository of CRI lesson plans, and conducted classroom observations. Two other schools implemented CRI with a moderate number of ninth grade teachers before the pandemic. **Student exposure to CRI before the pandemic was limited.**

4.1.4 Challenges and Facilitators

Challenges

By far, the greatest challenge to implementation was the interruption caused by the COVID-19 pandemic. The pandemic occurred in Year 2, when the schools had gained significant momentum but not yet achieved full implementation. With school building closures lasting a full year, an already tight period of implementation—3 years—became condensed to less than 2 years. Although some of the implementation tools (e.g., PBIS behavior matrix) were eventually adapted as feasible for virtual learning, school leadership and staff had their hands full dealing with the pandemic and teaching academic material virtually.

The complexity of the initiative was the next greatest challenge to implementing to fidelity within the study period. Individually, each of the interventions is challenging for schools to implement with fidelity. PBIS involves changing schoolwide systems, cultures, practices, and relationships with students. RP also involves changing organizational culture and practices, as well as relationships between staff and students and among students. CRI requires teacher self-reflection, openness to learning about student cultures and life experiences, and the skills to make lesson plans more culturally responsive. Teachers also need to develop skills and comfort to discuss more racially sensitive topics. Taken together, this ambitious initiative involved changing school organization and cultures, changing staff mindset, and building new skills in several areas. In focus groups, many teachers felt that they needed more skills, practice, and confidence. Before the launch of an initiative of this scope, the coaches also need several months to develop sufficient expertise in all three interventions, something that was not adequately planned for in the implementation timeline.

Any one of these interventions alone typically takes multiple years to fully implement. Although time varies by school and district context, it has been estimated that full implementation of PBIS can take up to 5–8 years in high schools (Flannery et al., 2009). Full implementation of RP in schools may take 3–6 years (Beckman et al., 2012). Asking staff to implement multiple interventions at the same time expands the challenges and load on the staff. In addition to this initiative, the school district also had other initiatives in some of the study schools that competed for staff training time and attention. Challenges in scheduling training meant that some schools were not able to complete all the training for the initiative, or that training—and thus implementation—was delayed.

The high school setting itself added to the complexity and the need for more time for implementation. Traditionally, high school teachers are expected to be subject area experts focused on academic outcomes, with less emphasis on the social-emotional and behavioral dimensions of student education. Compared with elementary schools and middle schools, high schools are also larger and more complex organizations, with more students and staff and separate, compartmentalized academic departments (Flannery et al., 2018). The larger

size also requires buy-in from more stakeholders (Flannery et al., 2014). Because of these structural and organizational cultural factors, PBIS typically takes longer to implement in high schools than in elementary and middle schools (Flannery et al., 2014; Flannery, et al., 2018).

Vincent et al. (2015) evaluated fidelity of implementation of PBIS in 35 middle schools in Oregon. Using a tool similar to the TFI (Prevention Practices Assessment, the Institute on Violence and Destructive Behavior, 2008) the researchers found that after 4 years of implementation, the average subscale score (percentage of possible points) increased from 20% in year 1 to 54% in year 4; none of the subscale scores reached the full fidelity score of 80%. In a randomized controlled trial testing PBIS in 31 Maryland high schools, the schools began with a relatively high average baseline SET score of 60% (meaning they already had some of the necessary infrastructure in place); this increased to 71% in 1 year and 83% in 2 years (Bradshaw et al., 2015). In light of these studies, the levels of implementation achieved before the COVID-19 pandemic by many of the district's treatment high schools seems reasonable.

Facilitators

In spite of these challenges, the quantitative fidelity data and qualitative interview data with school staff demonstrated considerable progress in a short time. Several factors facilitated the progress that was achieved, including the following:

- Intensive coaching, once the coaches were fully established, in the first three quarters of Year 2 (2019–2020)
- Commitment and engagement of the administrators, including the assignment of an assistant principal to the PBIS leadership team in most schools
- Engagement of teacher champions and broader staff participation in the initiative
- Promoting staff buy-in through open two-way communication that included data to demonstrate the needs in each school and the potential benefits of these interventions, and ongoing forums to hear and try to address teacher concerns
- Having structures of support at multiple levels—for leadership teams, professional learning communities, and teachers—to facilitate and monitor implementation. After the initial training, teachers needed follow-up professional development and coaching to continue to build skills and self-efficacy in these new practices.

In addition, with a multi-component initiative such as this, timing and sequencing must be carefully considered. Because PBIS was intended to serve as the foundation, ideally schools would have achieved fidelity or close to fidelity of implementation of PBIS before adding RP on a large scale. In this initiative, some schools tried to manage the load by focusing RP only on the ninth grade or in some ninth grade classrooms. In addition, CRI was only rolled out in ninth grade. Timing in the school calendar also matters. Having training before or early in the school year facilitated teachers' integration of new practices into their plans and

routines. Given the three interventions, competing district initiatives, and timeline for this study, this timing was often not possible.

Despite the aforementioned challenges of implementing three interventions, integrating the interventions (where feasible) could facilitate implementation and sustainability. In this initiative, the three components shared the goal of increasing equity and creating a positive school climate, positive relationships, and strategies to engage students. Conversely, staff did not always understand how the three interventions related to each other, and some school leaders wanted help in integrating them more. Although the district mandated MTSS in all the schools, the level of MTSS implementation in the high schools varied. In one school where MTSS was not fully operational, an administrator was initially confused about integrating PBIS with MTSS. Based on anecdotal evidence, PBIS likely supported the treatment schools' advancement of their overall MTSS implementation because PBIS addresses the behavioral side of MTSS and shares many of the same structural elements as the academic MTSS. District supports for MTSS could help facilitate continuation of PBIS in these schools.

Finally, the national social context increased awareness of the need for this initiative. Because the pandemic drew attention to the social-emotional needs of students, the district instituted a mandatory weekly SEL period. In addition, national and local events related to racial justice increased awareness of the need to increase equity in schools, including disparities in discipline. As one member of a school leadership team said,

I will say that there was some pushback [initially].... I think the other elements in society that have brought the reality of systemic racism to the forefront, that now people are less likely to sort of say, "I don't understand why this is important." ...We've...made larger scale structural shifts [toward equity], and [the] NIJ [grant] has been one of the elements to help move the dial.

In conclusion, although the COVID-19 pandemic severely impeded implementation, all the treatment schools made some progress in implementing at least some of the interventions. However, none of the schools were able to fully implement the three-intervention model in the time period before the pandemic. Therefore, our outcome evaluation did not test the effects of the conceptual model as designed but instead assessed early effects of partial implementation of PBIS in most schools and initial limited implementation of RP or CRI in some schools.

4.2 Outcomes

In this study, we used administrative data to examine an array of disproportionality metrics, extending important recent advances in studying disciplinary disproportionalities by Girvan et al. (2019) and Curran (2020). We also analyzed changes in disproportionality metrics and staff and student survey responses to assess differential change in treatment and control

group schools and, in treatment schools, whether changes were associated with implementation levels. Finally, focus groups with Black male students from three treatment schools allowed us to explore their perspectives related to the interventions and study outcomes.

4.2.1 Administrative Data

Disproportionality metrics

Our examination of various disproportionality metrics confirmed, as expected, that staff gave Black male students substantially more ODRs and OSSs than non-Black male students. Disproportionalities in ODRs and OSSs were seen in risk ratios, RDR (i.e., the estimated number of Black male students who experienced them but who would not have if Black male students were disciplined at the same rate as non-Black male students), and the discipline rate (per 100 students, scaled to the number of school days).

For point of reference beyond this study, the baseline year risk ratio for ODR in the study high schools was slightly higher (2.28 in treatment schools, 1.98 in control schools) than the 1.80 ODR risk ratio that Girvan et al. (2019) reported for a national sample of middle schools in the 2012–2013 school year. The OSS risk ratio was higher for treatment schools (3.33) and slightly lower for control schools (2.40) than the OSS risk ratio reported by Girvan et al. (2.51) and by Curran (2020) for schools in Maryland (2.50). Girvan et al. and Curran presented other disproportionality metrics, but they were not as directly comparable as the risk ratios; for example, Girvan et al. presented RDR, but not per 100 students.

In our study’s treatment schools, the RDR-100 for ODRs during the baseline year was 8.36, indicating that for every 100 Black male students, on average more than 8 of them received an ODR more than would be expected if Black male students and non-Black male students received ODRs at the same rate. For OSS, this metric was somewhat lower at 4.55 (as expected, because OSS is less frequent than ODRs) but still substantial and troubling. This metric helps one understand the scale of the problem that discipline disproportionality poses to actual individual Black male students and the need for initiatives to address the problem.

Although this study was focused on Black male students and disproportionalities relative to non-Black male students, we reviewed disproportionality metrics for female students as well. The disproportionality metrics showed that disparities between Black and non-Black male students and Black and non-Black female students were very similar. One difference between genders was that female students tended to have fewer ODRs and OSS than male students, but Black and non-Black *disproportionalities* in ODRs and OSSs were similar between genders. We focused analyses for this report on disparities between Black and non-Black male students.

Outcomes by treatment condition

To test for intervention effects on ODRs, OSS, and absences, we examined the three-way interaction between treatment group, time point, and race to test for changes in disparities over time as a function of the intervention. Using these statistical models allowed us to avoid some of the pitfalls and potentially misleading results associated with analyzing, for example, changes in risk ratios themselves, as described by Girvan et al. (2019) and Curran (2020). Our models examined the same underlying phenomena—disparities in ODRs, OSSs, and absences—but in a more mathematically rigorous manner. Girvan et al. and Curran provided insights and recommendations that benefited our study, which was planned before those publications. We appreciate the advances made in those articles and their role in improving this study.

Findings from our statistical analysis of year-to-year changes in ODRs, OSSs, and absences among Black male students and non-Black male students in treatment schools and control schools offered mixed or limited support for the hypothesis that the intervention would lead to more favorable changes in outcomes for Black male students in treatment schools (relative to non-Black male students in those schools) than for Black male students in control schools. This limited support for the intervention’s effects is not surprising, given the implementation challenges and limitations discussed above.

In treatment schools, ODRs dramatically increased from Year 0 to Year 1 for both Black and non-Black male students, though increases were less for non-Black male students. These increases were not seen in control schools where the ODR binary decreased from Year 0 to Year 1 for Black male students and was relatively flat for non-Black male students. The increases from Year 0 to Year 1 in the ODR binary seen in treatment schools did not persist during Year 2. It is tempting to ascribe this leveling off to a lagged effect of PBIS training in Year 1, but that interpretation begs the question of why the treatment schools experienced the initial sharp increases for Black and non-Black male students during the school year in which PBIS training took place.

The trendlines for the ODR count outcome resembled the ODR binary trendlines, but the three-way interaction was not quite statistically significant for either interval (Year 0 to Year 1 and Year 1 to Year 2). For the OSS binary outcome, the three-way interaction again fell short of significance, and for the OSS count outcome it was far from significant.

Results involving the absences outcome initially supported intervention effects more strongly. A statistically significant three-way interaction was driven by a substantial decrease from Year 0 to Year 1 for treatment group Black male students’ absences: a much steeper decrease than that for treatment group non-Black male students. That is, in treatment group schools, Year 0 absences for Black male students were substantially higher than for non-Black male students, but in Year 1 the difference in absences was greatly reduced. In control group schools, absences did not change substantially from Year 0 to

Year 1 for either Black male students or non-Black male students. These results suggest that PBIS training and implementation in Year 1 helped to reduce absences for Black male students in particular but also, to a lesser degree, for non-Black male students. However, when Year 1 was compared with Year 2, the three-way interaction term was not significant. Absences decreased for all four groups (treatment group Black male students and non-Black male students, and control group Black male students and non-Black male students), which is most likely due to the suspension of in-person school in March 2020 because of the COVID-19 pandemic. Importantly, the relative decrease in absences from Year 1 to Year 2 did not differ among the four groups, suggesting that the greater reduction in absences from Year 0 to Year 1 for treatment group Black and non-Black male students than for their control group counterparts did not persist in Year 2.

Outcomes by implementation level

The implementation challenges discussed above weakened any differences we would have expected to see between the treatment and control groups. Therefore, we conducted additional analyses that took advantage of the variation in levels of implementation among treatment schools. Simply put, did treatment schools with higher levels of implementation show better outcomes than other treatment schools with lower levels of implementation? If so, such findings would offer some support for the benefits of implementing the interventions, albeit without the level of rigor provided by the school-randomized treatment-and-control design.

We looked first at the association between PBIS implementation (as measured by SET/TFI scores, which we categorized into low, average, and high levels of implementation) and outcomes. Of 10 models tested for changes in each of five outcomes for two year-to-year intervals, only one model had implementation level significantly associated with the change in outcome. A second model's association was very close to statistically significant and also warrants discussion.

Regarding the statistically significant finding, in Year 1, Black male students in schools with low SET/TFI scores in Year 2 (i.e., lower implementation of PBIS) were more likely to receive an OSS than students in schools with average SET scores, and Black male students in schools with average SET scores were more likely to receive an OSS than those in schools with high SET/TFI scores. In Year 2, the differences in the OSS binary outcome for Black male students in schools with differing levels of SET scores remained, but somewhat lessened. Non-Black male students were similarly more likely to receive an OSS if they attended a school with low SET scores and less likely to receive an OSS if they attended a school with high SET scores.

The near-significant effect involved the change in ODR count from Year 0 to Year 1. ODR counts were routinely higher for Black male students than non-Black male students. Surprisingly, Black male students in schools with higher SET scores in Year 1 had higher

ODR counts than those in schools with average SET scores; the latter had higher ODR counts than Black male students in schools with lower SET scores. A similar pattern was seen for non-Black male students, though with smaller differences among schools with low, average, and high SET scores. This pattern of results suggests no clear interpretation, but the results do not suggest that higher levels of PBIS implementation led to reduced ODRs or less disproportionality.

For models using the percentage of teachers who reported leading RP circles as the implementation measure, there were no statistically significant interactions of implementation level, year, and race. That is, there was no evidence that implementation of RP circles, as reported by teachers, affected levels of discipline or absences by race.

4.2.2 Staff Survey Results

Findings by treatment condition

Our analyses of the staff survey data measured four primary constructs rated by staff: school climate, authoritative school climate, perceived safety, and culturally responsive self-efficacy. Our analysis of school climate showed significant interactions between baseline measures of school climate and measures at Wave 3 and, separately, Wave 4, such that treatment schools showed increases in positive school climates to a larger degree than control schools. This positive finding of treatment schools showing improvements at Wave 3 suggests that the interventions in those schools appear to improve the impressions of school staff about the climates of their schools. Results of comparisons of baseline to Wave 4 should be interpreted with caution due to significant changes in school climate wrought by the COVID-19 pandemic, rather than treatment condition.

For authoritative school climate representing both a supportive and structured school climate, the only significant interaction between wave and treatment condition was between Wave 1 and Wave 4. This suggests that treatment schools, more than control schools, increased authoritativeness (characterized by support and structure for students) in school climate. It could be that these results show that treatment schools, which had a foundation of PBIS and RP prior to the pandemic, were able to build upon that foundation to improve school climate once in-person learning returned in spring 2021.

Neither perceived safety nor culturally responsive self-efficacy measures on the staff survey showed significant interactions over time with respect to treatment condition. It could be that both of these measures were already toward the top of their variability ranges and that there was little room for improvement. Indeed, staff reported feeling very safe overall, as well as feeling that they had high self-efficacy in their ability engage in culturally competent teaching.

Findings by implementation level

Given the mixed results in the analyses of the staff survey between treatment and control groups, we conducted analyses to examine whether levels of implementation of PBIS and, separately, RP moderated the change in staff survey outcomes over time. We did not find any significant effects of the implementation levels of PBIS and RP. These nonsignificant findings mostly aligned with findings that did not show many significant effects of implementation levels on administrative data outcomes. This could be, in part, because the small numbers of schools in the treatment condition provided insufficient statistical power to detect small to moderate effects.

4.2.3 Student Survey Results

The most striking finding from analysis of the student survey data was that at Wave 3, administered in spring of the 2020–2021 school year, after in-person instruction resumed and following implementation of aspects of the intervention in the preceding 2 school years. Black male students in treatment group schools showed a large increase in their reports of their teachers' level of CRI. This finding suggests benefits of the targeted intervention.

However, a similar level of increase was seen in schools with high, average, and low levels of teachers reporting that they had conducted RP circles and, in separate analysis, of schools with high, average, and low SET/TFI ratings of PBIS implementation. We did not have reliable data on teachers' actual implementation of CRI, which would be the most relevant. Although the implementation of CRI was curtailed due to the pandemic, after schools returned to in-person learning, teachers who had been exposed to the CRI training may have started to use CRI more in the classroom. Many schools had developed implementation plans before the pandemic. In addition, treatment schools received training in implicit bias, which could also have helped to increase CRI. Finally, the initiative emphasized the need for CRI, which could also have affected teachers' behavior.

4.2.4 Qualitative Outcomes from Student Focus Groups

Student focus groups with Black male students from three treatment schools allowed us to explore context and detail about their experiences in the treatment schools. We asked students about their direct experiences of elements of PBIS, RP, and CRI in their schools. For example, almost all students described having a clear understanding of the rules and expectations of them, which is a central tenet of PBIS. They noted that expectations were posted, repeated by school staff, and that, most importantly, students generally knew what they should do. We cannot be certain that this is the result of PBIS, but these qualitative findings align with what we would expect in a school that was implementing PBIS to fidelity. Focus group participants in two out of the three schools reported that staff applied discipline generally equally to students regardless of race and gender. Participants from one school, however, reported that discipline was harsher for Black students; in contrast to the other

two schools, this school did not have improvements in PBIS fidelity scores during the study. Nonetheless, given the small number of students participating in each school focus group, findings may not be representative of Black male students from a specific school.

Many students reported that they felt emotionally safe and able to confide in at least one staff member in their school. One student noted that many of his teachers share his same race, which added to his comfort level.

For RP, students reported no current restorative circles, but were able to identify being engaged in restorative circles in at least one of their classrooms prior to the pandemic. They expressed that engaging in restorative circles had the intended effect of improving their relationships with their peers and allowing sharing perspectives between different types of peers. These indications strongly support the benefit of restorative circles for some students, which should not be overlooked by relying solely on quantitative evidence of effects on outcomes.

Lastly, the Black male students in the focus groups provided rich data about their daily perceptions of equity in their schools. Findings were mixed. In two schools, students reported that the teachers treated all the students the same; in one school, some students noted that some teachers and school staff made assumptions about them because of their skin color. Students were also asked whether they could relate to class lessons and the extent to which they reflected their culture and background. Overall, students reported being able to relate to some lessons, but not most of them. Importantly, students from one school reported that an African American history class was offered that they really enjoyed and related to. A student from this school noted that their U.S. history class did not really focus on African American history and the African American experience was underrepresented. However, a student from another school reported that he connected with their weekly current events writing assignment in their U.S. history class and that sometimes those assignments focused on issues related to African Americans. This provides an important example of how relevant material can be integrated into coursework.

4.3 Cost Evaluation

We conducted a cost and net benefit analysis of three interventions in a group of randomly assigned treatment high schools in comparison to control schools. Our cost analysis found the combined interventions of PBIS, RP, and CRI cost \$60 per student over 2 years of treatment, or about \$30 per student annually. Hollands et al. (2022) conducted a cost analysis of RP that was implemented in elementary and middle schools in Jefferson County, KY, and found the program had cost about \$145 per student in its first year (adjusted to 2021 dollars). Despite differences in setting, student age, and the current study being inhibited by implementation delays and the COVID-19 pandemic, the finding from their study is quite higher than ours. Hollands et al. also compiled costs of seven other education

interventions that were similarly estimated through rigorous means and found their estimate for RP to be relatively low cost in comparison. By extension, the intervention in the current study is especially low cost. We outlined previously in this discussion section how various factors led to implementation difficulties. These difficulties inevitably played a role in our low costs compared to other interventions, as lower implementation means lower levels of resource use and thus lower costs. This lower level of expected implementation at least partially explains why our costs were much lower than that of Hollands et al. and the other studies they outlined.

A low-cost intervention is worthwhile only if it improves the intended outcome(s), and absences (used to predict probability of students dropping out; the primary outcome for the net benefit analysis) did not significantly change. For completeness of the proposed research study, we continued with the net benefit analysis and monetized benefits as a function of absences, showing that the intervention generated an overall negative benefit for the treatment group; however, these findings were not significant. Therefore, the overall finding is that the intervention cost about \$30 per student annually with no perceivable positive or negative monetized economic benefit. Although Hollands et al. (2022) did not analyze absences, they too found either insignificant or negative effects among their five outcomes. But perhaps more strikingly, they also found that schools that implemented RP longer saw better improvements in suspensions among Black students. It may be the case that our study was too short—because of both implementation delays and the interruption from COVID-19—to show improvements in absences. Future research should address these limitations of the Discipline Disparities study.

4.4 Interpretation of the Study

Overall, our undertaking of an ambitious study of combining PBIS, RP, and CRI into one initiative encountered many challenges for implementation caused by the COVID-19 pandemic. We think this was the primary reason that we did not detect significant improvements in discipline disparities for Black male students in the treatment condition. Despite the limitation, we had several significant results that pointed to the benefit of our treatment. First, absences appeared to go down for Black male students in our treatment schools during Year 1 of treatment, though this effect was not observed in Year 2. It could be that PBIS or simply engaging in a study related to equity engaged Black male students more at school and, therefore, they were more likely to attend. Importantly, our metric included excused and unexcused absences. Another important finding was improvements in two measures of school climate (as rated on staff surveys) within treatment schools. This suggests that a schoolwide intervention, such as PBIS, may have impacts directly on the perceptions of teachers. Those perceptions may take more time than the study had to flow down to actions or student behaviors.

Year 1, when PBIS was the primary intervention being implemented, had one unexpected result during the school year (2018–2019). Treatment school staff increased giving ODRs to Black male students, but control school staff decreased giving Black male students ODRs. There are a few possible reasons. First, the introduction of an equity-focused intervention could have drawn attention to the specific group of Black male students. This was intentional, to reduce disparities, although this did not appear to happen in Year 1. It is possible that control school staff, knowing they were not chosen for equity intervention and knowing their disparities metrics would be monitored, felt that they needed to improve their disparities metrics independently. Alternatively, treatment schools that were beginning to implement PBIS could have been more stringent in their discipline practices Year 1, yet, having not engaged in RP or CRI, inadvertently exacerbated existing biases in behavioral interventions. On a positive note, this iatrogenic trend did not continue in Year 2. Still, it is important that researchers and practitioners implementing PBIS be aware of potential differential impacts on students, especially as the intervention is just getting started.

4.5 Implications

This study has important implementation for whole school interventions. We learned that they are quite challenging, especially in the high school setting. Moreover, we learned that integrating multiple interventions in high schools can take much longer than expected and requires more training and follow-up than were possible during this study. Overall, we experienced successes in using intervention coaching to provide implementation support. The result was an improvement in some metrics for staff and students across our study. More research on the interventions of PBIS, RP, and CRI at the high school level, both in isolation and in combination, is needed to understand how they affect students and disproportionate discipline.

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.
- Algozzine, B., Barrett, S., Eber, L., George, H., Horner, R., Lewis, T., Putnam, B., Swain-Bradway, J., McIntosh, K., & Sugai, G. (2014). School-wide PBIS Tiered Fidelity Inventory. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. www.pbis.org.
- American Psychological Association (APA) Zero Tolerance Task, F. (2008). Are zero tolerance policies effective in the schools?: an evidentiary review and recommendations. *The American Psychologist, 63*(9), 852-862. PMID: 19086747.
- Augustine, Catherine H., John Engberg, Geoffrey E. Grimm, Emma Lee, Elaine Lin Wang, Karen Christianson, and Andrea A. Joseph. (2018). *Can Restorative Practices Improve School Climate and Curb Suspensions? An Evaluation of the Impact of Restorative Practices in a Mid-Sized Urban School District*, RAND Corporation, RR-2840-DOJ, 2018. Retrieved April 27, 2023: https://www.rand.org/pubs/research_reports/RR2840.html
- Beckman, K., McMorris, B., & Gower, A. (2012). Restorative interventions toolkit. University of Minnesota, Healthy Youth Development—Prevention Research Center. <http://restorativeschoolstoolkit.org/sites/default/files/Minnesota%20Dept%20of%20Ed%20RP%20tool%20kit.pdf>
- Bjerk, D. (2012). Re-examining the impact of dropping out on criminal and labor outcomes in early adulthood. *Economics of Education Review, 31*(1), 110–122.
- Bradshaw, C. P., Pas, E. T., Debnam, K. J., & Lindstrom Johnson, S. (2015). A focus on implementation of positive behavioral interventions and supports (PBIS) in high schools: Associations with bullying and other indicators of school disorder. *School Psychology Review, 44*(4), 480-498.
- Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130*(5), e1136-e1145.
- Caldarella, P., Shatzer, R. H., Gray, K. M., Young, K. R., & Young, E. L. (2011). The effects of school-wide positive behavior support on middle school climate and student outcomes. *RMLE Online, 35*(4), 1-14.
- Cohen, M. A., & Piquero, A. R. (2009). New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology, 25*(1), 25–49.
- Cohen, M., Miller, T., & Rossman, S. (1994). *Understanding and Preventing Violence, Volume 4: Consequences and Control*. (Vol. 4) National Academies Press.

- Curran (2020). A Matter of Measurement: How Different Ways of Measuring Racial Gaps in School Discipline Can Yield Drastically Different Conclusions About Racial Disparities in Discipline. *Educational Researcher*. Vol. 49, No 5, pp. 382-387.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., Torrance, G. W., et al. (2015). *Methods for the economic evaluation of health care programmes*. Oxford University Press.
- 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. *New York: Council of State Governments Justice Center*.
- Flannery, K. B., Fenning, P., Kato, M. M., & McIntosh, K. (2014). Effects of School-Wide Positive Behavioral Interventions and Supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly*, 29(2), 111-124.
- Flannery, K. B., Hershfeldt, P., & Freeman, J. (2018). *Lessons learned on implementation of PBIS in high schools: Current trends and future directions*. Center for Positive Behavioral Interventions and Supports (funded by the Office of Special Education Programs, US Department of Education). Eugene, Oregon: University of Oregon Press.
- Flannery, K. B., Sugai, G., & Anderson, C. M. (2009). School-wide positive behavior support in high school: Early lessons learned. *Journal of Positive Behavior Interventions*, 11, 177-185.
- Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, 18(1), 41-51.
- Gay, G. 2000. *Culturally responsive teaching: Theory, research, & practice*, New York, NY: Teachers College Press.
- Girvan, E. J., McIntosh, K., & Smolkowski, K. (2019). Tail, tusk, and trunk: What different metrics reveal about racial disproportionality in school discipline. *Educational Psychologist*, 54(1), 40-59.
- González, T. (2014). Socializing schools: Addressing racial disparities in discipline through restorative justice. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion*. New York: Teachers College Press.
- Gregory, A., & Weinstein, R. S. (2008). The discipline gap and Blacks: Defiance or cooperation in the high school classroom. *Journal of School Psychology*, 46(4), 455-475.
- 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.

- Gregory, A., Gerewitz, J., Clawson, K., Davis, A., Green, C., Korth, J., Schotland, M., & Roderick, T. (2017). RP-Observe, Version 5, Unpublished measure. Rutgers University.
- Hamilton, A. B., & Finley, E.P. (2019). Qualitative methods in implementation research: an introduction. *Psychiatry Research*, 280;112516.
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., ... & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of biomedical informatics*, 95, 103208.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics*, 42(2), 377-381.
- Heifetz, R. A., Heifetz, R., Grashow, A., & Linsky, M. (2009). *The practice of adaptive leadership: Tools and tactics for changing your organization and the world*. Harvard Business Press.
- Hollands, F. M., Leach, S. M., Shand, R., Head, L., Wang, Y., Dossett, D., ... & Hensel, S. (2022). Restorative Practices: Using local evidence on costs and student outcomes to inform school district decisions about behavioral interventions. *Journal of School Psychology*, 92, 188-208.
- Hollands, F.M., Hanisch-Cerda, B., Levin, H. M., Belfield, C.R., Menon, A., Shand, R., Pan, Y., Bakir, I., & Cheng, H. (2015). *CostOut - the CBCSE cost tool kit*. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. Retrieved from: www.cbcsecosttoolkit.org
- Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-Wide Positive Behavior Support in elementary schools. *Journal of Positive Behavior Interventions*, 11(3), 133- 144. PMID: WOS:000267184300002.
- Horner, R., Sugai, G., & Lewis, T. (2007). Is school-wide positive behavior support an evidence based practice. Retrieved March 20, 2017, from <http://www.pbis.org/research/default.aspx>
- Huang, F. L., & Cornell, D. G. (2017). Student attitudes and behaviors as explanations for the Black-White suspension gap. *Children and Youth Services*, 73(C), 298-308.
- Institute on Violence and Destructive Behavior (2008). Prevention Practices Assessment. University of Oregon, Eugene: C. G. Vincent.
- International Institute for Restorative Practices. (2017). International Institute for Restorative Practices home page. Retrieved March 20, 2017, from <http://www.iirp.edu/>
- Kirksey, J. J. (2019). Academic harms of missing high school and the accuracy of current policy thresholds: Analysis of preregistered administrative data from a California school district. *American Educational Research Association (AERA Open)*, 5(3), <https://doi.org/10.1177/2332858419867692> .

- Ladson-Billings (1994): Ladson-Billings, G. (1994). What we can learn from multicultural education research. *Educational leadership*, 51(8), 22-26.
- Lodi, E., Perrella, L., Lepri, G. L., Scarpa, M. L., & Patrizi, P. (2021). Use of restorative justice and restorative practices at school: A systematic literature review. *International Journal of Environmental Research and Public Health*, 19(1), 96.
- McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: new crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence*, 108(1-2), 98-109. <https://doi.org/10.1016/j.drugalcdep.2009.12.002>
- Monahan, K. C., VanDerhei, S., Bechtold, J., & Cauffman, E. (2014). From the school yard to the squad car: School discipline, truancy, and arrest. *Journal of Youth and Adolescence*, 43(7), 1110-1122.
- Muennig, P. A. (2007). Consequences in health status and costs. Columbia University Libraries, Academic Commons. <https://doi.org/10.7916/D87W6CJ5>
- Nowicki, J. M. (2018). K-12 Education: Discipline Disparities for Black Students, Boys, and Students with Disabilities. Report to Congressional Requesters. GAO-18-258. *US Government Accountability Office*.
- Robison, S., Jagers, J., Rhodes, J., Blackmon, B. J., & Church, W. (2017, February). Correlates of educational success: predictors of school dropout and graduation for urban students in the Deep South. *Children and Youth Services Review*, 73, 37-46.
- Simson, D. (2012, May 12). *Restorative justice and its effects on (racially disparate) punitive school discipline*. Paper presented at the 7th Annual Conference on Empirical Legal Studies. Available at <https://ssrn.com/abstract=2107240>.
- Skiba, R. J. (2013). Reaching a critical juncture for our kids: The need to reassess school-justice practices. *Family Court Review*, 51(3), 380-387.
- Skiba, R. J., Arredondo, M. I., & Rausch, M. K. (2014). *New and developing research on disparities in discipline*. The Equity Project at Indiana University.
- Skiba, R. J., Arredondo, M. I., Gray, C., & Rausch, M. K. (2018). Discipline disparities: New and emerging research in the United States. *The Palgrave international handbook of school discipline, surveillance, and social control*, 235-252.
- Staats, C. (2014). Implicit racial bias and school discipline disparities. Columbus, OH: Ohio State University, Kirwan Institute Special Report.
- Sugai, G., Lewis-Palmer, T. L., Todd, A. W., & Horner, R. H. (2001). School-wide evaluation tool (SET). Eugene, OR: Center for Positive Behavioral Supports, University of Oregon.
- Todd, A. W., Lewis-Palmer, T., Horner, R.H., Sugai, G., Sampson, N. K., & Phillips, D. (2012). Schoolwide Evaluation Tool (SET) Implementation Manual, Version 2.0: Positive Behavioral Interventions & Supports, University of Oregon.

- U.S. Commission on Civil Rights (July 23, 2019). *Beyond suspensions: Examining School Discipline Policies and Connections to the School-to-Prison Pipeline for Students of Color with Disabilities*. <https://www.usccr.gov/files/pubs/2019/07-23-Beyond-Suspensions.pdf>
- Vavrus, F., & Cole, K. (2002). "I didn't do nothin'": The discursive construction of school suspension. *The Urban Review*, 34(2), 87-111.
- Vincent, C. G., & Tobin, T. J. (2011). The relationship between implementation of School-Wide Positive Behavior Support (SWPBS) and disciplinary exclusion of students from various ethnic backgrounds with and without disabilities. *Journal of Emotional and Behavioral Disorders*, 19(4), 217-232.
- Vincent, C. G., Randall, C., Cartledge, G., Tobin, T. J., & Swain-Bradway, J. (2011). Toward a conceptual integration of cultural responsiveness and schoolwide positive behavior support. *Journal of Positive Behavior Interventions*, 13(4), 219-229.
- Vincent, C. G., Swain-Bradway, J., Tobin, T. J., & May, S. (2011). Disciplinary referrals for culturally and linguistically diverse students with and without disabilities: Patterns resulting from School-Wide Positive Behavior Support. *Exceptionality*, 19(3), 175-190. PMID: WOS:000299232800004
- Vincent, C. G., Sprague, J. R., Pavel, M., Tobin, T. J., & Gau, J. M. (2015). Effectiveness of schoolwide positive behavior interventions and supports in reducing racially inequitable disciplinary exclusion. In D. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 207-221). New York, NY: Teachers College Press.