

The Promise of Restorative Practices to Transform Teacher-Student Relationships and Achieve Equity in School Discipline

ANNE GREGORY, KATHLEEN CLAWSON, ALYCIA DAVIS, and
JENNIFER GEREWITZ

Rutgers, The State University of New Jersey

Restorative approaches to school discipline are increasingly being implemented throughout the United States in an attempt to reduce reliance on suspension and eradicate the racial discipline gap. Yet, little is known about the experience of students in classrooms utilizing restorative practices (RP). This study draws on student surveys (N = 412) in 29 high school classrooms. Hierarchical linear modeling and regression analyses show that high RP-implementing teachers had more positive relationships with their diverse students. Students perceived them as more respectful and they issued fewer exclusionary discipline referrals compared with low RP implementers. In addition, the findings demonstrate some initial promise of well-implemented RP for narrowing the racial discipline gap. The study found that higher RP implementers issued fewer discipline referrals to Latino and African American students compared with lower RP implementers. The study findings have implications for equity-focused consultation in schools that honor student experience of new programming.

Suspensions remain a widely utilized approach to school discipline despite a lack of evidence that they prevent future misbehavior or make schools safer (American Psychological Association Task Force, 2008). The American Academy of Pediatrics (2013) recently issued a statement describing the effectiveness of exclusionary discipline as “increasingly questionable.” Their statement reflects a growing body of evidence demonstrating the harmful effects of suspension (e.g., Hemphill, Toumbourou, Herrenkohl, McMorris, & Catalano, 2006). For instance, after accounting for demographics, attendance,

Correspondence should be sent to Anne Gregory, Rutgers Graduate School of Applied and Professional Psychology, 152 Frelinghuysen Road, Piscataway, NJ 08854. E-mail: anne.greg@rutgers.edu

and course performance, each additional suspension further decreases a student's odds of graduating high school by 20% (Balfanz, Byrnes, & Fox, 2014). Outcry over the negative correlates of suspension also reflects concern about the racial discipline gap (e.g., Losen & Gillespie, 2012). African Americans and, in many regions, Latino and American Indians are disproportionately overrepresented in school discipline (Gregory, Skiba, & Noguera, 2010; Wallace, Goodkind, Wallace, & Bachman, 2008). A recent longitudinal study followed students in the Texas public school system (Fabelo et al., 2011). African American students (26.2%) were more likely to receive out-of-school suspension in response to a first infraction compared with Latinos (18%) and Whites (9.9%). This disparity held when accounting for other risk factors. For instance, African American ninth graders were 31% more likely to receive a discretionary discipline referral compared with White students when student characteristics were taken into account (e.g., socioeconomic status, academic test scores, and number of days absent). Discipline encounters were also not uncommon for Latino students. Over the 6 years they were followed, almost 65% of Latino students encountered some type of disciplinary action.

This suggests that high schools need to rethink their approach to preventing conflict, handling rule infractions, and re-engaging students after an infraction has occurred. As a result, policymakers are seeking alternatives to current discipline practices that (a) reduce the reliance on school exclusion and (b) reduce the overrepresentation of ethnic minorities in the discipline system. Yet, teachers and policymakers at the high school level have few empirically based, developmentally appropriate school discipline interventions at their disposal.

Review of the classroom management literature shows that promising interventions are not likely to be stand-alone curricula but, rather, are integrated into daily instructional practices (Osher, Bear, Sprague, & Doyle, 2010). One such promising intervention, restorative practices (RP), needs further systematic examination to understand its full potential at the high school level. Trainers from the International Institute for Restorative Practices (IIRP) implement RP as a 2-year whole-school change program (SaferSanerSchools). With a prevention and intervention focus, RP aims to transform how students and adults interact with one another thereby creating a more positive school climate.

The 3-year RP program has been implemented in a diverse range of U.S. schools. School record data in RP high schools have shown a promising drop in the use of punitive school discipline (Lewis, 2009). For example, in an urban largely African American high school, violent acts and serious incidents were reduced by 52% compared with the year before. In a rural high school, there was a 50% reduction in suspensions. Finally, in a large suburban high school, the number of incidents of "disrespect to teacher" and "classroom disruption" reduced by 70% after 1 year of the intervention. Other models that primarily focus on restoring relationships after a negative incident has

occurred (i.e., restorative justice [RJ]) have been implemented widely in states such as Minnesota, California, Colorado, and Florida and internationally in countries such as Australia, Scotland, Wales, England, Canada, and Hong Kong (González, 2011; McCluskey et al., 2008a; for a summary see Schiff, 2013). Many schools using RJ report reduced use of out-of-school suspension (Karp & Breslin, 2001; Schiff, 2013; Stinchcomb, Bazemore, & Riestenberg, 2006). For example, González (2014) recently showed that in a district implementing restorative approaches the percentage of students issued one or more suspensions dropped by 7% for African Americans, 5% for Latinos, and close to 4% for White students. Recently, Simson (2012) conducted one of the few comparison studies of RJ and non-RJ schools. In schools across two states, he found that RJ schools had a slightly greater decrease in suspension rates and a slightly smaller African American-White gap in suspension rates compared with non-RJ schools (significant at the $p < .10$ level). It is important to note that the study accounted for school size, poverty, and grade level (e.g., elementary, high school), which increases the confidence that sociodemographic differences in the schools did not account for the change.

Despite the existing case studies and emerging comparison studies, the research on the RP whole-school change approach is in its nascent stages. Specifically, as of yet, no studies have examined the link between RP and diverse students' relationships with their teachers. This is important given positive teacher-student relationships among all racial groups are key to creating a supportive and equitable school climate that does not rely on punitive approaches to behavior (Gregory, Cornell, & Fan, 2011). Moreover, few studies have considered the link between quality of RP implementation in classrooms and student outcomes (McCluskey et al., 2008a).

The current study addresses this need for additional research with its examination of teacher and student reports of RP implementation in two high schools. Namely, we identify whether higher RP implementation in high school classrooms is associated with positive teacher relationships for students of all racial and ethnic groups as seen through (a) student experience of their teachers as respectful and (b) infrequent use of teacher-issued referrals for misconduct/defiance across racial and ethnic groups (a discipline referral category accounting for large racial discipline gaps; Gregory & Weinstein, 2008). The focus on implementation of RP (a new innovation in the schools) also provides broader implications for the value of school consultants focusing on the initial uptake of RP to improve students' experience in the classroom and school.

DEFINING RESTORATIVE PRACTICES

In restorative justice (RJ), those affected by an infraction or crime come together to identify how people were affected by the incident (Coates, Um-

breit, & Vos, 2003; Gal & Moyal, 2011; McGarrell & Hipple, 2007; Rodriguez, 2007). Together, they decide how to repair the harm *after* an infraction has occurred. RJ embodies a philosophical approach to wrongdoing. Namely, it arises from a humanist tradition in which the victim and the disputant's subjective experiences of the wrongdoing are highlighted along with a belief in the need for collaborative problem solving. A focus on mending relationships is central. Zehr and Toews (2004) contend that, fundamentally, RJ's core underlying value is respect. Zehr (2002) has traced RJ's roots back to a range of diverse cultures (e.g., American Indian, Maori) and religious traditions (e.g., Judaism).

For many years, RJ has been operationalized in school settings outside of the United States through an expanded set of practices that include methods for *preventing* infractions in the first place (e.g., Blood & Thorsborne, 2005; McCluskey et al., 2008a). In the U.S. context, Amstutz and Mullet (2005) describe restorative school environments as prevention oriented when they emphasize an ethos of care and social and emotional learning. Thus, the notion of "restorative" encompasses more than a set of procedures that occur after a rule infraction. Wachtel, Costello, and Wachtel (2009) at the IIRP, and other interventionists implementing a range of RJ programming in schools, have helped to spread this encompassing approach in the U.S. school setting. Grounded in the same philosophical traditions of earlier RJ methods (Zehr, 2002), RP—as it is called by IIRP—includes strategies to both *prevent* rule infractions before they occur and to *intervene* after an infraction has occurred. RP trainers teach school staff what they call "11 essential elements" (See Program Overview at <http://www.safersanerschools.org/>). Many of the RP elements can serve prevention or intervention functions, yet several are specifically focused on reducing the likelihood of student rule breaking (e.g., proactive circles) and others on intervening after rule breaking has occurred (e.g., restorative conferences; see Table 1).

Many RP elements provide community-building opportunities. For instance, one of the essential RP elements is the "Proactive Circle" in which teachers use structured group discussion and meaningful exchanges while sitting in a circle (Costello, Wachtel, & Wachtel, 2010). Facing one another, they have frank and open discussions about academic topics (e.g., their academic goals for the day or the semester), emotional topics (e.g., their experiences being the target of teasing), and classroom-specific topics (e.g., what norms of respect they would like to establish in the classroom). The types of topics and specific content are limitless, yet the goal is similar: provide an opportunity for students and teachers to learn about one another (and thus respond more appropriately to one another).

RP attempts to strengthen social connection and responsibility for one another by increasing opportunities for affective communication—one technique used is called "Affective Statements" in which both teachers and students express their emotional reactions to both positive and negative

TABLE 1 Elements of Restorative Practices

Domain	Elements	Description
Prevention (building relationships and developing community)	1. Affective Statements	Use in response to negative or positive events in the classroom and school
	2. Proactive Circles	Run on daily or weekly basis (e.g., students sit in a circle and discuss a topic that helps build community)
	3. Fair Process	Engage students in decisions, explain the rationale
	4–5. Restorative Staff Community/ Restorative Approach with Families	Model and use restorative practices among school staff and with student families
	6. Fundamental Hypothesis Understandings	Provides a framework to guide daily interactions with the appropriate mix of control and support
	Intervention (repairing harm and restoring community)	7. Restorative Questions
8. Responsive Circles		After a moderately serious incident, students sit in a circle and address who has been harmed and what needs to be done to make things right
9. Small Impromptu Circles		Address negative behaviors by asking the wrongdoer and those harmed to answer restorative questions in front of each other.
10. Restorative Conference Circles		Respond to a serious incident using a scripted approach to facilitate accountability and repair harm
11. Reintegrative Management of Shame		Acknowledge the emotions of the wrongdoers and those impacted by the wrongdoing

events (Wachtel, Costello, & Wachtel, 2009). A teacher may also use “Affective Questions” to encourage students to reflect on how their behavior has affected others (Mirsky, 2011). Theorists argue that relationships are made more supportive when people engage in free and appropriate expression of emotion (Tomkins, 1991; Wachtel, 2012).

As students and teachers learn about one another in proactive circles, they can also develop a sense of shared authority/ownership over the classroom climate to increase accountability. Teachers may use the circles to have students jointly develop behavioral expectations for behavior, classroom rules, and consequences for breaking those rules (Costello et al., 2010). In response to a breach of trust, teachers implement “Responsive Circles” in which the classroom as a whole discusses an incident with the hopes of

restoring community. Responsive circles engage students in the management of conflict that is affecting many students or adults in the classroom. Students discuss feelings, identify who has been affected, and develop a plan to repair the harm and prevent future conflict. All people involved in the wrongdoing are expected to participate. This process aims to hold students accountable for breaching trust in the community.

More serious incidents are addressed in “Restorative Conferences” with the goal of developing joint solutions to repair the harm (Braithwaite, 2001). Administrators or teachers use a structured and scripted meeting protocol (Wachtel, O’Connell, & Wachtel, 2010). Accountability for wrongdoing is central to the conferences. Yet, it is important to note that there is also a focus on reintegrating the wrongdoer into the community rather than stigmatizing him or her (Braithwaite, 1989). Students are able to bring a supportive person with them to the conference, which is part of the process of restoring their sense of community. Also noteworthy is the focus on the emotions of all involved, including a process to help the wrongdoer resolve shame by making amends (Nathanson, 1997; Wachtel et al., 2010).

Throughout all RP elements, student opinion and emotional reaction are mindfully integrated into all procedures. RP emphasizes fair process and its three guiding principles: engagement (involve students in decision making), explanation (provide rationale for decisions), and expectation clarity (widespread understanding of behavioral expectations and consequences for infractions; Wachtel, 2012). A restorative classroom should evidence this participatory form of decision making whenever possible.

RESTORATIVE PRACTICES AND POSITIVE TEACHER-STUDENT RELATIONSHIPS

RP elements, as a whole, may be effective at eliciting teacher-student cooperation, fostering constructive conflict resolution, and working toward equitable disciplinary practices given three broad underlying processes. Specifically, the elements may (a) promote interpersonal support and connection, (b) uphold structure and fair process, and (c) integrate student voice. This is in keeping with an authoritative style to socializing adolescents (Baumrind, 1968, 1991). Theory and research on adolescents suggests that adolescents may be most responsive to authority when schools have an authoritative disciplinary climate (Gregory, Cornell et al., 2010; Gregory et al., 2011). In such a climate, student voice is honored and adults express care yet remain firm in shared expectations for behavior. This is particularly important for adolescents as they seek greater control in decision making (Smetana & Gaines, 1999) and expect fair and legitimate adult authority (Turiel, 2005). An authoritative approach to African American, Latino, and American Indian students may nurture trusting and positive teacher-student interactions. In

other words, support, structure, and student voice may be key ingredients that have the potential to “humanize” teacher interactions with historically stigmatized groups. With a focus on becoming sensitive to the individual needs of students and fostering genuine interest in students, individualizing student support may disrupt negative stereotyping or implicit bias about stigmatized groups of students of color (Devine, Forscher, Austin, & Cox, 2012; Simson, 2012). Increasing structure and accountability in the context of respect for student input and fair implementation of rules may legitimize teacher and administrator authority. Past research has shown that African American students who feel fairly treated by their teachers tend to be perceived as less defiant and more cooperative by their teachers (Gregory & Thompson, 2010).

SUMMARY

Taken together, the “11 Essential Elements” of RP aim to increase support, structure, and student voice in the classroom. Theory on authoritative approaches to adolescents suggests that teachers who implement RP well will in fact have more positive relationships with their students and ultimately will less frequently draw upon punitive approaches to school discipline (e.g., Gregory et al., 2011). This assertion, however, has yet to be tested. Given the nascent stage of empirical research on restorative approaches to school discipline, this is the first study to examine whether teachers with higher (compared to lower) implementation of RP, as reported by teachers and students, tend to have more positive relationships with their students. Positive relationships were measured from two sources and in two ways: (a) student perceptions of teachers as respectful and (b) teachers’ low use of exclusionary discipline for perceived misconduct and defiance. Given the racial discipline gap that has been well documented for African American students for decades (e.g., Fabelo et al., 2011) and is of increasing concern for Latino and American Indian students (Krezmien, Leone, & Achilles, 2006; U.S. Department of Education, National Center for Education Statistics, 2003), the study focuses on whether the link between RP implementation and positive teacher relationships are experienced similarly by students of varying racial and ethnic groups. Given its examination of RP implementation, the study also offers implications for best practices in teacher consultation during the adoption of new programming. Namely, the study sheds light on the utility of collecting teacher and student reports of program implementation. Two central research questions guide the study:

Research Question 1: Is greater implementation of RP, as perceived by students and teachers, associated with higher student-reported teacher respect? Does this association hold across student racial/ethnic groups?

Research Question 2: Is greater implementation of RP, as perceived by students and teachers, associated with teachers issuing fewer misconduct/defiance discipline referrals to Latino/African American and Asian/White students?

It was anticipated that teachers with higher implementation of RP would have more positive relationships with their students from all racial and ethnic groups. Specifically higher RP teachers would be perceived as more respectful than teachers with lower implementation of RP. RP's association with teacher respect, we believed, would be similar for Asian, White, Latino, African American, and American Indian students. It is important to note that, according to the school records, no American Indian students were issued misconduct/defiance discipline referrals by the teachers in our study. Thus, we did not include them in the analyses for the second research question. Accordingly, we hypothesized that high RP teachers would issue fewer exclusionary discipline referrals for perceived misbehavior—their reduced rates of referral would be similar across Asian, White, Latino, and African American students.

METHOD

Participants

High schools. Two large and diverse high schools in a small city on the East Coast of the United States participated in the research during their first year of implementing RP (2011–2012). Based on school records, enrollment across both high schools at the time of the research consisted of 9 American Indian students (<1%), 149 Asian students (3%), 2,444 White students (54%), 1,428 Latino students (31%), and 522 African American students (11%).

The year before the RP program was brought into the schools (2010–2011), referrals related to misconduct/defiance comprised almost 30.3% of all discipline incidents. This was the second most common reason students received a discipline referral, following reasons related to missing class time (e.g., truancy, tardiness). In the 2010–2011 school year, greater percentages of Latino and African American students were issued misconduct/defiance referrals than Asian and White students. Specifically, close to a third of Latino and African American students (34% and 38%, respectively) compared with 5% and 11% of Asian and White students (respectively) were issued referrals for misconduct/defiance.

Implementation of RP. RP trainers led two full-day workshops with teachers, administrators, and staff at the start of the school year. IIRP trainers also provided two full days of consultation in each school. The days included

the following: (a) Observation: RP trainers spent a majority of each consultation day observing teachers in their classrooms and providing feedback on RP implementation. Teachers also had opportunities to seek out consultants during “drop-in” hours. (b) Modeling: RP trainers also modeled practices by engaging with youth, conducting classroom circles, and participating in meetings with youth and families. (c) Targeted Planning: RP trainers asked school leadership to identify key areas in need of immediate or intensive focus. Topics included restorative leadership skills, developing a restorative staff community, and supporting ongoing growth and learning.

Survey respondents. Thirty-one teachers agreed to participate in taking the surveys during the 2011–2012 school year, the first year in which RP was implemented in both high schools. We used a random number generator to select a single classroom from each teacher’s daily course schedule (e.g., class Periods 1–5). From each of the teachers’ course schedules, one classroom was randomly selected (herein called the focal classroom). Two teachers returned surveys that were significantly incomplete, thus the final teacher sample was slightly reduced ($N = 29$). The teachers had a wide range of experience (Min: 3 years; Max: 32 years) with an average of 13 years ($SD = 9$). Almost three quarters of the teachers were women. With the exception of one self-identified Puerto Rican teacher, all teachers identified as White.

Within the 29 classrooms, 412 students had consent to participate and completed surveys. Students without parent/guardian consent did not fill out the surveys. On average, 60% of students in each class participated. Participation across the two schools was uneven given the sample was comprised of 55 students from one school (3% of the total enrollment) and 357 students from the other school (13% of the total enrollment).

The student sample was comprised of slightly more male (53%) than female students (47%). Thirty-eight percent of the students reported that one or both of their caregivers (e.g., parents or guardians) had a high school diploma or less, whereas 62% of students reported that one or both of their caregivers had completed some higher education (community college or beyond). The diverse level of caregivers’ education suggests the students came from families with a range of economic resources.

The sample was racially and ethnically diverse with students self-reporting the following: 44% White, 21% Latino, 3% American Indian, 2% Asian, 5% African American, and 25% Mixed Race. Of the 106 Mixed Race students, 45% reported they were partially of African American descent and 73% reported they were partially of Latino descent. The survey sample was significantly different from the racial composition of the enrolled students ($\chi^2(5) = 35.7$, $p < .001$) with proportionally fewer White, Latino, and African American students and more students identifying as Mixed Race. The differences may be a measurement artifact—unlike on our surveys, it appears that parents were unable to select “Mixed Race” on the school records.

Research Procedures

Members of the research team introduced the study aims during the initial RP school staff training and subsequent faculty meetings. Once a teacher consented to participate, members of the team randomly selected a classroom from each teacher's weekly schedule. During late fall and early winter of the first year of RP implementation, the team members then presented the study to students in the randomly selected classrooms and provided student assent and parent/caregiver consent forms. Consented students and teachers completed 30- to 40-min surveys once during the school year. Teachers turned in their completed surveys the same day members of the research team collected the student surveys. Thus, teachers did not have access to the confidential student surveys. Classrooms and teachers received a small monetary gift as a thank-you for their participation.

Measures

Student self-reported race/ethnicity. Given the small sample size of teachers and the parsimony required in our data analysis, we made a number of theoretically grounded decisions to reduce the number of racial and ethnic categories. Given the research that Latino, African American, and American Indian students can be overrepresented in discipline referrals depending on the geographic region (Gregory, Skiba, & Noguera, 2010), we decided to compare the experience of these three groups with the experience of Asian and White students—two groups typically underrepresented in discipline referrals, as was confirmed in the enrollment and discipline referral data in the participating schools. Thus, when using the student self-reported race/ethnicity for the first research question, we coded Latino, African American, and American Indian as 1 and Asian and White as 0. It is noteworthy that 25% of the student sample self-reported they were of mixed descent. Some of the disparities in discipline referrals may relate to how teachers “read” their students based on phenotype (Simson, 2012). Thus, we included multiracial students in the Latino, African American, and American Indian group if they reported they were members of any of these three groups. We recognize this oversimplifies the complexity of racial and ethnic experience given the range of racial/ethnic phenotypic expression (Monroe, 2013). Yet, we believe important insights can be gained provided we recognize the limits of this approach. In sum, for Research Question 1, the sample of Latino, African American, and American Indian students includes students who were of mixed descent. In total they comprised 54% of the sample. The Asian/White category comprised 46% of the sample.

For the second research question using school records of discipline referrals, we coded Latino and African American as 1 and Asian and White as 0. No American Indians were indicated in the school discipline records and

thus, analyses using the records did not shed light on this group. In the 2010–2011 school year, compared with White students, the relative risk ratios for receiving a misconduct/defiance referral were similar for Latino and African American students—Latinos were 3.07 times and African Americans were 3.43 times as likely to receive a referral for misconduct/defiance compared with White students. This similar degree of disproportionally Latino and African American students in discipline referrals provides further rationale for combining them into a single group in the analyses.

Implementation of restorative practices. IIRP developed the teacher and student RP implementation survey scales as part of their training materials. They were designed for teachers to use as self-assessment on their quality of implementation. From a face validity perspective, they aimed to link the behaviors described in the scale items with the behaviors and proficiency standards identified in the 11 essential elements. (See Whole School Change program overview: <http://www.iirp.edu/pdf/WSC-Overview.pdf>). Given some underutilization of survey results over the past 3 years of use, IIRP consultants are now developing ways to better automate the data collection so that timely feedback reports can be provided to staff (J. Bailie, personal communication, July 30, 2013). As of yet, no research has tested their reliability and concurrent/predictive validity.

For this study, students answered all items on a 5-point scale, rating the degree to which the teacher engaged in the particular RP element (i.e., *not at all*, *rarely*, *sometimes*, *often*, and *always*). The Affective Statements Scale (3 items, $\alpha = .59$) included, “My teacher is respectful when talking about feelings.” The Restorative Questions Scale (4 items, $\alpha = .81$) included, “When someone misbehaves, my teacher responds to negative behaviors by asking students questions about what happened, who has been harmed and how the harm can be repaired.” The Proactive Circles Scale (4 items, $\alpha = .75$) included, “My teacher uses circles to provide opportunities for students to share feelings, ideas and experiences.” The Fair Process Scale (4 items, $\alpha = .73$) included, “Asks students for their thoughts and ideas when decisions need to be made that affect the class.” The Responsive Circles Scale (6 items, $\alpha = .72$) included, “My teacher uses circles to respond to behavior problems and repair harm caused by misbehavior.” The Management of Shame Scale (4 items, $\alpha = .71$) included, “My teacher acknowledges the feelings of students when they have misbehaved.” We calculated Cronbach’s alpha for each scale to assess internal consistency of the items. The alphas ranged from fair (.59) to good (.81).

We found no statistical differences across racial and ethnic groups on RP implementation scales. This suggests that Latino/African American and White/Asian students experienced similar levels of RP implementation. Given the similarities across racial/ethnic group ratings, we decided it was appropriate to average student scales for each teacher to obtain an overall classroom average.

Teachers completed RP implementation scales that were parallel to the student RP implementation scales. Teachers rated the degree to which they implemented a range of RP elements including the Affective Statements Scale, (8 items, alpha = .80), Restorative Questions Scale (7 items, alpha = .90), Proactive Circles Scale (8 items, alpha = .59), Fair Process Scale (6 items, alpha = .93), Responsive Circles Scale (10 items, alpha = .76), and the Management of Shame Scale (7 items, alpha = .93). The alphas suggest adequate internal consistency.

For data reduction purposes given the small number of teachers in our sample, we conducted principal components analysis with the student-completed RP scales and the teacher-completed RP scales. In the factor analyses, we used four of the six scales (Affective Statements, Restorative Questions, Proactive Circles, and Fair Process). This decision was based on the desirability of extracting a student-reported and teacher-reported RP factor that was comprised of the same scales across informants. Given that 2 teachers had not completed any of the items on the Responsive Circles scale and 3 teachers had not completed any items on the Management of Shame scale, we decided to exclude those scales. When we factor analyzed the four *student-reported* RP scales, all the scales loaded onto one factor (factor loading greater than .654) and accounted for 69% of the variance. Given the loading onto one factor, we extracted a single factor score for each teacher, and together the scores were normally distributed. As with the student scales, we conducted a principal component factor analysis with the four *teacher-reported* RP scales. The scales fell on one factor (factor loading greater than .707) and accounted for 62% of the variance. The teacher-perceived RP factor scores were normally distributed (see Table 2). It is important to note that there were no significant differences on the RP implementation factor scores for teacher and student surveys collected in the late fall versus early winter, $t(29) = .31, p > .05$, and $t(29) = .66, p > .05$, respectively.

Quality of teacher-student relationship. The quality of teacher-student relationships was measured using two different sources—student surveys and school discipline records. Specifically, students completed four items on the Teacher Respect scale using a 4-point Likert scale (*not at all true, somewhat*

TABLE 2 Component Loadings From Principal Component Analysis

	Teacher survey	Student survey
Affective Statements	.843	.867
Restorative Questions	.877	.882
Proactive Circles	.709	.654
Fair Process	.707	.929
Eigenvalues	2.48	2.77
% of total variance	62.07	69.27

true, *true*, and *very true*). They indicated whether the teacher “liked them,” “interrupted them when they had something to say” (reverse scored), “did not enjoy having them in class” (reverse scored), and “never listened to their side” (reverse scored). The scale uses items from Belmont, Skinner, Wellborn, and Connell’s (1992) teacher care and respect scales, which had good reliability in a previous sample ($\alpha = .71, .77$, respectively). The Teacher Respect scale in this study had adequate reliability with the current sample ($\alpha = .67$). In terms of validity, Skinner, Furrer, Marchand, and Kindermann (2008) used the items as part of their teacher support measure, which predicted improvements in emotional and behavioral engagement and declines in behavioral and emotional disaffection in the classroom over time.

The second way we measured the quality of the teacher-student relationship was through teachers’ use of discipline referrals (as found in the school records). The school had over 120 reasons for discipline referral. We were interested in examining reasons that likely reflect some degree of adult-student conflict given research indicates this may be a substantial driver of the racial discipline gap (Gregory & Weinstein, 2008; Skiba, Michael, Nardo, & Peterson, 2002). Similar to previous research (Gregory & Weinstein, 2008), we grouped the following reasons into a “misconduct/defiance” category: disrespect, insubordination, profanity/obscenity, misconduct, and disorderly conduct. The issuing of a referral for any of these reasons typically reflects the culmination of a series of negative interactions between teachers and students—suggesting the disputants were not able to diffuse the conflict with a constructive resolution (Vavrus & Cole, 2002). Typically, when a teacher issues a discipline referral, the student leaves the classroom and meets with an administrator who determines the consequence (e.g., out-of-school suspension). Numerous research studies have used teachers’ office discipline referrals as reliable indicators of the classroom and school climate (for a summary see Irvin, Tobin, Sprague, Sugai, & Vincent, 2004). Irvin and colleagues (2004) synthesized empirical studies and found that higher levels of schoolwide use of office discipline referrals were associated with classroom disorderliness and with student and teacher perceptions of unsafe school conditions. They concluded that reductions in the use of such classroom discipline are a valid indicator of intervention success.

We obtained all referral records on participating teachers for the school year 2011–2012, the same year the student and teacher RP surveys were collected. The discipline referrals were extracted from a schoolwide database and, thus, included referrals issued to any students the participating teachers encountered during the school year (not just the students in the focal classroom in which the surveys were collected). In addition, we were unable to link the discipline referral data with the identities of the student survey respondents given our parent/guardian consent forms did not explicitly request permission to do so. As mentioned previously, none of the students to

whom our participating teachers issued discipline referrals were identified as “American Indian” students. Thus, analyses using the school discipline records included four of the five racial/ethnic groups.

Covariate when predicting teacher respect. Whether a student perceives respect from his or her teacher may in fact be confounded by his or her own behavior. In other words, some students who are unmotivated, aggressive with peers, or oppositional to the teacher might perceive minimal respect from adults in general (regardless of the adult’s approach to discipline in the classroom). This assertion is supported by research on hostile attribution. Students who tend to attribute more hostility to others in ambiguous situations also tend to exhibit more aggressive behavior (e.g., Dodge, 2006). Thus, we included a scale of teacher-reported student cooperation as a covariate to help isolate the effect of RP implementation on teacher respect. Teachers rated every participating student in their focal classroom on the degree to which the student “pays attention,” “tries hard,” “defies or refuses” teacher requests (reverse scored), and “has trouble” working with peers (reverse scored). The 4-point Cooperation Scale ranged from *not at all* to *very much*. The items were selected from previously used scales measuring student engaged and disruptive behavior (Swanson, 1992; Wellborn, 1991). In the current sample, it demonstrated good internal consistency ($\alpha = .80$).

Data-Analytic Plan

The varying structure of our dependent variables required two different data-analytic plans. For the first research question, the dependent variable was at the student level (student-perceived teacher respect). In addition, the student survey respondents were “nested” in teacher classrooms. Thus, we conducted multilevel analyses using HLM 7.0 to account for the student groupings within classrooms (Raudenbush & Bryk, 2002). We examined the unconditional model, which had no predictors, and found that the intraclass correlation ($ICC = .14$) indicated 14% of the variance in teacher respect was between teachers. This significant variation between teachers justified our comparison of RP implementation across classrooms. In our first model using HLM, we entered two Level 1 predictors, whether a student was Asian/White (0) or Latino/African American (1), and student cooperative behavior, as reported by the teacher. In the next model, we entered factor scores of student- and teacher-reported RP implementation at Level 2. Finally, we tested two cross-level interactions—the Level 2 RP implementation variables were entered to help explain the random slope variance of race/ethnicity at Level 1 (Snijders & Bosker, 1999). This addition to the model ascertained whether the link between RP implementation and teacher respect was similar no matter the student race/ethnicity. Given the nonsignificant cross-level interactions, our final hierarchical linear model (HLM) equation was as follows:

Level 1 Model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Race/Ethnicity}_{ij}) + \beta_{2j}(\text{Student Cooperation}_{ij}) + e_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Student-reported RP Implementation}_j) + \gamma_{02}(\text{Teacher-reported RP Implementation}_j) + u_{0j},$$

where *i* refers to student level, *j* refers to teacher level, *e* refers to residual at the student level, and *u* refers to residual at the teacher level.

For the second research question, all of our data were at the teacher level given the school discipline data were not linked to individual student identifiers. We were, thus, unable to conduct multilevel analyses and instead used multiple linear regression in SPSS 20. In two separate blocks, we entered RP implementation as reported by the teachers, followed by RP implementation as reported by the students.

RESULTS

Descriptives

Student-reported, but not teacher-reported, RP implementation was related to the two indicators of teacher-student relationships (teacher respect and discipline referrals; see Table 3). Specifically, higher student-reported RP was associated with greater teacher respect ($r = .58, p < .01$) and fewer misconduct/defiance referrals issued to Latino/African American students ($r = -.45, p < .05$). It was also associated with fewer Asian/White referrals yet the correlation approached significance ($r = -.36, p < .10$). The correlations also show that student-reported and teacher-reported implementation were

TABLE 3 Descriptive Statistics and Correlations Among Restorative Practices (RP) Implementation, Teacher Respect, and Referrals

	<i>M</i>	<i>SD</i>	2	3	4	5	6
1. TR RP implementation ^a	0	1	.23	.08	-.07	-.14	-.18
2. SR RP implementation	0	1	—	.34†	.58**	-.45*	-.36†
3. Student cooperation	3.5	0.27		—	.35†	-.54**	-.44*
4. Teacher respect	3.6	0.27			—	-.07	-.05
5. Referrals to African American/Latino ^b	6.3	9.3				—	.89***
6. Referrals to Asian/White	1.3	2.0					—

^aTR (teacher-reported) RP and SR (student-reported) RP implementation. ^bReferrals = discipline referrals for reasons related to misconduct/defiance.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

not significantly related to one another. This was unexpected given that the teachers and students rated the degree to which RP elements were present in the same classroom around the same time.

RP Implementation Predicting Teacher Respect

In our HLM analyses, Model 1 shows that teacher reports of behavior were associated with student-perceived teacher respect (see Table 4). More specifically, when a teacher reported a student was more cooperative, then the student tended to see the teacher as more respectful ($\beta = .20, p < .01$). Noteworthy is that racial group membership was not associated with teacher respect. In other words, the degree to which the student found the teacher respectful was not related to whether the student was in the Latino/African American/American Indian or Asian/White group.

Model 2 shows that, after accounting for student race/ethnicity and cooperative behavior, student-reported RP implementation ($\beta = .12, p < .01$), but not teacher-reported RP implementation ($\beta = -.05, ns$), was associated with teacher respect. Students reporting greater implementation of the RP elements tended to perceive those teachers as more respectful. The addition of the RP implementation factors explained 17% of the between-teacher variance in teacher respect, yet the variation between teachers remained significant. The cross-level interactions with student- and teacher-reported RP implementation did not significantly explain the random slope variance of race/ethnicity. This suggests the link between RP implementation and teacher respect did not vary by student race/ethnicity.

TABLE 4 HLM Analysis With Student-Reported Teacher Respect as Level 1 Outcome

Measure	Model 1 estimate (SE)	Model 2 estimate (SE)
Level 1 Student-level predictors		
Race (1: Latino/Black; 0: Asian/White) β_{1j}	-.02 (.05)	-.02 (.05)
Student cooperation _{ij} β_{2j}	.20** (.06)	.19** (.06)
Level 2 Teacher-level predictors		
Student report_RP_Implement _j γ_{01}		.12** (.04)
Teacher report_RP_Implement _j γ_{02}		-.05 (.03)
Random effects		
Student level σ^2	.18	.18
Teacher level σ^2	.58*	.48*
Reduced variance between students ^a		0%
Reduced variance between teachers ^a		17%

HLM = hierarchical linear modeling; RP = restorative practices.

^aProportion of unexplained variance reduced from Model 1.

* $p < .05$. ** $p < .01$.

RP Implementation Predicting Teacher Use of Discipline Referrals

We found that teachers issued Asian and White students fewer ($M = 1.28$) misconduct/defiance referrals than they issued to Latino and African American students ($M = 6.34$), as tested in a paired sample t test ($t(29) = 3.63$, $p = .001$). The regression analyses show that student-reported, but not teacher-reported, RP implementation was a significant predictor of misconduct/defiance referrals issued to Latino and African American students ($\beta = -.44$, $p < .05$) and approached significance for referrals issued to Asian and White students ($\beta = -.34$, $p < .10$; see Table 5). Higher implementation of RP, as perceived by students, was associated with lower use of misconduct/defiance referrals. In terms of effect size, the student-reported RP implementation measure explained 11% of the variance in Asian/White referrals and 18% of the variance in Latino/African American referrals.

To illustrate the student-reported RP implementation finding, we split the teachers into those who scored above the mean (High RP) on the student-perceived RP factor and below the mean on the factor (Low RP). Please note that the decision to split the teachers at the mean of RP implementation was based on the conventional use of splitting at the mean to reflect high or low given there has been no prior research establishing an empirically and theoretically derived cut point. Figure 1 shows that the gap in misconduct/defiance referrals between Asian/White ($M = 1.69$ referrals) and Latino/African American ($M = 9.13$ referrals) was wide for those teachers perceived by students as having low RP implementation, as demonstrated by a paired-sample t test ($t(15) = 3.21$, $p = .006$). The gap was smaller when teachers were perceived by their students as having high RP implementation. Yet, a paired sample t test showed a significant difference in referrals remained for this group as well: Asian/White ($M = .77$ referrals) versus Latino/African American ($M = 2.92$ referrals; $t(12) = 2.69$, $p = .02$). This suggests higher RP implementers (above the mean) narrowed the racial discipline gap but did not eradicate it in their referral patterns.

TABLE 5 Regression Models for Number of Defiance Referrals

	White/Asian referrals	African American/Latino referrals
R ²	.11	.18*
Standardized betas		
Teacher-reported RP implementation	-.01	-.04
Student-reported RP implementation	-.34†	-.44*

† $p < .10$. * $p < .05$.

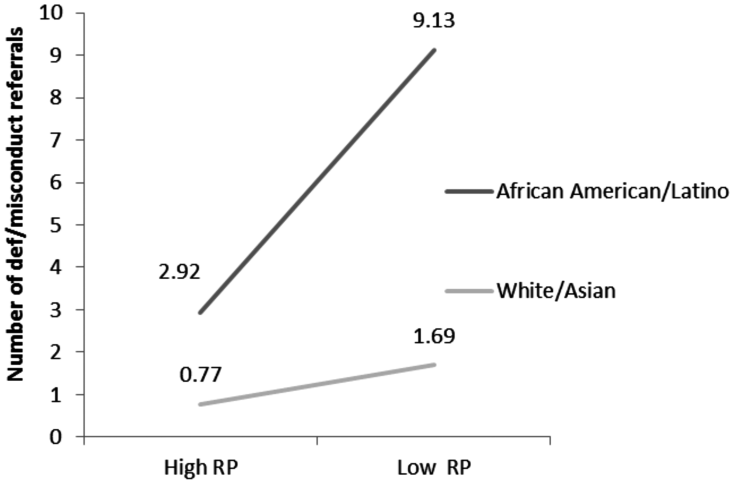


FIGURE 1 Teachers above ($n = 13$) and below ($n = 16$) the mean on student-perceived restorative practices (RP) implementation and number of misconduct/defiance referrals by race/ethnicity.

DISCUSSION

The results suggest that greater RP implementation levels were associated with better teacher-student relationships as measured by student-perceived teacher respect and teacher use of exclusionary discipline. The strength of the findings is the corroboration across source (student survey and school records). The variability of RP implementation across classrooms coupled with the salience of student reports of implementation for positive outcomes suggest there are key functions teacher consultants can play in assessing and improving teachers' uptake of RP in their classrooms.

In addition, the findings offer some initial promise that high-quality RP implementation may be associated with more equitable disciplinary practices. Namely, higher RP implementation predicted greater teacher respect—a relationship that held for students across varying racial and ethnic groups. In addition, teachers who were perceived as implementing more RP elements by their students tended to have fewer differences in the number of misconduct/defiance referrals issued to Asian/White and Latino/African American student groups compared with the large discipline gap for teachers perceived as low on RP elements.

The Need for High-Quality Implementation

This study found a wide range of RP implementation in the participating teachers' classrooms. This corroborates research on RP implementation at the school level. In a study of 18 schools in Scotland, McCluskey et al.

(2008a) documented that RP implementation across schools varied substantially, which they argued may have been somewhat due to fundamental ideological differences between RP programming and more traditional beliefs and practices about how to manage student behavior. The authors suggest that many school administrators and teachers in their study held a more authoritarian perspective of adult power, in which “getting tough” through the use of exclusionary discipline practices was seen as the most effective response to student misbehavior (McCluskey et al., 2008b). This contrasts to the underlying assumptions of RP about the need to reintegrate the wrongdoer into the community. Stinchcomb and colleagues (2006) also noted that RP implementation can falter given the conflict of values underlying a zero tolerance approach versus RP’s flexible negotiation and problem-solving approach. Another potential obstacle to implementation includes teacher perceptions that RP is too time intensive and will interfere with instruction. Given the current climate around teacher evaluation, many teachers may consider any time taken away from conveying course content a serious threat to student academic progress. Given the potential for clashing values and fear of lost instructional time, RP consultation may need to include additional techniques to effectively engage teachers in RP. For instance, motivational interviewing (MI) techniques may be needed with some school staff to develop authentic willingness to change from a more punitive to a more restorative approach—in fact, MI techniques have been used to help facilitate change through school-based consultation (Blom-Hoffman & Rose, 2007; Gueldner & Merrell, 2013; Reinke, Lewis-Palmer, & Merrell, 2008). Consultants may also need to demonstrate to teachers that RP approaches can be integrated into everyday interactions and instruction with their students. This would show how lost instructional time can be minimized and potentially increase teachers’ adoption of the program.

This study found that degree of RP implementation was linked to the quality of teacher-student relationships, which confirms the well-established relationship between fidelity of implementation and student outcomes (Durlak & Dupre, 2008). Our findings also add to the growing recognition that a systematic focus on how to implement programs well is crucial to the dissemination of school-based programming (for a review, see Forman et al., 2013). Scholars of implementation science emphasize that single training workshops are not enough (Forman et al., 2013). Forman, Olin, Hoagwood, Crowe, and Saka (2009) noted that staff need ongoing support and iterative feedback to improve their implementation—a need that could be filled by teacher consultants. Research has confirmed that performance feedback effectively increases teachers’ use of novel programs or approaches (e.g., Noell, 2008). In the current study, we are unable to explain why some teachers implemented RP more than other teachers. Explanatory factors to consider in future research include conflict or correspondence between the underlying values of RP and teachers’ approach to discipline as well as

the degree to which teachers perceive RP impedes or facilitates instruction. Future research needs to examine the link between teachers' utilization of on-site consultation with RP trainers (i.e., observation, modeling, assistance, and problem solving) and the quality of RP implementation in their classrooms.

Honoring Student Voice in Consultation

Student report, but not teacher report, of RP implementation was associated with teacher respect and teacher use of discipline referrals. This suggests we need to seriously consider student perspective on implementation integrity—a perspective that can often be overlooked when consulting with teachers. Pearrow and Pollack (2012) suggested youth should be engaged in a critical examination of their school conditions and offered collaborative roles in affecting change. The findings in the current study echo the need to take the voices of youth seriously in schoolwide implementation of RP. Including youth in the implementation process is in direct alignment with the RP principle of authentically integrating student emotions, opinions, and suggestions when solving disputes. According to McClusky et al. (2008a), a marker of higher fidelity of RP implementation in their sample of 18 Scottish schools included students indicating that they “felt heard.” In other words, on a schoolwide basis, the well-implemented RP schools tended to integrate student voice. A similar process of including student voice could occur when implementing a new RP initiative. This might entail providing feedback to teachers based on a regular collection of student surveys on RP implementation.

Race, Ethnicity, and RP Implementation

This was the first study to examine RP implementation in relation to equity in school discipline as measured by teachers' differential use of office disciplinary referrals for disruption and defiance with students in varying racial and ethnic groups. This line of research is needed given the enduring nature of the racial discipline gap and the dearth of documented interventions that narrow or eradicate the gap. Few studies, as far as authors are aware, have examined whether changing from a more punitive approach to discipline to a more restorative approach helped reduce the racial discipline gap. Simson (2012) found that the disparity in percentage of African American versus White student suspensions was slightly lower (approaching significance at $p < .10$) in RJ schools ($n = 13$) compared with a matched set of non-RJ schools ($n = 45$) in Denver and Santa Fe. The findings are somewhat limited given there was no systematic assessment of the quality of RP implementation in the RJ schools.

Our study findings indicate that students of varying race/ethnicity experienced RP implementation similarly. For instance, within the same class-

room, when White and Asian students reported the teacher frequently employed RP elements, African American, Latino, and American Indian students tended to concur. In addition, with higher implementing RP teachers, both Asian/White and Latino/African American/American Indian groups reported feeling respected by the teacher. That the relationship between RP implementation and respect held across racial and ethnic groups suggests that RP may be culturally appropriate or culturally congruent with varying groups. This is promising given interventions need to address what may be called “a relationship gap” between some student groups and their teachers—for example, as a group, African American students report less fairness and support compared with White students in schools (Wald & Kurlaender, 2003). Interventions that can equally improve the quality of teacher-student relationships across racial and ethnic groups may have potential to reduce the racial discipline gap (Gregory, Allen, Mikami, Hafen, & Pianta, 2014).

With improved relationships, distrust, implicit bias, and cultural misunderstanding may be reduced between teachers and students historically overrepresented in school discipline (Gregory et al., 2011; Simson, 2012). RP’s focus on developing an authoritative climate in the classroom through a range of practices (e.g., Proactive Circles, Affective Statements, Restorative Questions) may elicit trusting teacher-student interactions in which students feel supported and treated fairly. A sensitivity to individual student perspectives and the collective voice of students accompanied by consistent and fair accountability for jointly developed classroom rules may reduce the likelihood that students in stigmatized groups will be excluded from the classroom for discipline reasons. In fact, this study showed that high-implementing RP teachers rarely used exclusionary discipline for misconduct/defiance, and they had a narrower gap in referrals between White/Asian and African American/Latino students compared with low-implementing RP teachers. Future research should explore possible mediating processes that can help explain why well-implemented RP was associated with reduced reliance on exclusionary discipline, especially among African American and Latino students. In addition, the racial discipline gap was not totally eradicated in classrooms with high RP implementation, which raises many questions about whether future consultants can further boost these teachers’ RP implementation or whether they need to offer other approaches to improving teacher-student relationships (Gregory et al., 2014).

Limitations and Future Directions

Some limitations should be noted when drawing conclusions from the research. The study examined student and teacher surveys and discipline referrals during the first-year of RP implementation in the school. We did not have outsider observers verify quality of RP implementation, which

many consider the “gold standard” for measuring fidelity of implementation (Forman et al., 2013). Also, given that IIRP implements its RP programs across 2 years, this study may not have captured the full effect of RP implementation on student outcomes. Schoolwide interventions typically require a minimum of 2 years to take hold, thus follow-up studies are warranted.

Shared method variance is a limitation of the methodology used to examine the link between RP implementation and teacher respect (Research Question 1). Students reported on RP implementation (independent variable) and teacher respect (dependent variable). Thus, rater bias likely inflated the relationship between the two variables. That said, by including teacher ratings of student behavior as a covariate we provide additional rigor to the findings—we likely accounted for some of the rater bias. Another limitation to consider is that despite having explained a substantial amount of variance in teacher respect across the classrooms (17%), variance between teachers remained significant suggesting systematic differences were left unexplained. Thus, additional research is needed to understand why some teachers were experienced as more respectful than others. Future research may include an examination of teachers’ instructional practices—some of which may be experienced as more or less respectful (e.g., differential treatment of higher and lower achievers with the classroom; Weinstein, 2002). Other student explanatory characteristics might be related to perceptions of teacher respect as well. For instance, students’ past experience managing conflict with teachers and receiving discipline referrals may relate to the degree to which they read hostility into current teachers’ ambiguous behaviors (Dodge, 2006).

By randomly selecting *one* of the classrooms of our participating teachers at a single point in time, we took a “snapshot” of RP implementation. That snapshot was linked to teachers’ use of exclusionary discipline with all the students they encountered in the school year. The analyses were thus based on an assumption that our snapshot is reflective of RP implementation across the teachers’ instructional schedule. This assumption needs to be tested in future research. Namely, it is unknown whether teachers implement RP differently as they encounter different constellations of students in their classrooms. That we explained 11% to 18% of the variance in use of exclusionary discipline for disruption and defiance, however, suggests that our snapshot tapped into ways the teachers differ in their approach to school discipline. That said, future research should link teachers’ RP use in specific classrooms with referrals in those same classrooms. With such a link, future multilevel research would address the limits inherent in our use of multiple regression when predicting teachers’ use of discipline referrals (Research Question 2). A multilevel model could differentiate the degree to which the teacher level (e.g., high-quality RP implementation) and student level (e.g., observed student behavior) explain variance in student disciplinary referral rates. This would provide more precise information for consultants

to educate teachers about the relative contribution of teacher versus student characteristics to referral patterns. Learning about the teacher contribution to such patterns could help teachers understand the degree to which they have agency in reducing the racial discipline gap.

In addition, future research should systematically track RP consultants' activities throughout the implementation process and measure the quality of RP implementation at multiple time points in the school year. Our single snapshot did not capture the possible improvements (or decrements) in implementation across the school year. Identifying patterns in implementation over time, such as typical "fade-out" points, would provide specific times in the school year when consultants might do check-ins with teachers to prevent predictable declines in implementation.

We believe our collection of student perceptions of RP implementation is a strength of the study. Yet, student surveys are not without limitations. Student responses on the RP implementation surveys may have related to overall perceptions of their teachers and not specifically to what they observed as new disciplinary practices in the classroom (e.g., students may be inclined to report their teachers are doing more RP simply because they want to be positive about teachers with whom they feel connected). Following this logic, teachers with high RP implementation, according to students, may simply be better at providing social, emotional, and academic support to students. This might suggest that even without RP these teachers would tend to rely less on exclusionary discipline and be adept at preventing or diffusing conflict. As such our student-report RP implementation factor would need to be corroborated by systematic observations of RP in classrooms. Observations would also help us understand the lack of correspondence between teacher-reported and student-reported RP implementation. Further scale development would also ascertain whether this study's teacher-reported scale did not correspond with student report simply because it lacked validity—which means it might not have measured what it claimed to measure. In other words, further research on reliable and valid ways to capture teacher-reported RP implementation is needed.

This study examined RP implementation in a single RP factor score, which combined reports of four of the RP elements: Affective Statements, Restorative Questions, Proactive Circles, and Fair Process. Future research might examine the quality of implementation and associated outcomes of other RP elements, which were not included in this factor score (e.g., Management of Shame). Another important line of future research might compare the effectiveness of the prevention-oriented elements (e.g., proactive circles) and the intervention-oriented elements (e.g., restorative conferences) in terms of building community, deterring future rule breaking, and reducing the racial discipline gap.

A specific focus on the prevention-oriented RP elements would also enable a comparison study with other prevention-oriented programming such

as School-Wide Positive Behavior Intervention Supports (SWPBIS). This kind of comparison study would help discern whether or not a humanist/systems approach inherent to RP or a more behavioral approach inherent to programs such as SWPBIS would be the most developmentally sensitive and culturally sensitive approach to working with adolescents. Noteworthy is that despite being well disseminated, SWPBIS, as of yet, has not shown a reduction in the racial discipline gap (despite an overall reduction in use of exclusionary discipline sanctions; Kaufman et al., 2010; Vincent et al., 2013; Vincent & Tobin, 2011).

In terms of the racial discipline gap, the current findings need to be replicated given this is a correlational and nonexperimental study that cannot support causal conclusions (i.e., RP caused the gap to reduce). Moreover, student-reported RP implementation was significantly related to Latino and African American referrals for disruption/defiance but was only a trend for Asian and White referrals. This discrepancy may be due to the low occurrence of referrals for Asian and White students, which would make the finding a measurement artifact. Keeping that caveat in mind, however, it would be informative to ascertain whether RP has a more robust effect on interactions between teachers and their Latino and African American students. If this is found in future research, RP would have tremendous potential to reduce disparate use of exclusionary discipline with these groups.

Conclusion

The study contributes to a growing body of research that demonstrates the potential of RP for improving schools. It found that teachers who were perceived by their students as frequently implementing many of the RP elements tended to have better relationships with their students compared with infrequent implementers of RP. This was seen in the degree to which students felt respected by their teachers and teachers' use of disruption/defiance disciplinary referrals. The findings also have implications for the potential of RP in terms of reducing the racial discipline gap. Higher RP implementation was associated with lower use of disruption/defiance disciplinary referrals with Latino and African American students. Finally, the study also confirms the need for consultation that integrates student perspectives on implementation and systematically addresses implementation challenges for novel practices in schools.

ACKNOWLEDGMENTS

We are grateful to the educators and students who participated in the current research.

FUNDING

We thank the Atlantic Philanthropies and the Novo Foundation for their grant support of this research.

REFERENCES

- American Academy of Pediatrics. (2013). Policy statement: Out-of-school suspension and expulsion. *Pediatrics*, *131*, 1000–1007.
- American Psychological Association Task Force. (2008). Are zero tolerance policies effective in schools? An evidentiary review and recommendations. *American Psychologist*, *63*, 852–863.
- Amstutz, L., & Mullet, J. H. (2005). *The little book of restorative discipline for schools*. Intercourse, PA: Good Books.
- Balfanz, R., Byrnes, V., & Fox, J. (2014). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. In D. Losen (Ed.), *Closing the school discipline gap*. New York, NY: Teachers College Press.
- Baumrind, D. (1968). Authoritarian vs. authoritative parental control. *Adolescence*, *3*, 255–272.
- Baumrind, D. (1991). Parenting styles and adolescent development. In R. Lerner, A. C. Petersen, & J. Brooks-Gunn (Eds.), *The encyclopedia of adolescence* (pp. 746–758). New York, NY: Garland.
- Belmont, M., Skinner, E., Wellborn, J., & Connell, J. (1992). *Teacher as social context (TASC): Two measures of teacher provision of involvement, structure, and autonomy support* (Tech. Rep.). Rochester, NY: University of Rochester.
- Blom-Hoffman, J., & Rose, G. S. (2007). Applying motivational interviewing to school based consultation: A commentary on “Has consultation achieved its primary prevention potential?” *Journal of Educational and Psychological Consultation*, *17*, 151–156.
- Blood, P., & Thorsborne, M. (2005). *The challenges of culture change: Embedding restorative practice in schools*. Paper presented at the Sixth International Conference on Conferencing, Circles and other Restorative Practices: Building a Global Alliance for Restorative Practices and Family Empowerment, Sydney, Australia.
- Braithwaite, J. (1989). *Crime, shame and reintegration*. New York, NY: Cambridge University Press.
- Braithwaite, J. (2001). Restorative justice and a new criminal law of substance abuse. *Youth and Society*, *33*, 227–248.
- Coates, R. B., Umbreit, M., & Vos, B. (2003). Restorative justice circles: An exploratory study. *Contemporary Justice Review*, *6*, 265–278.
- Costello, B., Wachtel, J., & Wachtel, T. (2010). *Restorative circles in schools: Building community and enhancing learning*. Bethlehem, PA: International Institute for Restorative Practices.
- Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*, 1–12.

- Dodge, K. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*, 791–814.
- Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.
- Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., III, & Booth, E. A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement*. New York, NY: Council of State Governments. Retrieved from <http://justicecenter.csg.org/resources/juveniles>
- Forman, S. G., Olin, S., Hoagwood, K., Crowe, M., & Saka, N. (2009). Evidence-based intervention in schools: Developers' views of implementation barriers and facilitators. *School Mental Health, 1*, 26–36.
- Forman, S. G., Shapiro, E. S., Coddling, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., . . . Stoiber, K. C. (2013). Implementation science and school psychology. *School Psychology Quarterly, 28*, 77–100.
- Gal, T., & Moyal, S. (2011). Juvenile victims in restorative justice findings from the reintegrative shaming experiments. *British Journal of Criminology, 51*, 1014–1034.
- González, T. (2011). Restoring justice: Community organizing to transform school discipline policies. *U.C. Davis Journal of Juvenile Law & Policy, 15*(1), 9–10.
- González, T. (2014). Addressing racial disparities in school discipline: Restorative alternatives to exclusionary practices in schools. In D. Losen (Ed.), *Closing the school discipline gap*. New York, NY: Teachers College Press.
- Gregory, A., Allen, J., Mikami, A., Hafen, C., & Pianta, R. (2014). The promise of a teacher professional development program in reducing the racial disparity in classroom exclusionary discipline. In D. Losen (Ed.), *Closing the school discipline gap*. New York, NY: Teachers College Press.
- Gregory, A., & Cornell, D. (2009). “Tolerating” adolescent needs: Moving beyond zero tolerance policies in high school. *Theory into Practice, 48*, 106–113.
- Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for black and white high school students. *American Educational Research Journal, 48*, 904–934.
- Gregory, A., Cornell, D., Fan, X., Sheras, P. L., Shih, T., & Huang, F. (2010). Authoritative school discipline: High school practices associated with lower student bullying and victimization. *Journal of Educational Psychology, 102*, 483–496.
- Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39*, 59–68.
- Gregory, A., & Thompson, A. R. (2010). African American high school students and variability in behavior across classrooms. *Journal of Community Psychology, 38*, 386–402.
- Gregory, A., & Weinstein, S. R. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology, 46*, 455–475.
- Gueldner, B., & Merrell, K. (2013). Evaluation of a social emotional learning program in conjunction with the exploratory application of performance feedback incorporating motivational interviewing techniques. *Journal of Educational and Psychological Consultation, 21*, 1–27.

- Hemphill, S. A., Toumbourou, J. W., Herrenkohl, T. I., McMorris, B. J., & Catalano, R. F. (2006). The effect of school suspensions and arrests on subsequent adolescent antisocial behavior in Australia and the United States. *Journal of Adolescent Health, 39*, 736–744.
- Irvin, L. K., Tobin, T. J., Sprague, J. R., Sugai, G., & Vincent, C. G. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status and effects of school-wide behavioral interventions. *Journal of Positive Behavior Interventions, 6*, 131–147.
- Karp, D. R., & Breslin, B. (2001). Restorative justice in school communities. *Youth & Society, 33*, 249–272.
- Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N., & Hernandez-Brereton, M. (2010). Patterns in office discipline referral data by grade, race/ethnicity, and gender. *Journal of Positive Behavior Interventions, 12*, 44–54.
- Krezmien, M. P., Leone, P. E., & Achilles, G. M. (2006). Suspension, race, and disability: Analysis of statewide practices and reporting. *Journal of Emotional and Behavioral Disorders, 14*, 217–226.
- Lewis, S. (2009). *Improving school climate: Findings from schools implementing restorative practices*. Bethlehem, PA: International Institute of Restorative Practices. Retrieved from <http://www.iirp.edu/pdf/IIRP-Improving-School-Climate.pdf>
- Losen, D. J., & Gillespie, J. (2012). *Opportunities suspended: The disparate impact of disciplinary exclusion from school*. Los Angeles, CA: The Center for Civil Rights Remedies. Retrieved from <http://civilrightsproject.ucla.edu>
- McCluskey, G., Lloyd, G., Kane, J., Stead, J., Riddell, S., & Weedon, E. (2008a). Can restorative practices in schools make a difference? *Educational Review, 60*(4), 405–417.
- McCluskey, G., Lloyd, G., Stead, J., Kane, J., Riddell, S., & Weedon, E. (2008b). ‘I was dead restorative today’: From restorative justice to restorative approaches in school. *Cambridge Journal of Education, 38*, 199–216.
- McGarrell, E. F., & Hipple, N. K. (2007). Family group conferencing and re-offending among first-time juvenile offenders: The Indianapolis experiment. *Justice Quarterly, 24*, 221–246.
- Mirsky, L. (2011). Building safer, saner schools. *Educational Leadership, 69*, 45–49.
- Monroe, C. (2013). Colorizing educational research: African American life and school as an exemplar. *Educational Research, 42*, 9–19.
- Nathanson, D. (1997). Affect theory and the compass of shame. In M. Lansky & A. Morrison (Eds.), *The widening scope of shame* (pp. 339–354). Hillsdale, NJ: Analytic Press.
- Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315–334). New York, NY: Erlbaum.
- Osher, D., Bear, G. G., Sprague, J. R., & Doyle, W. (2010). How can we improve school discipline? *Educational Researcher, 39*, 48–58.
- Pearrow, M. M., & Pollack, S. (2012). Youth empowerment in oppressive systems: Opportunities for school consultants. *Journal of Educational and Psychological Consultation, 19*, 45–60.

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Reinke, W. M., Lewis-Palmer, T., & Merrell, K. W. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*, 315–332.
- Rodriguez, N. (2007). Restorative justice at work: Examining the impact of restorative justice resolutions on juvenile recidivism. *Crime and Delinquency, 53*, 355–379.
- Schiff, M. (2013, January). *Dignity, disparity, and desistance: Effective restorative justice strategies to plug the "school-to-prison pipeline."* Paper presented at the Center for Civil Rights Remedies National Conference, Closing the School to Research Gap: Research to Remedies, Washington, DC.
- Simson, D. (2012, May). *Restorative justice and its effects on (racially disparate) punitive discipline.* Paper presented at the Seventh Annual Conference on Empirical Legal Studies, Los Angeles, CA. Retrieved from <http://ssrn.com/abstract=21072420>
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *Urban Review, 34*, 317–342.
- Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*, 765–781.
- Smetana, J., & Gaines, C. (1999). Adolescent-parent conflict in middle-class African American families. *Child Development, 70*, 1447–1463.
- Snijders, T. A., & Bosker, R. L. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Stinchcomb, J. B., Bazemore, G., & Riestenberg, N. (2006). Beyond zero tolerance: Restorative justice in secondary schools. *Youth Violence and Juvenile Justice, 4*, 123–147.
- Swanson, J. M. (1992). *Assessment and treatment of ADD*. Irvine, CA: K.C. Press.
- Tomkins, S. S. (1991). *Affect imagery consciousness: Vol. 3. The negative affects: Anger and fear*. New York, NY: Springer.
- Turiel, E. (2005). The many faces of parenting. In J. G. Smetana (Ed.), *Changing boundaries of parental authority during adolescence: New directions in child and adolescent development* (pp. 79–88). San Francisco, CA: Jossey-Bass.
- U.S. Department of Education, National Center for Education Statistics. (2003). *Status and trends in the education of Hispanics* (NCES 2003-008). Washington, DC: Author.
- Vavrus, F., & Cole, K. M. (2002). "I Didn't Do Nothin'": The discursive construction of school suspension. *The Urban Review, 34*, 87–111.
- Vincent, C. G., Sprague, J. R., CHiXapkaid (Pavel, M.), Tobin, T. J. & Gau, J. (2014). The Effectiveness of school-wide positive behavior interventions and supports in reducing racially inequitable disciplinary exclusion. In D. Losen (Ed.), *Closing the school discipline gap: Research for policymaker*. New York, NY: Teachers College Press.
- Vincent, C. G., & Tobin, T. J. (2011). An examination of the relationship between implementation of school-wide positive behavior support (SWPBS) and exclusion of students from various ethnic backgrounds with and without disabilities. *Journal of Emotional and Behavioral Disorders, 19*, 217–232.

- Wachtel, T. (2012). *Defining restorative*. Retrieved from <http://iirpds.pointinspace.com/pdf/Defining-Restorative.pdf>
- Wachtel, T., Costello, B., & Wachtel, J. (2009). *The restorative practices handbook for teachers, disciplinarians and administrators*. Bethlehem, PA: International Institute of Restorative Practices.
- Wachtel, T., O'Connell, T., & Wachtel, J. (2010). *Restorative justice conferencing: Real justice and the conferencing handbook*. Bethlehem, PA: International Institute for Restorative Practices.
- Wald, J., & Kurlaender, M. (2003). Connected in Seattle? An exploratory study of student perceptions of discipline and attachments to teachers. *New Directions for Youth Development, 2003*, 35–54.
- Wallace, J. M., Jr., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among U.S. high school students: 1991–2005. *Negro Educational Review, 59*, 47–62.
- Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*. Cambridge, MA: Harvard University Press.
- Wellborn, J. G. (1991). *Engaged and disaffected action: The conceptualization and measurement of motivation in the academic domain*. Rochester, NY: University of Rochester.
- Zehr, H. (2002). *The little book of restorative justice*. Intercourse, PA: Good Books.
- Zehr, H., & Toews, B. (2004). *Critical issues in restorative justice*. Monsey, NY: Criminal Justice Press.

Anne Gregory, PhD, is an Associate Professor in the Graduate School of Applied and Professional Psychology at Rutgers, the State University of New Jersey. Her research has focused on the persistent trend that African American adolescents are issued school suspensions and expulsions at higher rates than adolescents from other groups. Her research interests also include teacher professional development and characteristics of restorative communities in schools. Through program development, implementation, and evaluation, she aims to address this trend by strengthening characteristics of teachers, classrooms, and schools associated with the successful schooling of diverse students.

Kathleen Clawson, MEd, is a doctoral candidate in the School Psychology Program in the Graduate School of Applied and Professional Psychology at Rutgers, the State University of New Jersey. Her research addresses students' experience of schooling with a focus on how they experience discipline practices.

Alycia Davis, PsyM, is a doctoral candidate in the Clinical Psychology Program at the Graduate School of Applied and Professional Psychology at Rutgers, the State University of New Jersey. Her research addresses preventive interventions for youth at risk for antisocial behaviors, especially in regards to underserved populations.

Jennifer Gerewitz, PsyM, is a doctoral candidate in the School Psychology Program in the Graduate School of Applied and Professional Psychology at Rutgers, the State University of New Jersey. Her research addresses interventions that effectively decrease discipline referrals and have the potential to narrow the racial discipline gap.

Note: The authors report that to the best of their knowledge neither they nor their affiliated institutions have financial or personal relationships or affiliations that could influence or bias the opinions, decisions, or work presented in this article.