AERA Open
October-December 2019, Vol. 5, No. 4, pp. 1–15
DOI: 10.1177/2332858419897274
Article reuse guidelines: sagepub.com/journals-permissions

Focused Classroom Coaching and Widespread Racial Equity in School Discipline

Anne Gregory
Rutgers University
Erik A. Ruzek
Jamie DeCoster
University of Virginia
Amori Yee Mikami
University of British Columbia
Joseph P. Allen
University of Virginia

We examined the effects of a teacher coaching program on discipline referrals using records from 7,794 U.S. classrooms in secondary schools. Some classroom teachers took part in a trial: They were randomized to receive intensive coaching in a focal classroom or to form a business-as-usual control group. The remaining teachers taught in the same schools as the teachers in the trial. Previous research suggested that the coaching program was associated with increasing equity in discipline referrals in focal, coached classrooms. The current study addressed whether effects found in the teachers' focal, coached classrooms generalized to diverse classrooms in their course load. Results suggested that the coaching program had no generalized effects on reducing referrals with African American students or racial referral gaps in classrooms with coached teachers, relative to the control teachers and the other teachers in the schools. We offer implications for coaching programs and directions for equity-oriented efforts to reduce racial discipline gaps.

Keywords: *teacher coaching, professional development, racial equity, discipline referrals*

My Teaching Partner–Secondary (MTP-S) was the first teacher professional development program in middle and high schools shown to reduce the racial discipline gap in *focal* classrooms in which teachers received intensive coaching (Gregory, Allen, Mikami, Hafen, & Pianta, 2015). The MTP-S coaching targets a single instructional period (i.e., the focal classroom), out of the four to six periods in a typical secondary school teacher's daily schedule (Allen, Pianta, Gregory, Mikami, & Lun, 2011). Little is known about generalizable effects of teacher professional development programming. The current study investigated the extent to which MTP-S coaching resulted in *widespread* effects on disciplinary practices.

Racial Disparities in School Discipline

The most recently released national discipline data from 2015 to 2016 showed that African American males and females each constituted 8% of student enrollment in the United States, but 25% and 14% of those issued one or more suspensions, respectively (U.S. Department of Education, 2018). Racial differences in discipline rates are significant, whether statistically controlling for a measure of socioeconomic status at the school level (e.g., Anyon et al., 2014) or at the student level (e.g., Wallace, Goodkind, Wallace, & Bachman, 2008). Similar results have been found in studies accounting for student differences in achievement and in teacher-, parent- and self-reported behavior (e.g., Bradshaw, Mitchell, O'Brennan, & Leaf, 2010; Owens & McLanahan, 2018; Skiba & Williams, 2014). In other words, African American students remain overrepresented in school discipline sanctions after accounting for a range of potential explanatory factors. Reducing the use of exclusionary discipline is therefore imperative from a civil rights perspective of ensuring access to equitable schooling (U.S. Department of Education, 2014).

Further underscoring this imperative, the American Academy of Pediatrics (2013) issued a statement that "the adverse effects of out-of-school suspension and expulsion on the student can be profound" and results in students being "separated from the educational process" (p. e1001). When students are removed from class due to discipline sanctions, they miss valuable instructional time, making it subsequently harder to keep up with coursework (Noltemeyer, Ward, &

Mcloughlin, 2015). This may explain consistent research findings that exclusionary discipline, such as suspension, is associated with worsening student academic and behavioral trajectories (Hwang, 2018; Mittleman, 2018; Noltemeyer et al., 2015).

Most negative discipline encounters originate in the classroom (McIntosh, Frank, & Spaulding 2010). Teacher-issued office discipline referrals result in students being sent to an administrator where they are typically assigned consequences such as in-school or out-of-school suspension (McIntosh et al., 2010). This underscores the importance of interventions that support the *classroom teacher* to address racial disparities and reduce exclusionary approaches to handling student misconduct (Bradshaw et al., 2018). Existing reforms include teacher professional development workshops, revisions to discipline policy, and tiered behavioral supports (e.g., Restorative Practices, Schoolwide Positive Behavior Intervention Supports; Anyon et al., 2016; Deakin & Kupchik, 2018; Gregory, Skiba, & Mediratta, 2017; McCluskey, 2018; Swain-Bradway, Lindstrom Johnson, Bradshaw, & McIntosh, 2017). In some reform initiatives, coaches work one-on-one with teachers (e.g., Bradshaw et al., 2018).

MTP-S is one such teacher coaching program, where coaches partner with teachers over a 2-year intervention period to improve teacher-student interactions. MTP-S has been found to result in improved student academic achievement (Allen et al., 2011), engagement (Gregory, Allen, Mikami, Hafen, & Pianta, 2014), and classroom peer relationships (Mikami, Gregory, Allen, Pianta, & Lun, 2011). In our recent experimental trial, the coaches did not explicitly focus on equitable disciplinary practices. Yet, with an eye toward finding downstream outcomes, Gregory et al. (2016) examined MTP-S for its promise in reducing racial disparities in classroom-issued referrals. In focal classrooms (to which the coaching intervention was targeted), MTP-S significantly reduced teachers' use of discipline referrals with their African American students (Gregory et al., 2015). The findings held when accounting for student gender, prior achievement, and low-income status. The reduced racial discipline gap among MTP-S teachers remained significant in focal classrooms the year after the coaching was discontinued (Gregory et al., 2016).

MTP-S in Focal, Coached Classrooms

MTP-S is a 2-year professional development program designed to improve the emotional, organizational, and instructional features of classrooms through coaching the teacher to display high-quality interactions with students (Hafen, Ruzek, Gregory, Allen, & Mikami, 2015; Hamre et al., 2013; for a recent summary of MTP-S theory, see Gregory, Ruzek, et al., 2017). In the MTP-S program for secondary school, teachers pick one focal classroom out of their course load each academic year to be the target of the coaching, based on teacher perception that this classroom is the most academically or behaviorally challenging. Every 2 weeks teachers submit recordings of their teaching in the focal classroom. Coaches view the video and reflect with the teacher about ways to improve their teacher-student interactions in that focal classroom. In a follow-up year, the coaching was stopped but teachers continued to send video recordings of their focal classroom.

Coaches use the Classroom Assessment Scoring System–Secondary (CLASS-S) to guide the coaching process (*My Teaching Partner Consultancy Manual*, 2010; Pianta, Hamre, Hayes, Mintz, & LaParo, 2008). The CLASS-S is a reliable and valid observation tool that measures teacher-student interaction quality (e.g., Allen et al., 2013; Hafen et al., 2015). Coaches look for illustrative examples of one or more (from a total of 10) dimensions of the CLASS-S (Pianta et al., 2008), which reflect the emotional support, instructional support, and organization of the classroom (Pianta & Hamre, 2009), until all dimensions are covered in a school year.

Effects of MTP-S on teachers' disciplinary practices indicated that improvements in two CLASS-S dimensions were possible mediators to explain the reduced racial disparities in discipline referrals in coached teachers' focal classrooms (Gregory et al., 2016). Accounting for student, classroom, and teacher covariates, in classrooms where teachers showed greater improvement in Teacher Sensitivity and Analysis and Inquiry, African American students in focal classrooms were less likely to be issued a disciplinary referral than their peers (Gregory et al., 2016). We speculate that when MTP-S teachers increased their sensitivity to students' social and emotional needs, they individuated and personalized their relationships with African American students. By doing so, they may have slowed down their disciplinary decision making and disrupted unconsciously held racial implicit bias (Girvan, Gion, McIntosh, & Smolkowski, 2017). In terms of Analysis and Inquiry, teachers' facilitation of higher level thinking skills through challenging problems may have been somewhat novel for African American students who had been tracked into more remedial and understimulating classrooms (Oakes, 1985). This positive dynamic may have reinforced mutually held trust and high expectations (Kuklinski & Weinstein, 2001; Okonofua, Walton, & Eberhardt, 2016; Yeager et al., 2014).

Despite these promising findings, to date, empirical research about MTP-S has only considered outcomes in the focal classrooms that were the target of coaching during the 2-year intervention phase, and the year after the coaching was discontinued. Research has not examined generalization of treatment effects across classrooms taught by the same teacher within the same academic year. Just like dropping a stone in water produces ripple effects, MTP-S-elicited change may ripple (or cross over) to other classrooms in the teacher's course load. For the current study, videotaped instruction from teachers' full course load was not available. Thus, we

were unable to examine whether Teacher Sensitivity and Analysis and Inquiry (CLASS-S dimensions) improved in all of the MTP-S teachers' classrooms. Yet we were able to test whether benefits of MTP-S on equitable discipline referrals were found in teachers' full schedule of classes, not just in their most challenging classroom where they were being coached or were collecting data for the research project in the follow-up year.

Crossover Effects of Teacher Professional Development

To our knowledge, few studies of secondary school teachers have examined whether the benefits of a teacher professional development intervention generalize from one classroom setting to other classrooms taught by the same teacher across the school day. This is an interesting question for theoretical and practical reasons. Understanding generalization helps determine the ingredients that allow teachers to apply skills learned in one setting to another and from one year to the next year with new groups of students (i.e., sustainability).

Specific to teacher professional development interventions, a recent meta-analysis of 60 experimental studies showed that teacher coaching improved student achievement (Kraft, Blazar, & Hogan, 2018), but did not examine the degree to which effects crossed over to other classrooms if teachers taught changing groups of students across the day and school year. Notably, the majority of the studies in this meta-analysis were at the elementary level, restricting the examination of generalization across classrooms taught by the same teacher. Generalization of intervention effects from one class period to another, taught by the same teacher, is highly pertinent at the secondary school level where teachers instruct four to six class periods each day. The various classes taught by the same teacher may be diverse. One class may be predominantly composed of students with a history of low (or high) academic achievement, while another class may have heterogeneous grouping. Moreover, given that academic tracks can fall along racial and socioeconomic lines (Oakes, 1985), a remedial class may differ in student demographics relative to an advanced class. Classrooms may also differ in the number of special education students and whether the classroom is co-taught (Friend, 2015; Jackson, Willis, Giles, Lastrapes, & Mooney, 2017; McLeskey, Landers, Williamson, & Hoppey, 2012). Such diversity across periods taught by the same teacher in a school day may present a challenge for generalization of intervention effects across classrooms. Furthermore, it is important to consider whether any intervention effects hold when accounting for variability in classroom characteristics.

Current Study

Given the lack of prior research on generalization of teacher coaching effects, we offer a set of exploratory hypotheses. We anticipate that, just like dropping a stone in water produces ripple effects, MTP-S-elicited change in the focal classroom selected by the teacher to be the most challenging and to be the target of coaching may generalize to other classrooms taught by the same teacher. Thus, the research questions address whether intervening at the classroom level in one course leveraged *widespread change* in reducing racial disparities in discipline. The current study considered (a) whether there was a positive, discipline ripple effect that spread from MTP-S teachers' focal classroom into their other classrooms during the 2-year active coaching period and (b) whether such change also spread from the focal classroom to the other classrooms in the 1-year follow-up period after the intervention ended.

The current study builds on an aforementioned experimental trial of MTP-S (Allen, Hafen, Gregory, Mikami, & Pianta, 2015; Gregory et al., 2016). The district supplied de-identified records of students enrolled in all classrooms of teachers in the randomized trial during the 2 years of MTP-S intervention and the follow-up year. They also supplied the records of students taught by all other teachers at the same schools for those 3 years. Thus, we were able to compare MTP-S coached teachers' discipline referrals to (a) the control teachers (who volunteered for the study but were randomly assigned to a business-as-usual condition) and (b) the nonstudy teachers (who did not volunteer to participate or who did not meet inclusion criteria for the study). Having these two comparison groups increased the scope of the analyses. We were able to expand the comparison of the treatment group to all teachers in the schools, not just those who (a) were motivated to participate in a coaching program, (b) met inclusion criteria such as having an end-of-course standardized exam to assess student learning, and (c) may have altered their behavior given they were aware of being observed in the study (i.e., the Hawthorne effect; McCambridge, Witton, & Elbourne, 2014). Thus, the nonstudy group, relative to the control group, can be conceptualized as a "business-as-usual" group that is more closely aligned with the realities of everyday teaching. However, note that the comparison of MTP-S coached and nonstudy teachers no longer utilizes random selection and, therefore, is not within the experimental paradigm.

In recognition of the heterogeneity of classroom composition and drawing on prior school discipline research (Anyon et al., 2014), the analyses covaried classroom and student characteristics. Our research questions were as follows:

Research Question 1 (RQ1): During the 2 years of MTP-S coaching in focal classrooms and in the year after the intervention ended, did *African American students* enrolled in MTP-S teachers' classrooms receive fewer office discipline referrals in comparison with *African American students* enrolled in the classrooms of control teachers and in the classrooms of nonstudy teachers?

Research Question 2 (RQ2): During the 2 years of MTP-S coaching in focal classrooms and in the year after the intervention ended, was the *discipline referral gap between African American and non–African American students* enrolled in MTP-S teachers' classrooms smaller in comparison with *referral gap between these groups* enrolled in the classrooms of control teachers and in the classrooms of nonstudy teachers?

Method

Participants

The participating public school district was in a small city in the southeastern United States. Students in this district were predominantly African American (60%), with a smaller percentage who were White (28%), and Asian, Latinx, Pacific Islander, American Indian, or two or more races (all between 0.2% and 5.2%). In 2011, African American students in this district received 73% of the suspensions relative to the 19% received by White students (https://ocrdata.ed.gov/). Passing rates in the district for statewide achievement tests were 83% for English and 60% for Math, which were below state averages. The current study used student records from four high schools and one middle school in the district. The schools ranged in size from 1,120 to 1,900 students and from 74 to 126 staff. Across the five schools, 20% to 40% of students were eligible for free and reduced-priced meals and 40% to 79% of students were African American.

We obtained school records from the five schools, which included 613 teachers who instructed 7,794 classrooms. Of these teachers, 69 of them participated in the aforementioned experimental trial, with attrition to 58 teachers in the follow-up year after the coaching had ended (Table 1; Gregory et al., 2016); teachers were split between the intervention and control conditions. The rest of the 613 teachers had not volunteered to participate in the randomized trial or were not eligible to do so given they did not teach a course with an end-of-course state achievement exam (see online Supplemental Table A1, for more detailed sample description). Among teachers in the experimental trial, we found no group differences between intervention and control teachers' race, gender, years of teaching experience, and attainment of higher degrees (Gregory et al., 2016). Moreover, the groups were evenly distributed across course content areas with 45% teaching English Language Arts, 30% teaching Science, and 25% teaching Mathematics. We did not have access to this sociodemographic information from the nonstudy teachers.

Procedures

All procedures for the current study were approved by our university institutional review board and the participating school district.

MTP-S Randomized Trial. In presentations at school staff meetings, we requested volunteers to participate in an intervention to support teachers in classroom interactions that enhance students' motivation and engagement. Eligible teacher participants instructed a course in which student learning was assessed through end-of-course state achievement exams.

Teachers who consented to participate were stratified within course content area (Language Arts/Social Studies/History vs. Math/Science) and assigned randomly to the MTP-S coaching condition or to a business-as-usual control condition. We then asked teachers to select their most academically and/or behaviorally challenging class. For those assigned to the intervention, this became their focal classroom that was the target of the coaching. For those assigned to the control condition, this became their focal data collection classroom. Teachers repeated the selection of a focal classroom in the second year of the intervention. About 77% of teachers described the academic level of their focal classrooms as "remedial" or "average to below average."

The teachers assigned to the MTP-S condition were paired with a coach who was a veteran teacher. Every 2 weeks, coaches watched teachers' video recordings of their instruction in focal classrooms and conferenced with each teacher to improve teacher-student interactions in that focal classroom (Pianta et al., 2008). See online Supplemental Material for additional details about the intervention. During the 2-year active coaching period, the MTP-S teachers received coaching and the control teachers engaged in whatever professional development they would normally experience. Both MTP-S and control teachers submitted videotapes of their instruction in focal classrooms. After the 2-year coaching period ended, all teachers (intervention and control) were asked to participate in a follow-up study year. For this year, teachers selected another focal classroom and submitted videotapes of their instruction in that classroom.

Additional School Records. Because we wished to expand the sample to include teachers and students who did not elect to participate in the MTP-S randomized trial, we requested data from the district for all teachers and students in the five schools that had taken part in the trial, during the 2-year coaching period and the follow-up year. We received all students' office discipline referral records, achievement test scores, and demographic data, as well as student and teacher course schedules, over this period. The district created pseudo-identifiers to de-identify the records, which enabled us to link the schedule/course enrollment, achievement, and discipline data. Therefore, we were able to access these data for all students at the participating schools, regardless of whether the student was enrolled in a focal class taking part in the randomized trial, and regardless of whether the parent and student had consented/assented to the research in the original trial. Because our primary

research questions examined racial disparities in discipline referrals at the classroom level, we retained the classrooms in which at least 10% of students were African American students *and* 10% of the students were non–African American; this resulted in 7,794 classrooms, taught by a total of 613 teachers.

Measures

Office Discipline Referrals. The records included the date of the referral, the student who was referred, and the teacher who made the referral. Using course enrollment data, we established whether the referred student was enrolled in a course taught by the referring teacher. This established a classroom connected between the referring teacher and the student. It is important to note, however, that the class period in which the referral was issued was not recorded, so it cannot be assured that the referral was received when the student was physically in that referring teacher's classroom. (That is, the teacher could have issued the referral to the student in the hall.)

Office discipline referrals are a meaningful dependent variable given evidence of their concurrent and predictive validity; students' receipt of one or more referrals is associated with negative teacher behavioral ratings (Pas, Bradshaw, & Mitchell, 2011) and, years later, with being off track for graduation (Tobin & Sugai, 1999). Classroom referrals may be a part of the negative events that "snowball," cluminating in disparate life trajectories for African American and White students (McIntosh et al., 2010; Okonofua et al., 2016).

Student and Classroom Characteristics. Given the small number of students coded as Asian, Latinx, Pacific Islander, American Indian, or two or more races, we grouped these students together with White students into a non–African American group (37%), to be compared with the African American students (63%). The decision to cluster the non–African American groups was further supported by patterns in the discipline data: African American students were the only group overrepresented in discipline referrals relative to their enrollment. For more details, see Supplemental Appendix Table A2.

For each class period, we calculated the number of enrolled students who were males, who were eligible for free or reduced-price meals, and who were receiving special education services. We also indicated whether there was a co-teacher instructing that class. Finally, for each classroom we calculated a mean of the statewide subject matter standardized tests for the enrolled students across all course subject areas from the prior school year.

Data Analytic Strategy

Research Question 1. RQ1 tested whether across teachers' course loads, African American students in MTP-S teachers' classrooms received fewer office discipline referrals in comparison with African American students in the classrooms of control teachers and nonstudy teachers during the 2 years of MTP-S coaching and in the year after the intervention ended.

Our dependent variable for RQ1 was a count of the number of African American students in the classroom who received at least one referral from that teacher, which necessitated the use of a Poisson regression model. However, in 54% of classrooms, no African American students received referrals. An excess number of zeros is not uncommon in count data but often leads to a violation of the Poisson regression assumption that the mean of the outcome variable is equivalent to its variance. This was the case in our data, and accordingly, we employed a negative binomial regression model, which adds a gamma-distributed random effect to account for overdispersion in the outcome. This strategy was supported by nested likelihood ratio tests favoring the less parsimonious negative binomial model over the basic Poisson model.

The structure of the data was hierarchical, with whole classrooms, including the focal classroom (Level 1), nested in teachers (Level 2), and nested in 3 study years (Level 3). Because of the small number of years, we utilized a two-level negative binomial model with classrooms (Level 1) nested in teachers (Level 2) and added indicator variables for study year at Level 2. The random intercept at Level 2 was assumed to be normally distributed with a mean of 0 and an estimated variance. Standard errors of parameters were estimated using the Huber-White (or sandwich) estimator of variance in Stata 15.1 (StataCorp, 2017), which are robust to heteroskedasticity and autocorrelation in the Level 1 errors (Rabe-Hesketh & Skrondal, 2012).

At Level 1, the units of which are unique classrooms in a teacher's schedule, we controlled for the number of students who were African American to account for variable African American enrollment size in each classroom when predicting African American referrals. We also controlled for number of students who were male, who received free or reduced-price meals, and who received special education services, as well as for the average student score on the prior year's standardized state achievement test, and whether a co-teacher was assigned to the classroom. Additionally, at the teacher level (Level 2) we included the mean of each of these classroom composition variables across all of a teacher's classrooms over the 3-year period to help account for Level 2 (teacher) endogeneity of Level 1 (classroom) covariates (McNeish & Kelley, 2019; Rabe-Hesketh & Skrondal, 2012). Coefficients are reported as incident rate ratios (IRRs); thus Level 1 coefficients indicate the percent increase or decrease in the expected number/incidence rate of African American student referrals for a classroom that differs by one unit on the Level 1 covariate relative to the other classrooms taught by that same teacher in a given year.

We created a three-category predictor variable for the MTP-S status of each teacher. The reference group (0)

consisted of teachers who received MTP-S coaching, a value of 1 was given to teachers who were part of the control group in the randomized trial, and a value of 2 was given to the nonstudy teachers who were not part of the trial. This variable was treated as categorical and interacted with categorical indicators for study year (0 = first year of coaching, 1 = second year of coaching, 2 = follow-up year). To answer RQ1, the main effect coefficient for MTP-S status tested whether MTP-S coached teachers' referral rates of African American students were equivalent to referral rates of control teachers and nonstudy teachers across the three study years on average. Then, in a second model, we added the interaction term between MTP-S status and study year. In this second model, the coefficient for MTP-S status is a test of whether MTP-S coached teachers' referral rates of African American students were equivalent to the rates of teachers in the comparison groups *in the first year of coaching*. The interaction terms for MTP-S status by study year tested for group differences in referral rates *in the second year of coaching* and *the follow-up year*.

Research Question 2. RQ2 examined whether across teachers' course loads, MTP-S teachers had a smaller discipline referral gap between African American students relative to non–African American students, compared with control teachers and to nonstudy teachers. Our dependent variable was the discipline referral gap, calculated as the number of referred African American students minus the number of referred non–African American students in each classroom. Because this outcome variable was continuous (as opposed to count data), we tested RQ2 using two-level linear random effects models.

As in RQ1, we created a two-level model with intact classrooms (Level 1) nested in teachers (Level 2), including study year as an indicator at Level 2. We added the same Level 1 classroom composition factors as covariates as was done in RQ1 (including the number of African American students in the class) and each variable's mean for a teacher across all their classrooms over the 3 years. Similarly, using the three-category predictor variable indicating the MTP-S status of each teacher as a main effect in one model and in interaction with the variable indicating study year in a second model, we tested whether MTP-S coached teachers showed a smaller discipline referral gap in each of the 2 years of coaching and in the follow-up year.

Supplementary Analyses. Bottiani, Bradshaw, and Gregory (2018) note a "lack of consensus on an efficient, valid, and reliable method of identifying and monitoring disproportionality" (p. 112). For this reason, we conducted additional analyses to better understand our results. First, we conducted similar analyses as those in RQ1 with the outcome variable the number of *non*–African American student referrals. We did this to address the possibility that any potential reduction in a discipline referral gap was due to iatrogenic effects (or increase in referrals) for non–African American students. Second, we conducted sensitivity analyses for RQ1 in which we converted our dependent variable from the number of African American students in the classroom receiving at least one referral to the *percentage* of African American students in the classroom receiving at least one referral. Third, we conducted sensitivity analyses for RQ2 with the dependent variable gaps in referrals calculated as the *percentages* of referred African American students minus the *percentages* of referred non–African American students. These analyses are in the Supplemental Appendix.

TABLE 1
Descriptive Statistics of Classroom Variables Used in Multilevel Modeling

Classroom Characteristic	Mean	Standard Deviation	Intraclass Correlation Coefficient
Number of AA referrals by classroom teacher[a]	1.16	1.87	0.48[a]
Number of non-AA referrals by classroom teacher[b]	0.38	0.88	0.46[b]
Number of AA students	13.05	6.46	0.50
Number of non-AA students	7.87	5.22	0.52
Average prior achievement[c]	433.14	34.69	0.53
Number of male students	10.72	5.57	0.42
Number of low-income students[d]	9.93	5.35	0.45
Number of special education students	2.59	2.71	0.39
Co-teacher assigned	0.15	0.36	0.76

Note. N = 7,794 classroom observations within 588 teachers. AA = African American.
[a]Estimated from an ordinal multilevel model with an ordered categorical variable for AA referrals (0, 1, or 2+). [b]Estimated from an ordinal multilevel model with an ordered categorical variable for non-AA referrals (0, 1, or 2+). [c]Range of achievement scores was 282.78 to 588.33. [d]Low-income students were eligible for free or reduced-priced meals.

Results

Descriptive Statistics

Table 1 shows that the average classroom was composed of more African American (M = 13.05) than non–African American students (M = 6.46). The intraclass correlations (ICC) show the percentage of variance at Level 2 (the teacher level) for each variable. Most ICCs suggest considerable differences between teachers in the sociodemographic composition of the classrooms in their course load (ICCs range from .39 to .53). That is, a teacher tended to instruct courses

with a certain student demographic profile. The ICCs also show some degree of within-teacher consistency in disciplinary practices with 48% of the variance in referrals to African American students at the teacher level and 52% of the variance within teachers at the classroom level. Finally, most of the variance in whether a classroom had a co-teacher was at the teacher level (ICC = .76). This means that those teachers who were paired with a co-teacher tended to have a co-teacher throughout their course schedule.

Also, important to note was that in about half of the classrooms the teacher had not issued an office discipline referral to any enrolled student (3,682 of 7,794 classrooms). In the remaining 4,112 classrooms where the teacher had issued at least one referral, 73% of these referrals went to African American students. In 2,169 of those 4,112 classrooms, all the referrals were issued to African American students. For more details, see Supplemental Appendix Table A3.

Bivariate correlations in Table 2 indicate that more African American students received referrals in classrooms with lower mean-level achievement ($r = -.15$, $p < .001$), and with more male students ($r = .20$, $p < .001$), low-income students ($r = .20$, $p < .001$), and students receiving special education services ($r = .14$, $p < .001$). These correlations should be viewed cautiously as they assume a linear association between the count of African American referrals and other variables whereas the true association is nonlinear, requiring a negative binomial model to appropriately estimate the associations. The classrooms of the MTP-S coached teachers were also less likely to have a co-teacher relative to the control teachers and more likely to have a co-teacher than nonstudy teachers ($r = .08$, $p < .001$ and $r = -.10$, $p < .001$, respectively). When teachers referred more African American students, they also tended to refer more non–African American students ($r = .38$, $p < .001$).

Number of Referrals to African American Students

RQ1 asks whether the MTP-S teacher referrals to African American students enrolled in their classrooms were lower, relative to the control teachers and the nonstudy teachers. These analyses are displayed in Table 3.

Model 1 shows that MTP-S participation was unrelated to the incident rate of referrals to African American students in comparison with the control group (IRR = 0.67, $p > .05$) and the nonstudy group (IRR = 0.72, $p > .05$), averaging across the 3 study years and accounting for number of African American students in the classroom. Model 2 adds the interactions between MTP-S status and study year, to test effects of MTP-S coaching in each year. As evidenced by the MTP-S versus control and MTP-S versus nonstudy teacher "main effect" coefficients in Model 2, in the first year of coaching, there were no differences in the expected number of referrals to African American students in classrooms of MTP-S intervention teachers relative to control teachers (IRR = 0.66, $p > .05$) and nonstudy teachers (IRR = 0.68, $p > .05$). The MTP-S by study year interaction coefficients in Model 2 show whether MTP-S differences in referral incident rates of African American students were observed in the second year of the intervention and the follow-up year. Across the multiple years and comparison groups (i.e., control and nonstudy teachers), no coefficient was significant (IRRs ranging from 0.94 to 1.20, $p > .05$). Notably, the addition of the interaction terms did not improve model fit relative to a model without the interaction terms according to likelihood ratio tests; $\chi^2 = 2.93(4)$, $p = .57$.

Despite not finding effects of MTP-S on this outcome, some classroom characteristics were significant predictors of the number of referrals made to African American students. With the inclusion of the average number of African American students enrolled across teachers' classrooms in our models, the male IRR of 1.03 in Model 1 is interpreted as such: compared with the average number of male students in their other classrooms, having one additional male student in a teacher's classroom was associated with a 3% increase in the number of referrals for African American students in that classroom, adjusting for all the other covariates (IRR = 1.03, $p < .001$). Similar findings held for the number of low-income students (IRR = 1.04, $p < .001$). When teachers had an additional male student in their classroom or low-income student compared with the average number of male or low-income students in their other classrooms, their expected number of referrals to African American students increased between 3% and 4%, adjusting for other variables in the model.

Three Level 1 classroom characteristics were significant predictors of lower African American referral rates (Models 1–2). Relative to the average achievement in their other classrooms, teachers had a slight decrease in the number of referrals issued to African American students enrolled in classrooms where the average achievement was higher (IRR = 0.99, $p < .001$). Noteworthy was that relative to their other classrooms that did not include a co-teacher, when a teacher had a co-teacher in a classroom, the number of referrals issued to African American students enrolled in that classroom declined by 24% (IRR = 0.76, $p < .001$).

Disparities in Referrals

RQ2 asks whether the gap in referrals between African American and non–African American students enrolled in their classrooms were smaller for MTP-S coached teachers, relative to the control teachers and the nonstudy teachers. These analyses are presented in Table 4.

Model 1 shows that differences in referral gaps between MTP-S teachers relative to control ($b = -0.21$, $p > .05$) and nonstudy teachers ($b = -0.14$, $p > .05$), averaging across the 3 study years, were not significant. In Model 2, the first two coefficients test whether MTP-S teachers showed differences

TABLE 2
Correlations Among Discipline Referrals and Classroom Characteristics

	No. of AA Referrals	No. of Non-AA Referrals	Gap in No. of AA and Non-AA Referrals	No. of AA Students	No. of Non-AA Students	Average Prior Achievement	No. of Male Students	No. of Low-Income Students	No. of Special Education Students	Co-teacher (1)	MTP-S (0) vs. Control Teacher (1)	MTP-S (0) vs. Nonstudy Teacher (2)
#AA referral	1											
No. of non-AA referral	.38***	1										
Gap in no. of AA and non-AA referrals	.87***	−.12***	1									
No. of AA students	.27***	0	.29***	1								
No. of non-AA students	.02	.31***	−.14***	.05**	1							
Average prior achievement	−.15***	.04	−.18***	−.20***	.34***	1						
No. of male students	.20***	.16***	.13***	.63***	.47***	−.06***	1					
No. of low-income students	.20***	.07***	.18***	.75***	.19***	−.29***	.57***	1				
No. of special education students	.14***	.09***	.10***	.30***	.09***	−.46***	.38***	.39***	1			
Co-teacher (1)	.06***	0	.06***	.08***	−.06***	−.26***	.12***	.15***	.40***	1		
MTP-S (0) vs. control teacher (1)	.02	0	.02	.06***	.01	−.02	.03	.04	.02	.08***	1	
MTP-S (0) vs. nonstudy teacher (2)	−.05***	−.02	−.05**	−.08***	−.01	.01	−.03	−.06***	−.01	−.10***	−.67***	1

Note. $N = 7{,}794$ classroom observations. AA = African American; MTP-S = My Teaching Partner–Secondary.
*$p < .05$. **$p < .01$. ***$p < .001$; Bonferroni correction employed.

TABLE 3
Multilevel Negative Binomial Models Predicting Number of Referrals to African American Students (Incident Rate Ratios)

	(1)	(2)
	Main Effects	Interaction Model
	IRR (*SE*)	IRR (*SE*)
Level 1 (within teacher within year)		
MTP-S (0) vs. control teacher (1)	0.67 (0.16)	0.66 (0.17)
MTP-S (0) vs. nonstudy teacher (2)	0.72 (0.13)	0.68 (0.14)
School year 2010 (0) vs. 2011	0.82*** (0.04)	0.70** (0.09)
School year 2010 (0) vs. 2012	0.81** (0.05)	0.80 (0.16)
No. of AA students	1.04*** (0.01)	1.04*** (0.01)
No. of non-AA students	0.96*** (0.01)	0.96*** (0.01)
Average prior achievement	0.99*** (0.00)	0.99*** (0.00)
No. of male students	1.03*** (0.01)	1.03*** (0.01)
No. of low-income students	1.04*** (0.01)	1.04*** (0.01)
No. of special education students	0.98* (0.01)	0.98* (0.01)
Classroom with a co-teacher	0.76*** (0.05)	0.76*** (0.05)
Level 2 (between teachers across school)		
Average no. of AA students	1.06* (0.03)	1.06* (0.03)
Average no. of non-AA students	1.17*** (0.02)	1.17*** (0.02)
Average prior achievement	1.00 (0.00)	1.00 (0.00)
Average no. of male students	0.94** (0.02)	0.94* (0.02)
Average no. of low-income students	0.90*** (0.02)	0.90*** (0.02)
Average no. of special education students	1.03 (0.04)	1.03 (0.04)
Percentage of classes with a co-teacher	1.63 (0.43)	1.62 (0.43)
MTP-S × school year interaction		
MTP-S/control teacher × 2011		1.12 (0.18)
MTP-S/control teacher × 2012		0.94 (0.23)
MTP-S/nonstudy teacher × 2011		1.20 (0.16)
MTP-S/nonstudy teacher × 2012		1.02 (0.21)
α (conditional overdispersion)	0.27*** (0.03)	0.27*** (0.03)
Variance (between teacher)	4.11*** (0.50)	4.11*** (0.50)
N	7,794	7,794

Note. AA = African American; MTP-S = My Teaching Partner–Secondary; IRR = incidence rate ratio. Sample includes classrooms with five or more AA students. MTP-S teachers are the reference group and they received MTP-S intervention, 1 = control teachers in original study, 2 = teachers not in original study.
*$p < .05$. **$p < .01$. ***$p < .001$.

in referral gaps relative to control and nonstudy teachers in the first year of the intervention. As in RQ1, we found no differences in the referral gaps in classrooms led by MTP-S intervention teachers relative to control teachers ($b = -0.10$, $p > .05$) and nonstudy teachers ($b = -0.14$, $p > .05$). The interaction terms in Model 2 test for differences for the second year of the intervention and the follow-up year for MTP-S teachers relative to control and nonstudy teachers. Across the multiple years and comparison groups, again, no coefficients were significant (all ranged from −0.28 to 0.14, $p > .05$). The addition of the interaction terms did not improve model fit relative to a model without the interaction terms according to a likelihood ratio test; $\chi^2 = 5.08(4)$, $p = .28$.

Supplementary Analyses

We ran multilevel models examining whether MTP-S coached teachers, relative to control and nonstudy teachers, had (a) different number referrals issued to *non*–African American students (Supplemental Appendix Table A4), (b) lower *percentage* of African American students in the classroom receiving at least one referral (Supplemental Appendix Table A5), (c) smaller racial gaps in referral based on the *percentages* of referred African American students minus the *percentages* of referred non–African American students (Supplemental Appendix Table A6). Similar to results above, receiving the MTP-S coaching versus being in the control condition was not associated with number of referrals issued

TABLE 4
Multilevel Linear Models Predicting Gap in Referrals Between African American Students and Non–African American Students

	(1)	(2)
	Main Effects	Interaction Model
	b (*SE*)	*b* (*SE*)
Level 1 (within teacher within year)		
MTP-S (0) vs. control teacher (1)	−0.21 (0.20)	−0.10 (0.29)
MTP-S (0) vs. nonstudy teacher (2)	−0.14 (0.17)	−0.14 (0.25)
School year 2010 (0) vs. 2011	−0.03 (0.05)	0.11 (0.24)
School year 2010 (0) vs. 2012	0.00 (0.05)	−0.11 (0.24)
No. of AA students	0.05*** (0.01)	0.05*** (0.01)
No. of non-AA students	−0.07*** (0.01)	−0.07*** (0.01)
Average prior achievement	−0.00*** (0.00)	−0.00*** (0.00)
No. of male students	0.02** (0.01)	0.02** (0.01)
No. of low-income students	0.03*** (0.01)	0.03** (0.01)
No. of special education students	−0.02 (0.01)	−0.02 (0.01)
Classroom with a co-teacher	−0.17* (0.07)	−0.18* (0.07)
Level 2 (between teachers across school)		
Average no. of AA students	0.04* (0.02)	0.04* (0.02)
Average no. of non-AA students	0.06*** (0.01)	0.06*** (0.01)
Average prior achievement	−0.00 (0.00)	−0.00 (0.00)
Average no. of male students	−0.03 (0.02)	−0.03 (0.02)
Average no. of low-income students	−0.09*** (0.02)	−0.09*** (0.02)
Average no. of special education students	0.01 (0.04)	0.01 (0.04)
Percentage of classes with a co-teacher	0.60** (0.22)	0.61** (0.22)
MTP-S × school year interaction		
MTP-S/control teacher × 2011		−0.28 (0.28)
MTP-S/control teacher × 2012		−0.07 (0.30)
MTP-S/nonstudy teacher × 2011		−0.13 (0.24)
MTP-S/nonstudy teacher × 2012		0.14 (0.24)
Intercept	2.80** (0.90)	2.78** (0.92)
Variance (within teacher)	0.66** (0.09)	0.66** (0.09)
Variance (between teacher)	1.80*** (0.09)	1.80*** (0.09)
N	7,794	7,794

Note. AA = African American; MTP-S = My Teaching Partner–Secondary. Sample includes classrooms with at least five non-AA students *and* at least five AA students. MTP-S teachers are the reference group and they received MTP-S intervention, 1 = control teachers in original study, 2 = teachers not in original study.
*$p < .05$. **$p < .01$. ***$p < .001$.

to non–African Americans. However, in the first year of the study, receiving the MTP-S coaching versus not being in the randomized trial at all was associated with a lower number of referrals issued to non–African Americans (IRR = 0.60, $p < .05$, Supplemental Appendix Table A3, Model 2). Also, noteworthy was that when a teacher had a co-teacher in a classroom, the number of referrals issued to non–African American students enrolled in that classroom declined by 31% (IRR = 0.69, $p < .001$), relative to their other classrooms that did not contain a co-teacher. In the supplementary analyses, we also found that teachers in the coached condition did not have smaller racial referral gaps (based on differences in percentages referred) relative to the comparison teachers (Supplemental Appendix Table A6).

Discussion

Previous research has demonstrated that secondary school teachers receiving a coaching professional development intervention, MTP-S, reduced their racial gaps in discipline referrals to students in their focal, coached classrooms (Gregory et al., 2016). This was true in focal classrooms during both the active 2-year coaching period and in a follow-up year after coaching ended (Gregory et al., 2016). However,

the current study failed to find evidence of any discipline ripple effects for these findings in all the class periods taught by the same teacher over the school year. That is, compared with teachers randomly assigned to a business-as-usual control condition and to teachers who did not participate in the randomized trial, those receiving MTP-S showed no differences in their racial gaps in disciplinary referrals or the number of referred African American students in their full complement of classrooms during the 2 years of intervention and the postintervention year.

In summary, the anticipated discipline ripple effects of MTP-S from coaching in teachers' self-selected challenging classroom into the diverse classrooms across teachers' instructional schedules did not occur. One possible explanation for the nonsignificant findings in this study is that teachers chose focal classes with the largest academic and/or behavioral challenges for MTP-S coaching. As such, there may have been more room to improve in those focal classrooms, as a result of MTP-S, than in their other classes. Or, teachers may have been more motivated to try new strategies that could lead to improvement in their focal classes because they were unhappy with the status quo. Another explanation relates to differences between the current sample, which used all students, compared with the sample in Gregory et al. (2016), which was restricted to students who had consented to the research procedures. Perhaps research-consented students (and their parents/guardians) tended to have more institutional trust in the school than their non-research-consented peers. Once MTP-S teachers engaged these research-consented African American students, perhaps it was easier to build mutually held positive expectancies (Okonofua et al., 2016). Or, with these students, teachers may have been more receptive to disconfirming information that reduced implicit racial bias. A third possibility, which is discussed in greater detail below and has implications for practice, is that the experience of videotaping themselves heightened teachers' awareness about their behaviors in that classroom and reminded coached teachers to use the strategies they had learned in coaching. This may explain why the positive discipline effect of having received MTP-S coaching persisted in focal classrooms during the follow-up year after coaching was discontinued but was not evident in nonfocal classrooms.

Alternatively, our null findings may simply indicate that coaching in a focal classroom does not lead to widespread shifts across a school day in discipline referral practices with African American students. This underscores other work suggesting that generalization of interventions cannot be assumed and is likely more challenging to achieve than many researchers assume (Kraft et al., 2018). There are numerous implications for practice as detailed below, regarding what changes to interventions may be needed to bolster generalization of effects.

Despite a lack of intervention effects, several findings offer novel contributions to the field. The study was the first to identify considerable within- *and* between-teacher variability in African American referral patterns across classrooms. This suggests that there is more to understand about the classroom-specific versus classroom-independent differences in teachers' disciplinary practices. The study pointed to one classroom-specific finding. Within their instructional schedules, if teachers had a co-teacher (relative to when they did not), they issued fewer discipline referrals to their African American and non–African American students, adjusting for the sociodemographic and achievement differences in classrooms. This finding offers a new direction for understanding classroom processes associated with discipline equity.

Utility of Explicit Generalization Strategies

While recognizing that sampling of focal classrooms or research-consented students may have influenced our lack of results, we also speculate about the possible underlying reasons we did not find a wider generalization effect. As referenced above, coaching heightens the teacher's awareness about enacting the strategies in the setting in which the teacher is coached (i.e., the focal classroom). This level of attention and reminders may be needed, realistically, for teachers to enact behavior change. This would explain why we found MTP-S related effects on increasing equity of discipline referrals in the focal classrooms during the active coaching period. In the follow-up year when coaching was discontinued, we speculate that when MTP-S teachers video-recorded their instruction to send it to the coaches, their coaches' support and guidance from the previous 2 years may have been cued. This reminder of the coaching process may have helped carry forward the prior learnings. Although control teachers also recorded their instruction in focal classrooms during all 3 study years, there was no coaching to be cued.

An implication for practice is that focal and intensive professional development may benefit from incorporating explicit strategies to increase generalization, as opposed to hoping generalization will happen by itself. Perhaps interventions might use low-cost cues or boosters to help teachers recall newly acquired approaches in a different classroom setting. Interventions might also consider using explicit generalization strategies to help teachers consider how they might apply a strategy learned in one course (e.g., Algebra 1) to a different course (e.g., honors geometry). For example, MTP-S coaches could build in reflection questions such as, "How can you apply what we just talked about to the other classes that you teach?"

Equity Indirect and Direct Teacher Professional Development

We also speculate that the equity indirect nature of MTP-S may have limited its generalization. MTP-S coaching focused

on increasing the motivating and engaging quality of instruction within a climate of social and emotional support (Hamre et al., 2013). The primary (and expected) outcomes of the intervention were students' increased academic performance and engagement. As such, greater discipline equity was a downstream effect and not the explicit target of the intervention.

The lack of ripple effects on discipline in the current study may suggest a need for interventions to have a more explicit focus on racial dynamics, including strategizing for how teachers build trust with more historically marginalized students (Yeager et al., 2014) and become aware of how racial implicit bias may affect their decision making (Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014). Indeed, scholars have called for equity explicit or direct interventions that confront dynamics of race, power, and privilege (Carter, Skiba, Arredondo, & Pollock, 2015: Gregory, Skiba, et al., 2017), although it is largely unknown whether such approaches result in teacher behavior change (Forscher, Mitamura, Dix, Cox, & Devine, 2017; Singleton & Linton, 2006). Future research might examine whether coaching programs need a hybrid of equity direct and indirect approaches that offer concrete strategies for teacher behavioral change in the classroom and aim to increase bias awareness and cultural fluency when interacting with marginalized student groups (e.g., Bradshaw et al., 2018).

Schoolwide Equity Initiatives

A third potential reason why our study failed to find generalization of coaching effects on discipline is that focal and intensive professional development may need to occur in the context of schoolwide equity initiatives. The MTP-S intervention was conducted with individual teachers and applied to those teachers' focal classrooms. It was not synchronized with schoolwide equity-oriented or discipline reform initiatives. Without schoolwide shifts in policies, teacher change may be siloed with insufficient reinforcement from administrators or peers. Future research might examine intervention effects when the strengths of a focused and intensive teacher coaching program is paired with widespread school reform efforts (Bradshaw et al., 2018; Fenning & Johnson, 2016; Manassah, Roderick, & Gregory, 2018).

Within- Versus Between-Teacher Differences in Discipline Referrals

The current study found some degree of between-teacher differences in disciplinary practices (48% of the variance in referrals to African American students was between-teacher). At the same time, teachers varied in their referrals across their course load (52% of the variance in referrals to African American students were within-teacher). Variability in instructional practice during the school day has been found in prior studies of elementary school teachers (e.g., Curby et al., 2011). Yet few studies have the strength of our longitudinal data and analyses with rigorous Level 1 within-teacher estimates, meaning that teachers served as their own controls. In addition, our sample of secondary teachers enabled us to examine referrals across shifting classroom contexts given middle and high school teachers instruct different classrooms across the day and year. Thus, we were able to examine the classroom characteristics that predicted variability in disciplinary referrals.

We found that accounting for the achievement level and number of students in special education, when teachers had more low-income or male students enrolled in a given class (relative to the composition of their other classes), they had a higher African American student referral incidence rate. This could reflect teachers simply responding to differences in peer dynamics or student behavior in classrooms with more low-income or male students. However, it also may reflect teachers becoming more reactive or punitive in classrooms with groups that they perceive as more threatening. This speculation draws on the minority threat hypothesis (Welch & Payne, 2010, 2018). Future research might use observational methods to identify the degree to which teachers rely on exclusionary discipline in their classrooms comprising a greater number of students from marginalized groups relative to their other classrooms.

The current study also found that when there was a co-teacher in the classroom, relative to when there was not, teachers' gap in the number of referrals issued to African American versus non–African American was smaller. In their co-taught classrooms, teachers also issued fewer referrals to students from both groups. Given that it is common practice to include a special education teacher as the co-teacher (Friend, 2015; Jackson et al., 2017), it is striking that these findings held when accounting for the number of students in special education. This raises questions about the underlying processes that explain why teachers' referral patterns differ between their classes with versus without co-teachers. Potentially, (a) teachers draw on the behavioral and relational expertise of co-teachers in the room (Gallo-Fox & Scantlebury, 2016), (b) teachers who are struggling with classroom management gain additional adult supervision when they have a co-teacher, and (c) teachers reduce their own stress, dysregulated reaction to student misconduct, or racial implicit bias and vulnerable decision making when they have a co-teacher observing them or supporting them.

Limitations

A number of limitations should be considered. We were unable to test whether Teacher Sensitivity and Analysis and Inquiry (CLASS-S dimensions) improved in all of the MTP-S teachers' classrooms. Thus, we could not examine generalization of gains that were proximal targets of the coaching

program (e.g., student engagement, quality of interactions). This means MTP-S may have generalizing effects that have yet to be identified.

The discipline data also have some limitations. The district did not reliably record the exact time or class period in which the discipline referrals were issued. Thus, it was unknown whether the referral was received during the class period when the student was enrolled in that referring teacher's classroom. This means that some referrals may have occurred outside of the classroom, suggesting that some disciplinary interactions may be only tenuously linked to teachers' instructional approach. Another limitation is that the district had a small number of students identified as Asian, Latinx, American Indian, and Two or more races, which resulted in our grouping them with White students in our analyses. With a more diverse sample, future research should examine the within-teacher discipline patterns separately (Peguero & Shekarkhar 2011; Wallace et al., 2008).

In addition, due to our not being able to link information from our archived MTP-S data (Gregory et al., 2016) to the district-issued classroom data, we were not able to identify the focal classroom in which the MTP-S coaching occurred. This meant that we could not test whether the MTP-S teachers, relative to the control teachers, had eradicated the racial discipline gaps in their focal classrooms using the full enrollment of the class, as opposed to the smaller consented sample in the Gregory et al. (2016) study. This precluded us from the seemingly logical next step of corroborating findings from the prior study using the full focal classroom enrollment. Finally, the sample of teachers in the original randomized trial was small and had attrition over time. To detect intervention effects with small samples, programs must have large effect sizes.

Summary

Despite the urgency of addressing racial disparities in school discipline, few rigorously tested and efficacious programs target the early classroom precursors of students' entry into the discipline system. Many positive outcomes have been associated with the MTP-S program in focal, coached classrooms (e.g., Allen et al., 2015; Mikami et al., 2011), including increased equity in office discipline referrals (Gregory et al., 2016). However, the current study did not find evidence that the promising discipline equity findings in coached classrooms associated with MTP-S occurred across teachers' instructional schedules. Future research is needed to ascertain under what conditions discipline gains from intensive coaching in single classrooms generalize beyond the specificities of the coached context (Kraft et al., 2018). A novel contribution of this study is the identification of within-teacher differences in referral patterns. Having a co-teacher in the classroom was associated with teachers' lower likelihood of referring African American students and smaller racial discipline gaps. This co-teacher finding suggests a promising new direction for research and intervention to increase equity in classroom discipline.

Acknowledgments

This study and its write-up were supported by a grant from the William T. Grant Foundation and draws on data collected through the support of the Institute of Education Sciences (R305A100367). The views in this article do not necessarily reflect policies or recommendations of the funding agencies. The authors are grateful to the teachers and students who participated in this study and the many staff who contributed to this project, including Judith Wasserman, Sharon Deal, Marla Capper, and Kathy Neesen.

References

Allen, J. P., Gregory, A., Mikami, A. Y., Lun, J., Hamre, B., & Pianta, R. C. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the Classroom Assessment Scoring System Secondary. *School Psychology Review*, *42*, 76–98.

Allen, J. P., Hafen, C. A., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2015). Enhancing secondary school instruction and student achievement: Replication and extension of the My Teaching Partner–Secondary Intervention. *Journal of Research on Educational Effectiveness*, *8*, 475–489.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *19*, 1034–1037.

American Academy of Pediatrics. (2013). Out of school suspension and expulsion. *Pediatrics*, *131*, e1000–e1007. Retrieved from http://pediatrics.aappublications.org/content/131/3/e1000.full.pdf+html

Anyon, Y., Gregory, A., Stone, S. I., Farrar, J., Jenson, J. M., McQueen, J., . . . Simmons, J. (2016). Restorative interventions and school discipline sanctions in a large urban school district. *American Education Research Journal*, *53*, 1663–1697.

Anyon, Y., Jenson, J., Altschul, I., Farrar, J., McQueen, J., Greer, E., . . . Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review*, *44*, 379–386.

Bottiani, J. H., Bradshaw, C. P., & Gregory, A. (2018). Nudging the gap: Introduction to the special issue Closing in on Discipline Disproportionality. *School Psychology Review*, *47*, 109–117. doi:10.17105/SPR-2018-0023.V47-2

Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of Black students in office disciplinary referrals. *Journal of Educational Psychology*, *102*, 508–520.

Bradshaw, C. P., Pas, E. T., Debnam, K. J., Bottiani, J., Reinke, W., Herman, K., & Rosenberg, M. S. (2018). Promoting cultural responsivity and student engagement through Double Check Coaching of classroom teachers: An efficacy study. *School Psychology Review*, *47*, 118–134.

Carter, P., Skiba, R., Arredondo, M., & Pollock, M. (2015). *You can't fix what you don't look at: Acknowledging race in addressing racial discipline disparities* (Discipline Disparities Research-to-Practice Briefing Paper Series #4). Retrieved from

http://www.indiana.edu/~atlantic/wp-content/uploads/2014/12/Acknowledging-Race_121514.pdf

Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Downer, J., . . . Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade: Implications for children's experiences and conducting classroom observations. *Elementary School Journal*, *112*, 16–37.

Deakin, J., & Kupchik, A. (2018). Managing behaviour: From exclusion to restorative practices. In J. Deakin, E. Taylor, & A. Kupchik (Eds.), *Palgrave international handbook of school discipline* (pp. 511–527). London, England: Palgrave Macmillan.

Fenning, P., & Johnson, M. B. (2016). Developing prevention-oriented discipline codes of conduct. *Children's Legal Rights Journal*, *36*, 107–136.

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T. L., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology*, *72*, 133–146.

Friend, M. (2015). Welcome to co-teaching 2.0. *Educational Leadership*, *73*(4), 16–22.

Gallo-Fox, J., & Scantlebury, K. (2016). Coteaching as professional development for cooperating teachers. *Teaching and Teacher Education*, *60*, 121–202.

Girvan, E. J., Gion, C., McIntosh, K., & Smolkowski, K. (2017). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School Psychology Quarterly*, *32*, 392–404. doi:10.1037/spq0000178

Goff, P. A., Jackson, M. C., Di Leone, B. A., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing black children. *Journal of Personality and Social Psychology*, *106*, 526–545.

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools*, *51*, 143–163.

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2015). The promise of a teacher professional development program in reducing racial disparity in classroom exclusionary discipline. In D. J. Losen (Ed.), *Closing the discipline gap: Equitable remedies for excessive exclusion* (pp. 166–179). New York, NY: Teachers College Press.

Gregory, A., Hafen, C. A., Ruzek, E. A., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the racial discipline gap in classrooms by changing teacher practice. *School Psychology Review*, *45*, 171–191.

Gregory, A., Ruzek, E., Hafen, C. A., Mikami, A. Y, Allen, J. P., & Pianta, R. C. (2017). My Teaching Partner-Secondary: A video-based coaching model. *Theory Into Practice*, *56*, 38–45. doi:10.1080/00405841.2016.1260402

Gregory, A., Skiba, R. J., & Mediratta, K. (2017). Eliminating disparities in school discipline: A framework for intervention. Special issue on equity in school. *Review of Research in Education*, *41*, 253–278.

Hafen, C. A., Ruzek, E. A., Gregory, A., Allen, J. P., & Mikami, A. Y. (2015). Focusing on teacher-student interactions eliminates the negative impact of students' disruptive behavior on teacher perceptions. *International Journal of Behavioral Development*, *39*, 426–431. doi:10.1177/0165025415579455

Hamre, B., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . Aki Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal*, *113*, 461–487.

Hwang, N. (2018). Suspensions and achievement: Varying links by type, frequency, and subgroup. *Educational Researcher*, *47*, 363–374.

Jackson, K., Willis, K., Giles, L., Lastrapes, R., & Mooney, P. (2017). How to meaningfully incorporate co-teaching into programs for middle school students with emotional and behavioral disorders. *Beyond Behavior*, *26*, 11–18. doi:10.1177/1074295617694408

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*, 547–588. doi:10.3102/0034654318759268

Kuklinski, M., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, *72*, 1554–1578.

Manassah, T., Roderick, T., & Gregory, A. (2018). A promising path toward equity: Restorative circles develop relationships, build communities, and bridge differences. *Learning Forward*, *39*(4), 36–40.

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*, 267–277. doi:10.1016/j.jclinepi.2013.08.015

McCluskey, G. (2018). Restorative approaches in schools: Current practices, future directions. In J. Deakin, E. Taylor, & A. Kupchik (Eds.), *Palgrave international handbook of school discipline* (pp. 573–593). London, England: Palgrave Macmillan.

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

McLeskey, J., Landers, E., Williamson, P., & Hoppey, D. (2012). Are we moving toward educating students with disabilities in less restrictive settings? *Journal of Special Education*, *46*, 131–140. doi:10.1177/0022466910376670

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, *24*, 20–35. doi:10.1037/met0000182

Mikami, A. Y., Gregory, A., Allen, J. P., Pianta, R. C., & Lun, J. (2011). Effects of a teacher professional development intervention on peer relationships in secondary classrooms. *School Psychology Review*, *40*, 367–385.

Mittleman, J. (2018). A downward spiral? Childhood suspension and the path to juvenile arrest. *Sociology of Education*, *91*, 183–204, doi:10.1177/0038040718784603

My Teaching Partner Secondary Consultancy Manual. (2010). [Unpublished manuscript]. University of Virginia, Charlottesville, VA.

Noltemeyer, A., Ward, R. M., & Mcloughlin, C. S. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*, 224–240.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social–psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, *11*, 381–398. doi:10.1177/1745691616635592

Owens, J., & McLanahan, S. (2018). *Unpacking the drivers of racial disparities in school suspension and expulsion* (Working Papers wp18-04-ff). Retrieved from https://ideas.repec.org/p/pri/crcwel/wp18-04-ff.html

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

Peguero, A. A., & Shekarkhar, Z. (2011). 'Latino/a student misbehavior and school punishment. *Hispanic Journal of Behavioral Sciences*, *33*, 54–70.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*, 109–119.

Pianta, R. C., Hamre, B. K., Hayes, N., Mintz, S., & LaParo, K. M. (2008). *Classroom assessment scoring system–Secondary (CLASS-S)*. Charlottesville: University of Virginia.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling using Stata* (Vol. 1). College Station, TX: Stata Press.

Singleton, G., & Linton, C. (2006). *Courageous conversation about race: A field guide for achieving equity in schools*. Thousand Oaks, CA: Corwin.

Skiba, R. J., & Williams, N. T. (2014). *Are Black kids worse? Myths and facts about racial differences in behavior: A summary of the literature*. Equity Project at Indiana University. Retrieved from http://www.indiana.edu/~atlantic/wp-content/uploads/2014/03/African-American-Differential-Behavior_031214.pdf

StataCorp. (2017). *Stata Statistical Software: Release 15*. College Station, TX: Author.

Swain-Bradway, J., Lindstrom Johnson, S., Bradshaw, C., & McIntosh, K. (2017). *What are the economic costs of implementing SWPBIS in comparison to the benefits from reducing suspensions?* Retrieved from https://www.pbis.org/evaluation/evaluation-briefs/economic-costs

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

U.S. Department of Education (2014). *Guiding principles: A resource guide for improving school climate and discipline*. Washington, D.C.

U.S. Department of Education. (2018). *2015-2016 Civil Rights data collection: School climate and safety*. Washington, DC: Author. Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/school-climate-and-safety.pdf

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.

Welch, K., & Payne, A. A. (2010). Racial threat and punitive school discipline. *Social Problems*, *57*, 25–48.

Welch, K., & Payne, A. A. (2018). Latino/a Student Threat and School Disciplinary Policies and Practices. *Sociology of Education*, *91*, 91–110. doi:10.1177/0038040718757720

Yeager, D. S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Brzustoski, P., Master, A., . . . Cohen, G. L. (2014). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *Journal of Experimental Psychology: General*, *143*, 804–824.

Authors

ANNE GREGORY, PhD, Rutgers University. Her work addresses the persistent trend that Black adolescents are issued school suspensions and expulsions at higher rates than adolescents from other racial-ethnic groups. Through research and intervention, she aims to address this trend by strengthening the characteristics of teachers, classrooms, and schools associated with the successful schooling of Black students.

ERIK A. RUZEK, PhD, University of Virginia. His research examines the relations between classroom contexts and students' academic motivation, engagement, and achievement across prekindergarten to high school.

JAMIE DECOSTER, PhD, University of Virginia. His personal research focuses on discovering ways to make the methods practiced by scientists more accurate, flexible, and efficient.

AMORI YEE MIKAMI, PhD, University of British Columbia. Her work focuses on ways in which a supportive classroom or home environment can help children to make friends, with a special interest in designing and evaluating interventions that train teachers or parents in strategies to assist children with peer problems.

JOSEPH P. ALLEN, PhD, University of Virginia. He has performed extensive research in the area of adolescent social development, with an emphasis on using observational techniques to identify the critical facets of adolescents' relationships with others in their social world (both adults and peers). He has also conducted randomized trials of interventions targeting school failure, school dropout, and other social difficulties (e.g., teen pregnancy) among adolescents.