

Closing the Racial Discipline Gap in Classrooms by Changing Teacher Practice

Anne Gregory
Rutgers University

Christopher A. Hafen
University of Virginia

Erik Ruzek
University of Virginia

Amori Yee Mikami
University of British Columbia

Joseph P. Allen
University of Virginia

Robert C. Pianta
University of Virginia

Abstract. Black students are issued school discipline sanctions at rates higher than members of other racial and ethnic groups, underscoring the need for professional development that addresses this gap. In 86 secondary school classrooms, a randomized controlled trial examined the effects of a 2-year teacher-coaching program, My Teaching Partner Secondary (MTP-S). Results from the second year of coaching and from the year after coaching was discontinued replicated previous findings from the first year of coaching—intervention teachers had no significant disparities in discipline referrals between Black students and their classmates, as compared with teachers in the control condition, for whom racial discipline gaps remained. Thus, MTP-S effects were replicated in the second year of coaching and maintained when coaching was withdrawn. Mediation analyses identified mechanisms for these effects; Black students had a low probability of receiving disciplinary referrals with teachers who increased skills to engage students in high-level analysis and inquiry.

Racial disparities in school discipline have garnered recent attention in national reports issued by the U.S. Departments of Education and Justice (U.S. Department of Education, 2014), the Council of State Governments Justice Center (Morgan, Salomon,

This study and its write-up were supported by a grant from the Institute of Education Sciences (R305A100367). The views in this article do not necessarily reflect policies or recommendations of the funding agency. 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.

Correspondence concerning this article should be addressed to Anne Gregory, Rutgers University, Graduate School of Applied and Professional Psychology, 152 Frelinghuysen Rd, Piscataway, NJ 08854; e-mail: annegreg@rutgers.edu

Copyright 2016 by the National Association of School Psychologists, ISSN 0279-6015, eISSN 2372-966x

Plotkin, & Cohen, 2014), and the Discipline Disparities Research to Practice Collaborative (Carter, Fine, & Russell, 2014). Whereas national attention to this problem has increased recently, these disparities (particularly for Black students) have been documented for decades (Children's Defense Fund, 1975). Across the United States, suspension rates of Black students are two to three times higher than those of other racial and ethnic groups (Fabelo et al., 2011; Losen & Martinez, 2013; Wallace, Goodkind, Wallace, & Bachman, 2008). Moreover, rigorous research has documented that Black students remain overrepresented in school discipline sanctions after accounting for their achievement, socioeconomic status, and teacher- and self-reported behavior (e.g., Bradshaw, Mitchell, O'Brennan, & Leaf, 2010; Fabelo et al., 2011; Finn & Servoss, 2015).

Keeping students in the classroom and minimizing their referrals to the office for misconduct could interrupt negative trajectories whereby students receive suspensions, lose instructional time, fall behind on coursework, become discouraged, and ultimately drop out. Recent research has shown each suspension decreases a student's odds of graduating high school by an additional 20% (Balfanz, Byrnes, & Fox, 2015). Moreover, compared with their peers, suspended youth have a higher likelihood of subsequent interactions with the criminal justice system (Shollenberger, 2015). The strategies used to prevent discipline referrals might strengthen positive processes—student engagement and motivation, teacher–student relationships, and attendance—aspects of student experience shown to promote achievement (e.g., Fredricks, Blumenfeld, & Paris, 2004; Roorda, Koomen, Spilt, & Oort, 2011).

Despite the widespread call for prevention programming in school discipline (e.g., U.S. Department of Education, 2014), few experimental studies have demonstrated that teacher professional development can reduce racial disparities in discipline referrals. An exception is a study of the teacher-coaching program My Teaching Partner Secondary

(MTP-S). The first randomized controlled trial (RCT) of MTP-S showed that teacher participation in coaching resulted in student achievement gains (Allen, Pianta, Gregory, Mikami, & Lun, 2011), increased behavioral engagement (Gregory, Allen, Mikami, Hafen, & Pianta, 2014), and improved peer interactions (Mikami, Gregory, Allen, Pianta, & Lun, 2011). The second and more recent RCT with a new sample of teachers and students showed after 1 year of coaching, MTP-S teachers reduced the racial discipline gap between Black students and students from other racial and ethnic groups (Gregory, Allen, Mikami, Hafen, & Pianta, 2015). In control teachers' classrooms, Black students were issued referrals at two times the rate of other groups. The findings held after accounting for a range of student characteristics (gender, prior achievement, low-income status), teacher characteristics (race, years of teaching experience), and classroom characteristics (course content, percentage of Black students).

The current study extends these findings by addressing the degree to which program effects on classroom discipline were (a) replicated in a second year of coaching (with a new group of students) and (b) maintained after coaching had been terminated the following school year (also with a new group of students). Specifically, it addresses the unanswered question about whether benefits of the program continue during MTP-S teachers' second year of coaching when the teachers instruct a new cohort of students and are maintained when the teachers no longer have their coaches in the subsequent school year. The current study also is the first to examine why the MTP-S program reduces the racial disparities, focusing on the role of changes in teachers' observed emotional, instructional, and behavioral supports in the classroom. Identifying the mediating mechanisms of change through theory-driven evaluation is key in advancing an understanding of core ingredients that cut across diverse training programs to improve classroom settings (Mercer, Idler, & Bartfai, 2014).

Table 1. Theoretical Model of CLASS-S

Domain	Dimension	Description
Emotional Support	Positive Climate	The emotional tone of the classroom (e.g., warmth and connection among teachers and students)
	Teacher Sensitivity	The teacher's responsiveness to academic and social-emotional needs of students
	Regard for Adolescent Perspectives	The extent to which the teacher offers leadership, autonomy, and content relevance to students
Classroom Organization	Behavior Management	The teacher's use of effective methods to encourage desirable behavior and redirect misbehavior
	Productivity	The teacher's management of time to maximize instruction
	Negative Climate	The level of expressed negativity (e.g., irritability, frustration, anger)
Instructional Support	Instructional Learning Formats	The teacher's provision of interesting, varied lessons and materials
	Content Understanding	The depth of lesson content and integration of facts, skills, concepts, and principles
	Analysis and Inquiry	The degree to which the teacher facilitates higher level thinking skills, problem solving, and metacognition
	Quality of Feedback	The provision of feedback that expands or extends learning and understanding

Note. CLASS-S = Classroom Assessment Scoring System–Secondary.

RIGOROUS AND SYSTEMATIC TEACHER COACHING

Schools tend to rely on brief, expert-led workshops for teacher professional development; however, there is increasing consensus that such workshops are not likely to effect change in teachers' everyday practice (Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009). Instead, individualized coaching and performance feedback are considered more promising approaches to teacher professional development (e.g., Reinke, Lewis-Palmer, & Merrell, 2008). Typical coaching models include teachers learning new skills and practicing them under the supervision of their coaches (Stormont & Reinke, 2012). Coaches observe the teachers'

utilization of the skills and provide feedback on observed areas of strength and challenge (Stormont & Reinke, 2012).

As coaches observe instruction and provide feedback, their insights need to be anchored in an empirically grounded theory of best practice (Pianta & Hamre, 2009). MTP-S coaches use the Classroom Assessment Scoring System–Secondary (CLASS-S; Pianta, Hamre, Haynes, Mintz, & LaParo, 2008), which was first developed as an observational system to measure teacher–student interactions in a reliable and valid manner (e.g., Allen et al., 2013; Hamre et al., 2013). The CLASS-S is composed of classroom behaviors that fall into three domains (Emotional Support, Classroom Organization, and Instruc-

tional Support), each of which is made up of a handful of dimensions (see Table 1). The CLASS-S has been systematically integrated into the MTP-S coaching process (*My Teaching Partner Secondary Consultancy Manual*, 2010). MTP-S coaches use the CLASS-S to explicitly target the quality of teacher–student interactions and guide teachers in creating emotionally positive, motivating, and cognitively challenging classrooms characterized by sensitivity to students’ socioemotional and academic needs.

POSSIBLE MECHANISMS OF ACTION TO REDUCE DISCIPLINARY REFERRAL

Classrooms are dynamic settings with a host of interactive, mutually reinforcing influences on student behavior. The MTP-S program recognizes this complexity and aims to improve classrooms through strengthened teacher emotional supports (e.g., CLASS-S dimensions such as Positive Climate and Teacher Sensitivity) and instructional supports (e.g., CLASS-S dimensions such as Instructional Learning Formats, as well as Analysis and Inquiry). Improvements on these CLASS-S dimensions may serve as possible mechanisms to explain the program effects. Theory and research suggest that classrooms with strong emotional supports will engender positive teacher–student relationships and cooperative and engaged students (Cornelius-White, 2007; Roorda et al., 2011). With positive, trusting relationships and engaged students, teachers may be more likely to prevent misunderstandings or misconduct in the first place and diffuse it when it arises. Furthermore, when students and teachers trust one another, they may give each other the benefit of the doubt when intentions are not clear. Seeing benign intentions (and not hostile intentions) in classroom behavior may prevent or resolve conflict (Dodge & Frame, 1982).

Instructional supports also matter in preventing discipline referrals. Recent research on the MTP-S program showed teachers who interacted with students using varied instructional formats and problem-solving activities

tended to have more engaged students, as rated by outside observers (e.g., CLASS-S dimensions such as Instructional Learning Formats, as well as Analysis and Inquiry; Gregory et al., 2014). This is corroborated by basic research on classroom process. In a meta-analysis of 119 studies, teachers’ encouragement of higher order thinking ($r = .29$) and learning ($r = .23$) was associated with positive behavioral outcomes (Cornelius-White, 2007).

Focusing teachers on building relationships and providing engaging instruction may be especially important for the welfare of students who are in groups more vulnerable to negative interactions with teachers. One study showed that, across the elementary grades, teachers tend to report less warmth in their relationships with Black students compared with White students (Hughes, 2011). The largest discipline gaps between Black and White students occur for reasons related to “defiance,” “disrespect,” and “uncooperative behavior” (Fabelo et al., 2011; Gregory & Weinstein, 2008; Losen, Martinez, & Okelola, 2014). Numerous underlying reasons have been posited for these trends, including misunderstanding and distrust that can arise because of teachers’ lack of cultural sensitivity or responsiveness (Townsend, 2000; Weinstein, Tomlinson-Clarke, & Curran, 2004), racial tension (Stevenson, 2014), implicit bias (Skiba et al., 2014), or perceived racial threat (Welch & Payne, 2010). In addition, by the time some Black students reach secondary school, their histories of discipline and negative exchanges with teachers may set up patterns of interactions that result in office discipline referrals. In a sense, emotionally and instructionally supportive teachers may be a “breath of fresh air” for many Black students.

SOCIAL EQUITY RESEARCH IN CLASSROOMS

Four issues need to be considered when conducting rigorous, quantitative research on social equity. First, when examining racial disparities in classrooms, scholars need to include numerous covariates in statistical models to help pinpoint the link between a stu-

dent's race and the outcome. In school discipline research, covariates include students' gender, achievement level, and low-income status (e.g., Balfanz et al., 2015; Petras, Masyn, Buckley, Ialongo, & Kellam, 2011; Losen & Martinez, 2013; Wu, Pink, Crain, & Moles, 1982). Second, classroom research using individual student outcomes (in this case, likelihood of discipline referral) should analyze data that disaggregate individual student- and teacher-level variance using multilevel modeling. Such modeling helps to identify teacher effects while accounting for variability at the student level (Bryk & Raudenbush, 1992).

Third, as noted elsewhere, research needs to consider absolute levels of outcomes and socially equitable outcomes (Gregory, Cornell, & Fan, 2011; Lee & Bryk, 1989). It would not be desirable if interventions closed racial gaps by dampening the positive outcomes of some student groups so that they match the negative outcomes of other student groups. As such, researchers need to consider whether interventions reduce the discipline gap by improving the outcomes of the most adversely affected groups (in this case, Black students) and at the same time do not worsen the already positive outcomes of groups with low discipline referral rates. This is particularly relevant to research on school discipline given that discipline referrals can be rare with some groups, indicating a possible "floor effect." For student groups with low base rates of referral at the start of an intervention, there may be minimal room for improvement.

Fourth, research on programs that reduce disparities might also consider whether the program goals and content are more or less equity implicit or explicit. Equity-explicit teacher professional development programs unequivocally forefront the aim of reducing historical disparities. For example, in the Double-Check program, teachers reflect with their coaches about the cultural proficiency of their interactions with diverse students and aim to improve the outcomes of underserved students (Bottiani et al., 2012; Hershfeldt et al., 2009). In equity-explicit teacher professional development programs, the aim of reducing historical disparities is unequivocally at the fore-

front. In other words, the goal to reduce disparities is not necessarily at the forefront in the delivery of the intervention but it is a desirable outcome. For example, many social and emotional interventions train teachers to implement curricula with all of their students in the classroom (e.g., Jones, Brown, & Aber, 2011). The delivery of the universally administered curricula can result in reduced disparities given the larger effects on students at risk of later difficulties (Jones et al., 2011).

PURPOSE OF THE CURRENT STUDY

The current study extends previous findings by testing whether the positive effects of the MTP-S teacher professional development program on discipline referrals in the first year of coaching continue in a second year of coaching (with a new group of students) and are maintained with another new group of students the year after coaching support is removed. A previous study using an experimental design showed 1 year of teacher coaching resulted in the closing of the gap in office discipline referrals between Black students and students from all other racial and ethnic groups whereas the gap was maintained in the control teachers' classrooms (Gregory et al., 2015). The aim of the current study is to examine whether the positive effects will continue in teachers' second year of coaching when they instruct new students, as well as whether the effects will be maintained in the year after the intervention stops (again with a new group of students) when the coaches are no longer available. If this is the case, it would suggest durable and robust shifts in teacher behavior—shifts that could possibly be sustained as teachers encounter new cohorts of students in the future.

The study also examined another unknown, namely, why MTP-S results in teachers' lowered use of office discipline referrals. The underlying theory of the program suggests classrooms characterized by high emotional and instructional supports would have trusting teacher–student relationships and engaged, on-task students, both of which are likely to prevent negative interactions that culminate in office discipline referrals. Specifically, we ex-

amed dimensions of the CLASS-S that may mediate the link between the program and its effect on disciplinary practices.

METHOD

Teachers from five middle and high schools were recruited to participate in the intervention study. The schools were located in a district in Virginia serving a predominantly low- to middle-income, ethnic minority community. The median household income for the schools' catchment area was \$35,000 to \$49,999. Participating schools had a sizable enrollment (ranging from 1,120 to 1,900 students), and 71% of enrolled students were from racial-ethnic minority groups. The 4-year graduation rate for the schools was 80.5%, which was significantly lower than the comparable statewide average. The pass rate for the Virginia Standards of Learning examinations was 83% for English and 60% for mathematics, also significantly below statewide averages. At the time of the research, none of the schools were implementing whole-school discipline initiatives such as multitiered systems of supports.

The research study was presented to teachers in the spring prior to the academic year in which the intervention commenced via presentations at faculty meetings held at the schools. The teachers were told that the study aimed to learn more about how to best support teachers in classroom interactions, lessons, and activities that enhance motivation and engagement of their students. Thus, they were unaware that we were examining whether the intervention led to reduced use of office discipline referrals. To meet study inclusion criteria, teachers were required to work in a secondary school, agree to randomization, and be the primary instructor of a course that had an end-of-course standardized examination to assess student learning. Teachers provided written consent, and study procedures were approved by a university institutional review board. A majority of focal courses (approximately 77%) were described by the teachers as "remedial" or "average to below average" in their academic level. Parents of students in the

focal courses were invited to provide written consent, and students were also asked to provide written assent. In the second year of the intervention, 64% of students were consented to participate.

Teachers were stratified within grade level (high school versus middle school) and within course content area (language arts, social studies, and history versus math and science). Then, they were assigned randomly to 2 years of MTP-S coaching or to a control group that received business-as-usual professional development (50% probability of being assigned to each condition). All teachers also agreed to 1 year of follow-up after coaching discontinued (the postintervention year, i.e., Year 3).

Participants

During initial spring recruitment, 97 teachers were selected to participate in the study. Of these, 86 completed 2 years of the study. The Year 2 teacher sample included 86 secondary school teachers (30 men and 56 women) from five schools who participated for 2 years. Participating teachers had an average of 9.4 years of teaching experience ($SD = 6.5$). The teacher racial-ethnic composition was 56% White, 33% Black, 7% mixed ethnicity, 1% Asian, 1% Latino, and 2% other. Twenty percent of teachers had a terminal BA degree, and 80% had advanced education beyond the BA degree. Teachers in this study were primarily at the high school level (89%), with fewer at the middle school level (11%). Randomization was effective in producing equivalent samples given the absence of statistically significant differences between the intervention-group and control-group teachers on any of the aforementioned teacher characteristics—a finding that held each of the 3 study years.

In Year 2 of the study, the student sample ($n = 1,195$) participated in the focal classrooms instructed by teachers in their second year of MTP-S coaching or the control condition. The student sample was composed of 52% female students and, by race and ethnicity, was 58% Black, 31% White, 9% Latino, and 2% Asian American. Furthermore, 37% of students qualified for free and reduced-price lunch and 13% of

students fell below proficiency on their prior year's achievement test. On average, we consented 14 students in each classroom, with an average of 8 Black students and 6 students from other racial-ethnic groups in each classroom.

In Year 3 of the study, the student sample ($n = 1,163$) was composed of 53% female students and, by race and ethnicity, was 59% Black, 29% White, 9% Latino, and 3% Asian American. Similar to Year 2, 41% of students qualified for free and reduced-price lunch and 16% of students fell below proficiency on their prior year's achievement test. On average, we consented 15 students in each classroom, with an average of 9 Black students and 6 students from other racial-ethnic groups in each classroom.

MTP-S Coaching Cycles

My Teaching Partner (MTP) was originally developed for prekindergarten and early elementary classrooms (MTP Pre-K; Pianta et al., 2003; Pianta, Mashburn, Downer, Hamre, & Justice, 2008). The secondary version of the program (MTP-S) shares similar intervention procedures and strategies with the MTP Pre-K program. Both programs offer teachers ongoing, personalized coaching and feedback. MTP-S coaches focus teachers' attention on each of the CLASS-S dimensions through "coaching cycles" held across the school year.

Each coaching cycle was composed of five steps (Pianta, Hamre, et al., 2008): In Step 1, the teacher video-records instruction from his or her focal classroom. Splicing the video, the coach isolates illustrative examples of one or more dimensions of the CLASS-S. The coach submits a "nice work" clip and "consider this" clip, in which the interactions reflect a CLASS-S dimension upon which the teacher needs to improve. In Step 2, the coach sends the clips back to the teacher along with written prompts, in which the coach clearly describes the observed teacher-student interaction in CLASS-S terms, along with how the teacher's interaction with the student directly affects student responses. The prompts are also intended to promote teacher skills in observation and self-reflection. In Step 3, the teacher reviews the clips and responds in writ-

ing to the prompts. In Step 4, the teacher and the coach meet via telephone or computer for a conference to discuss the clips, written prompts, and responses. In Step 5, the coach summarizes the conference in writing including an action plan for future improvement.

The coach integrates considerations about CLASS-S dimensions throughout the coaching cycle. For example, the coach uses "prompts" to spur teachers' reflection on a CLASS-S dimension as it relates to their behavior. A coach's "consider this" prompt might read as follows:

At the high end of the dimension of Analysis and Inquiry, the teacher promotes his students' use of higher-level thinking, and he keeps the focus on his students doing the thinking. In this clip, you present an interesting story that has some challenging issues. As you view this clip, what do you hear your students say? What else might you have asked them that would have pushed them to do even more thinking?

Coaching cycles conclude with summaries of the discussion and action plans to implement new behaviors reflective of a CLASS-S dimension in upcoming instruction. A coach's written conference summary would explicitly refer to teacher behaviors as they relate to one or more CLASS-S dimensions. For instance, a coach might highlight teacher behaviors that led to students having an opportunity to play an integral role in class activities (CLASS-S dimension Regard for Adolescent Perspectives). The coach might point out that by giving students opportunities for leadership and autonomy, the students appeared to take a high level of responsibility for their own learning. The coach might suggest that in upcoming instruction the teacher ask students to apply their thinking to a real-world situation. The coach might also point out teacher behaviors that helped set a warm, "up-beat" tone in the classroom (CLASS-S dimension Positive Climate, Gregory et al., 2014). As part of the coaching process, the coach also has the option of encouraging the teacher to access a video library on the MTP-S website to learn more about a given CLASS-S dimension. The video clips are drawn from authentic classrooms and include written explanations

that elucidate the teacher practices related to a given CLASS-S dimension.

According to Dane and Schneider's (1998) five guidelines, MTP-S was implemented with fidelity in the second year of coaching. First, exposure to the program was high for teachers assigned to MTP-S, with 100% of teachers attending the workshop and teacher completion of an average of 12 coaching cycles ($SD = 5$; range = 2–19) and viewing of an average of 16 clips of video exemplars ($SD = 12$; range = 0–48) on the MTP-S website. Second, adherence to treatment was demonstrated by the coaching content being similar to that in the manual, measured by 86% of teachers reporting that they agreed they had learned something new about the CLASS-S dimensions after the coaching cycles. Third, 100% of MTP-S teachers said the coaching cycle was “worth the time it took,” an indicator of high participant responsiveness. Fourth, 95% of the teachers agreed that the cycles were “productive,” an indicator of good quality of delivery. Fifth, regarding diffusion of MTP-related experiences to control teachers, the program was set up such that control teachers were given no access to the coaching video library or the coaches. In fact, a majority of control teachers (>52%) did not receive even routine mentoring.

Measures

We collected a range of teacher, student, and classroom measures. Teachers completed surveys, the district provided student records, and observers coded videotaped instruction.

Characteristics of Teachers

At the start of the intervention, teachers reported their sociodemographic characteristics, including their race and years of teaching experience. Such characteristics are important to include as covariates in statistical analyses to determine if the effects of the professional development held across varying teacher groups. Moreover, covarying teacher characteristics is needed given prior research. A handful of studies have shown that Black teachers tend to perceive Black students in a more positive light compared with White

teachers (Downey & Pribesh, 2004; Pigott & Cowen, 2000; Zimmerman, Khoury, Vega, Gil, & Warheit, 1995). In addition, one study found female teachers and teachers with fewer years of experience tended to see more negative interactions among students compared with their male or more experienced colleagues (Gregory et al., 2010).

Characteristics of Students and Classrooms

Covariates also included a range of student and classroom characteristics. School records were used to identify the gender and race or ethnicity of consented students. Records also indicated whether students came from low-income families (coded based on student eligibility for free and reduced-price lunch, which is offered to families with incomes up to 185% of the federal poverty line). Given the link between achievement and school discipline (Balfanz et al., 2015), we also covaried a measure of students' prior achievement. Specifically, we obtained consented students' scores on the Virginia Standards of Learning end-of-course examinations, which are validated, standardized achievement tests given across the state (Commonwealth of Virginia, 2005). We selected the prior examination result from a course subject most comparable to the subject taught in the focal classroom in the study. The prior year's performance in a comparable course was highly correlated with performance on the end-of-course examination in the focal classroom ($r = .77, p < .001$) and thus was considered appropriate to use as a baseline level of student achievement.

The current study also covaried the course content area to test whether the effect of MTP-S on office discipline referrals held for teachers no matter their course subject area given that past research has shown MTP-S benefitted students regardless of the classroom subject area (Allen et al., 2011). We grouped focal classrooms across two broad content areas (e.g., English, history, and social studies versus math and science). Finally, we calculated the percentage of Black students in the classroom to use as a covariate given findings that racial composition has been linked to rates

of suspension (Gregory et al., 2011; Skiba et al., 2014; Welch & Payne, 2010).

Office Discipline Referrals

We obtained school records of the participating students' receipt of office discipline referrals in the teachers' focal classrooms. Teachers issued referrals mostly for reasons related to disrespect, disruption, and fighting or bullying that, according to the school code of conduct, fell into Severity Categories 1 through 3, reflecting minor to more serious infractions. No referrals were issued for Category 4 reasons (e.g., assault).

Observations of Teachers and Classrooms

Intervention and control teachers submitted videotaped instruction, which was then CLASS-S coded by coders shielded to teachers' condition in the study. For the purposes of this study, analyses used coding of one 40- to 60-min video-recording of instruction from the first months of teachers' fall courses (before the first cycle of coaching in the first year of the intervention) and one from the last months of their instruction in the spring of the second year. Each teacher's videotaped instruction was divided into two 20-min segments. Each segment was assigned randomly to two coders (two segments \times two coders per segment = four sets of CLASS-S scores). Coders rated each of the CLASS-S dimensions on a 7-point scale placing the rating in the low range (1, 2), midrange (3, 4, 5), or high range (6, 7). For the fall and spring, their four scores were then averaged to maximize the reliability of observation scores (Raudenbush, Martinez, Bloom, Zhu, & Lin, 2008). The reliability of the coding was tested using intraclass correlation coefficients (ICCs), which were in the good to excellent range (ranging from .62 to .77), based on Cicchetti and Sparrow's (1981) standards for interpreting ICCs. Interrater agreement was acceptable—codes based on the same observations were within 1 point of each other 80.3% of the time.

Initial psychometric work on the CLASS-S found links to academic outcomes (Allen et al., 2013), which allowed us to identify which specific dimensions appeared to be the most pow-

erful and salient aspects of the system to use in analyses going forward. We used these a priori identified scales in the current analyses given the strong links between discipline and academic outcomes (Noltemeyer, Ward, & McLoughlin, 2015) and the desire for a parsimonious approach to analyses using a modest number of dimensions. Thus, five CLASS-S dimensions were examined as mediators, despite the coaches' use of all CLASS-S dimensions in their feedback to teachers.

The five CLASS-S dimensions used in these analyses comprised three dimensions of the Emotional Support domain, i.e., Positive Climate (respectful and warm communications, shared positive affect), Teacher Sensitivity (teacher responsiveness to student needs), and Regard for Adolescent Perspectives (opportunities for students' active, leadership roles and exposure to relevant course content); as well as two dimensions from the Instructional Support domain, i.e., Instructional Learning Formats (varied use of instructional modalities and strategies) and Analysis and Inquiry (engagement in higher order thinking skills and novel application of knowledge; see Table 1). The selection of the five dimensions was also theoretically justified given the link between the positive affective quality of teacher–student relationships and student behavior (Cornelius-White, 2007; Roorda et al., 2011) and between cognitively stimulating and engaging instruction and student behavior (Cornelius-White, 2007; National Research Council, 2005).

At the start of the study, there were no significant differences between control teachers and intervention teachers on any of the five CLASS-S dimensions. In addition, teachers were not told that their submitted videotapes were being CLASS-S coded. Despite the coaches using the framework of the CLASS-S, the MTP-S teachers never received any CLASS-S ratings (1–7).

Data Analytic Plan

We undertook “intent-to-treat” analyses, which prioritize the randomization process and ignore nonadherence—a degree of which is to be expected in all intervention studies

(Hollis & Campbell, 1999). In other words, we compared all teachers randomly assigned to each condition, without excluding those who did not meet expected levels of participation. This analytic approach incorporates teachers with variable levels of participation in the intervention, as incomplete or imperfect adherence is likely to occur in real-world program implementation.

Program Effects

A series of statistical models were run to address whether the MTP-S intervention reduced the racial discipline gap in the second year of the intervention and the postintervention year. We used hierarchical generalized linear modeling (HGLM), which accounts for nested data with dichotomous outcomes (O'Connell, Goldstein, Rogers, & Peng, 2008). Specifically, we used two-level HGLM models (students nested in classrooms), after confirming there was minimal between-school variability in the number of referrals.

We created a dichotomous outcome for each student (no office discipline referral from the participating teacher = 0 and one or more disciplinary referrals = 1). The decision to dichotomize referral data was based on the small sample size and ease of interpreting results using odds ratios. That said, we recognize that analyzing the referral data in their continuous form is also informative as it indicates whether teachers issued fewer referrals to the same students. This is important to examine given that Black students tend to have high rates of chronic exposure to harsh discipline (Skiba, Michael, Nardo, & Peterson, 2002) and students with two or more referrals are at higher risk of receiving future referrals than their peers (McIntosh, Frank, & Spaulding, 2010). Thus, we also reran the HGLM models, described later, using referral data in their continuous form using a robust estimator to address nonnormality.

The HGLM models covaried student, teacher, and classroom characteristics, including the percentage of Black students in the classroom, course subject area, teacher race, and years of teaching experience, as well as student gender, low-income status, and prior

achievement. We also included whether the student was Black (1) or not (0). It is important to note that we decided to examine Black student receipt of a disciplinary referral compared with all other students, which combined White, Latino, and Asian student groups. This decision was based on (a) the small percentage of Latinos (9%) and Asians (2%) in the sample and (b) neither group having disproportionately high rates of in- or out-of-school suspension in the study schools (U.S. Department of Education, 2013).

In the HGLM models, we then included teacher status in the intervention or control condition of MTP-S and a random intercept at the classroom level. Finally, we examined a cross-level interaction—specifically, whether teachers' program condition (intervention or control) was a significant predictor of the slope of the association between student race and the likelihood of disciplinary referral (the dependent variable). This tested whether the intervention or control condition of the teacher moderated the link between student race and disciplinary referral. In other words, we examined whether the probability of a Black student versus a non-Black student being given a disciplinary referral was less in the intervention teachers' classrooms compared with the control teachers' classrooms.

Mediational Analyses

We then conducted a series of analyses to identify promising mediators of the MTP-S program effects for Black students. We ran separate multilevel models examining the link between each of the five aforementioned CLASS-S dimensions (Positive Climate, Teacher Sensitivity, Regard for Adolescent Perspectives, Instructional Learning Formats, and Analysis and Inquiry) and the likelihood a Black student received one or more disciplinary referrals. Similar to the analyses described earlier, we accounted for teacher race, course subject area, students' achievement at the start of the course, and students' low-income status. We entered the fall observed rating of the CLASS-S from Year 1, which was based on video-recorded instruction before the first cycle of coaching in the first year of the inter-

vention. This provided a true baseline for each teacher. We then entered the proposed mediators into the models—the ratings of the CLASS-S dimension from the last months of instruction in the spring of Year 2 of the program. The analyses identified the CLASS-S dimensions most likely to be mediating mechanisms. Using HGLM, we then conducted formal mediational analyses and tested the significance of indirect paths using the MODEL INDIRECT option in Mplus (Muthén & Muthén, 2010). We used bootstrapping procedures to determine the proper confidence intervals (CIs) around the indirect effects. Finally, to eliminate any potential overweighting of individual students' disciplinary patterns in classrooms with few consented students, we excluded classrooms that had fewer than five consented students (three classrooms). The findings remained similar; thus we presented results using all the classrooms.

Attrition and Missing Data

Of the 97 teacher participants in Year 1, 86 completed both years of the intervention. Regarding the 11 teachers not available by the end of the second year of the study, virtually all attrition was a result of factors unrelated to program participation.¹ Formal attrition analyses of Year 2 indicated no differences in levels of attrition across treatment and control groups. Moreover, there were no significant differences in teacher characteristics (e.g., years of teaching experience, gender, education level, or racial–ethnic minority group) between those who did participate in the final evaluation and those who did not. Furthermore, in the postintervention year (Year 3) after the coaching ended, an additional seven teachers did not continue with the study.² Attrition analyses of Year 3 did not indicate any differences between these seven teachers and the teachers who provided data.

Our handling of missing data is also important to note. In total, 86 teachers were included in the analyses for Year 2 of the intervention, and 79 were included in the postintervention analyses. There were no missing student referral data in either year. Of

the 86 teachers in Year 2 of the intervention, 4 did not provide an eligible CLASS-S observation for the fall and spring. Another four teachers did not provide a fall observation, and five teachers did not provide a spring observation. Given the missing data and given that there were no detectable differences between those who did provide observational data and those who did not, we used full information maximum likelihood methods through Mplus software version 6 (Muthén & Muthén, 2010), which enabled us to conduct analyses with a more complete dataset. Assumptions that missingness was completely random and not based on characteristics of the teacher were met: We found no statistical differences in the use of discipline referrals and the CLASS-S dimensions for teachers who did or did not submit video in the spring. This finding suggests that the probability of missing data in the spring was not a function of the values of the data in the fall. In addition, for ease of interpretation of the interaction term estimates, all variables in the HGLM models were grand mean standardized with a mean of zero and a standard deviation of one for all analyses (Aiken & West, 1991).

RESULTS

During the entire second year of the intervention, teachers in the MTP-S group issued between 0 and 8 referrals ($M = 0.95$ referrals) to participating students in their focal classroom. That same year, teachers in the control group issued 0 to 12 referrals ($M = 2.21$ referrals) to participating students in their focal classroom.

Consistent and Maintained Effects of MTP-S

Results from the statistical models answered whether, after accounting for teacher, student, and classroom covariates, the intervention condition (MTP-S teacher versus control teacher) predicted the slope of association between the student ethnicity variable (Black versus non-Black student) and disciplinary referral (a test of moderation). As depicted in Table 2, the significant interaction (Student

Table 2. MTP-S Intervention and Disciplinary Referral in Year 2 of Intervention

	β^b	SE	95% CI
Main effects ^a			
Teacher race (0 = Black, 1 = not Black)	-.01	.02	[-.038, .022]
Years teaching	.00	.01	[-.002, .003]
Subject (0 = math or science, 1 = English or humanities)	.00	.02	[-.032, .035]
Percentage of Black students in classroom	.04	.03	[-.007, .063]
Student gender (0 = female, 1 = male)	.03*	.01	[.003, .062]
Student prior achievement	-.00	.00	[-.004, .002]
Student free and reduced-price lunch (0 = not qualified, 1 = qualified)	.02	.02	[-.019, .048]
Student race (0 = not Black, 1 = Black)	.04**	.01	[.004, .069]
MTP-S intervention group (0 = control, 1 = intervention)	-.07**	.02	[-.079, -.013]
Interaction: Student race \times MTP-S intervention group ^c	-.06**	.02	[-.107, -.032]

Note. CI = confidence interval; MTP-S = My Teaching Partner Secondary.

^aN = 86 teachers, with 44 MTP-S and 42 control; N = 1,195 students, with 659 MTP-S and 536 control.

^bStudent outcome is defined as a disciplinary referral by the teacher (0 = no referral; 1 = one or more disciplinary referrals). The estimates are for predictors when they were entered as a block.

^cThe estimate is a cross-level interaction term in hierarchical generalized linear modeling.

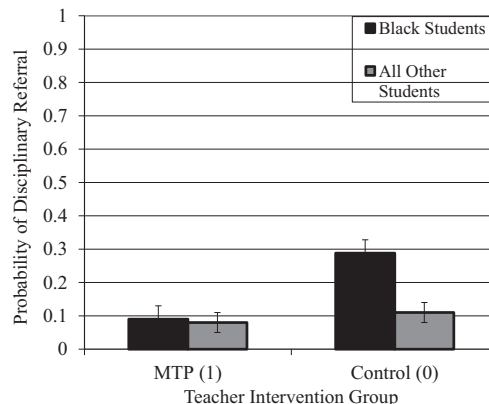
*p < .05. **p < .01.

race \times MTP-S intervention group) suggested that the effect of the intervention on teacher use of discipline referrals varied by race. A probing of the interaction showed that intervention teachers, overall, had a lower use of disciplinary referrals, especially with Black students ($d = 0.24$). The reduction for non-Black students was not quite large enough to reach statistical significance ($d = 0.11$). In contrast to teachers not receiving MTP-S coaching, those who were coached had no racial discipline gap in their classrooms, whereas Black students in the control teachers' classrooms were over two times more likely to be issued a referral compared with non-Black students (see Figure 1). In addition, there appeared to be no detrimental disciplinary effects of the intervention on non-Black students given that their referral rate was similar to their peers taught by control teachers.

Results from the postintervention year (once coaching was withdrawn) also indicated that the probability of a Black student versus a non-Black student being given a disciplinary referral was less in the intervention teachers'

classrooms compared with the control teachers' classrooms (see Table 3). A probing of the interaction showed that intervention teachers, overall, had a lower use of disciplinary refer-

Figure 1. Probability of Disciplinary Referral in Year 2



Note. The probability of disciplinary referral as a function of student race by intervention group is shown for Year 2 of the teacher-coaching program My Teaching Partner Secondary (MTP-S). The probability estimates were hand calculated based on the unstandardized coefficients in Table 2.

Table 3. MTP-S Intervention and Disciplinary Referral in Postintervention Year

	β^b	SE	95% CI
Main effects ^a			
Teacher race (0 = Black, 1 = not Black)	-.02	.02	[-.041, .033]
Years teaching	.00	.01	[-.004, .003]
Subject (0 = math or science, 1 = English or humanities)	.00	.01	[-.024, .025]
Percentage of Black students in classroom	.05	.03	[-.008, .059]
Student gender (0 = female, 1 = male)	.03*	.01	[.001, .058]
Student prior achievement	-.01	.01	[-.017, .011]
Student free and reduced-price lunch (0 = not qualified, 1 = qualified)	.02	.02	[-.011, .049]
Student race (0 = not Black, 1 = Black)	.04**	.01	[.003, .067]
MTP-S intervention group (0 = control, 1 = intervention)	-.05**	.02	[-.074, -.011]
Interaction: Student race \times MTP-S intervention group ^c	-.06**	.03	[-.092, -.009]

Note. CI = confidence interval; MTP-S = My Teaching Partner Secondary.

^a $n = 79$ teachers, with 40 MTP-S and 39 control; $n = 1,163$ students, with 630 MTP-S and 533 control.

^bStudent outcome is defined as a disciplinary referral by the teacher (0 = no referral; 1 = one or more disciplinary referrals). The estimates are for predictors when they were entered as a block.

^cThe estimate is a cross-level interaction term in hierarchical generalized linear modeling.

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

als, especially with Black students ($d = 0.21$). The reduction for non-Black students was not quite large enough to reach statistical significance ($d = 0.08$). Similar to the effects in Year 2 of the intervention, teachers who had received the intervention had no evidence of a racial discipline gap in their classrooms whereas Black students in the control teachers' classrooms were over two times more likely to be issued a referral compared with their peers.

We reran the HGLM models using the referral data in their continuous form and found the same pattern of statistically significant results. This suggests that the MTP-S teachers did not simply reduce their referral use with students who might have received only one referral anyway. Instead, these additional analyses suggest the MTP-S teachers tended not to issue multiple referrals to Black students, thereby reducing their repeated exposure to exclusionary sanctions.

Explaining the Effects of MTP-S on Black Disciplinary Referrals

We then examined whether changes in the CLASS-S dimensions from the start of the

intervention to the end of Year 2 predicted the likelihood Black students received one or more disciplinary referrals (see Table 4). None of the student, teacher, and classroom covariates were significantly associated with likelihood of referral when the CLASS-S dimensions were in the models. Moreover, we found that only two of the CLASS-S dimensions (Teacher Sensitivity, as well as Analysis and Inquiry) were linked to likelihood of disciplinary referral during the second year of MTP-S. More specifically, accounting for student, classroom, and teacher covariates, in classrooms where teachers showed greater improvement in Teacher Sensitivity and in Analysis and Inquiry across 2 years of the study, Black students were less likely to be issued a disciplinary referral than their peers in classrooms where teachers showed less improvement on these two CLASS-S dimensions.

In Mplus, we then tested whether Teacher Sensitivity and Analysis and Inquiry accounted for the MTP-S program effect on Black disciplinary referrals. We tested the significance of the indirect paths using the MODEL INDIRECT option. As Figure 2

Table 4. CLASS-S Dimensions and Disciplinary Referrals

	Predicting Classroom Discipline Referrals, β (SE) ^c				
	PC	TS	RAP	ILF	AI
Step 1 ^{a,b} : Year 1 fall observation	-.03* (.01)	-.02 (.02)	-.04* (.02)	-.01 (.02)	-.01 (.02)
Step 2: Year 2 spring observation	-.01 (.02)	-.04* (.02)	-.03 (.02)	-.03 (.02)	-.06* (.03)

Note. AI = Analysis and Inquiry; CLASS-S = Classroom Assessment Scoring System–Secondary; ILF = Instructional Learning Formats; MLM = Multilevel Models; PC = Positive Climate; RAP = Regard for Adolescent Perspectives; TS = Teacher Sensitivity.

^aRegression coefficients are from each step of the MLMs.

^bAll models accounted for teacher race, course subject, student’s low-income status, and student’s achievement at the start of the course.

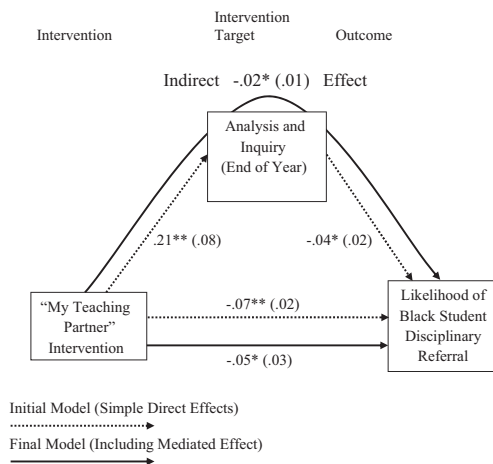
^cBeta estimates are taken from each step.

* $p < .05$.

shows, accounting for covariates, the Analysis and Inquiry dimension partially mediated the effects of MTP-S on likelihood of Black student referral. The findings suggest that the intervention helped teachers increase the degree to which they incorporated higher level thinking and hypothesis generation into their

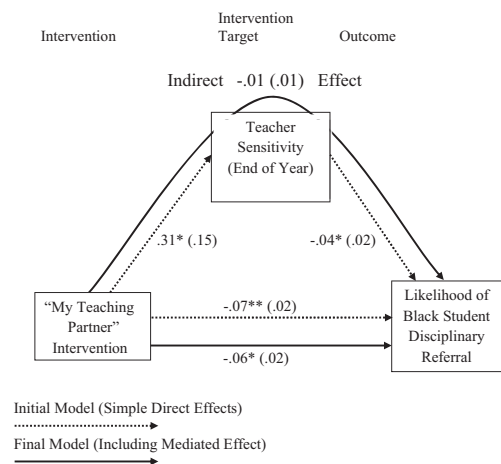
classrooms. This change in instructional practice was accompanied by a lower use of disciplinary referrals with Black students (Indirect Effects, IND = $-.02$, 95% CI [$-.039$, $-.011$]). As Figure 3 shows, Teacher Sensitivity was not statistically confirmed as a partial mediator of the effects of MTP-S on likeli-

Figure 2. Analysis and Inquiry as a Mediator



Note. Mediation of the effects of My Teaching Partner Secondary (MTP-S) on the Year 2 likelihood of black student disciplinary referral is shown for the Analysis and Inquiry dimension. Covariates included teacher’s race, class subject, student’s low-income status, and student’s achievement at the start of the course. * $p < .05$. ** $p < .01$.

Figure 3. Teacher Sensitivity as a Mediator



Note. Mediation of the effects of My Teaching Partner Secondary (MTP-S) on the Year 2 likelihood of black student disciplinary referral is shown for the Teacher Sensitivity dimension. Covariates included teacher’s race, class subject, student’s low-income status, and student’s achievement at the start of the course. * $p < .05$. ** $p < .01$.

hood of Black student referral given no significant indirect effect (IND = $-.01$, 95% CI $[-.022, .011]$), despite the significant link between MTP-S and changes in this CLASS-S dimension ($\beta = .31$, $p < .05$). Similarly, for teachers' discipline referrals in Year 3 (the postintervention year), Analysis and Inquiry was the only CLASS-S dimension that partially mediated the effect of MTP-S on likelihood of Black student referral (IND = $-.02$, 95% CI $[-.036, -.008]$).

DISCUSSION

The current study indicates consistency and maintenance of effects of a teacher-coaching program in reducing the racial discipline gap. Results showed intervention effects in the second year of coaching were consistent with effects shown in the first year of coaching (Gregory et al., 2015) and were maintained the subsequent year when coaching was discontinued. The findings suggest that effects were not transient but were robust as teachers instructed new cohorts of students in the second year of the program and in the year after the program was ended. The findings also provide new knowledge about underlying program processes that can help explain why MTP-S contributed to reducing the racial discipline gap. From the start to the end of the coaching 2 years later, improved instruction in the areas of problem solving and higher level thinking helped explain the program effects.

Understanding the Program Effects on Teacher Practices

Improvements in teacher instruction, as defined by the CLASS-S, were found to partially mediate the program effects—analyses suggest that the program greatly closed the racial discipline gap in the intervention teachers' classrooms when teachers made strides in the Analysis and Inquiry dimension. As a caveat, given the correlational nature of the mediational analyses, no claims of causation can be made. That said, the findings are intriguing and worthy of additional consideration. The degree to which teachers were observed as facilitating higher level thinking skills, prob-

lem solving, and metacognition was significantly linked to their equitable and infrequent use of discipline referrals. In classrooms observed as high on Analysis and Inquiry, students appeared to carry the "cognitive load." Teachers created opportunities for students to evaluate, synthesize, or engage in challenging problems. Examples include teachers (a) asking students to formulate the history of an imaginary country based on its geography, (b) supporting small groups in developing a plan to reduce pollutants from entering the stream behind the school, (c) facilitating student experiments to test how plants respond to different environmental stimuli, and (d) helping students apply existing knowledge to new applications such as writing out word problems based on their solved mathematical equations (Pianta, Hamre, et al., 2008).

The findings confirm prior research on characteristics of high-quality instruction. Studies have shown that students are more engaged when they have cognitively challenging tasks (Stodolsky, 1988) and opportunities to solve meaningful problems (Newmann, Wehlage, & Lamborn, 1992). In fact, adolescents tend to learn more when they are asked to demonstrate higher order thinking skills (not simply memorizing discrete facts) and apply knowledge and procedures to new problems (National Research Council, 2005). Engaged students tend to be perceived as cooperative (National Research Council, 2004).

MTP-S and Racial Disparities in Classroom Referrals

The robust effects across multiple years of coaching and when the coaching was discontinued strengthen claims about program effectiveness in reducing racial disparities in classroom disciplinary referrals (Gregory et al., 2015). Noteworthy is that MTP-S resulted in specific gains for a vulnerable group. Thus, the study demonstrated its potential as an equity-oriented intervention. This is especially intriguing given its past evidence producing student-wide gains in achievement and engagement in diverse classrooms (Allen et al., 2011). It suggests that the very same pro-

gram can have targeted effects on a vulnerable student group (thereby reducing disparities) while also raising competencies in other domains for all student groups. In other words, a single program can have both differential effects (e.g., greater impacts on vulnerable groups) and widespread impacts depending on the student outcome. This is especially interesting given that the program is not equity explicit in its mission—for example, it does not explicitly focus on raising teacher consciousness about implicit bias or institutional racism. Moreover, it does not draw teachers' attention in a targeted manner to the quality of interactions they have with Black students in particular; rather, it focuses on skills in effectively interacting with any student. This focus on attunement to individual student needs and provision of engaging instruction to all suggests MTP-S is equity implicit in its mission. The aim is to improve the quality of interactions across all students. As the field continues to develop programs that aim to reduce discipline disparities, it may be useful to understand the extent to which equity-explicit approaches may in fact be relevant for strengthening positive outcomes and processes for all teachers and students and when equity-implicit approaches have greater effects on vulnerable populations as a consequence of improving basic fundamental classroom processes.

The mediational analyses found that instruction characterized by high-level problem solving might be particularly beneficial for Black students. This issue is worthy of attention, particularly because so many attempts to address the discipline gap presume that perceived classroom misbehavior is best addressed by improved classroom management; the present study suggests a different path. Cognitively demanding instruction might be experienced as somewhat novel by Black students in lower level classrooms composed of predominantly low-achieving students where they can be subjected to understimulating "sheet work" (Oakes, 1985). The creative problem-solving tasks themselves may elicit Black students' active engagement, which prevents negative interactions that culminate in disciplinary referrals. It would be informative

to identify whether teachers engaging students in problem solving tended to use a group format and peer-mediated learning. If this were found to be the case, it would be relevant to identify whether Black students need greater access to this form of learning, which is beneficial to all students (Cohen, 1994; Frisby, 2013), or whether some Black students, given the heterogeneity of the group (O'Connor, Lewis, and Mueller, 2007), experience group work as culturally congruent in its communal orientation to learning (e.g., American Psychological Association Task Force, 2008).

Alternatively, classrooms high on Analysis and Inquiry may reflect mutually held positive beliefs and attributions between teachers and Black students. Teachers who participated in the MTP-S coaching and created opportunities for independent problem solving may have learned to view their Black students in a positive manner and defy unconsciously held negative racial stereotypes of them as less academically capable, as academically disinclined, or as prone to aggression (e.g., Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014). Given the opportunity to engage in cognitively demanding problem-solving tasks, Black students may detect their teachers' high expectations and confidence in them as scholars. Expectancy research has shown that students live up to the positive beliefs through a self-fulfilling prophecy (Kuklinski & Weinstein, 2001). Further research is needed to identify processes that can help explain why classrooms with cognitively demanding activities shift the disciplinary patterns between teachers and their Black students.

Findings regarding the Teacher Sensitivity dimension were equivocal. On the one hand, the overall mediated effect was not statistically reliable. However, each of the two paths of the model (from MTP-S to Teacher Sensitivity and from Teacher Sensitivity to disciplinary referral) was significant. In other words, teachers' responsiveness to students' social and emotional needs improved through the program and was also related to a lower likelihood of Black student referral. These paths had enough unreliability that when com-

bined, the overall mediated path was not statistically reliable. From this, we cannot conclude that we have established mediation; however, results suggest that further consideration of this variable, perhaps with a larger sample size, may well be warranted. Furthermore, greater precision may be needed in measurement to clarify the role of Teacher Sensitivity. The CLASS-S ratings were based on observations across all students; thus, they were not sensitive to differential treatment (or improvements) across student groups (Weinstein, 2008). In other words, the class average or global nature of the observational codes may have obscured improvements in teachers' emotionally attuned relationships with Black students. It might be the case that teachers who improved in their sensitivity to Black students' social and emotional needs, in particular, may have issued them few disciplinary referrals. Given the research on the protective role of warm and close teacher relationships for Black students (Gregory & Ripski, 2008; Meehan, Hughes, & Cavell, 2003; Sabol & Pianta, 2012), further research is needed to confirm whether improving how teachers respond to Black students' social and emotional needs is a potential remedy for discipline disparities.

Limitations and Future Research

A handful of limitations should be considered when interpreting the study's findings. The MTP-S program had no significant positive or negative effect on discipline for students who were not Black. This is likely because of their low rates of discipline referral and a floor effect with little room for their improvement. Future research might examine processes that occur with more regularity across all racial and ethnic groups to determine if the program has more subtle behavior-related effects. For example, observers might code, by racial group, cooperative teacher-student exchanges and students' on-task behavior. This would help determine whether the program results in improved quality of interactions across all racial groups. It would be especially fruitful to ascertain whether stu-

dents detect shifts in teacher practices across racial groups through student surveys and interviews. In addition, future research might include more than two observation points to strengthen claims that the sampled teacher practices adequately reflected somewhat stable or consistent behavior (Mashburn, Meyer, Allen, & Pianta, 2014).

Corroborative research is needed given that we were unable to rule out differences in referral rates between intervention and control teachers prior to the start of the intervention. We did not have access to student discipline records before the start of the intervention. However, it is likely the case that teachers in each study condition were similar on this one indicator (use of discipline referrals) given that random assignment effectively equated them on all other measured individual or classroom characteristics (e.g., years teaching, percentage of students eligible for free and reduced-price lunch). That said, future intervention research should account for teachers' prior referral patterns.

Corroborative research is also needed given the small sample of students in some of the classrooms, which limited a more nuanced analysis differentiating program effects by varying reasons for discipline referral. It would be informative to identify whether the MTP-S program resulted in fewer referrals in categories in which Black students have been historically overrepresented (e.g., defiance, disruption, insubordination)—categories that have been described as subjective and thus more prone to implicit bias (Skiba et al., 2014). Finally, given the small sample size and demographic characteristics of the students, the analyses were not able to parse out the potential of the program to address other well-documented disparities, including the disproportionate sanctioning of students in special education, American Indian students, and Latino students (Fabelo et al., 2011; Wallace et al., 2008).

Summary

The current study confirmed the durable effect of a teacher professional development

program in reducing the racial discipline gap in classrooms. It found that teachers who participated in a second year of MTP-S coaching had no significant gap between Black students and their peers in receiving disciplinary referrals—a gap present in the control teachers’ classrooms. The more equitable use of referrals was maintained by MTP-S teachers the year after the coaching was discontinued. Furthermore, the findings held when accounting for a range of student risk factors including students’ low achievement, male gender, and low-income status and teacher characteristics including teacher race and years of teaching experience. The study also showed that the quality of instruction mattered for disciplinary outcomes. Teachers who improved in using high-level problem solving in their classrooms tended to issue referrals at low and comparable rates across student racial and ethnic groups. This suggests MTP-S, via teacher training, is likely exposing students to rigorous, engaging curricula and to high expectations for engagement and achievement. A proactive, prevention-oriented approach to discipline, therefore, is a means to reduce racial disproportionality in exclusionary discipline.

FOOTNOTES

¹ Of the 11 teachers who left the study by the end of the second year, 3 had retired, 3 had moved out of the district, 3 were no longer teaching classes with end-of-year achievement tests, and 2 stopped participation prior to the beginning of the second year of the intervention (thus not identifying a target class for the evaluation).

² The attrition in the third year of the study was a result of three teachers moving out of the district, two teachers moving into nonteaching roles, one teacher retiring, and one teacher moving to an elementary school.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications.

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

American Psychological Association Task Force. (2008). *Resiliency and strength in Black children and adolescents*. Washington, DC: American Psychological Association. Retrieved from <http://www.apa.org/pi/families/resources/task-force/resilience-af-am.aspx>

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

Bottiani, J. H., Bradshaw, C. P., Rosenberg, M.S., Hershfeldt, P.A., Pell, K. L., & Debnam, K. J. (2012). Applying Double Check to response to intervention: Culturally responsive practices for students with learning disabilities. *Insights on learning disabilities: From Prevailing Theories to validated Practices, 9*(1), 93–107.

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

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.

Carter, P., Fine, M., & Russell, S. (2014). *Discipline disparities series: Overview*. Bloomington, IN: The Equity Project at Indiana University. Retrieved from <http://rtppcollaborative.indiana.edu/briefing-papers/>

Children’s Defense Fund. (1975). *School suspensions: Are they helping children?* Cambridge, MA: Washington Research Project.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127–137.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1–35.

Commonwealth of Virginia. (2005). *The Virginia standards of learning: Technical report*. Richmond, VA: Author.

Cornelius-White, J. (2007). Learner-centered teacher–student relationships are effective: A meta-analysis. *Review of Educational Research, 77*, 113–143.

Dane, A. V., & Schneider, B. A. (1998). Program integrity in providing primary and early secondary prevention. Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.

Darling-Hammond, L., Chung Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and Abroad*. Dallas, TX: National Staff Development Council. Retrieved from www.nsdc.org

Dodge, K. A., & Frame, C. L. (1982). Social cognitive biases and deficits in aggressive boys. *Child Development, 53*, 629–635.

Downey, D., & Pribesh, S. (2004). When race matters: Teachers’ evaluations of students’ classroom behavior. *Sociology of Education, 77*, 267–282.

- 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 Justice Center Publications. Retrieved from <http://justicecenter.csg.org/resources/juveniles>
- Finn, J. D., & Servoss, T. J. (2015). Security measures and discipline in American high schools. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 44–58). New York, NY: Teachers College Press.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109.
- Frisby, C. (2013). *Meeting the psychoeducational needs of minority students*. Hoboken, NJ: John Wiley & Sons.
- 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., Mikami, A., Hafen, C., & Pianta, R. (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., 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., 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., & Ripski, M. (2008). Adolescent trust in teachers: Implications for behavior in the high school classroom. *School Psychology Review, 37*, 337–353.
- 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.
- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*, 461–487.
- Hershfeldt, P. A., Sechrest, R., Pell, K. L., Rosenberg, M. S., Bradshaw, C. P., & Leaf, P. J. (2009). Double-Check: A framework of cultural responsiveness applied to classroom behavior. *Teaching Exceptional Children Plus, 6*(2), 1–18.
- Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomized controlled trials. *British Medical Journal, 319*, 670–674.
- Hughes, J. N. (2011). Longitudinal effects of teacher and student perceptions of teacher–student relationship qualities on academic adjustment. *Elementary School Journal, 112*, 38–60.
- Jones, S. M., Brown, J. L., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82*, 533–554.
- Kuklinski, M., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development, 72*, 1554–1578.
- Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*, 172–192.
- Losen, D. J., Martinez, T., & Okelola, V. (2014). *Keeping California's kids in school*. Los Angeles, CA: University of California, Los Angeles.
- Losen, D. J., & Martinez, T. E. (2013). *Out of school and off track: The overuse of suspensions in American middle and high schools*. Los Angeles, CA: University of California, Los Angeles.
- Mashburn, A. J., Meyer, J. P., Allen, J. P., & Pianta, R. C. (2014). The effect of observation length and presentation order on the reliability and validity of an observational measure of teaching quality. *Educational and Psychological Measurement, 74*, 400–422.
- 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.
- Meehan, B. T., Hughes, J. N., & Cavell, T. A. (2003). Teacher–student relationships as compensatory resources for aggressive children. *Child Development, 74*, 1145–1157.
- Mercer, S. H., Idler, A. M., & Bartfai, J. M. (2014). Theory-driven evaluation in school psychology intervention research: 2007–2012. *School Psychology Review, 43*, 119–131.
- Mikami, A., 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.
- Morgan, E., Salomon, N., Plotkin, M., & Cohen, R. (2014). *The school discipline consensus report: Strategies from the field to keep students engaged in school and out of the juvenile justice system*. New York, NY: Council of State Governments Justice Center.
- Muthén, L. K., & Muthén, B. (2010). *Mplus: Statistical analysis with latent variables, user's guide*. Los Angeles, LA: Author.
- My Teaching Partner Secondary Consultancy Manual*. (2010). Unpublished manuscript, University of Virginia, Charlottesville, VA.
- National Research Council. (2004). *Engaging schools: Fostering high school students' motivation to learn*. Washington, DC: National Academies Press.
- National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom*. Washington, DC: National Academies Press.
- Newmann, F. M., Wehlage, G. G., & Lamborn, S. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achieve-*

- ment in American secondary schools (pp. 11–39). New York, NY: Teachers College Press.
- 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.
- O’Connell, A. A., Goldstein, J., Rogers, J., & Peng, C. J. (2008). Multilevel logistic models for dichotomous and ordinal data. In A. A. O’Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 199–242). Charlotte, NC: Information Age.
- O’Connor, C., Lewis, A., & Mueller, J. (2007). Researching “black” educational experiences and outcomes: Theoretical and methodological considerations. *Educational Researcher, 36*, 541–552.
- Petras, H., Masyn, K. E., Buckley, J. A., Ialongo, N. S., & Kellam, S. (2011). Who is most at risk for school removal? A multi-level discrete-time survival analysis of individual- and context-level influences. *Journal of Educational Psychology, 103*, 223–237.
- 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, VA: University of Virginia.
- Pianta, R. C., Kinzie, M., Justice, L., Pullen, P., Fan, X., & Lloyd, J. (2003). *Web training: Pre-K teachers, literacy, and relationships. Effectiveness of early childhood program, curricula, and interventions*. Washington, DC: National Institute of Child Health and Human Development.
- Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431–451.
- Pigott, R. L., & Cowen, E. L. (2000). Teacher race, child race, racial congruence, and teacher ratings of children’s school adjustment. *Journal of School Psychology, 38*, 177–196.
- Raudenbush, S. W., Martinez, A., Bloom, H., Zhu, P., & Lin, F. (2008). *An eight-step paradigm for studying the reliability of group-level measures*. Report from the William T. Grant Foundation. Retrieved from http://www.wtgrantfoundation.org/publications_and_reports/featured_reports
- Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*, 315–332.
- Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students’ school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81*, 493–529.
- Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher-child relationships. *Attachment & Human Development, 14*, 213–231.
- Shollenberger, T. L. (2015). Racial disparities in school suspension and subsequent outcomes: Evidence from the National Longitudinal Survey of Youth. In D. J. Losen (Ed.), *Closing the school discipline gap: Equitable remedies for excessive exclusion* (pp. 31–43). New York, NY: Teacher’s College Press.
- Skiba, R. J., Chung, C., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal, 51*, 640–670.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and disproportionality in school punishment. *Urban Review, 34*, 317–342.
- Stevenson, H. C. (2014). *Promoting racial literacy in schools: Differences that make a difference*. New York: Teachers College Press.
- Stodolsky, S. S. (1988). *The subject matters: Classroom activity in math and social studies*. Chicago, IL: University of Chicago Press.
- Stormont, M., & Reinke, W. M. (2012). Using coaching to support classroom-level adoption and use of interventions within school-wide positive behavioral interventions and support systems. *Beyond Behavior, 21*, 11–19.
- Townsend, B. L. (2000). The disproportionate discipline of African American learners: Suspensions and expulsions. *Exceptional Children, 66*, 381–391.
- U.S. Department of Education. (2013). *Office for Civil Rights: Civil rights data collection*. Retrieved from <http://ocrdata.ed.gov/DistrictSchoolSearch>
- U.S. Department of Education. (2014). *Guiding principles: A resource guide for improving school climate and discipline*. Washington, DC: Author.
- 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, C. S., Tomlinson-Clarke, S., & Curran, M. (2004). Toward a conception of culturally responsive classroom management. *Journal of Teacher Education, 55*, 25–38.
- Weinstein, R. (2008). Schools that actualize high expectations for all youth: Theory for setting change and setting creation. In M. Shinn & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 81–101). Oxford: Oxford University Press.
- Welch, K., & Payne, A. (2010). Racial threat and punitive school discipline. *Social Problems, 57*, 25–48.
- Wu, S. C., Pink, W. T., Crain, R. L., & Moles, O. (1982). Student suspension: A critical reappraisal. *The Urban Review, 14*, 245–303.
- Zimmerman, R. S., Khoury, E. L., Vega, W. A., Gil, A. G., & Warheit, G. J. (1995). Teacher and parent perceptions of behavior problems among a sample of African American, Hispanic, and non-Hispanic white students. *American Journal of Community Psychology, 23*, 181–197.

Date Received: August 29, 2014

Date Accepted: May 31, 2015

Associate Editor: Lisa Bowman-Perrott ■

Anne Gregory is currently an associate professor at Rutgers, The State University of New Jersey. She received her PhD from the University of California at Berkeley. 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.

Christopher A. Hafen is currently an assistant professor of psychology at Northern Virginia Community College. He received his PhD from Florida Atlantic University and was the project manager for the My Teaching Partner project from 2010–2015 at the University of Virginia. His research interests include understanding the function and development of close relationships in adolescence.

Erik Ruzek is currently a research scientist at the Center for Advanced Study of Teaching and Learning at the University of Virginia. He received his PhD in education from the University of California, Irvine in 2012. His research examines the relations between classroom contexts and students' academic motivation, engagement, and achievement across prekindergarten to high school.

Amori Yee Mikami is currently an associate professor at the University of British Columbia. She received her PhD from the University of California at Berkeley. 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 is currently the Hugh P. Kelly Professor of Psychology at the 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.

Robert C. Pianta is currently Dean and Novartis Professor of Education for the Curry School of Education at the University of Virginia. He received his PhD from the University of Minnesota. His work focuses on policy and practice that enhance children's outcomes, school readiness, and later achievement. He has served on the steering committee and was a study investigator for the National Institute of Child Health and Human Development Early Child Care Research Network.