Running Head: COACHING IN PAIRS VERSUS INDIVIDUALLY

Is It More Effective or Efficient to Coach Teachers in Pairs or Individually?: A Comparison of Teacher and Student Outcomes and Coaching Costs

Elise T. Pas[1]

Chelsea A. K. Duran[2]

Katrina J. Debnam[3,2]

Catherine P. Bradshaw[2]

[1]Johns Hopkins University, Bloomberg School of Public Health

[2]University of Virginia, School of Education and Human Development

[3]University of Virginia, School of Nursing

Pas, E. T., Kaihoi, C. A., Debnam, K. J., & Bradshaw, C. P. (2022). Is it more effective or efficient to coach teachers in pairs or individually? A comparison of teacher and student outcomes and coaching costs. *Journal of School Psychology*, *92*, 346-359.

ACKNOWLEDGEMENT: Support for this article comes from the Institute of Education Sciences, U.S. Department of Education (Grant #s: R305A150221 and R324A110107) and the Spencer Foundation awarded to the University of Virginia (PI: C. Bradshaw). The opinions expressed here are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or the Spencer Foundation. We would like to thank the study coaches, Sandy Rouiller, Lana Bates, Sheilah Myers, and Lauren Kaiser, as well as the research assistants who supported data collection. We also thank Yifan Zhu for her consultation on the cost estimates.

Abstract

A growing body of research documents the positive impacts of teacher coaching, but research contrasting the effectiveness of different coaching approaches is limited. This study contrasted paired coaching – delivered to two teachers simultaneously – with traditional coaching for individual in-service teachers. We examined the effects of these two approaches on observations of teachers' classroom management practices and student behavioral outcomes, relative to non-coached conditions, over the course of a single school year and at a 1-year follow-up. We also explored the relative time and cost efficiency of the two approaches. Two hundred fifty-two teachers from 18 elementary and middle schools participated in the randomized controlled trial. Hierarchical linear modeling was used to account for repeated measures nested within teachers and teachers within schools. Combined effects indicated improved global ratings by observers of teacher behavior management ($\Delta = .29$) after a multiple comparison adjustment. Paired coaching was less effective than individual coaching at improving some observed student behaviors, though these did not remain significant after multiple comparison adjustments. Neither model demonstrated sustained effects after one year. Although the paired coaching was significantly more efficient for coach time, it represented just a modest overall cost savings per school. Results indicate a need for more research to identify feasible coaching approaches yielding sustainable effects.

KEYWORDS: randomized controlled trial; culturally responsive classroom management; sustainability of coaching impacts; time logs**;** cost analysis

Is It More Effective or Efficient to Coach Teachers in Pairs or Individually? A Comparison of Teacher and Student Outcomes and Coaching Costs

Tailored, data-driven coaching shows promise for enhancing in-service teachers' knowledge and skills through modeling, observation, and problem-solving (Denton & Hasbrouck, 2009; Kraft et al., 2018; Pianta et al., 2008). Recent studies have demonstrated that coaching improves teacher dosage (Pas, Bradshaw et al., 2015) and quality of implementation of evidence-based practices and interventions (Sutherland et al., 2015; also see Kraft et al., 2018), as well as instructional practices (Bradshaw et al., 2018; Neuman & Cunningham, 2009; Reinke et al., 2008). Coaching also supports improvements in student outcomes including reduced discipline referrals of Black students (Bradshaw et al., 2018; Gregory et al., 2016), reduced student disruption (Reinke et al., 2008), and improved academic achievement (Kraft et al., 2018). Although these findings are indeed promising, much coaching research has focused exclusively on individual coaching. The feasibility and sustainability of this resource-intensive approach as a standard school practice is questionable (Cappella et al., 2012) and effectiveness in real-world practice has fallen short of effects detected in efficacy trials (Kraft et al., 2018).

The current study aimed to address these gaps by contrasting two delivery approaches for a coaching model provided by an external (i.e., research team-hired) coach to teachers in two arms of a randomized controlled trial (RCT). We compared traditional *individual coaching* (i.e., one teacher) to *paired coaching*, whereby the coach worked with two teachers simultaneously. Both focused on behavior management using an adapted version of the evidence-based Classroom Check-Up coaching model (CCU; Reinke et al., 2011), which, in this study, also focused on culturally responsive behavior management (i.e., Double Check coaching; Bradshaw et al., 2018). Prior to this study, the Classroom Check-Up had only been utilized to coach

individual teachers and had generally demonstrated positive effects on teacher classroom management and disruptive student behavior (Bradshaw et al., 2018; Reinke et al., 2008). The current study builds on prior CCU research by focusing on two implementation approaches to coaching, in comparison to no coaching, to determine impacts on teachers' use of behavior management strategies and student behavioral outcomes as assessed by external observers. The focus on sustainability, teacher and coach time efficiency, and relative effectiveness of the two coaching delivery methods is a unique feature of this study (Kraft et al., 2018).

Effectiveness of In-Service Coaching

The teacher coaching literature has grown considerably in both volume and sophistication since the early work of Showers and Joyce (1996) demonstrated its promise relative to traditional professional development. As noted earlier, extant research demonstrates effects across a range of outcomes, including improved dosage (Pas, Bradshaw et al., 2015) and quality of implementation of evidence-based practices and interventions (Sutherland et al., 2015; also see Kraft et al., 2018), instructional practices (Bradshaw et al., 2018; Neuman & Cunningham, 2009; Reinke et al., 2008), and student behavior and discipline referrals (Bradshaw et al., 2018; Gregory et al., 2016; Reinke et al., 2008). For example, a recent meta-analysis indicated an average effect size (Cohen's d) of coaching of 0.49 on instruction and 0.18 on student achievement (Kraft et al., 2018). Notably, this research base is for individual coaching.

The specific mechanisms of change involved in the coaching process have received limited attention (see Johnson et al., 2016). It is theorized that coaching impacts teachers by (a) leveraging social interactions to promote modeling (Vygotsky, 1978), (b) encouraging reflection on teaching practices (Garet et al., 2001), and (c) providing the opportunity to map out implementation intentions, an important precursor for behavioral change (Gollwitzer & Sheeran,

2006; Gregory et al., 2017). Toward that end, the available coaching research suggests that tailored support facilitates changes in teacher practice (Showers & Joyce, 1996), including performance feedback based on classroom observations (e.g., Reinke et al., 2008).

The Promise and Challenges of Two Teacher Coaching Models

Although the impact of one-on-one coaching has been well established, effectiveness in real-world practice has fallen short of effects detected in efficacy trials (Kraft et al., 2018) and feasibility and capacity within schools to one-on-one coaching is limited (Cappella et al., 2012). Coaching two teachers at once should be more time efficient, at least from the coach perspective, and therefore, more feasible (Allen et al., 2011; Capella et al., 2012; Pianta et al., 2008). Unfortunately, extant research on coaching two teachers simultaneously has weak causal evidence due to non-experimental designs, small sample sizes, and/or lack of a focus on key outcomes. For these reasons, these studies were excluded from the Kraft and colleagues (2018) coaching review.

Whereas much of the pre-service literature has focused on a supervisor or trainer coaching teachers in pairs or small groups (e.g., Miller et al., 1991), most in-service two-teacher coaching has used a peer coaching approach, where two teachers coach one another without a third-party coach (e.g., Scheeler et al., 2010). We are not aware of any research focused on paired coaching delivered by a third-party coach, as we examined in the current study. Extant in-service research of peer coaching has often used single-subject designs and has typically paired co-teachers (e.g., Ottley et al., 2017; Scheeler et al., 2010; Strother, 1989). Furthermore, findings from these studies have been mixed. Some research has shown promise that teachers both learned to provide constructive feedback as peer coaches and improved their practices (e.g., Miller et al., 1991; Scheeler et al., 2010; Stichter et al., 2006), particularly when coaching

targeted a specific skill or strategy (e.g., Ottley et al., 2017; Stichter et al., 2006). Yet other studies have demonstrated no teacher behavioral changes (e.g., Murray et al., 2009) or had mixed findings across the outcomes measured (e.g., Johnson et al., 2017). These findings suggest that peer coaching may not be effective when teachers coach one another. To our knowledge, only one study has utilized a form of random assignment (i.e., purposive sample) to compare peer coaching versus coaching delivered by online expert coaches. In this sample of 20 teachers, peer coaching was more effective at improving teachers' planning and execution of instruction (Ma et al., 2018), highlighting the potential value of peer engagement in coaching.

Process studies of peer coaching have identified additional challenges to incorporating the needed level of structure and training into peer coaching. Studies have concluded that the value of peer coaching is dependent upon an emphasis on observations (Jenkins et al., 2005) and neutral feedback blended with assistance in setting clear goals (Thurlings et al., 2012). However, such observational data collection, interpretation of data, and structured problem solving are not typically part of teacher training (see Marsh et al., 2015). Not only might peer coaches struggle with these tasks, requiring rigorous training to attain these skills, but peer coaches likely also have limited time to carry these tasks out.

Optimizing Paired Coaching

Given above-raised concerns about training challenges for teachers to coach one another, third-party coaching may optimize paired approaches and is of interest here. For example, supporting teachers in pairs may reinforce specific mechanisms of change; by leveraging pre-existing, personal relationships between teachers, social learning may be optimized (see Johnson et al., 2016). Two teachers working in the same school are likely to interact more frequently than they do with a coach, especially if the coach is external to the school. Greater frequency and

duration of interactions may also increase the likelihood of mutually providing supports and performance feedback, in turn boosting effectiveness. When coached by a third-party, paired coaching could shift collegial relationships to become more effective, providing mutual support between teachers. An intentionally structured pairing approach may also promote power sharing and enhance opportunities for learning, achieved through regular check-ins, observing the other teacher's classroom, and ongoing performance feedback to promote more frequent use of the evidence-based strategies (Boudah et al., 2001; Erchul & Raven, 1997).

The Current Study

The purpose of this study was to examine the effects of coaching when delivered either to teachers individually or in pairs, relative to business-as-usual (i.e., comparison) conditions. In addition to determining the average and differential impacts of these two coaching approaches on teacher and student outcomes, we were also interested in the relative time efficiency, and thus cost effectiveness, of these approaches. We leveraged a prior RCT of Double Check (see Bradshaw et al., 2018) within 18 elementary and middle schools; the coaching is a culturally adapted version of the research-based Classroom Check-Up (Reinke et al., 2011). In a first arm of the teacher randomized RCT, teachers were coached individually (see Bradshaw et al., 2018), whereas in a separate arm, paired coaching was conducted.

Our first research aim was to determine whether there were main effects of coaching (using either approach) on teacher classroom management practices and student behavior (measured through classroom observations and office disciplinary referrals), as well as teacher self-reported attitudes and beliefs (i.e., efficacy and stress), from pre- to post-test (i.e., beginning to end of the school year) and at a 1-year follow-up. For this aim, we combined both teacher coaching conditions and contrasted with comparison teacher outcomes. Coached teachers were

expected to demonstrate improvements in classroom management, efficacy, and stress, from pre- to post-test relative to non-coached teachers. We also anticipated improved student behavior. These effects were expected to be sustained through the follow-up, one year later.

Our second aim was to determine if one of the two delivery approaches was more effective from pre- to post-test and again at follow-up. Based on the premise that the ongoing support of coaching underlies its impacts on teacher practices (Showers & Joyce, 1996), and the possibility that paired coaching could result in additional ongoing support between peer teachers, we hypothesized that the paired coaching would be as effective, if not more so, at improving teacher practices, attitudes, and beliefs as well as student behavior at post-test, and more sustainable (i.e., more effective at follow-up) than the individual coaching.

The third aim examined whether the paired coaching was at least, if not more, time and cost efficient than individual teacher coaching. Each of the coaching meetings were intended to occupy the same length of time between the two conditions, regardless of whether one or two teachers were present. Thus, we expected efficiency for coaches' time, and thus costs, for the paired coaching condition compared to the individual coaching condition (Pas et al., 2020). Teacher time was expected to be comparable in both. This aim was important for school divisions interested in scaling coaching supports while maximizing reach, impact, and efficiency.

Method

Participants

Participants in this study were 252 teachers (i.e., 152 intervention, 100 comparison) in 18 elementary and middle schools (nine each, elementary and middle) in one school district (see Figure 1). The sample for the paired coaching study arm comprised 94 teachers in six schools (i.e., 52 were coached) whereas the individual coaching study arm comprised 158 teachers in 12

schools (i.e., 100 were coached). A slightly higher proportion of teachers was randomized to the coached condition to enable the examination of factors associated with quality coaching (see Pas et al., 2016). The combined sample largely comprised female teachers and White teachers (i.e., greater than 80% in each category, ~70% were both White and female). Approximately one-third of teachers were 30 years old or younger and close to two-thirds taught in middle schools. The schools had implemented positive behavior supports for about 8 years, on average, with high fidelity at baseline (i.e., average scores over 90% on the School-wide Evaluation Tool; Sugai et al., 2001). Schools comprised a diverse student body. See Table 1 for additional teacher and school-level demographics.

All coaching was provided in person by four research team-hired coaches who were external to the participating schools. Coaches read the book outlining the CCU model (Reinke et al., 2011), viewed training videos, and participated in didactic trainings led by a trained CCU coach and the CCU developers. The trained CCU coach provided bi-weekly supervision, which was conducted in consultation with a CCU developer. Two coaches had an education master's degree and the other two had a school psychology doctorate. All coaches had prior coaching training and school experience. All coaches were female and two each were African American and White. School assignments were made based on the schools' relative location and size to ensure an equitable coaching workload and to maximize travel time efficiency.

Coaching Intervention

The coaching model was an adaptation of the Classroom Check-Up (CCU; Reinke et al., 2008, 2011) and included a stage-based, problem-solving process of five steps: (a) interview to build rapport, (b) data collection from the teacher via a Classroom Ecology Checklist survey and three coach-conducted classroom visits, (c) feedback regarding relative strengths and

weaknesses, (d) collaborative goal setting, and (e) implementation with progress monitoring. The CCU coach used motivational interviewing (Miller & Rollnick, 2002) to empower teachers to overcome ambivalence about change and to adopt and sustain use of new or improved classroom behavior management strategies (Reinke et al., 2011). Select data elements captured in the original CCU were trimmed back to allow for the addition of culturally responsive practices for the Double Check version of the CCU. Specifically, the Double Check CCU data collection and feedback form included basic positive behavioral supports (e.g., praise, reprimands) and instructional strategies (e.g., opportunities to respond) but also included the five Double Check CARES domains (i.e., Connection to the Curriculum, Authentic Relationships, Reflective Thinking, Effective Communication, and Sensitivity to Students Culture). These CARES domains were the focus of the five school-wide Double Check Professional Development (DCPD) sessions offered to all teachers in the study schools.

The intent of the Double Check CCU was to build teacher skills and efficacy in culturally responsive behavioral management, with the goal of addressing the disproportionate exclusionary discipline of students of color. See Bradshaw et al. (2018) for a more complete summary of the Double Check adaptations to the original CCU model and the DCPD. In a second year, coaches provided two additional schoolwide DCPD booster sessions in all participating schools, for a total of seven DCPD sessions across the 2 years. Coaches also sent quarterly emails to all coached teachers to check in on their progress, utilizing motivational interviewing techniques (e.g., open-ended questions) to elicit further engagement. No additional in-person coaching was conducted during the second year.

The paired and individual teacher Double Check coaching utilized the same five Double Check CCU steps outlined above; the only differences between the two modalities were that

interviews (one meeting with the coach), feedback, and goal setting (typically conducted during one meeting with the coach) were conducted for both teachers simultaneously. As an additional component to Step 2, each teacher in the pair was also trained (during a separate meeting) and encouraged to observe their paired peer's classroom, using the coach's data collection form. Rarely done in regular practice, peer classroom visits were meant to facilitate each teacher's understanding of data collection and their peer's classroom context, as well as to provide practice ideas and an objective perspective of classroom dynamics. Coaches again emailed quarterly check-ins to teachers in the follow-up year but sent these emails to teacher pairs together and encouraged pairs to continue to support one another.

Procedure

A total of 18 schools were recruited in a two-arm RCT, with random assignment of teachers (within school) to condition. The first arm of the trial involved 12 schools within which teachers were randomly assigned to either the individual coaching or comparison (no coaching) condition (see Bradshaw et al., 2018). The second arm of the trial involved six schools within which teachers were randomized to either paired coaching or the comparison condition. Across both arms of the trial, the school district hosted principal meetings to obtain interest and school-level commitment to join the project. Interested principals volunteered to attend and those who were also willing to have their school participate signed commitment letters. Classroom teachers were provided information about enrolling in the voluntary study and those willing to participate provided written consent to be randomized to either receive coaching or not. The researchers' Institutional Review Board approved this study. The overall teacher consent rate in the 18 project schools was 46.0% (i.e., 252 out of 548 eligible teachers consented); for the six schools where paired coaching versus comparison status was randomized, the teacher consent rate was 53.4%.

Our analytic sample of 248 teachers represented 98.4% of the 252 consented teachers who had baseline covariates for inclusion in the models. For the different outcomes, there was slight variability in coverage (i.e., 94.4%–98.4%; see missing data section and Figure 1).

Teachers randomly assigned to receive paired coaching were asked if there was any teacher that they preferred to work with, or not work with, and when they wanted to begin coaching. Coaches then paired teachers based on their preferences, when their joint availability would allow them to observe one another's classrooms, and any known commonalities or relationships between teachers. In middle schools, teachers who were also department chairs were paired either with another department chair or with a teacher outside of their content area to avoid power differentials within pairs. Regardless of randomized condition, all participating teachers were also expected to attend five, 1 hr whole-school Double Check Professional Development (DCPD) sessions provided by the coach at the school during regular school hours (see Bradshaw et al., 2018, for a full description of the DCPD sessions).

Data collection

In both study arms, data were collected from coached and non-coached teachers at three time points: (a) the beginning of the first school year of participation (i.e., pre-test, in the fall), (b) at the end of the first school year of participation (i.e., at post-test, in the spring), and (c) again one year following post-test (i.e., 1-year follow-up).

Surveys. Teacher self-report surveys were administered using a secure online survey. A $10 gift card was provided after completion of each of up to three surveys. The completion rate among for the full sample of teachers in the 18 schools was 95% at baseline, 90% at post-test, and 46% at the 1-year follow-up. For the teachers in the individual coaching arm only, it was

94% at baseline, 94% at post-test, and 42% at the 1-year follow-up, whereas it was 97% at baseline, 82% at post-test, and 52% at the 1-year follow-up for the teachers in the paired arm.

Classroom Observations. External observers were recruited using a position listing on various online job boards. The positions were part-time, temporary, and for hourly pay. We specifically targeted college and graduate students with experience working in schools or with school-aged children; most observers were female. All observers were hired by one of the research partner institutions and required to pass a criminal background check and human subjects training before receiving training in the *Assessing School Settings: Interactions of Students and Teachers* (ASSIST; Rusby et al., 2001, 2011) measure, as described below. Observers were first provided didactic instruction using a manual with observation procedures and code definitions. Then, observers engaged in coding practice in non-project schools and reliability assessments during which they had to reach 80% reliability with an expert coder in three classrooms to continue as an observer. Finally, once reliability was reached, they observed in schools, and there was an additional recalibration during data collection to monitor for observer drift. Average inter-observer agreement during recalibration was 87% and was calculated as the agreements in tallied behaviors divided by the total agreements and disagreements (Barlow & Hersen, 1984). Observers were blind to the intervention status of teachers. The ASSIST was conducted during a single, approximately 20-min, live classroom observation of each teacher at each of the three time points. Observations occurred during regular classroom instruction.

Measures

The outcomes collected mapped to key coaching focus areas and aligned both with prior research on the Double Check CCU as well as original CCU and other coaching model studies. Measures collected included observational, school archival, and survey data.

Observations of Teacher Classroom Management Practices

ASSIST observations were used to assess both teacher classroom management skills and student behavior. These were proximal and therefore primary outcomes of interest. The observations began with observers spending 3 min acclimating to the environment and recording information such as the number of teachers and students present and subject taught. Observers then recorded event-based tallies for 15 min before leaving the classroom to complete global ratings of teacher and student behaviors. The tallied teacher behaviors included the use of (a) proactive behavioral management, which included verbal (e.g., explaining, reminding, commanding, prompting) and physical (e.g., modeling) demonstrations of behavioral expectations (i.e., not as a reaction to behavioral issues); (b) opportunities to respond (OTRs; i.e., behavioral or instructional prompts seeking student responding either to the teacher, a peer or peers, or written responses that are publicly displayed); (c) approval (i.e., instances of teacher provision of a tangible item, verbal praise, approving gestures, or physical contact like a pat on the back); and (d) disapproval in the classroom (i.e., the threat or actual use of a punitive consequence, verbal criticism, or gestural or physical contact to demonstrate discontent with a student behavior). Instances of student non-cooperation (i.e., when a student failed to respond to a teacher directive) and disruptions (i.e., any behavior that interfered with the activity of another student or students, the entire classroom, or the teacher as indicated by the target person or group being taken off task) were also tallied. These tallies were from the original ASSIST measure.

The global rating subscales comprised 5-point Likert items (i.e., 0 for *never* to 4 for *almost continuously* observed). Here we analyzed seven subscales: (a) *teacher direction and influence* (5 items) including items such as, "There is evidence of classroom routines – students know what they're supposed to be doing"; (b) *teacher anticipation and responsiveness* (6 items), for example, "Teacher anticipates when students may have problems behaviorally"; (c) *teacher proactive behavior management* (4 items), for example, "Teacher gives clear instructions and directives to students"; (d) *culturally responsive teaching strategies* scale (7 items), which includes, "Teacher integrates cultural artifacts reflective of students' interests into learning activities"; (e) *teacher and student meaningful participation* (8 items), including, "Students are provided opportunities to contribute to discussion"; (f) *student cooperation* (7 items), for example, "Students handle transitions well"; and (g) *student socially disruptive behavior* (3 items), for example, "Students argue with peers". This latter scale was coded as 0 for *never occurred* to 4 for *often occurred (6+ times)*.

Global scale scores were the average of all scale item responses. Higher scores reflected more frequent engagement in the behaviors. Prior studies of the ASSIST demonstrated adequate reliability of this measure; for example, ICCs across three observations demonstrated little variability across three cycles and ranged from 0.72 to 0.81, with an average of 0.75 (see Gaias et al., 2019). For more psychometric data, including predictive validity, see Pas, Cash, et al. (2015), Debnam et al. (2015), and Bottiani et al. (2020).

Office Disciplinary Referrals

The School-Wide Information System (SWIS; Irvin et al, 2004; May et al., 2003) routinely tracked office disciplinary referrals (ODRs) in all schools throughout the 2 years of data collection. The school district downloaded the SWIS data for the respective school years

and provided the total number of ODRs by incident type, for all students (i.e., total) and disaggregated by race/ethnicity (i.e., White and Black only, as these were the two largest student groups in the schools). Disproportionate use of ODRs by student race (i.e., specifically the over-referral of Black students to the office) was an outcome of interest, given gaps in the use of potentially harmful exclusionary disciplinary practices between White and Black students (Losen et al., 2015; also see Bradshaw et al., 2018). This was included as a measure reflecting the complex interaction of teacher (and administrator) practice with student behavior, rather than solely as an indicator of student behavior (Irvin et al., 2004). The goal of the coaching was to improve teachers' culturally responsive behavior management so that they would not over-rely on and over-refer Black students to the office; therefore, ODRs served as a primary outcome of interest.

Teacher Beliefs and Attitudes

Teachers completed self-report surveys online about their perceived self-efficacy (3 scales) and work-related stress on items using a 6-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Scale scores were the average of all item responses. Higher scores indicated more efficacy (desired), stress, and social desirability (undesired). These outcomes served as secondary outcomes and are described below.

Culturally Responsive Practices Self-Efficacy. The teacher survey assessed self-efficacy to implement culturally responsive practices using two measures: (a) the *Multicultural Efficacy Scale* (Guyton & Wesche, 2005; $\alpha = .80–.82$) assessed teachers' self-efficacy to provide culturally responsive instruction using 15 items (e.g., "I can help students take on the perspective of ethnic and cultural groups different from their own", "I can develop activities that increase the self-confidence of diverse students") and the (b) *Culturally-Responsive Teaching Self-Efficacy*

Scale (Siwatu, 2007; α = .78–.84) measured culturally responsive self-efficacy and the perceived ability to connect with diverse students using 15 items (e.g., "I implement strategies to minimize the effects of mismatches between my students' home culture and the school culture" and "I use my students' cultural background to create a meaningful learning experience"). These were included given the focus of coaching on culturally responsive teaching.

Classroom Behavior Management Self-efficacy. The Hoy and Woolfolk (1993) efficacy scale assessed teachers' ability to handle students with behavior problems (e.g., "I can manage almost any student behavior problem"; α = .80–.81) using 5 items. This was included given the focus of coaching on behavior management.

Stress. The work-related stress scale (Hurrell & McLaney, 1988) included 5 items (e.g., "In my job, I feel I am under great stress"; "I am unable to cope with the stress of my job on a daily basis"; α = .82–.83) and was included to determine whether coaching decreased stress or had an iatrogenic effect on stress.

Social Desirability Bias. We controlled for social desirability bias in our analyses as it is a common concern in self-report measures of cultural responsivity (Larson & Bradshaw, 2017). We utilized an abbreviated, 10-item version of the social desirability bias scale (Crowne & Marlowe, 1960; e.g., "I have never intensely disliked any of my students"; "I always try to practice what I preach"; α = .61–.68).

Demographics. Teachers also provided basic demographic data, including their race/ ethnicity, age, and gender (see Table 1). These variables were dichotomized (i.e., White = 1, all other races = 0; age ≤30 = 1, >31 years of age = 0; and female gender = 1, male = 0) and served as control variables, as did the school level in which the teacher taught (i.e., middle = 1, elementary = 0).

Time Spent in Coaching

After each coaching activity conducted with or for a teacher or teacher pair, coaches documented their time spent using an electronic time log. Time was logged for core components of coaching (i.e., interview, data collection, feedback, action planning, and follow-up observations and feedback), relationship building activities, and preparing for the coaching case. For paired coaching, coaches also documented time spent training the teachers to conduct data collection. All time logged was counted as coach time; for the paired coaching, time in joint meetings was divided in half to indicate coach time allocated to each teacher. Only time spent in meetings (i.e., interviews, feedback, and action planning; and, for paired coaching, data collection training) was counted as teacher time. Time logs were utilized to estimate costs (see Pas et al., 2020, for another example of using such logs for cost analyses).

Coaching Fidelity

Coaches also completed an adherence checklist indicating whether they had excluded (0), partially implemented (1), or fully implemented (2) each part of the interview, feedback, and goal setting (Pas et al., 2016). Coaches conducting paired coaching also responded whether they had implemented each part of the data collection training. The checklist included items reflecting what the coach was supposed to cover during each session. For the individual interview, this included 7 items for individual coaching and 9 items for paired coaching (i.e., additional items asked about allotting equal time for both teachers and scheduling the CCU data collection training). There were 8 items for CCU data collection training (paired only), 8 items for the feedback, and 10 items for the goal setting session. Descriptive analyses of fidelity indicated a high level of adherence to coaching procedures in both conditions. In the paired condition, coaches reported implementing 94% of the interview elements on average, 87% of the data

collection training, 99% of the feedback, and 89% of the goal setting. In the individual coaching condition, coaches reported implementing an average of about 98% of the interview elements, 99% of the feedback, and 92% of the goal setting.

Analyses

Change in each outcome across three time points was modeled using three-level, hierarchical linear models in the HLM 7 software (Raudenbush et al., 2011). The data time points were fall pre-intervention, spring post-intervention, and follow-up 1 year after post-intervention. In all models, repeated observations (Level 1) were nested within teachers (Level 2) who were nested within schools (Level 3). As indicated in Equation 1 (EQ 1) below, time variables at Level 1 were dummy coded for post-test (i.e., post-test = 1; all other time points = 0) and the follow-up time point (i.e., follow-up = 1; all other time points = 0) to model non-linear change. As shown in EQ 2 below, Level 2 covariates were: teacher age (i.e., < 30 years old = 1; > 31 years old = 0), teacher race/ethnicity (1 = White; 0 = all other), and gender (1 = female; 0 = male). As shown in EQ 5 below, Level 3 included school level (1 = middle; 0 = elementary) and study cohort (0 = individual coaching; 1 = paired coaching) as a covariate. As indicated in EQs 6–8 below, teacher-level covariates were modeled as fixed effects (i.e., not freed to differ across schools). As indicated in EQ 9 and EQ 11 below, cross-level interactions between study arm (i.e., cohort) and the time dummy codes were also included to address sample differences in comparison group change over time.

Main Effects Model (Aim 1, Model 1)

To estimate the main effects of coaching in both coaching conditions relative to all comparisons, Model 1 included intervention condition (1 = coached in either condition, 0 = comparison) at Level 2 as a predictor of timing effects (i.e., fixed effects for post and follow-up;

see EQs 3 and 4 below). Intervention group coefficients for post (i.e., $\beta_{11j} = \gamma_{110}$ in EQs 3 and 10a) and follow-up fixed effects (i.e., $\beta_{21j} = \gamma_{210}$ in EQs 4 and 12a) were the parameters of interest for this aim.

Level-1 Model – Observation-Level Models (Main and Differential Effects Models)

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}*(POST_{ij}) + \pi_{2ij}*(FOLLOWUP_{tij}) + e_{tij} \quad \text{(EQ1)}$$

Level-2 Model – Teacher-Level Models (Main and Differential Effects Models)

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}*(UNDER\ 30_{ij}) + \beta_{02j}*(WHITE_{ij}) + \beta_{03j}*(FEMALE_{ij}) + r_{0ij} \quad \text{(EQ 2)}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j}*(COACHED_{ij}) \quad \text{(EQ 3)}$$
$$\pi_{2ij} = \beta_{20j} + \beta_{21j}*(COACHED_{ij}) \quad \text{(EQ 4)}$$

Level-3 Model – School-Level Model for Main Effects Model

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(MS_j) + \gamma_{002}(COHORT_j) + u_{00j} \quad \text{(EQ 5)}$$
$$\beta_{01j} \text{ through } \beta_{03j} = \gamma_{010} \text{ through } \gamma_{030} \quad \text{(EQs 6–8)}$$
$$\beta_{10j} = \gamma_{100} + \gamma_{101}(COHORT_j) \quad \text{(EQ 9)}$$
$$\beta_{11j} = \gamma_{110} \quad \text{(EQ 10a)}$$
$$\beta_{20j} = \gamma_{200} + \gamma_{201}(COHORT_j) \quad \text{(EQ 11)}$$
$$\beta_{21j} = \gamma_{210} \quad \text{(EQ 12a)}$$

Differential Post-test Effects and Sustainability (Aim 2, Model 2)

To examine differential effects between coaching approaches, Model 2 differed from Model 1 only in that it included a cross-level interaction between cohort (i.e., 1 = paired and 0 = individual coaching) and the intervention effect at each time point (see EQs 10b and 12b). The cross-level interaction in EQ 10b provided the estimate of differential effects at post-test (i.e., γ_{111}); the interaction in EQ 12b provided an estimate for differential sustainability (i.e., γ_{211}; follow-up effect).

Level-3 Model – School-Level Model for Differential Effects Model

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(MS_j) + \gamma_{002}(COHORT_j) + u_{00j} \quad \text{(EQ 5)}$$
$$\beta_{01j} \text{ through } \beta_{03j} = \gamma_{010} \text{ through } \gamma_{030} \quad \text{(EQs 6–8)}$$
$$\beta_{10j} = \gamma_{100} + \gamma_{101}(COHORT_j) \quad \text{(EQ 9)}$$
$$\beta_{11j} = \gamma_{110} + \gamma_{111}(COHORT_j) \quad \text{(EQ 10b)}$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}(COHORT_j) \quad \text{(EQ 11)}$$
$$\beta_{21j} = \gamma_{210} + \gamma_{211}(COHORT_j) \quad \text{(EQ 12b)}$$

Distributional Assumptions

In all models, continuous scale scores (i.e., ASSIST global ratings and teacher self-reports) were modeled with the normal distribution. A positive coefficient was considered desirable for all global ratings and self-reports, except the socially disruptive student behavior (ASSIST global rating) and stress (self-report) scales, for which negative coefficients were desirable. We report effect sizes (i.e., Spybrook's delta [Δ]; Spybrook, 2008) for the timing and intervention effects; these represent effects in terms of standard deviation, and so are similar in interpretation to Cohen's *d* (Cohen, 1992). Thus, we considered effects up to .20 as small, those from 0.20 to 0.50 as moderate, and those above 0.50 as large.

Because ODRs and ASSIST tally outcomes were counts, they were modeled using a Poisson distribution, thereby accounting for the boundedness on the lower-end of the distribution (i.e., zero and above) and unboundedness on the higher-end of the distribution (Cameron & Trivedi, 1998). ODRs and ASSIST tally outcomes exhibited overdispersion (i.e., variance greater than the mean), which we accounted for as suggested by Cameron and Trivedi (1998). Poisson regression coefficients were exponentiated to calculate incidence rate ratios (IRRs), for which values less than 1 indicated reduced rates and greater than 1 indicated increased rates. Therefore, IRRs less than 1 were desirable for ODRs and the negatively worded ASSIST tally measures (i.e., disapprovals, student non-cooperation, and student disruptions), whereas IRRs greater than 1 indicated desirable effects for all other ASSIST tally measures. Significance of IRRs was ascertained and reported using 95% confidence intervals (CI).

Missing Data Handling for Impact Analyses

Four teachers (i.e., 1.6% of the sample) were excluded from all analyses because of missing teacher-reported covariates (i.e., race, gender, and age; $n = 248$). Two of these excluded teachers were from the coached groups and two were from comparison groups. This represents the observational and ODR data samples. An additional 10 teachers were excluded from the teacher-survey analyses only (i.e., efficacy and stress), nine of whom did not have baseline social desirability (covariate) data available. Aside from these exclusions, all other teachers were retained in the final analyses (see Figure 1). Rates of missing data did not significantly differ between randomized treatment groups at any time point. Thus, missing data in outcomes was handled using full information maximum likelihood.

Time/Cost Efficiency for Coaching Approach (Aim 3)

To assess relative time and cost efficiency between coaching approaches, independent samples *t*-tests were used to calculate and compare the average time the coach dedicated to each teacher in the individual and paired coaching. To compare the cost between the two models, an ingredients-based cost approach was applied from a government perspective, which calculated the labor cost by multiplying the time spent and wages for teachers and coaches (Levin et al., 2017). Net savings for paired coaching was also calculated as the total cost per paired coaching school minus the total cost per individual coaching school, including the average teachers per school sample size (for additional coaching costs calculation details, see Pas et al., 2020).

Results

Descriptive data for all outcomes are presented in Table 2. Given that this was a small RCT to test the promise of the intervention, we present effect sizes and IRRs for all significant effects prior to a multiple test (Benjamini-Hochberg) adjustment as well as adjusted *p*-values.

Aim 1: Main Effects in Both Samples at Post-test and Follow-up

Intervention effects at post-test for the overall (combined) intervention effect are shown in Table 3 under Model 1 ("Combined Effects") and post-test. There are significant effects for 4 of 20 outcomes examined, and 3 of these 4 significant effects were in the expected/desired direction. Relative to comparison teachers, tallied approvals were higher for intervention teachers (IRR = 1.33, 95% CI [1.04, 1.70]), tallied student non-cooperation was lower for intervention teachers (IRR = 0.51, 95% CI [0.31, 0.88]), and global ratings of intervention teachers' behavior management were higher ($\Delta = .29$). Unexpectedly, tallied post-test disapprovals were also higher for intervention teachers (IRR = 1.82, 95% CI [1.00, 3.32]). To adjust for multiple comparisons, we applied a Benjamini-Hochberg adjustment to account for a false discovery rate. In this adjustment, only the detected improvements in teachers' behavior management remained significant ($p < .05$), whereas teachers' use of approvals and disapprovals and student non-cooperation were no longer statistically significantly different between groups.

Model 1 also estimated intervention effects for the combined sample at follow-up (see Table 3) as compared to all comparison teachers. This model indicated that effects were not sustained. The only significant effect detected was a 61% increase in ODRs for White students among intervention teachers (IRR = 1.61) but did not remain significant after the Benjamini-Hochberg adjustment. No other effects were evident for any other measure (i.e., ASSIST observations or teacher self-report scales).

Aim 2: Differential Effects of Coaching Approach

Differential effects between the two coaching approaches are reported in Table 3 under Model 2. For differential effects at post-test, comparing the individual to paired coaching were significant for only two outcomes, as depicted under Model 2, "Differential Intervention Effects" for post-test. Paired coaching was less effective than individual coaching in increasing student

cooperation (Δ = -.62) and decreasing student disruptive behaviors (Δ = .53) at post-test, although neither effect remained statistically significant following the Benjamini-Hochberg adjustment. A model identical to that described here, but positing the paired coaching, rather than the individual coaching, as the reference group yielded no significant effects for the paired coaching (results not tabled). In other words, the individual coaching was the only approach yielding significant and desirable effects on the measured outcomes. No other effects were detected for other ASSIST measures, ODRs, or teacher self-report measures at post-test.

As shown under the "Differential Intervention Effects" follow-up column in Table 3, the sustainability of intervention effects was no more or less desirable in either coaching approach. Although there was a non-significant sustained effect in student non-cooperation overall for the individual coaching, when examining the interaction between cohort and follow-up, there was a statistically significant difference in the ratios of IRRs between these two groups in student non-cooperation (IRR = 4.11, 95% CI: [1.65, 10.25]) that remained significant ($p < .01$) after the Benjamini-Hochberg adjustment. Furthermore, the rates of student non-cooperation were about 2.7 times higher (in the classrooms of teachers coached in pairs) at follow-up as compared to their within-school comparison classrooms (effect not tabled). No other effects were detected for other ASSIST measures, ODRs, or teacher self-report measures at follow-up.

Aim 3: Time/Cost Efficiency

For individual coaching, the coach time spent on the teacher represented the sum of all coach-teacher time in coaching sessions, whereas in the paired session, this was 50% of the coach's time spent in each session (e.g., if the full paired interview was 60 min, 30 min were allocated as "coach time" for each paired teacher). Based on this metric, the paired coaching was more time-efficient than the individual coaching, from the coaches' perspective. Coaches

conducting paired coaching spent an average of 446.9 min (*SD* = 130.2) per teacher, whereas individual coaches spent an average of 522.3 min (*SD* = 195.1). Based on Levene's test for equality of variances, variances were not assumed to be equal ($F = 6.723$, $p = .01$). The independent samples *t*-test yielded a significant means difference of 75.4 min ($p < .01$). The total time that teachers spent with coaches was calculated by summing total individual coaching time or the total paired coaching time. Based on this metric, the two coaching models did not differ in terms of efficiency: teachers spent 230.5 min (*SD* = 108.0) with coaches in the individual coaching and 213.0 min (*SD* = 56.7) with their coaches in paired coaching; this equates to a difference of 17 min, which was not statistically significant ($p = .49$). See Table 4.

According to Maryland State Department of Education, statewide average teacher salary in public schools in Maryland was $64,546 in 2013[1], which is $71,533 in 2020 USD. Based on an FTE of 2080 hr/year, teachers made $34.40/hr in 2020 USD. Coaches' salaries were derived from the researchers and reflect actual paid wages (i.e., $77,637 in 2020 USD) plus fringe benefits for coaches (i.e., 33.54% FTE, or $26,039), totaling $49.84/hr in 2020 USD value. The average costs per teacher for each coaching delivery approach were quite comparable (i.e., about $132 for individual and $122 for paired). The coach costs were more notably different per teacher (i.e., about $434 and $371 for individual and paired, respectively). This results in a total cost per teacher of $565.63 for individual coaching and $493.02 for paired coaching. The net savings for the paired coaching was $72.61 per teacher and about $440 per school. See Table 4.

Discussion

[1] Analysis of professional Salaries Maryland Public Schools (October 2013) http://www.marylandpublicschools.org/about/Documents/DCAA/SSP/20132014Staff/2014_Analysis_of_Prof_Sal.pdf

This study contrasted in-service coaching as delivered by one coach to either one teacher versus two teachers simultaneously. The impetus for developing and testing this paired coaching approach was several school principals involved in a study testing individual coaching (i.e., provided by researcher-hired coaches) describing a need to address teachers' desire for additional feedback, time for coaching, and modeling of skills, which is consistent with a community-based participatory research approach (Israel et al., 2005). Moreover, concerns have been raised within the field about whether coaching tested in research and provided by external (i.e., research-hired) coaches is transportable to schools, from a cost and feasibility standpoint (Cappella et al., 2012) or as effective as efficacy trials suggest (Kraft et al., 2018). There has been limited research examining the coaching of teachers simultaneously, which is an approach that may theoretically optimize coaching mechanisms and yield cost-savings by way of reducing coaches' time and thus may be more feasible and sustainable. To our knowledge, this is the first study to contrast these two coaching delivery approaches. Furthermore, few studies have examined the sustainability of effects after direct coaching ends (Kraft et al., 2018). This novel study contributes to our understanding of the immediate and sustained effects of coaching delivered to teacher pairs, relative to individual coaching, as well as contrasts in time and cost efficiency, with direct implications for feasibility and costs.

Consistent with the prior study of the coaching of one teacher (Bradshaw et al., 2018), comparing pre- and post-test, the overall effects of both coaching approaches (as compared to non-coached comparisons) on observed behaviors were positive but indicated weaker effects for the combined sample than among the individual coaching sample only. A robust finding across both coaching arms of the trial was on teacher use of proactive behavior management, where effect sizes and *p*-values were very similar in this and the original study (Bradshaw et al., 2018).

Conversely, teachers who received individual coaching had a higher likelihood of use of approvals than the combined (paired and individual) sample; similarly, individual coaching also had significantly positive effects on teacher anticipation and responsiveness and students' socially disruptive behaviors (Bradshaw et al., 2018), but were only marginally significant in the combined sample. In the combined sample, there were no changes in office discipline referrals for coached teachers; however, there was evidence of reductions in ODRs for Black students in the original two-armed Double Check RCT (Bradshaw et al., 2018). Notably, no impacts on teacher efficacy (i.e., for behavior management or culturally responsive teaching) were observed in the current study; in the original two-armed RCT of Double Check, a pre-post study of all teachers indicated improvements in efficacy following DCPD, but no value added on efficacy for coached teachers (Bradshaw et al., 2018). It is likely the same phenomenon occurred in the current study; there also may be a ceiling effect on these self-report efficacy measures given the high baseline ratings provided. Importantly, there were no significant effects on teacher-reported stress, which indicates that coaching did not create added stress for teachers.

In addition to contrasting the effects of the two coaching approaches, from pre- to post-test, we also examined sustainability of effects 1 year after the coaching concluded. Follow-up in coaching studies is quite rare; a recent review (Kraft et al., 2018) reported that just 8% of studies included such a follow-up data point (i.e., five studies) and findings indicated that changes to teacher practices were not sustained (see Garet et al., 2008; Teemant, 2014). Unfortunately, like prior research, in the current study there was no evidence of teacher practice changes being sustained 1 year later for the pooled sample of coached teachers. This may in part be because changes in the student population from year to year requires not just sustained knowledge and practices but generalizing skills to new groups of students. This may present teachers with novel

challenges and complicates the study and interpretation of sustained coaching effects on outcomes. Conversely, sustaining meaningful outcomes may require the social accountability that a physically present coach provides, or ongoing motivational interviewing to address ambivalence and empowerment to continue making practice changes. Additional research on tapered coaching models to determine the threshold of support needed to promote sustained changes are needed.

The analytic models examining differential effects between the two coaching delivery approaches yielded a consistent, but non-significant, advantage of the individual coaching at post-test, and basically no significant differences 1 year later. It is possible that the time required to engage in peer support outside of the coaching meetings was too burdensome given that teachers often lack adequate planning time. Also, the personal relationships between the teachers may have made it awkward to provide candid or constructive feedback separate from what the coach provided. Although teachers were trained to collect the data, we did not collect data about whether they engaged in this. Teachers also were not trained to coach one another, and this may have been a disadvantage. A possibly promising alternative would be to have the coach transition into providing training to the teachers to coach one another in an ongoing fashion, thereby allowing the teachers to take over the coaching responsibilities (e.g., as described in Barrett & Pas, 2020). However, this approach would likely take more time than the coaching studied here. Finally, although coaches dedicated less time to each teacher in paired coaching, it is possible that having the same teacher time without the full-undivided coach attention was not beneficial. The relative time savings for the paired coaching may, in part, be why there were less potent effects for the paired model.

Interestingly, the paired coaching saved coaches', but not teachers', time. Although this may not have been a clinically significant cost, personnel time often comprises most intervention costs (Belfield et al., 2015) and coaching costs specifically (Pas et al., 2020); therefore, it is important to examine further. The overall costs of both approaches were quite modest, representing less than one-tenth of an average teacher salary FTE in the United States. A drawback may be that the overall narrow time commitments for both types of coaching could explain why the coaching effects were not sustained. The cost analyses suggest that more time should be feasible and is worth the additional investment. To promote the sustainability of teacher practices, there is likely a longer-term investment needed; the precise time needed for such changes is an area in need of further research. Notably, Garet et al. (2008) examined the effects of coaching requiring much more time (i.e., > 60 hr) and still did not observe sustained effects into the next year.

Limitations

Although this study represents a rigorous test and comparison of two coaching delivery approaches, there are notable limitations to consider. Randomization of teachers to a coached or non-coached condition occurred within school. Schools were not randomly assigned to their respective arm of the study (i.e., individual vs. paired coaching) because funding for the second paired coaching arm was not secured at the time of recruitment into the original study, but rather augmented the original trial. A more rigorous comparison would contrast participants from the full set of schools, randomizing to 1 of 3 conditions: one comparison condition and the two intervention conditions. We included a broad set of outcomes, which allows for a thorough examination of outcomes but also resulted in multiple tests; when this was adjusted for, there were even fewer short-term effects detected. Due to the relatively small sample size and

developmental nature of this study, we did report some of the non-significant effects, but we place the greatest emphasis on the statistically significant effects in our review of the results and discussion of findings. Furthermore, there was not balance across the two arms (i.e., the individual arm was roughly double the size) or between the coached and non-coached conditions; this may have impacted the findings. All teachers were exposed to the 5-session DCPD series, regardless of condition. Therefore, we cannot formulate any conclusions about the impact of the DCPD, but rather only the impact of or the valued added of the coaching (Bradshaw et al., 2018). Future studies are needed to determine the impact of the coaching and professional development relative to controls who receive no training. Finally, we lacked data about whether the teachers observed their coaching partner, although the coach fidelity form indicated a high rate of training teachers; teacher self-report about the acceptability of the paired coaching similarly suggested that many teachers did observe their partner and saw value in it.

Conclusions and Implications

The findings of this study illustrate that the Double Check coaching resulted in teacher practice and student behavioral changes; the most robust of the findings was for teacher proactive behavior management. The coaching delivered to one teacher individually was more effective than delivering to pairs of teachers. However, neither coaching condition demonstrated sustained effects by the end of the follow-up year. This suggests a need for focusing on coaching models and professional development that result in sustained effects, which is a concern for the field as no studies currently demonstrate sustained teacher practices over time (i.e., some indicate effects for student outcomes over time). Thus, more coaching studies need to incorporate such measurement (Kraft et al., 2018). Although both coaching approaches required a relatively short amount of time for coaches (i.e., < 10 hr), the lack of sustained effects indicates that both are

likely worth a greater investment. Additionally, although the findings suggest that coaching teachers individually is a more effective coaching model, the paired model may be attractive for some contexts where there are limited coaching resources. In both cases, additional work is needed to ensure that teachers can generalize practice changes beyond the time during which they are coached (e.g., coaching into a second year to explicitly support generalization of skills). From a policy standpoint, on-going coaching is likely a high need in schools that would presumably lead to cost savings in the long term (e.g., by improving student outcomes and potentially teacher turnover).

References

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, 333*(6045), 1034–1037. https://doi.org/10.1126/science.1207998

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). Allyn and Bacon.

Barrett, C., & Pas, E. T. (2020). The relative low cost of coaching: A comparative cost analysis of three common professional development approaches in schools. *Prevention Science, 21* (5), 604-614. https://doi.org/10.1007/s11121-020-01115-5

Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit to Cost Analysis, 6*(3), 508–544. https://doi.org/10.1017/bca.2015.55

Bottiani, J., Lindstrom Johnson, S., McDaniel, H., & Bradshaw, C. (2020). Triangulating school climate: Areas of convergence and divergence across multiple levels and perspectives. *American Journal of Community Psychology, 65* (3–4), 423–436. https://doi.org/10.1002/ajcp.12410

Boudah, D. J., Logan, K. R., & Greenwood, C. R. (2001). The research to practice projects: Lessons learned about changing teacher practice. *Teacher Education and Special Education, 24*(4), 290–303. https://doi.org/10.1177/088840640102400404

Bradshaw, C. P, Pas, E. T., Bottiani, J., Debnam, K. J., Reinke, W., Herman, K., & Rosenberg, M. (2018). Promoting cultural responsivity and student engagement through Double Check coaching of classroom teachers: An efficacy study. *School Psychology Review, 47*(2)*,* 118–134. https://doi.org/10.17105/SPR-2017-0119.V47-2

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge Press.

Cappella, E., Hamre, B. K., Kim, H. Y., Henry, D. B., Frazier, S. L., Atkins, M. S., & Schoenwald, S. K. (2012). Teacher consultation and coaching within mental health practice: Classroom and child effects in urban elementary schools. *Journal of Consulting and Clinical Psychology, 80*(4), 597–610. https://psycnet.apa.org/doi/10.1037/a0027725

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. https://psycnet.apa.org/doi/10.1037/h0047358

Debnam, K. J., Pas, E. T., Bottiani, J., Cash, A. H., & Bradshaw, C. P. (2015). An examination of the association between self-reported and observed culturally responsive teaching practices. *Psychology in the Schools, 52*(6)*,* 533–548. https://doi.org/10.1002/pits.21845

Denton, C. A., & Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational & Psychological Consultation, 19*(2), 150–175. https://doi.org/10.1080/10474410802463296

Erchul, W. P., & Raven, B. H. (1997). School power in school consultation: A contemporary view of French and Raven's bases of power model. *Journal of School Psychology, 35*(2), 137–171. https://doi.org/10.1016/S0022-4405(97)00002-2

Gaias, L. M., Lindstrom Johnson, S., Bottiani, J. H., Debnam, K. J., & Bradshaw, C. P. (2019). Examining teachers' classroom management profiles: Incorporating a focus on culturally responsive practice. *Journal of School Psychology, 76,* 124–139. https://doi.org/10.1016/j.jsp.2019.07.017

Garet, M. S., Porter, A., Desimone, L., Birman, B., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Education Research Journal, 38*(4), 915–945. https://doi.org/10.3102/00028312038004915

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A Meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, *38*, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

Gregory, A., Hafen, C. A., Ruzek, E., 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*(2), 171–191. https://doi.org/10.17105/SPR45-2.171-191

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*(1), 38–45. https://doi.org/10.1080/00405841.2016.1260402

Guyton, E. M., & Wesche, M. V. (2005). The multicultural efficacy scale: Development, item selection, and reliability. *Multicultural Perspectives, 7*(4), 21–29. https://doi.org/10.1207/s15327892mcp0704_4

Hoy, W. K., & A. E. Woolfolk (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal, 93*(4), 355–372. https://doi.org/10.1086/461729

Hurrell, J. J. Jr., & McLaney, M. A. (1988). Exposure to job stress - A new psychometric instrument. *Scandinavian Journal of Work, Environment, and Health, 14,* 27–28.

Irvin, L., Tobin, T., Sprague, J., Sugai, G., & Vincent, C. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status and effects of school-wide behavioral interventions. *Journal of Positive Behavior Interventions, 6*(3), 131–147. https://doi.org/10.1177/10983007040060030201

Israel, B. A., Eng, E., Schulz, A. J., & Parker, E. A. (2005). Introduction to methods in community-based participatory research for heath. In B. A. Israel, E. Eng, A. J. Schulz, & E. A. Parker (Eds.), *Methods in community-based participatory research for health* (pp. 3–26). Jossey-Bass.

Jenkins, J. M., Garn, A., & Jenkins, P. (2005). Preservice teacher observations in peer coaching. *Journal of Teaching in Physical Education, 24*(1), 2–23. https://doi.org/10.1123/jtpe.24.1.2

Johnson, S. R., Finlon, K. J., Kobaki, R., & Izard, C. E. (2017). Promoting student-teacher interactions. Exploring a peer coaching model for teachers in a preschool setting. *Early Childhood Education Journal, 45,* 461–470. https://doi.org/10.1007/s10643-016-0790-1

Johnson, S., Pas, E., & Bradshaw, C. P. (2016). Understanding and measuring coach-teacher alliance: A glimpse inside the 'black box'. *Prevention Science, 17*(4), 439–449. https://doi.org/10.1007/s11121-016-0633-8

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*(4), 547–588. https://doi.org/10.3102/0034654318759268

Larson, K., & Bradshaw, C. P. (2017). Cultural competence and social desirability among practitioners: A systematic review of the literature. *Children & Youth Services Review, 76,* 100–111. https://doi.org/10.1016/j.childyouth.2017.02.034

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (3rd ed.). Sage.

Losen, D., Hodson, C., Keith, M. A., Morrison, K., & Belway, S. (2015). *Are we closing the school discipline gap?* The Center for Civil Rights Remedies at the Civil Rights Project, University of California Los Angeles. http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/are-we-closing-the-school-discipline-gap/AreWeClosingTheSchoolDisciplineGap_FINAL221.pdf

Ma, N., Xin, S., & Du, J. Y. (2018). A peer coaching-based professional development approach to improving the learning participation and learning design skills of in-service teachers. *Educational Technology & Society, 21*(2), 291–304. http://www.jstor.org/stable/26388408

Marsh, J. A, Bertrand, M., & Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record, 117*(4), 1–40. http://www.tcrecord.org/Content.asp?ContentId=17849

May, S., Ard, W., Todd, A. W., Horner, R. H., Glasgow, A., Sugai, G., & Sprague, J. (2003). *Schoolwide Information System*. University of Oregon, Educational & Community Supports.

Miller, S. P., Harris, C., & Watanbe, A. (1991). Professional coaching: A method for increasing effective and decreasing ineffective teacher behaviors. *Teacher Education and Special Education*, *14*(3), 183–191. https://doi.org/10.1177/088840649101400305

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change*. (2nd ed.). Guilford Press.

Murray, S., Ma, X., & Mazur, J. (2009). Effects of peer coaching on teachers' collaborative interactions and students' mathematics achievement. *The Journal of Educational Research, 102*(3)*,* 203–212. https://doi.org/10.3200/JOER.102.3.203-212

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, *46*(2), 532–566. https://doi.org/10.3102/0002831208328088

Ottley, J. R., Grygas Coogle, C., Rahn, N. L., & Spear, C. F. (2017). Impact of bug-in-ear professional development on early childhood co-teachers' use of communication strategies. *Topics in Early Childhood Special Education*, *36*(4), 218–229. https://doi.org/10.1177/0271121416631123

Pas, E. T., Bradshaw, C. P., Becker, K., Domitrovich, C., Berg, J., Musci, R., & Ialongo, N. (2015). Trajectories for coaching dosage as a means for improving implementation of the Good Behavior Game. *School Mental Health, 7,* 61–73. https://psycnet.apa.org/doi/10.1007/s12310-015-9145-0

Pas, E. T., Cash, A. H., O'Brennan, L., Debnam, K., & Bradshaw, C. P. (2015). Profiles of high school student behavior in the classroom: Associations with teacher behavior management strategies and classroom composition. *Journal of School Psychology, 53*(2)*,* 137–148. https://doi.org/10.1016/j.jsp.2014.12.005

Pas, E. T., Larson, K., Reinke, W., Herman, K., & Bradshaw, C. P. (2016). Implementation and acceptability of an adapted Classroom Check-Up coaching model to promote culturally responsive classroom management. *The Education & Treatment of Children, 39*(4)*,* 467–491. https://doi.org/10.1353/etc.2016.0021

Pas, E. T., Lindstrom Johnson, S., Alfonso, Y. N., & Bradshaw, C. P. (2020). Tracking time and resources associated with systems change and the adoption of evidence-based programs: The "hidden costs" of school-based coaching. *Administration and Policy in Mental Health and Mental Health Services Research, 47,* 720–734. https://doi.org/10.1007/s10488-020-01039-w

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*(4), 431–451. https://doi.org/10.1016/j.ecresq.2008.02.001

Raudenbush, S. W., Bryk, A. S., Cheong, Y. K., Fai, Y. F., Congdon, R. T., Jr., & duToit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling.* Scientific Software International.

Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The Classroom Check-Up*. Guilford Press.

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The Classroom Check-Up: A classwide consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*(3)*,* 315–332. https://doi.org/10.1080/02796015.2008.12087879

Rusby, J. C., Taylor, T., & Milchak, C. (2001). *Assessing school settings: Interactions of students and teachers (ASSIST) observation system*. Unpublished manual. Oregon Research Institute.

Rusby, J. C., Crowley, R., Sprague, J., & Biglan, A. (2011). Observations of the middle school environment: Context for student behavior beyond the classroom. *Psychology in the Schools, 48*(4), 400–415. https://doi.org/10.1002/pits.20562

Scheeler, M. C., Congdon, M., & Stansbery, S. (2010). Providing immediate feedback to co-teachers through bug-in-ear technology: An effective method of peer coaching in inclusion classrooms. *Teacher Education and Special Education, 33*(1), 83–96. https://doi.org/10.1177/0888406409357013

Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership, 53*(6), 12–16. http://www.educationalleader.com/subtopicintro/read/ASCD/ASCD_351_1.pdf

Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, *23*(7), 1086–1101. https://doi.org/10.1016/j.tate.2006.07.011

Spybrook, J. (2008). *Power, sample size, and design*. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 273–311). Information Age Publishing.

Stichter, J. P., Lewis, T. J., Richter, M., Johnson, N. W., & Bradley, L. (2006). Assessing antecedent variables: The effects of instructional variables on student outcomes through in-service and peer coaching professional development models. *Education and Treatment of Children, 29*(4), 665–692. http://www.jstor.org/stable/42900558

Strother, D.B. (1989). Peer coaching for teachers: Opening classroom doors. *Phi Delta Kappan, 70,* 824–827. https://www.jstor.org/stable/pdf/20404040.pdf

Sugai, G., Lewis-Palmer, T., Todd, A., & Horner, R. H. (2001). *School-wide evaluation tool*. Educational and Community Supports. https://www.pbis.org/resource/school-wide-evaluation-tool-set

Sutherland, K. S., Conroy, M. A., Vo, A., & Ladwig, C. (2015). Implementation integrity of practice-based coaching: Preliminary results from the BEST in CLASS efficacy trial. *School Mental Health, 7,* 21–33. https://doi.org/10.1007/s12310-014-9134-8

Teemant, A. (2014). A mixed-methods investigation of instructional coaching for teachers of diverse learners. *Urban Education, 49*(5), 574–604. https://doi.org/10.1177%2F0042085913481362

Thurlings, M., Vermeulen, M., Bastiaens, T., & Stjinen, S. (2012). Investigating feedback on practice among teachers: Coherence of observed and perceived feedback. *Mentoring & Tutoring: Partnership in Learning, 20*(4), 473–490. https://doi.org/10.1080/13611267.2012.725981

Vygotsky, L. (1978). *Mind in society*. Harvard University Press.

Figure 1

Study Arms, Sample Sizes, and Consent and Data Completion Rates

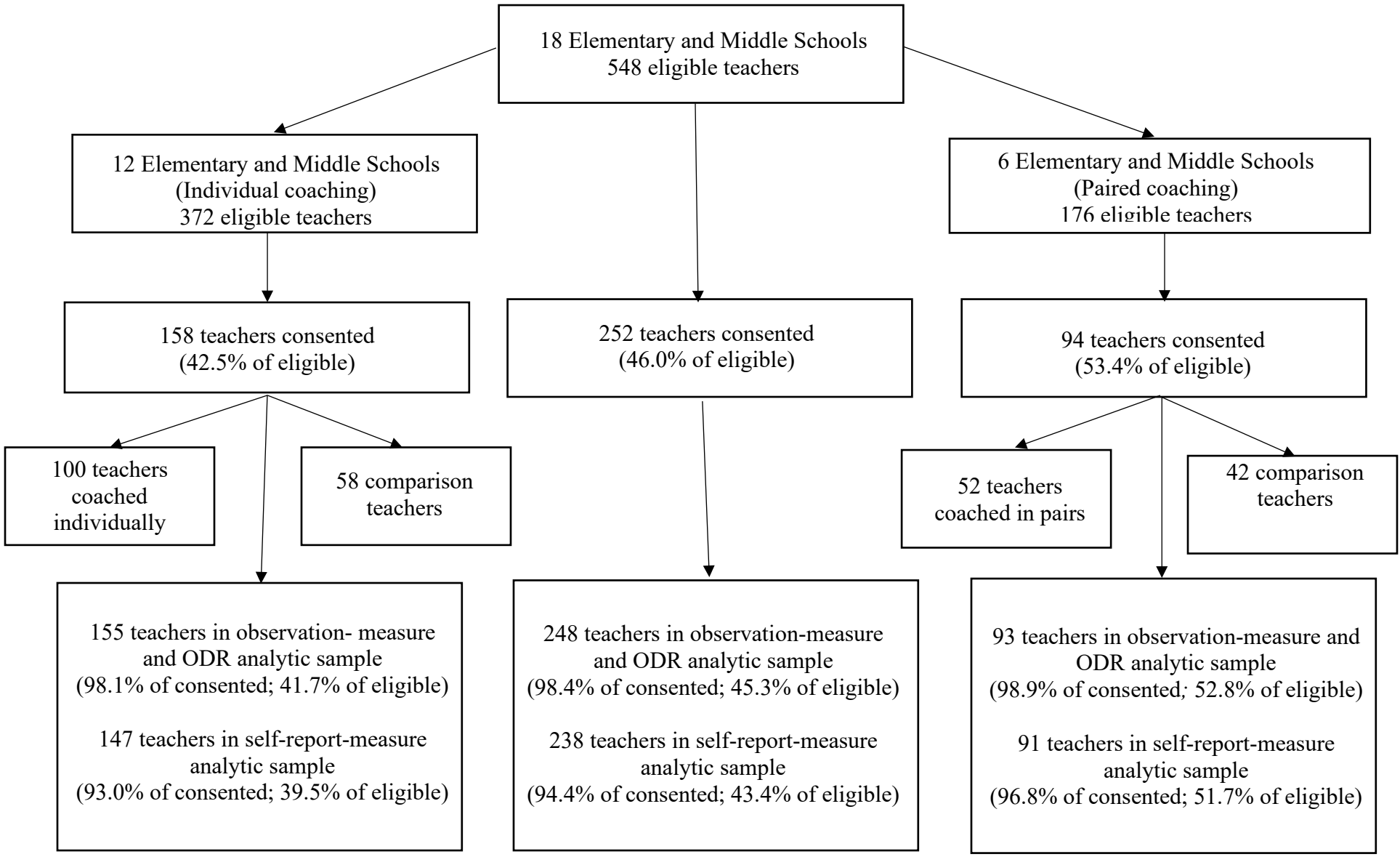

Table 1

School- and Teacher-Level Average Demographics at Baseline (n = 18)

	Individual Coaching[a]		Paired Coaching		Full Sample	
	(*n* [%] or *M* [*SD*])		(*n* [%] or *M* [*SD*])		(*n* [%] or *M* [*SD*])	
# Middle Schools	6	(50%)	3	(50%)	9	(50%)
Enrollment	594.3	(214.2)	725.3	(256.3)	638.0	(230.3)
Years since PBIS Training	7.7	(3.6)	8.8	(2.6)	8.1	(3.3)
SET Scores	94.0	(6.2)	98.3	(2.3)	95.4	(5.5)
Suspension Rate	14.0	(9.2)	5.6	(4.0)	11.2	(8.7)
Math Proficiency Rate	73.9	(11.1)	79.3	(9.1)	75.7	(10.5)
Reading Proficiency Rate	83.8	(4.6)	86.3	(6.1)	84.6	(5.1)
School-level Student Compositional Variables						
% Hispanic	14.0	(6.6)	9.6	(5.9)	12.5	(6.5)
% Black	37.9	(17.7)	27.4	(16.1)	34.4	(17.4)
% White	36.2	(17.0)	50.2	(23.8)	40.9	(20.0)
% Special Education	9.9	(1.6)	8.0	(2.2)	9.3	(2.0)
% FARMs	56.8	(19.2)	36.1	(24.5)	49.9	(22.7)
% ELL	6.5	(3.8)	5.1	(5.4)	6.0	(4.3)
Teacher Demographics (*n* = 252[b])						
Female	135	(86.0%)	81	(87.1%)	216	(86.4%)
White	122	(78.7%)	79	(85.0%)	201	(81.0%)
Black	19	(12.3%)	8	(8.6%)	27	(10.9%)
Other	14	(9.0%)	6	(6.4%)	20	(8.1%)
Early Career	56	(36.1%)	21	(22.6%)	77	(31.1%)
Middle	104	(65.8%)	47	(50.0%)	151	(59.9%)

Note. PBIS = Positive Behavioral Interventions and Supports; SET = School-wide Evaluation Tool (i.e., a PBIS implementation fidelity measure), FARMs = free and reduced-price meals, ELL = English language learners. [a]The results presented in this column also appear in Bradshaw et al. (2018) but are provided here for comparison. [b] Up to four cases were missing demographic data.

Table 2

Teacher-Level Univariate Descriptive Statistics for Analytic Variables

	Missingness		Full sample (n = 252)		Coached Teachers (Combined†) (n = 152)		Comparison Teachers (Combined††) (n = 100)		Diff Test
Demographics	*N*	%	*N*	%	*n*	%	*n*	%	*p*
Female	2	0.8%	216	86.4%	132	86.8%	84	85.7%	.85
White Race	4	1.6%	201	81.0%	119	79.3%	82	83.7%	.41
Black Race	4	1.6%	27	10.9%	18	12.0%	9	9.2%	.54
Other Race	4	1.6%	20	8.1%	13	8.7%	7	7.1%	.81
Young Age (30 or younger)	4	1.6%	77	31.0%	48	32.0%	29	29.6%	.78
Middle School Teacher	0	0.0%	151	59.9%	88	57.9%	63	63.0%	.43
Time 1 Means	*N*	%	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*
Office Disciplinary Referrals									
Total ODR	7	2.8%	3.48	4.75	3.54	4.95	3.39	4.47	.93
Black ODR	7	2.8%	1.57	2.43	1.47	2.22	1.73	2.72	.77
White ODR	7	2.8%	0.74	1.49	0.68	1.31	0.82	1.73	.67
ASSIST Tallies									
Proactive Behavior Management	5	2.0%	8.97	6.40	8.88	6.26	9.10	6.62	.95
Approvals	5	2.0%	6.85	6.72	6.84	6.58	6.86	6.97	.82
Disapprovals	5	2.0%	0.31	0.94	0.32	0.85	0.30	1.08	.53
OTR	5	2.0%	24.31	15.86	24.81	15.48	23.56	16.48	.47
Student Non-cooperation	5	2.0%	0.91	2.33	0.76	2.23	1.13	2.48	.04
Student Disruptive Behaviors	5	2.0%	12.61	9.36	12.47	9.69	12.82	8.87	.54
ASSIST Global Ratings									
Cultural Responsive Teaching Strategies	7	2.8%	1.23	0.95	1.20	0.96	1.28	0.95	.35
Direction and Influence	6	2.4%	3.48	0.52	3.50	0.54	3.45	0.50	.20
Responsiveness	6	2.4%	3.10	0.73	3.12	0.75	3.08	0.72	.75
Proactive Behavior Management	6	2.4%	3.00	0.63	3.02	0.63	2.97	0.63	.79
Meaningful Participation	7	2.8%	2.61	0.90	2.65	0.93	2.55	0.87	.48
Student Cooperation	8	3.2%	3.19	0.69	3.21	0.70	3.16	0.68	.49
Student Disruptive Behaviors	5	2.0%	0.85	0.45	0.84	0.44	0.88	0.46	.54
Teacher Self-Report									
CRT Self-Efficacy	13	5.2%	4.30	0.52	4.27	0.54	4.36	0.50	.30
Multicultural Efficacy Scale	12	4.8%	4.62	0.47	4.58	0.43	4.67	0.52	.14
Behavior Management Efficacy	12	4.8%	4.49	0.67	4.47	0.70	4.51	0.64	.77
Teacher Stress	13	5.2%	3.67	1.00	3.72	1.01	3.59	1.00	.28

Note. The *p*-values for the ODRs and ASSIST Tallies do not account for the non-normality of their distributions (i.e., these are generated from standard ANOVA/t-tests). CRT = culturally responsive teaching. † indicates the pooled sample of both individual and paired coaching teachers. †† indicates the pooled sample of teachers from each study arm who were *not* coached.

Table 3

Intervention Effect Estimates from Models 1 and 2: Main and Differential Effects Models

	Model 1 Combined Effects				Model 2 Differential Intervention Effects (Paired vs. Individual Coaching)			
	Post		Follow-up		Post		Follow-up	
	IRR	(95% CI)	IRR	(95% CI)	IRR	(95% CI)	IRR	(95% CI)
Office Disciplinary Referrals								
Total	0.89	(0.70,1.12)	0.96	(0.74,1.25)	1.22	(0.68, 2.19)	1.29	(0.76, 2.18)
Black Students	0.78	(0.57,1.05)	0.89	(0.59,1.36)	2.13	(0.79, 5.74)	1.26	(0.54, 2.91)
White Students	1.09	(0.69,1.70)	1.61*	(1.02,2.56)	0.38	(0.13, 1.12)	2.11	(0.76, 5.86)
ASSIST Tallies								
Proactive	1.05	(0.89,1.24)	1.04	(0.88,1.22)	1.02	(0.73, 1.43)	1.12	(0.81, 1.55)
Approvals	1.33*	(1.04,1.70)	1.19	(0.92,1.55)	0.74	(0.44, 1.22)	0.73	(0.42, 1.28)
Disapprovals	1.82*	(1.00,3.32)	1.60	(0.91,2.83)	0.68	(0.19, 2.40)	2.64	(0.80, 8.66)
OTR	1.03	(0.87,1.21)	1.02	(0.84,1.23)	1.11	(0.80, 1.54)	0.99	(0.66, 1.47)
Non-cooperation	0.53*	(0.31,0.88)	1.22	(0.79,1.90)	0.95	(0.34, 2.68)	**4.11***+	(1.65, 10.25)
Disruptives	0.97	(0.78,1.20)	1.25	(0.98,1.59)	1.00	(0.64, 1.57)	1.54	(0.95, 2.49)
	Δ	(*SE*)	Δ	(*SE*)	Δ	(*SE*)	Δ	(*SE*)
ASSIST Globals								
CR Teaching Strategies	0.04	(0.12)	0.01	(0.13)	-0.36	(0.27)	0.22	(0.30)
Direction/Influence	0.07	(0.07)	-0.11	(0.08)	-0.33	(0.28)	-0.28	(0.30)
Anticipation	0.18	(0.10)	-0.08	(0.11)	-0.41	(0.27)	-0.01	(0.28)
Beh. Mgmt.	**0.29*****+	(0.08)	-0.10	(0.09)	-0.29	(0.26)	0.26	(0.29)
Meaningful Part.	0.06	(0.10)	-0.07	(0.11)	-0.12	(0.24)	0.00	(0.27)
Student Cooperation	0.13	(0.09)	-0.13	(0.10)	-0.62*	(0.26)	-0.29	(0.29)
Student Disruptive Behaviors	-0.11	(0.06)	0.06	(0.07)	0.53*	(0.27)	0.51	(0.29)
Teacher Self-Report								
CRT Efficacy	-0.04	(0.06)	-0.05	(0.07)	-0.15	(0.23)	-0.06	(0.28)
Multicult. Efficacy	0.00	(0.05)	-0.04	(0.07)	0.00	(0.22)	0.02	(0.28)
Beh. Mgmt. Efficacy	0.01	(0.07)	-0.12	(0.10)	-0.29	(0.23)	-0.28	(0.31)
Stress	-0.12	(0.11)	0.16	(0.15)	0.16	(0.22)	0.11	(0.29)

Note. OTR = Opportunities to Respond; CR = Culturally Responsive; Beh. Mgmt. = Behavior Management; Part. = Participation; CRT = Culturally Responsive Teaching; Multicult. = Multicultural; IRR = Incident rate ratio; CI = Confidence interval; Δ = Spybrook's Delta. *$p < .05$, **$p < .01$, ***$p < .001$. **+ $p < .05$** after Benjamini-Hochberg adjustment for false discovery rate.

Table 4

Annual Total Labor Cost for Individual and Paired Coaching

Approach	*N* of Schools	Teacher *n*	Average teacher minutes (*SD*)	Cost per teacher for teacher time	Average coach minutes (*SD*)	Cost per teacher for coach time	Total Cost	Average Total Cost Per School	Net Saving Per Paired School
Individual	12	100	230.5 (108.0)	$132.12	522.3 (195.1)	$433.51	$56,563.00	$4,713.58	$440.74
Paired	6	52	213.0 (56.7)	$122.09	446.9 (130.2)	$370.93	$25,637.04	$4,272.84	

Note. The total cost is the number of teachers multiplied by the cost per teacher for teacher time, plus the number of teachers times the cost per teacher for coach time. The average total cost per school is the total cost divided by the number of schools.