School Psychology Review
2019, Volume 48, No. 1, pp. 31–45
DOI: 10.17105/SPR-2018-0003.V48-1

Outcomes of the BEST in CLASS Intervention on Teachers' Use of Effective Practices, Self-Efficacy, and Classroom Quality

Maureen A. Conroy
University of Florida

Kevin S. Sutherland
Virginia Commonwealth University

James Algina
Crystal Ladwig
Brittany Werch
Jose Martinez
University of Florida

Gabriel Jessee
Maria Gyure
Virginia Commonwealth University

Abstract. A growing body of research exists on the effectiveness of classroom-based intervention programs to prevent and ameliorate social, emotional, and learning difficulties demonstrated by young children at risk for emotional and behavioral disorders (EBD). Yet, little research has examined the influence of these targeted intervention programs on the teachers who are trained to deliver them. Impacts of the professional development associated with the intervention on teachers who implement the intervention are important to examine. Data from a 4-year study examining the efficacy of BEST in CLASS were used to examine the effect of BEST in CLASS on teachers' implementation of effective instructional practices, their sense of self-efficacy, and classroom quality. Using a multisite cluster randomized trial, a total of 186 early childhood teachers were included (92 assigned to BEST in CLASS and 94 assigned to a comparison group). Findings indicate BEST in CLASS positively impacted teachers' use of effective instructional practices, their sense of self-efficacy, and their overall classroom quality compared to teachers in the control condition. Future research and implications for professional development are discussed.

Keywords: early childhood, professional development, emotional and behavioral disorders

More than 2.5 million young children are served by federal and state-funded early childhood programs (Barnett et al., 2017). Unfortunately, many of these children demonstrate problem behaviors that negatively impact learning (Carter et al., 2010) and academic outcomes (Brennan, Shaw, Dishion, & Wilson, 2012; Bulotsky-Shearer, Bell, & Domínguez, 2012). Children who engage in elevated rates of problem behavior tend to develop coercive and negative

Authors' Notes: This research was supported by a grant (R324A110173) from the U.S. Department of Education, Institute for Education Sciences, with additional support from a National Institutes of Health/National Center for Advancing Translational Sciences Clinical and Translational Science Award to the University of Florida UL1 TR000064. The opinions expressed by the authors are not necessarily reflective of the position of or endorsed by the U.S. Department of Education.

Please address correspondence regarding this article to Maureen Conroy, Box 117050, School of Special Education, School Psychology, and Early Childhood Studies, University of Florida, Gainesville, FL 32611; e-mail: mconroy@coe.ufl.edu

interaction patterns with their teachers (Patterson, Reid, & Dishion, 1992), which have been linked to negative outcomes in school and beyond (Barnett & Boocock, 1998; Howes & Smith, 1995). In general, early childhood teachers are not well prepared to effectively address children's significant behavioral challenges (Hemmeter, Santos, & Ostrosky, 2008). Teachers' lack of training, knowledge, and skills to address the needs of young children exhibiting significant problem behaviors in their classrooms is particularly troublesome given the increasing number of children attending early childhood programs and the impact problem behavior can have on their future successes. Without preventive and targeted interventions implemented by early childhood teachers, the long-term outcomes for these young children is a significant concern.

Overview of BEST in CLASS

BEST in CLASS (Conroy & Sutherland, 2008) was developed to increase teachers' use of effective instructional practices that address problem behaviors and improve the quality of interactions with young children who demonstrate them. Founded in behavioral (Skinner, 1953) and transactional (Sameroff, 1983) theories, BEST in CLASS is a targeted intervention designed to increase teachers' intentional use of the BEST in CLASS practices during naturally occurring classroom activities with children who are demonstrating elevated rates of problem behavior and are at risk for developing emotional and behavioral disorders (EBD). Specifically, BEST in CLASS encourages embedding these practices into reciprocal interactions between teachers and target children (who are at risk for EBD) repeatedly during naturally occurring classroom-based activities. Over time, through professional development activities including a teacher training workshop and practice-based coaching, teachers master the BEST in CLASS practices and increase the quantity and quality of their use during times when focal children engage in high rates of problem behavior. The BEST in CLASS practices include rules, precorrection, opportunities to respond, behavior-specific praise, corrective feedback, and instructive feedback. Although these practices are effective teaching practices used by many teachers, BEST in CLASS trains teachers to implement these practices more intentionally and at a higher rate, directing them toward focal children in their classrooms. For example, if a focal child engages in high rates of disruption and noncompliance during morning circle, the teacher would remind the child about the classroom rules with a higher frequency, provide the child additional opportunities to respond, and deliver increased rates of behavior-specific praise when the child engages in appropriate classroom behavior. The teacher might also provide a precorrection statement before the activity to remind the child of the behavioral expectations of morning circle. Through increased direct use of the BEST in CLASS practices with focal children, negative and coercive interactions between the teacher and focal children decrease and positive teacher–child interactions increase; children also demonstrate fewer problem behaviors and increase their engagement (Sutherland et al., 2018). Additionally, we postulated that as interactions between teachers and focal children become more positive, teachers' sense of self-efficacy will increase (Pas, Bradshaw, & Hershfeldt, 2012), including their beliefs about their ability to manage their learning environments, deliver effective teaching practices, and accommodate individual differences. Finally, we posit that with more positive interactions between teachers and focal children, the overall quality of classroom atmosphere will increase. (For a more comprehensive overview of the BEST in CLASS intervention, including definitions of the practices and a theoretical model, see Conroy et al., 2015; Sutherland et al., 2018.)

Findings from a randomized controlled trial (Sutherland et al., 2018) investigating the effects of BEST in CLASS on child outcomes indicated that children who were enrolled in BEST in CLASS classrooms had significant increases in task orientation and task engagement, improvements in social skills, and decreases in externalizing problem behaviors, including disruption. Additionally, for children and teachers in BEST in CLASS classrooms, there was a significant increase in positive teacher–child interactions and closeness as well as a significant decrease in conflict compared to children and teachers in control classrooms. Although positive outcomes were found for children in this study, the influence of the BEST in CLASS professional development model on teacher outcomes has not been reported. As suggested by Fixsen, Naoom, Blase, Friedman, and Wallace (2005), it is important to examine the effectiveness of the intervention on child outcomes, but it is also important to distinguish between the outcomes of the intervention itself and the effectiveness of the professional development strategies employed to train those who implement the intervention (see Snyder, Hemmeter, & McLaughlin, 2011). Likewise, it is important to examine additional factors that might influence adoption and use of the intervention, such as a teacher's sense of self-efficacy (Han & Weiss, 2005) and any corollary effects of the intervention on other classroom variables, such as overall classroom quality.

Early Childhood Professional Development Research

Most classroom-based interventions include a professional development component to promote the adoption and use of the intervention by teachers. Recently, researchers have begun to examine professional development components that can facilitate teacher implementation of effective practices and interventions, suggesting that it is essential to provide teachers with high-quality, comprehensive professional development to facilitate actual use of interventions in the classroom. In a comprehensive review, Snyder et al. (2011) suggested the need to define, identify, and analyze the active ingredients of professional development related to teacher and child outcomes. Research suggests that when teachers receive high-quality professional development, including demonstrations in training, practice and feedback, and coaching, they are

more likely to implement intervention practices in their classrooms (Joyce & Showers, 2002; Snyder et al., 2012; Snyder, Hemmeter, & Fox, 2015).

In addition to increasing teachers' use of evidence-based practices, researchers found that when teachers receive high-quality professional development, there are a number of other benefits for teachers, including increases in their self-efficacy (Bradshaw, Koth, Thornton, & Leaf, 2009; Han & Weiss, 2005; Pas et al., 2012). Recent studies indicate early childhood teachers who receive training and professional development in implementing social–emotional programs experience improvements in classroom quality and the use of preventive behavior management strategies, lower levels of burnout, and higher levels of personal accomplishment (Domitrovich et al., 2016; Domitrovich et al., 2009). In a meta-analysis of targeted interventions in early childhood settings, Werner, Linting, Vermeer, and Van IJzendoorn (2016) found that individual-level training (i.e., coaching) moderated positive classroom and child effects such that programs including individual teacher-level training had greater effects.

In summary, providing young children who demonstrate chronic problem behavior with effective instruction and supports to be successful in school is critical during their early years. Several targeted interventions, including BEST in CLASS, have been found to improve social, emotional, and behavioral outcomes for young children; yet, limited research exists investigating the influence of these programs on teachers who act as intervention implementers. Building on the research of others and considering the potential impact on child outcomes, the influence of interventions on teacher implementation and corollary outcomes is important to consider. The current investigation used data from a randomized controlled trial examining the impact of the BEST in CLASS intervention on children's social, emotional, and behavioral outcomes. The aim of this study was to examine the influence of the BEST in CLASS professional development model on (a) teachers' implementation of effective instructional practices with young children who demonstrate chronic problem behaviors, (b) teacher self-efficacy, and (c) overall classroom quality.

METHOD

Data for this study were from a 4-year investigation examining the efficacy of the BEST in CLASS intervention conducted across two southeastern states, subsequently referred to as research sites. Research activities were conducted in federal and state-funded early childhood classrooms. At the two research sites, classrooms were located within public elementary schools ($n = 50$) or agency-based early childhood centers ($n = 28$) across five different school districts and three Head Start agencies. The mean number of children and adults per classroom was 17.55 and 2.21, respectively.

Randomization occurred at the teacher level with children nested in teachers and teachers in early childhood schools and centers (hereafter referred to collectively as schools). All teachers were randomly assigned from within schools to either the BEST in CLASS condition or the comparison (business-as-usual; BAU) condition (see Figure 1). Teachers in both the BEST in CLASS and comparison conditions continued to implement their existing early childhood curricula according to program requirements and policies, while the teachers in the BEST in CLASS condition also received the three core professional development components of the BEST in CLASS intervention (i.e., BEST in CLASS teacher manual, BEST in CLASS teacher workshop, and 14 weeks of practice-based coaching) described later. All research activities at both research sites were approved by their respective institutional review boards.

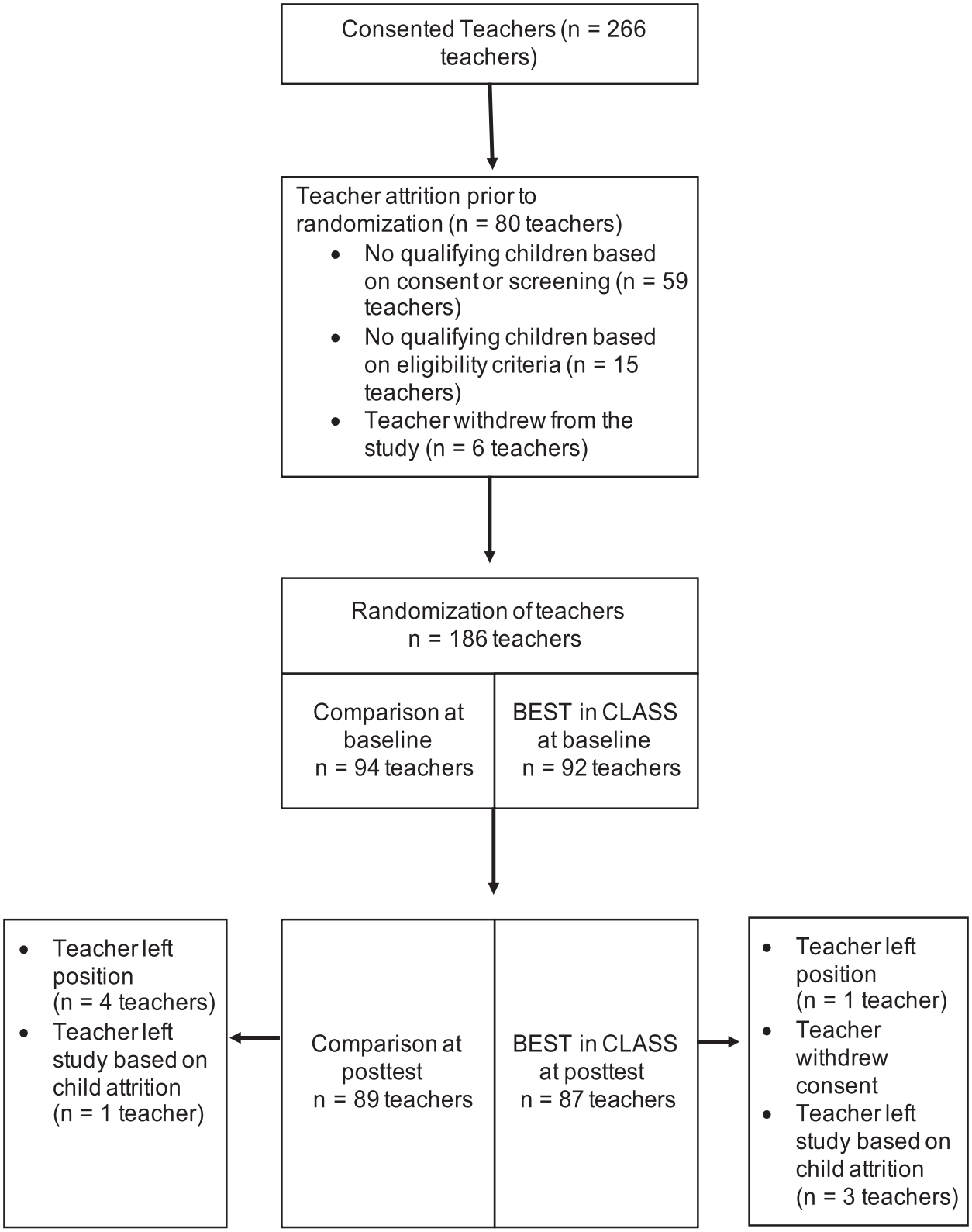
Participants

A total of 186 teachers participated in the study, with 92 teachers in the BEST in CLASS intervention group and 94 teachers in the comparison group. All teachers consented to participate and were eligible if they taught in early childhood classrooms and had children in their classrooms that were systematically identified as being at risk for EBD. Demographics across both the intervention and the comparison group were similar. Of the participating teachers, 181 were female, four were male, and one teacher did not report gender. Approximately 47.3% of teachers reported their race or ethnicity as African American, 46.2% as Caucasian, 2.7% as Hispanic, and 2.7% as other, and 1.1% did not report. Among all teachers, the average number of years of teaching experience was 12.09. Approximately 65% of teachers reported having a bachelor's degree or higher level of education. All teachers participated in the study for only 1 year and were compensated for their participation with a $100 stipend.

Within each classroom, teachers identified one to three children to participate based on the following criteria: (a) externalizing problem behavior that interfered with participation in the classroom; (b) risk for EBD using the Early Screening Project (ESP; Feil, Severson, & Walker, 1998); (c) no evidence of a cognitive delay based on the Battelle Developmental Inventory, Second Edition screener (BDI-II screener; Newborg, 2005); and (d) parental or guardian consent. A total of 465 children participated in the study (231 in the BEST in CLASS condition and 234 in the comparison condition).

In addition to teachers and children, a total of 26 coaches (24 female) provided practice-based coaching to the teachers in the BEST in CLASS condition. A majority of coaches ranged in age from 26–35 years old; 19 coaches were Caucasian, four were African American, and three were Hispanic. Most held a master's degree (57.7%) and had prior teaching experience (65.4%). Half (50%) were enrolled in a graduate degree program seeking a master's or doctoral degree in education or a related field (e.g., school psychology, special education, counselor education).

Figure 1. Consort Flow Diagram



Data Collection and Management

Data were collected on teacher implementation of instructional practices, classroom quality, and teacher self-efficacy in both the intervention and control conditions at pretest (i.e., prior to professional development) and posttest (i.e., after coaching ended). In addition, fidelity data were collected on the dosage and implementation of BEST in CLASS practice-based coaching and teachers' implementation of BEST in CLASS practices. Pretest data on teachers were collected within the first months of the school year (i.e., September and October). Posttest data were collected following the implementation of the full intervention and typically occurred several months prior to the end of the school year (e.g., April and May). Fidelity measures occurred at various time points throughout the study, as indicated.

Once collected, data were entered into the REDCap system (Harris et al., 2009), a secure, Web-based application designed to store and track research data, housed at one of the research sites. All data were entered twice into main and shadow databases by two separate data entry staff members and compared for accuracy. Upon conclusion of the study, data were exported from REDCap for data analysis.

Teacher Outcome Measures

The observational measure Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008) was used to measure classroom quality. Each CLASS observation consisted of four 10- to 20-min observation cycles during instructional time (e.g., circle time, centers). Classroom quality was rated along 10 dimensions using a scale from 1 to 7, with 7 representing quality indicators at the highest level. Dimension scores were averaged to obtain mean composite scores across three domains: emotional support, classroom organization, and instructional support. For the current sample, internal consistency was acceptable, with Cronbach's alpha equal to .88 for emotional support, .89 for classroom organization, and .85 for instructional support.

Certified observers conducted CLASS observations at pretest and posttest. To obtain certification, observers participated in a 2-day training workshop led by a certified CLASS trainer and completed the reliability test required for initial certification. To ensure reliability throughout the study, all certified CLASS observers recalibrated or recertified prior to pretest data collection and again prior to posttest data collection. Interobserver agreement (IOA) data were collected on 22.87% of all CLASS observations using a secondary observer. IOA was calculated using the formula for point-by-point agreement (i.e., the ratio between the total number of agreements [A] and the total number of agreements plus disagreements [A + D]). Agreement was defined as primary and secondary observer scores for the same construct that were within one point of difference. The mean IOA estimate was 93.11%.

To examine teacher implementation of the BEST in CLASS practices, a researcher-developed observational measure was used to examine teachers' implementation of BEST in CLASS instructional practices: Teacher–Child Interactions Direct Observation System (TCIDOS; Sutherland, Conroy, Vo, Abrams, & Ogston, 2013). The TCIDOS is a partial interval-based (i.e., 10-s observation interval, 5-s record interval) observational tool that measures teachers' use of BEST in CLASS practices with focal children during instructional times in the classroom. Using Lily software (Tapp, 2010) and tablets preprogrammed with TCIDOS codes, observers conducted 10- to 15-min observations of teachers' implementation of the BEST in CLASS practices directly with each focal child in the classroom during a planned classroom activity (e.g., circle time, small group) in both BEST in CLASS and comparison classrooms at pretest and posttest. Each observation of a teacher–child dyad occurred in a single day. During the observation, observers recorded the teachers' use of the BEST in CLASS practices as well the occurrence of focal child behaviors (i.e., engagement, problem behavior) and teacher–child interactions (i.e., positive or negative). Following each observation session, files were analyzed using INTMAN software (Tapp, 2003) to calculate the percentage of intervals in which target behaviors occurred. This procedure resulted in a score for each teacher–child pair on each TCIDOS variable.

To ensure reliability, observers were trained to a gold standard criterion and required to complete an IOA checkout process before using the TCIDOS. Additionally, recalibration occurred at least three times during the course of the study (i.e., prior to pretest, midpoint, and posttest). Calibration and recalibration required observers to code three 15-min master-coded video segments and reach at least 80% IOA across all codes. If agreement was not reached on the first three videos, the observer watched additional videos until this standard was reached. If a discrepancy occurred (less than 80% IOA) between the observer and the master code, a designated master coder at each research site clarified discrepancies and the observer conducted additional coding as needed until the observer met the 80% criterion. IOA was conducted on 23.88% of 1,202 TCIDOS observations during the study. IOA was calculated as the number of agreements divided by the total number of agreements and disagreements multiplied by100 ($A/A + D \times 100$). Both research sites maintained a mean IOA score of at least 80% across all codes, with the mean estimates of IOA across all codes ranging from 88.68% to 99.37%. Additionally, intraclass correlation coefficients (ICCs) were calculated. Across all teacher codes, the mean ICC was 0.75 ($SD = 0.17$).

Teacher self-efficacy was measured using two rating scales completed by teachers at pretest and posttest: Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Hoy, 2001) and Teachers' Efficacy Beliefs System (TEBS; Dellinger, Bobbett, Olivier, & Ellett, 2008). The TSES focuses on teachers' sense of efficacy in engaging students and managing classroom activities, whereas the TEBS examines teachers' personal beliefs regarding their ability to successfully conduct specific teaching tasks within the instructional context.

The TSES is a rating scale that includes 24 items with ratings ranging from 1 (*nothing*) to 9 (*a great deal*) across three subscales: Student Engagement (i.e., the extent to which a teacher perceives capacity for engaging students in the classroom and fostering their learning), Instructional Strategies (i.e., the extent to which a teacher perceives resourcefulness in providing individualized support to students), and Classroom Management (i.e., the extent to which a teacher perceives capacity for managing the classroom and children with disruptive behaviors). For the current sample, internal consistency was acceptable, with Cronbach's alpha equal to .86 for Student Engagement, .90 for Instructional Strategies, and .89 for Classroom Management.

The TEBS (Dellinger et al., 2008) is a rating scale that includes 31 items with competency ratings ranging from 1 (*weak*) to 4 (*very strong*). There are four subscales: Accommodating Individual Differences (i.e., teacher beliefs regarding ability to address individual child differences and maximize children's learning), Maintaining Positive Classroom Climate (i.e., teacher beliefs regarding ability to maintain a classroom climate that is fair, positive, and courteous), Monitoring and Feedback for Learning (i.e., teacher beliefs regarding ability to monitor students' understanding of learning content and provide students with clarifications, suggestions, and specific feedback), and Managing Learning Routines (i.e., teacher beliefs regarding ability to provide directions, set routines, and allocate time to maximize students' learning). For the current sample, internal consistency was acceptable, with Cronbach's alpha equal to .85 for Accommodating Individual Differences, .84 for Maintaining Positive Classroom Climate, .87 for Monitoring and Feedback for Learning, and .81 for Managing Learning Routines.

Fidelity Measures

During the intervention, coaching dosage was self-recorded by coaches following weekly coaching meetings with each teacher. Dosage data included the duration of the entire coaching meeting, the occurrence and duration of coaching components (i.e., modeling and prompting) within the coaching meeting, and the percentage of sessions in which required elements of the coaching protocol were completed. Coaches also recorded the frequency of follow-up emails, phone calls, and visits. The average coaching dosage per teacher was calculated using the duration (in minutes) and frequency of coaching components within coaching meetings.

To measure the extent to which BEST in CLASS practice-based coaching was implemented as designed, each coach was observed a minimum of two times over the 14-week coaching period of the intervention by independent observers using the BEST in CLASS Coaching Integrity form. The BEST in CLASS Coaching Integrity form is divided into five domains: (a) reflection and feedback, (b) strategy instruction/review, (c) shared goal setting and decision making, (d) general items, and (e) quality collaboration. Each domain includes quality indicators coded on a scale from 0 to 2 using the BEST in CLASS Coaching Integrity manual. A score of 0 means the indicator was not observed or was below the minimum quality standard acceptable. A score of 1 indicates that the coach met the minimum standard for the indicator, and a score of 2 indicates that the coach exceeded the minimum standards for all coaching behaviors within the indicator. Intermediate scores of 0.5 and 1.5 were coded when coaches demonstrated some but not all of the coaching behaviors required for a particular indicator. To ensure coaching fidelity prior to beginning the intervention, an intensive coach training and checkout process was completed (as described in the following paragraph). In addition, weekly coaching meetings were held with research staff to discuss coaching progress and review coaching procedures and skills.

Prior to measuring coaching fidelity, observers at each research site were trained on the procedures outlined in the BEST in CLASS Coaching Integrity manual. Training included memorizing and discussing the coding definitions as outlined in the manual and coding master videos of coaching meetings. Observers at each research site were required to achieve at least 80% IOA on each code prior to coding any coaching sessions. To sustain IOA, trained observers watched and coded an additional two master-coded videos in the fall and spring. Again, observers had to achieve at least 80% IOA on each code prior to coding any additional sessions. If a score of 80% was not attained, the observer and a master coder participated in a discrepancy discussion. The observer then reviewed the manual and coded an additional video, repeating this process until 80% IOA on two videos was achieved. IOA data were collected on 24% of all coaching fidelity observations and were calculated by the number of agreements divided by the number of agreements plus disagreements multiplied by 100. A mean IOA score of 90% was obtained (range = 80%–100%).

The BEST in CLASS Adherence and Competence Scale (BiCACS; Sutherland, McLeod, Conroy, Abrams, & Smith, 2014) was used to assess the delivery of BEST in CLASS practices by teachers in both intervention and control conditions. The BiCACS is an observational measure that assesses teachers' adherence and competence of delivery of the BEST in CLASS practices with focal children. The adherence scale measures extensiveness of delivery of the BEST in CLASS practices. Using a 7-point Likert type scale ranging from 1 (*not at all*) to 7 (*very extensive*), adherence is based on thoroughness, frequency, rate, and duration of the observed teacher behavior. The competence scale measures quality of delivery of the BEST in CLASS practices using a 7-point Likert type scale ranging from 1 (*very poor*) to 7 (*excellent*). Competence is based on the teacher's skillfulness, responsiveness, timing, and appropriateness in implementing the intervention practices.

Prior to using the BiCACS, observers were provided a manual and trained on procedures for administering and scoring the BiCACS. Following training, observers conducted 756 observations across both conditions (BEST in CLASS

n = 389; BAU n = 367) at pre- and posttest. Reliability checks were conducted on 27% (n = 204) of BiCACS observations. Throughout the study, ongoing calibration occurred between observers immediately following reliability observation sessions. Primary and secondary observers compared ratings on an item-by-item level. Ratings that were within ±1 point (e.g., a rating of 4 and a rating of 5) were considered reliable. Discrepancy discussions were conducted on any ratings that deviated from this rule. ICCs were computed for each item. The mean ICC on the adherence scale was .69 (SD = 0.08). The mean ICC on the competence scale was .44 (SD = 0.12).

Experimental Procedures

Coaches were initially trained in the BEST in CLASS intervention and the practice-based coaching procedures. Following coach training, pretest data collection, and randomization, the BEST in CLASS intervention was implemented with teachers. All coaches received manualized training in delivering BEST in CLASS practice-based coaching prior to working with teachers. This 2-day training included an overview of the BEST in CLASS intervention, coaching framework, and coaching manual; exemplar videos of practice-based coaching sessions; and practice conducting coaching sessions. To demonstrate mastery, coaches completed a six-step checkout process that involved pairing up with a coaching partner, role-playing a coaching session using BEST in CLASS coaching procedures and forms, recording the coaching session, completing a self-evaluation, and submitting the video and forms to be evaluated by a master coaching observer. In addition to demonstrating proficiency in implementation of the practice-based coaching process (i.e., shared goals and action planning, focused observation, reflection and feedback), successful completion of coaching checkout required coaches to demonstrate essential coaching skills including effective communication (e.g., open-ended questioning, avoiding judgment) and collaboration (e.g., joint problem solving). Master coaching observers used the BEST in CLASS Coaching Integrity form and manual to measure each coach's level of proficiency and determine whether or not each coach demonstrated integrity in implementing the coaching process and skills before allowing them to begin coaching.

BEST in CLASS includes the following three manualized professional development components for teachers: (a) BEST in CLASS teacher training workshop, (b) BEST in CLASS teacher manual, and (c) BEST in CLASS practice-based coaching. During the workshop, teachers were provided an overview of the BEST in CLASS intervention, each of the key instructional practices included in the intervention (i.e., rules, precorrection, opportunities to respond, behavior-specific praise, corrective feedback, and instructive feedback), and information about how the practice-based coaching process will occur. Teachers used the manual as a resource to gain further knowledge, understanding, and mastery of the practices. It provides a more in-depth explanation of each BEST in CLASS practice, how to link the practices together for optimal use, and how to share information about the practices with caregivers. Practice-based coaching was conducted weekly over 14 weeks and provided support to teachers in implementing the BEST in CLASS practices with focal children with high fidelity. For definitions of the practices and more detail about the professional development model, see Sutherland, Conroy, Vo, and Ladwig (2015).

The BEST in CLASS teacher training, a 6-hr workshop, helped teachers gain initial knowledge about BEST in CLASS and learn how the practices can be implemented with focal children during classroom activities. Teachers were also provided with information on the quality components of each BEST in CLASS practice and shown exemplar videos demonstrating the use of the BEST in CLASS practices during authentic classroom activities.

The BEST in CLASS teacher manual extended teachers' knowledge about each of the BEST in CLASS practices and how to implement them with high fidelity in their classrooms. The manual is comprised of seven modules (i.e., one module per practice and a final module that helps link practices together). To facilitate teacher–family communication, it includes strategies for communicating with children's caregivers about their child's behavior and how they might implement the BEST in CLASS practices at home.

Following the workshop, coaches worked with their assigned teachers for 14 weeks on the seven modules outlined in the manual (i.e., 2 weeks per module) to ensure that teachers were able to implement the BEST in CLASS practices with high quality during authentic classroom activities. According to the practice-based coaching model (for an overview of practice-based coaching, see Snyder et al., 2015), BEST in CLASS coaching was designed to encourage teachers to actively participate in the cyclical (i.e., weekly) coaching process and included three key components: (a) goal setting and action planning, (b) focused observations, and (c) reflection and performance feedback. During goal setting and action planning, coaches worked collaboratively with teachers to create specific, measurable, and reasonable weekly goals regarding teachers' implementation of the BEST in CLASS practices with focal children. Goal setting and action planning were followed by focused observations of teachers' implementation of the BEST in CLASS practices; coaches observed for approximately 45 min per week, and a practice-based coaching meeting was typically held on the same day as the observation. Coaches observed teachers' use of specific BEST in CLASS practices and teacher–child interactions. Observations were video-recorded and coaches collected data on teachers' use of the practices. Following observations, teachers and coaches held a coaching meeting at which the coaches reviewed teacher progress toward weekly goals, asked teachers to self-reflect on their implementation of the BEST in CLASS practices, provided teachers with feedback using the data collected during observations, and collaboratively planned teacher goals for the following week.

Design and Analysis

The design of the study was a multisite (i.e., multiple schools, not research sites) cluster randomized trial (Spybrook et al., 2011), with teachers randomly assigned to treatments from within schools. M*plus* 8.0 (Muthén & Muthén, 1998–2017) was used to analyze the data. To account for nesting of teacher–child pairs in teachers and teachers in schools, the two-level complex procedure was used to analyze TCIDOS variables, which were recorded for each teacher–child pair. The two-level complex procedure provides full information maximum likelihood estimates for a two-level model and cluster-robust standard errors (McNeish, Stapleton, & Silverman, 2017) that account for nonnormality and nesting in schools. The fixed effects included treatment, research site, Treatment by Research Site (T × S) interaction, a coefficient for a child-level covariate, a coefficient for a teacher-level covariate, and Covariate × Treatment (C × T) interactions at the child- and the teacher-levels. The child-level covariate was the class mean–centered pretest, and the teacher-level covariate was the grand mean–centered teacher-level mean pretest. The treatment factor was coded −.5 for the BAU comparison group and .5 for the BEST in CLASS group. The research site factor was coded −.5 for Research Site A and .5 for Research Site B. When either interaction was nonsignificant, it was deleted from the model and data were reanalyzed.

M*plus* 8.0 was also used to analyze TEBS, TSES, and CLASS data. To account for nesting of teachers in schools, the complex procedure was used. The complex procedure provides single-level full information maximum likelihood estimates and cluster-robust standard errors that account for nonnormality and nesting of teachers in schools. The fixed effects included treatment, research site, Treatment by Research Site (T × S) interaction, and coefficients for the teacher-level covariate and Covariate × Treatment (C × T). If the C × T interaction was not statistically significant, it was deleted from the model and the data were reanalyzed.

For each analysis, the covariate was brought into the model, and therefore the sample size for an analysis was the number of cases with pretest data, posttest data, or both. Sample sizes were 461 for TCIDOS, 185 for TEBS and TSES, and 186 for CLASS. The Benjamini–Hochberg false discovery rate (BHFDR) procedure with a false discovery rate of .05 was used to determine whether hypothesis tests were statistically significant. The false discovery rate is the expected proportion of a set of null hypotheses that are falsely rejected. Controlling the false discovery rate is more conservative than using a standard .05alpha level, for example, and should result in more powerful tests than using a procedure that controls the family-wise error rate like the Bonferroni procedure (see U.S. Department of Education, 2016 for additional information). The BHFDR procedure was applied to each set of scales for the TCIDOS, TEBS, TSES, and CLASS.

Attrition and Missing Data

During the course of the study, five teachers in each of the BEST in CLASS and comparison groups withdrew from the study. The total attrition rate was 5.38%, and the differential attrition rate (i.e., difference in attrition rate for the BEST in CLASS and comparison groups) was 0.12%. In addition, for each of the TEBS, TSES, and CLASS subscales, there were missing data for teachers who did not withdraw. Taking into account missing posttest data due to attrition and missing data for teachers who did not withdraw, the total missing data rate ranged from 5.91% to 8.06% across TEBS, TSES, and CLASS subscales, with a differential missing data rate that ranged from 0.90% to 2.31%. Using Table III.1 from the What Works Clearinghouse (WWC) *Procedures and Standards Handbook Version 3.0* (U.S. Department of Education, 2013), attrition and missing data rates were considered low.

According to the procedures for cluster designs presented by WWC (U.S. Department of Education, 2016), missing data rates should be assessed for children of teachers who did not withdraw from the study. Applying this procedure to TCIDOS data, there was a missing data rate of 15.37% and a differential missing data rate of 6.48%. Under WWC's liberal assumption about the relationship of posttest variables and the propensity to complete the study, these missing data rates are considered low. According to WWC, this may be a reasonable assumption when attrition arises, as it did in the current study, due to the movement of young children in and out of school districts as a result of family mobility and random absences on the days that assessments were conducted.

RESULTS

To examine baseline equivalence, chi-square tests comparing BEST in CLASS and control condition distributions for the demographic variables were conducted using PROC SURVEYFREQ in SAS 9.4, with school as the cluster. This procedure takes clustering of teachers in schools into account. None of the chi-square tests were significant using the raw *p* values or the BHFDR-adjusted *p* values. Descriptive statistics by intervention group, research site (i.e., state), occasion, and variable are presented in Table 1. The two-level complex procedure in M*plus* 8.0 was used to compare BEST in CLASS and control group means of the pretest TCIDOS variables. The complex procedure in M*plus* 8.0 was used to compare BEST in CLASS and control group means for TEBS, TSES, and CLASS pretest variables. Using the BHFDR procedure across all subscales, there were no significant differences between the BEST in CLASS and comparison group.

Estimates and Hypothesis Tests

In this section, findings are reported from the TCIDOS, TEBS, TSES, and CLASS.

Results of parameter estimation and hypothesis tests for the data obtained from the TCIDOS are presented in Table 2,

Table 1. Means and Standard Deviations for the TCIDOS, TEBS, TSES, and CLASS

			BEST in CLASS						Comparison Group					
			Pretest			Posttest			Pretest			Posttest		
Instrument	Scale	Site	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
TCIDOS	Rules	A	115	0.02	0.05	93	0.13	0.10	121	0.01	0.03	92	0.02	0.05
		B	112	0.01	0.02	103	0.08	0.07	111	0.01	0.02	92	0.01	0.01
	PC	A	115	0.01	0.02	93	0.07	0.09	121	0.01	0.02	92	0.01	0.06
		B	112	0.00	0.01	103	0.05	0.05	111	0.00	0.01	92	0.01	0.01
	OTR	A	115	0.36	0.20	93	0.54	0.20	121	0.34	0.19	92	0.31	0.20
		B	112	0.42	0.18	103	0.50	0.18	111	0.36	0.18	92	0.36	0.15
	BSP	A	115	0.01	0.02	93	0.07	0.06	121	0.01	0.02	92	0.01	0.02
		B	112	0.01	0.02	103	0.06	0.05	111	0.00	0.01	92	0.01	0.01
	CF	A	115	0.01	0.02	93	0.10	0.10	121	0.01	0.02	92	0.03	0.06
		B	112	0.01	0.02	103	0.05	0.04	111	0.01	0.01	92	0.01	0.02
	IF	A	115	0.01	0.02	93	0.11	0.13	121	0.01	0.02	92	0.02	0.05
		B	112	0.02	0.03	103	0.07	0.08	111	0.01	0.02	92	0.01	0.02
TEBS	AID	A	46	2.97	0.55	42	3.39	0.48	46	3.21	0.53	43	3.32	0.47
		B	46	2.71	0.51	45	3.44	0.47	46	2.90	0.46	45	3.07	0.61
	MPCC	A	46	3.38	0.44	42	3.63	0.42	46	3.38	0.58	43	3.42	0.61
		B	46	3.34	0.56	45	3.73	0.41	47	3.32	0.56	45	3.39	0.61
	MLR	A	46	3.08	0.46	42	3.50	0.46	46	3.32	0.56	43	3.43	0.55
		B	46	2.99	0.67	45	3.60	0.47	47	3.02	0.56	45	3.24	0.51
	MFL	A	46	3.00	0.47	42	3.37	0.50	46	3.22	0.50	43	3.26	0.51
		B	46	2.87	0.56	45	3.55	0.47	47	3.00	0.45	45	3.08	0.48
TSES	CM	A	46	6.86	1.04	41	7.63	0.85	45	7.08	1.06	43	7.33	0.76
		B	46	7.02	0.91	45	7.83	0.79	47	6.88	1.08	45	6.89	0.77
	ISt	A	45	7.15	0.95	42	7.83	0.67	46	7.33	1.08	43	7.63	0.81
		B	46	7.20	0.99	45	7.81	0.86	47	7.10	0.94	45	7.09	0.83
	SE	A	46	7.00	0.98	39	7.66	0.69	45	7.11	1.01	43	7.46	0.73
		B	46	7.06	0.99	45	7.55	0.87	46	7.09	1.03	45	6.92	0.83
CLASS	CO	A	44	4.63	1.00	40	5.17	0.95	44	4.45	0.92	42	4.32	0.97
		B	46	5.27	0.60	45	5.57	0.72	48	5.17	0.77	44	5.29	0.64
	ES	A	44	5.04	0.85	40	5.61	0.82	44	5.00	0.86	42	5.00	0.73
		B	46	5.46	0.65	45	5.62	0.88	48	5.30	0.92	44	5.41	0.64
	ISu	A	44	2.40	0.83	40	2.94	1.16	44	2.24	0.92	42	2.19	0.81
		B	46	2.18	0.70	45	2.94	0.95	48	2.24	0.84	44	2.45	0.78

Note. TCIDOS = Teacher–Child Interactions Direct Observation System; TEBS = Teachers' Efficacy Beliefs System; TSES = Teachers' Sense of Efficacy Scale; CLASS = Classroom Assessment Scoring System; PC = precorrection; OTR = opportunities to respond; BSP = behavior-specific praise; CF = corrective feedback; IF = instructive feedback; AID = Accommodating Individual Differences; Maintaining Positive Classroom Climate; MLR = Managing Learning Routines; MFL = Monitoring and Feedback for Learning; CM = Classroom Management; ISt = Instructional Strategies; SE = Student Engagement; CO = classroom organization; ES = emotional support; ISu = instructional support.

Table 2. Analysis Results for Teacher Outcome Variables: Teacher–Child Interactions Direct Observation System

	BSP			CF			IF		
Effect	*B*	*SE*	*z*	*B*	*SE*	*z*	*B*	*SE*	*z*
Treatment (T)	0.06	.01	9.85*	0.05	.01	6.00*	0.08	.01	6.96*
Site (S)	–0.01	.01	–1.03	–0.04	.01	–2.82*	–0.03	.01	–1.93
T × S	–0.01	.01	–1.05	–0.04	.02	–2.44	–0.03	.02	–1.13
Centered Covariate (CC)	–0.29	.20	–1.41	–0.43	.16	–2.64	0.09	.11	0.87
Mean Covariate (MC)	0.06	.20	–1.42	1.07	.61	1.75	–0.39	.29	–1.34
	OTR			PC			Rules		
	B	*SE*	*z*	*B*	*SE*	*z*	*B*	*SE*	*z*
Treatment (T)	0.18	.02	7.58*	0.05	.01	7.44*	0.09	.01	11.55*
Site (S)	0.00	.02	0.09	–0.02	.01	–1.96	–0.03	.01	–2.66*
T × S	–0.09	.05	–1.81	–0.02	.01	–1.23	–0.03	.02	–1.79
Centered Covariate (CC)	0.11	.07	1.52	–0.35	.02	1.45	0.01	.13	0.11
Mean Covariate (MC)	0.08	.07	1.10	–0.24	.39	–0.61	0.18	.26	0.69

Note. BSP = behavior-specific praise; CF = corrective feedback; IF = instructive feedback; OTR = opportunity to respond; PC = precorrection.
**Significant* controlling false discovery rate at .05.

with footnotes to indicate whether a test was significant using the BHFDR procedure. There were no significant C × T interactions for any of the TCIDOS variables. The distribution of all variables other than opportunity to respond (OTR) from TCIDOS were substantially nonnormal. Although the robust maximum likelihood method is designed to correct standard errors for nonnormality, a dichotomized version of each variable other than OTR was created in which scores greater than 0 were set equal to 1. The variables were analyzed using a two-level logistic regression model and the two-level complex procedure. The treatment effect was significant for all variables.

Results indicate there was a significant treatment effect for each of the TCIDOS variables and a significant research site effect for corrective feedback and rules. Adjusted means averaged across research sites, denoted by AM_B for BEST in CLASS and AM_C for the BAU comparison group, were $AM_B = .07$ and $AM_C = .01$ for behavior-specific praise, $AM_B = .08$ and $AM_C = .02$ for corrective feedback, $AM_B = .10$ and $AM_C = .02$ for instructive feedback, $AM_B = .53$ and $AM_C = .34$ for OTR, $AM_B = .06$ and $AM_C = .01$ for precorrection, and $AM_B = .11$ and $AM_C = .01$ for rules. OTR was observed during a substantial proportion of the observation intervals and was much larger for BEST in CLASS. In general, other TCIDOS variables were observed less frequently overall but relatively more frequently in BEST in CLASS classrooms compared to control classrooms. Although TCIDOS variables were observed infrequently, the magnitude of the difference between pretest and posttest for the BEST in CLASS condition was noteworthy. With the exception of OTR, which in general was observed to occur more often, teachers in the BEST in CLASS and comparison conditions displayed the practices during pretest at very low levels. However, at posttest, teachers in the BEST in CLASS condition increased their use of rules, precorrection, behavior-specific praise, corrective feedback, and instructive feedback from pretest levels compared to the control condition.

Results of parameter estimation, hypothesis tests, and effect size (Cohen's *d*) estimation are presented in Table 3 for TEBS, TSES, and CLASS. Across all variables, the only significant C × T interaction was for Maintaining Positive Classroom Climate on the TEBS. Results in Table 3 for variables other than Maintaining Positive Classroom Climate are for a model with the C × T interaction excluded.

Results for the TEBS are reported for each subscale. Accommodating Individual Differences and Monitoring and Feedback for Learning subscales indicate treatment effects that were moderated by research site. For the Accommodating Individual Differences subscale, adjusted means for BEST in CLASS and comparison groups were 3.38 and 3.19 in Research Site A and 3.57 and 3.09 and in Research Site B. For the Monitoring and Feedback for Learning subscale, adjusted means for BEST in CLASS and comparison groups were 3.38 and 3.17 in Research Site A and 3.63 and 3.10 and in Research

Table 3. Analysis Results for Teacher Outcome Variables: TEBS, TSES, and CLASS

	TEBS															
	AID				MFL				MLR				MPCC			
Effect	*B*	*SE*	*Z*	*d*	*B*	*SE*	*z*	*d*	*B*	*SE*	*z*	*d*	*B*	*SE*	*z*	*d*
Treatment (T)	.33	.06	5.47*	.64	.37	.06	5.85*	.75	.27	.06	4.29*	.53	.26	.06	4.32*	.50
Site (S)	.04	.07	0.59	.08	.09	.07	1.27	–.18	.03	.06	–0.52	–.06	.07	.05	1.29	.13
T × S	.30	.12	2.45*	.57	.32	.13	2.52*	.65	.23	.13	1.80	.45	.11	.12	0.94	.21
Covariate (C)	.52	.06	8.78*		.48	.07	7.14*		.38	.06	6.26*		.52	.05	10.76*	
C × T	–	–	–	–	–	–	–	–	–	–	–	–	–.41	.12	–3.54*	

	TSES											
	CM				ISt				SE			
Effect	*B*	*SE*	*z*	*d*	*B*	*SE*	*z*	*d*	*B*	*SE*	*z*	*d*
Treatment (T)	.63	.09	6.70*	.78	.48	.08	5.66*	.58	.43	.10	4.52*	.54
Site (S)	–.10	.09	–1.06	–.12	–.21	.10	–2.19*	–.26	–.30	.09	–3.22*	–.37
T × S	.52	.18	2.82*	.65	.42	.17	2.47*	.52	.42	.19	2.20*	.52
Covariate (C)	.39	.05	7.10*		.49	.05	9.09*		.41	.06	6.79*	
C × T	–	–	–	–	–	–	–	–	–	–	–	–

	CLASS											
	CO				ES				ISu			
Effect	*B*	*SE*	*z*	*d*	*B*	*SE*	*z*	*d*	*B*	*SE*	*z*	*d*
Treatment (T)	.53	.12	4.57*	.58	.37	.11	3.41*	.47	.60	.13	4.65*	.65
Site (S)	.54	.14	3.83*	.60	.12	.12	0.96	.15	.15	.14	1.10	.17
T × S	–.52	.23	–2.31	–.58	–.43	.21	–2.01	–.55	–.19	.25	–0.73	–.20
Covariate (C)	.22	.07	2.96*		.30	.07	4.04*		.20	.09	2.19*	
C × T	–	–	–	–	–	–	–	–	–	–	–	–

Note. TEBS = Teachers' Efficacy Beliefs System; TSES = Teachers' Sense of Efficacy Scale; CLASS = Classroom Assessment Scoring System; AID = Accommodating Individual Differences; MFL = Monitoring and Feedback for Learning; MLR = Managing Learning Routines; MPCC = Maintaining Positive Classroom Climate; CM = Classroom Management; ISt = Instructional Strategies; SE = Student Engagement; CO = classroom organization; ES = emotional support; ISu = instructional support; – = interaction not included in final model.
**Significant* controlling false discovery rate at .05.

Site B. Therefore, BEST in CLASS resulted in higher teacher self-efficacy for Accommodating Individual Differences as well as Monitoring and Feedback for Learning, with a larger effect for Research Site B. Results also indicated a significant treatment effect for the Managing Learning Routines subscale ($AM_B = 3.58$, $AM_C = 3.30$), suggesting that BEST in CLASS increased teacher self-efficacy for Managing Learning Routines. In addition, there was a significant treatment effect for the Maintaining Positive Classroom Climate (MPCC) subscale, with $AM_B = 3.67$ and $AM_C = 3.41$. This treatment effect was moderated by pretest. The estimated treatment effect was .26 when the MPCC pretest was held constant at the grand mean (3.36), and it was larger for teachers with MPCC pretest scores below the grand mean. Therefore, BEST in CLASS resulted in higher teacher self-efficacy for MPCC, and the effect was larger for teachers who had lower levels of self-efficacy at pretest. Effect sizes for treatment were .50 or larger for all four TEBS variables (see Table 3).

Results of the TSES indicate the treatment effect was significant for the Classroom Management subscale and was moderated by the T × S interaction. Adjusted means for BEST in CLASS and comparison groups were 7.65 and 7.28 in Research Site A and 7.82 and 6.92 in Research Site B. Treatment and research site effects and the T × S interaction were significant for the Instructional Strategies and Student Engagement subscales. For the Instructional Strategies subscale, adjusted means for BEST in CLASS and comparison groups were 7.82 and 7.56 in Research Site A and 7.82 and 7.14 and in Research Site B. For the Student Engagement subscale, adjusted means were 7.65 and 7.42 for Research Site A and 7.56 and 6.92 for Research Site B. For each of the TSES subscales, BEST in CLASS resulted in higher teacher self-efficacy, with larger effects at Research Site B than at Research Site A. Effect sizes for the treatment effect were .50 or larger for all three TSES variables.

Results of the CLASS are reported for each subscale. The treatment effect test was significant for the classroom organization domain. Adjusted means averaged across research sites were $AM_B = 5.34$ for BEST in CLASS and $AM_C = 4.82$ for the comparison group. In addition, there was a significant research site effect. The treatment effect was also significant for emotional support ($AM_B = 5.59$ and $AM_C = 5.22$) and instructional support ($AM_B = 2.93$ and $AM_C = 2.33$) domains. Therefore, BEST in CLASS increased teacher performance on all three CLASS domains. Effect sizes for all three CLASS domains were .50 or larger.

Fidelity Results

Outcomes of coaching dosage and fidelity indicate teachers in the BEST in CLASS condition each received 14 weeks of coaching. Out of 40 possible points, coaches' scores on the BEST in CLASS Coaching Integrity measure averaged 34.05 ($SD = 4.25$), which indicates that the components of the coaching meeting were mostly implemented as intended. Coaching meetings lasted an average of 30.8 min (range = 18.21–44.14). Additionally, teachers received an average of 2.08 follow-up contacts each week (range = 1.00–4.64).

Outcomes from the BiCACS indicate mean adherence for teachers in the BAU condition at pretest was 2.35 ($SD = 1.50$, range = 1.41–4.77). For teachers in the BEST in CLASS condition, mean adherence at pretest was 2.47 ($SD = 1.42$, range = 1.52–4.69). At posttest, mean adherence was 2.43 ($SD = 1.61$, range = 1.36–4.66) for teachers in the BAU condition. For teachers in the BEST in CLASS condition, mean adherence was 4.53 ($SD = 0.98$, range = 3.72–5.99) at posttest.

Mean competence of delivery for teachers in the BAU condition at pretest was 3.82 ($SD = 0.45$, range = 3.32–4.44). For teachers in the BEST in CLASS condition, mean competence of delivery at pretest was 3.78 ($SD = 0.37$, range = 3.31–4.37). In the BAU condition, mean competence of delivery was 4.30 ($SD = 0.42$, range = 3.95–5.04) at posttest. Mean competence of delivery for teachers in the BEST in CLASS condition was 5.51 ($SD = 0.33$, range = 5.13–5.97) at posttest.

DISCUSSION

The current study suggests that teachers who received BEST in CLASS professional development increased their use of effective instructional practices targeting young children at risk for EBD. Furthermore, teachers in the BEST in CLASS condition reported a greater sense of self-efficacy and improved classroom quality compared to teachers in the control condition. In light of recent emphasis on examining the influence of professional development on teacher and child outcomes, these findings are important. As such, BEST in CLASS appears to be a promising professional development intervention for increasing and improving teachers' use of instructional practices, particularly those focused on children who demonstrate chronic problem behaviors.

Research has also shown that professional development targeting improvements in teachers' implementation of instructional practices has positive effects on teacher self-efficacy (Bradshaw et al., 2009; Domitrovich et al., 2016). Results from the current study confirm these findings. As discussed earlier, Pas et al. (2012) suggested that teachers' beliefs that they can successfully teach children who may exhibit problem behavior is an important component of teacher self-efficacy. In the current study, the largest effect size was reported for the Classroom Management subscale of the TSES (0.78), suggesting that teachers in the BEST in CLASS condition felt more efficacious managing child behaviors in their classrooms. Additionally, findings from the current study suggest that teachers who had lower self-efficacy at pretest had even greater gains in the MPCC subscale of the TEBS after receiving BEST in CLASS practice-based coaching.

The link between teacher self-efficacy and child achievement is hypothesized to be indirect, with self-efficacy being related to teacher implementation of instructional

practices that promote child achievement (Guo, Connor, Yang, Roehrig, & Morrison, 2012). In the current study, teacher behaviors representing the core components of BEST in CLASS practices (e.g., rules, precorrection) were assessed, and teachers in the BEST in CLASS condition exhibited these behaviors more frequently from pretest to posttest compared to BAU teachers. It is possible that increased use of effective instructional practices associated with BEST in CLASS training and coaching contributed to increases in self-efficacy; that is, as teachers felt more competent delivering practices, they also felt more able to manage the learning environment in their classrooms and accommodate individual needs while increasing their monitoring and feedback to children.

Given the widespread use of the CLASS by Head Start as a tool for evaluating program quality and identifying teacher professional development needs, it is important to note that teachers who participated in BEST in CLASS scored higher on all three subscales of the CLASS compared to teachers in the control condition. Effect sizes were in the moderate range, highlighting the positive impact of BEST in CLASS on overall classroom quality. The levels of emotional support (effect size of .47) noted in BEST in CLASS classrooms may be particularly important, given that Curby, Brock, and Hamre (2013) found a relationship between the emotional support provided by teachers and children's social competence and academic-related outcomes. Further, effects at the classroom level were higher than those noted in the Werner et al. (2016) meta-analysis (effect size of .39) examining effects of targeted interventions in early childhood settings. One explanation for these larger effects may be the level of individualized training and coaching provided to BEST in CLASS teachers. Werner et al. (2016) found significant moderation of treatment effects with the presence of individual training, noting the potential added value of individual training such as coaching. These findings should not be surprising, as the importance of coaching in changing teacher behavior has been demonstrated repeatedly (see Snyder et al., 2011). The intensity of BEST in CLASS practice-based coaching, delivered weekly following training in the model components, may have contributed to the effects found across both classroom- and teacher-level outcomes.

Finally, it is important to note that the current study assessed the effect of BEST in CLASS across only two of the three levels (i.e., classroom, teacher, and child) examined in the Werner et al. (2016) meta-analysis. While moderate effects were found across self-report and observation measures at both the classroom and teacher levels, the impact on child outcomes was not included in the current study. However, previous studies of BEST in CLASS have found increases in child engagement and reductions in problem behavior (Conroy et al., 2015; Conroy, Sutherland, Vo, Carr, & Ogston, 2014) in addition to increased positive teacher–child interactions and fewer negative teacher–child interactions (Conroy et al., 2015). In a recent study (Sutherland et al., 2018), effect sizes for reductions in problem behavior ranged from .42 to .44. Therefore, improvements in teacher implementation of instructional practices, teacher self-efficacy, and classroom quality noted in the current study may ultimately be associated with improvements in behavior for young children at risk for EBD.

Limitations and Implications for Future Research

Several limitations should be kept in mind when interpreting the results of the current study. First, while conducting direct observations of teacher behavior was a strength of the study, observers were not blind to condition due to resource and logistical limitations. Although IOA was measured for direct observation data (i.e., CLASS and TCIDOS) and was determined to be acceptable, future research should attempt to utilize observers who are blind to condition. Second, while the short-term effect of BEST in CLASS on a variety of teacher and classroom outcomes is promising, the sustainability of these changes is unknown and was not assessed in the current study. Future work should evaluate whether teachers continue to implement practices associated with training and coaching, both within the same school year as well as in future years. Studies that examine necessary supports to help teachers maintain any improvements in instructional practice would also be helpful to ensure that initial effects of training and coaching are not lost. Third, while coaching is an important part of professional development, many early childhood programs may not have qualified individuals who can provide coaching or the resources available to provide the intensity of coaching (i.e., weekly) used during the BEST in CLASS intervention. In the current study, we did not examine the qualifications of an effective coach; however, most of the coaches in this study were graduate students or experienced teachers, which might indicate a high skill level. Future research should examine the qualifications of effective coaches and cost-effective models for delivering coaching supports to teachers, perhaps through Web-based programs that can increase access and sustainability. Additionally, fidelity data were collected on only a small percentage of the total coaching sessions, which is another limitation. Finally, the ICC for the teacher competence of delivery scale was fair (see Cicchetti, 1994), and care should be taken in interpreting these data. That said, the ICC for competence is consistent with previous studies (e.g., Barber, Crits-Christoph, & Luborsky, 1996; Hogue et al., 2008), and interrater reliability tends to be lower for competence than adherence (e.g., Carroll et al., 2000; Hogue et al., 2008).

Summary

Interventions that target improvements in early learning environments have been shown to have positive effects on both teacher and child outcomes (e.g., Domitrovich et al., 2016; Werner et al., 2016). The current study adds to this literature by highlighting the positive effects of BEST in CLASS on classroom quality, teacher instructional behavior, and teacher self-efficacy. Interventions such as BEST in CLASS

that impact multiple levels (i.e., teacher and classroom) within the early childhood context may be particularly promising in improving outcomes for children who are most vulnerable.

REFERENCES

Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology*, *64*, 619–622. doi:10.1037/0022-006X.64.3.619

Barnett, W. S., & Boocock, S. S. (Eds.). (1998). *Early care and education for children in poverty: Promises, programs, and long-term results* (pp. 35–63). Albany, NY: State University of New York Press.

Barnett, W. S., Friedman-Krauss, A. H., Weisenfeld, G. G., Horowitz, M., Kasmin, R., & Squires, J. H. (2017). *The state of preschool 2016*. New Brunswick, NJ: National Institute for Early Education Research.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide positive behavioral interventions and supports: Findings from a group-randomized effectiveness trial. *Prevention Science*, *10*, 100–115. doi:10.1007/s11121-008-0114-9

Brennan, L. M., Shaw, D. S., Dishion, T. J., & Wilson, M. (2012). Longitudinal predictors of school-age academic achievement: Unique contributions of toddler-age aggression, oppositionality, inattention, and hyperactivity. *Journal of Abnormal Child Psychology*, *40*(8), 1289–1300. doi:10.1007/s10802-012-9639-2

Bulotsky-Shearer, R. J., Bell, E. R., & Domínguez, X. (2012). Latent profiles of problem behavior within learning, peer, and teacher contexts: Identifying subgroups of children at academic risk across the preschool year. *Journal of School Psychology*, *50*, 775–798. doi:10.1016/j.jsp.2012.08.001

Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., … Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, *57*(3), 225–238. doi:10.1016/S0376-8716(99)00049-6

Carter, A. S., Wagmiller, R. J., Gray, S. A., McCarthy, K. J., Horwitz, S. M., & Briggs-Gowan, M. J. (2010). Prevalence of DSM-IV disorder in a representative, healthy birth cohort at school entry: Sociodemographic risks and social adaptation. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(7), 686–689. doi:10.1016/j.jaac.2010.03.018

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Conroy, M. A., & Sutherland, K. S. (2008). *Promoting social, emotional, and behavioral competence in young high-risk children: A preventative classroom-based early intervention model*. Unpublished manuscript, Virginia Commonwealth University, Richmond, VA.

Conroy, M. A., Sutherland, K. S., Algina, J. J., Wilson, R. E., Martinez, J. R., & Whalon, K. J. (2015). Measuring teacher implementation of the BEST in CLASS intervention program and corollary child outcomes. *Journal of Emotional and Behavioral Disorders*, *23*, 144–155. doi:10.1177/1063426614532949

Conroy, M. A., Sutherland, K. S., Vo, A. K., Carr, S., & Ogston, P. L. (2014). Early childhood teachers' use of effective instructional practices and the collateral effects on young children's behavior. *Journal of Positive Behavior Interventions*, *16*, 81–92. doi:10.1177/1098300713478666

Curby, T. W., Brock, L., & Hamre, B. (2013). Teachers' emotional support consistency predicts children's achievement gains and social skills. *Early Education and Development*, *24*(3), 292–309. doi:10.1080/10409289.2012.665760

Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. (2008). Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-Self. *Teaching and Teacher Education*, *24*(3), 751–766. doi:10.1016/j.tate.2007.02.010

Domitrovich, C. E., Bradshaw, C. P., Berg, J. K., Pas, E. T., Becker, K. D., Musci, R., … Ialongo, N. (2016). How do school-based prevention programs impact teachers? Findings from a randomized trial of an integrated classroom management and social–emotional program. *Prevention Science*, *17*(3), 325–337. doi:10.1007/s11121-015-0618-z

Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., & Jones, D. (2009). Fostering high-quality teaching with an enriched curriculum and professional development support: The Head Start REDI program. *American Educational Research Journal*, *46*, 567–597. doi:10.3102/0002831208328089

Feil, E. G., Severson, H. H., & Walker, H. M. (1998). Screening for emotional and behavioral delays: the Early Screening Project. *Journal of Early Intervention*, *21*, 252–266. doi:10.1177/105381519802100306

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualification, teacher self-efficacy, and classroom practices on fifth graders' literacy outcomes. *The Elementary School Journal*, *113*(1), 3–24. doi:10.1086/665816

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology*, *33*(6), 665–679. doi:10.1007/s10802-005-7646-2

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. doi:10.1016/j.jbi.2008.08.010

Hemmeter, M. L., Santos, R. M., & Ostrosky, M. M. (2008). Preparing early childhood educators to address young children's social-emotional development and challenging behavior: A survey of higher education programs in nine states. *Journal of Early Intervention*, *30*, 321–340. doi:10.1177/1053815108320900

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, *76*, 544–555. doi:10.1037/0022-006X.76.4.544

Howes, C., & Smith, E. W. (1995). Relations among child care quality, teacher behavior, children's play activities, emotional security, and cognitive activity in child care. *Early Childhood Research Quarterly*, *10*, 381–404. doi:10.1016/0885-2006(95)90013-6

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114–140. doi:10.1037/met0000078

Muthén, L. K., & Muthén, B. O. (1998–2017). M*plus* User's Guide (8th ed.). Los Angeles, CA: Muthén & Muthén.

Newborg, J. (2005). *Battelle Developmental Inventory* (2nd ed.). Itasca, IL: Riverside Publishing.

Pas, E. T., Bradshaw, C. P., & Hershfeldt, P. A. (2012). Teacher- and school-level predictors of teacher efficacy and burnout: Identifying potential areas for support. *Journal of School Psychology*, *50*, 129–145. doi:10.1016/j.jsp.2011.07.003

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System: Manual K–3*. Baltimore, MD: Paul H. Brookes Publishing.

Sameroff, A. J. (1983). Developmental systems: Contexts and evolution. In P. H. Mussen (Gen. Ed.) & W. Kessen (Vol. Ed.), *Handbook of child psychology: Vol 1. History, theory, and methods* (4th ed., pp. 237–294). New York, NY: Wiley.

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Snyder, P., Hemmeter, M. L., Meeker, K. A., Kinder, K., Pasia, C., & McLaughlin, T. (2012). Characterizing key features of the early childhood

professional development literature. *Infants and Young Children*, *25*(3), 188–212. doi:10.1097/IYC.0b013e31825a1ebf

Snyder, P. A., Hemmeter, M. L., & Fox, L. (2015). Supporting implementation of evidence-based practices through practice-based coaching. *Topics in Early Childhood Special Education*, *35*, 133–143. doi:10.1177/0271121415594925

Snyder, P., Hemmeter, M. L., & McLaughlin, T. (2011). Professional development in early childhood intervention: Where we stand on the silver anniversary of PL 99-457. *Journal of Early Intervention, 33*, 357–370. doi:10.1177/1053815111428336

Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. W. (2011). Optimal Design Plus empirical evidence: Documentation for the "Optimal Design" Software Version 3.0. Retrieved from http://wtgrantfoundation.org/resource/optimal-design-with-empirical-information-od.

Sutherland, K. S., Conroy, M. A., Algina, J., Ladwig, C., Jesse, G., & Gyure, M. (2018). Reducing child problem behaviors and improving teacher–child interactions and relationships: A randomized controlled trial of BEST in CLASS. *Early Childhood Research Quarterly*, *42*, 31–43. doi:10.1016/j.ecresq.2017.08.001

Sutherland, K. S., Conroy, M. A., Vo, A., Abrams, L., & Ogston, P. (2013). An initial evaluation of the Teacher–Child Interaction Direct Observation System: Measuring teacher–child interaction behaviors in classroom settings. *Assessment for Effective Intervention*, *39*(1), 12–23. doi:10.1177/1534508412463814

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*(1), 21–33. doi:10.1007/s12310-014-9134-8

Sutherland, K. S., McLeod, B. D., Conroy, M. A., Abrams, L. M., & Smith, M. M. (2014). Preliminary psychometric properties of the BEST in CLASS adherence and competence scale. *Journal of Emotional and Behavioral Disorders*, *22*(4), 249–259. doi:10.1177/1063426613497258

Tapp, J. (2003). *INTMAN interval observation analysis program*. Nashville, TN: Vanderbilt Kennedy Center.

Tapp, J. (2010). *Lily Data Collector*. Nashville, TN: Vanderbilt Kennedy Center.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805. doi:10.1016/S0742-051X(01)00036-1

U.S. Department of Education. (2013). *What Works Clearinghouse: Procedures and standards handbook, Version 3.0*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

U.S. Department of Education, Institute of Education Sciences. (2016). *What Works Clearinghouse: Cluster design standards*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_cluster_standards_030416.pdf

Werner, C. D., Linting, M., Vermeer, H. J., & Van IJzendoorn, V. (2016). Do intervention programs in child care promote the quality of caregiver-child interactions? A meta-analysis of randomized controlled trials. *Prevention Science*, *17*(2), 259–273. doi:10.1007/s11121-015-0602-7

Date Received: January 14, 2018
Date Accepted: August 3, 2018
Associate Editor: Timothy Curby ■

AUTHOR BIOGRAPHICAL STATEMENTS

Maureen A. Conroy is the Anita Zucker Endowed Professor in Early Childhood Studies and codirector of the Anita Zucker Center for Excellence in Early Childhood Studies, University of Florida. Her research focuses on the prevention and amelioration of social and emotional learning problems in young children at risk for school failure.

Kevin S. Sutherland is a professor in the School of Education at Virginia Commonwealth University. His research focuses on intervention development and implementation issues for students with or at risk for emotional–behavioral disorders.

James Algina is an emeritus professor of research and evaluation methodology at the University of Florida. He is a fellow of both the American Educational Research Association and Division 5 of the American Psychological Association. His research interests include classical test theory, item response theory, multilevel modeling, and structural equation modeling.

Crystal Ladwig is the BEST in CLASS project coordinator. Her research interests include intervention strategies supporting children with social–emotional learning difficulties and professional development for early childhood teachers.

Brittany Werch is a school psychologist in Columbia County, FL. Her research interests focus on early intervention, implementation science, and improving outcomes for children at risk for disabilities.

Jose Martinez is an assistant professor at the University of Kansas. His research interests include young children with social–emotional learning needs, including children with autism spectrum disorders.

Maria Gyure, MS, LCGC, is director of the Genetic Counseling Master's Degree Program at the University of Connecticut. Her research interests focus on community engagement and improving outcomes for children with and at risk for disabilities.

Gabriel Jessee, MS, is a data scientist at CapTech Ventures in Richmond, VA. His research interests include data usage and analyses and improving outcomes for children at risk for disabilities.