Research Report

Journal of Early Intervention
1–19

DOI: 10.1177/10538151211018662
journals.sagepub.com/home/jei
SAGE

Preliminary Study of the Effects of BEST in CLASS—Web on Young Children's Social-Emotional and Behavioral Outcomes

Maureen A. Conroy[1], Kevin S. Sutherland[2], Kristen L. Granger[2], Katerina M. Marcoulides[1,3], Ke Huang[1], and Alexandra Montesion[1]

Abstract

Many young children entering early childhood programs demonstrate challenging behaviors that place them at risk for emotional/behavioral disorders (EBDs), which impact their future success in school. The purpose of this study was to conduct a conceptual replication of BEST in CLASS by examining child outcomes from BEST in CLASS—Web, a web-based professional development intervention supporting teachers' use of effective practices for ameliorating young children's challenging behaviors. Participants included 29 early childhood teachers and 54 children (ages 3–5 years old) who were identified at-risk for EBD. Positive outcomes were found for both BEST in CLASS—Web and BEST in CLASS in reductions of children's problem behaviors and conflictual relationships with their teachers as well as increases in social skills, engagement, and closeness with teachers in comparison to children who did not receive the intervention. In some cases, positive child outcomes were greater when teachers received professional development in the web-condition as compared to the onsite condition. Future research directions and implications are discussed.

Keywords

early childhood, social-emotional learning, behavioral disorders, tier-2 intervention

High-quality early childhood education has the potential to positively impact social-emotional learning and prevent problem behaviors in young children (see Schindler et al., 2015). However, many preschool-age children entering early childhood programs engage in elevated rates of externalizing behavior (e.g., aggression, disruption; Blair & Raver, 2012; Briggs-Gowan &

[1]University of Florida, Gainesville, USA
[2]Virginia Commonwealth University, Richmond, USA
[3]University of Minnesota, Minneapolis, USA

Corresponding Author:
Maureen A. Conroy, Anita Zucker Center of Excellence in Early Childhood Studies, University of Florida, Box 117050, Gainesville, FL 32611, USA.
Email: mconroy@coe.ufl.edu

Carter, 2008; Kupersmidt et al., 2000), with internalizing behavior problems occurring at similar rates (see Ringeisen et al., 2017, for a discussion) that place them at risk for later development of emotional/behavioral disorders (EBDs) (Bulotsky-Shearer et al., 2012; Campbell et al., 2000; McCabe & Altamura, 2011). Over time, these early chronic problem behaviors lead to learning and behavioral difficulties when entering kindergarten and impact later academic outcomes and school success (Hetzner et al., 2011; McClelland et al., 2007; Yoshikawa et al., 2012). Although many early childhood programs have universal classroom intervention practices in place, some children require targeted interventions to address their social-emotional and behavioral needs. Without receiving early targeted intervention, their problem behaviors become more severe (Beyer et al., 2012), often resulting in chronic negative teacher–child interaction patterns, the use of coercive and harsh disciplinary practices, and potential identification of an EBD (Forness et al., 2012). Thus, the effects of persistent problem behavior on children's developmental outcomes make early intervention for these young children a high priority (Brown & Conroy, 2012).

Research has indicated that early childhood teachers who use effective intervention practices in their classrooms can reduce chronic problem behaviors and improve outcomes for young children who are at elevated risk for EBD (e.g., see Driscoll & Pianta, 2010; Hemmeter et al., 2016; Sutherland et al., 2018). Over the past several decades, a number of evidence-based intervention programs targeting young children at risk for EBD are available for use in classroom settings (for a review, see Conroy et al., 2020). One such program, BEST in CLASS (Conroy & Sutherland, 2008) is a targeted intervention designed to be delivered by early childhood teachers during authentic learning activities in their classrooms. The overall aim of BEST in CLASS is to increase positive interactions and relationships between early childhood teachers and children, increase children's social-emotional learning and engagement, and reduce negative teacher–child interactions and children's problem behavior. Although social-emotional learning intervention programs exist, many teachers do not receive training in these programs due to limited availability based on geographical location or limited programmatic resources (Dede et al., 2009; Skyhar, 2020; Wei et al., 2010). The adaptation of intervention programs, such as BEST in CLASS, to be delivered virtually is likely to increase accessibility and availability for early childhood teachers (Snyder et al., 2018).

Overview of BEST in CLASS

Based on transactional theory (Sameroff, 1983), which suggests that social exchanges between partners are bi-directional and the behavior of each partner in the exchange impacts the other, and the behavioral prinicples outlined by Skinner (1953), BEST in CLASS was designed to support early childhood teachers' use of evidence-based practices with young children who engage in high rates of problem behaviors. Following an introductory teacher workshop, which provides general information about the intervention practices, teachers receive 14-weeks of practice-based coaching (adapted from Snyder et al., 2015) and a BEST in CLASS resource manual. Over the 14-weeks of practice-based coaching, teachers gain skills in using the BEST in CLASS practices with select children during authentic instructional activities embedded throughout the school day (for a comprehensive description of the BEST in CLASS intervention see Conroy et al., 2019). Findings from a randomized controlled trial (RCT) examining the efficacy of BEST in CLASS indicated that early childhood teachers who received training and practice-based coaching on the BEST in CLASS intervention increased their use of the BEST in CLASS practices with high levels of fidelity, increased their self-efficacy, and improved their overall classroom quality in comparison to teachers in the control group (Conroy et al., 2019). Positive findings for children who received the BEST in CLASS intervention were also found. Children in teachers' classrooms who implemented the BEST in CLASS intervention included increased social skills, positive teacher–child interactions and relationships, and decreased problem behaviors and conflictual teacher–child interactions in

comparison to children in the control condition (Sutherland et al., 2018). This evidence suggests that BEST in CLASS is effective at improving practitioner's use of the BEST in CLASS practices while also reducing chronic problem behaviors of young children.

The professional development components (i.e., teacher training workshop, resource manual, and practice-based coaching) of the BEST in CLASS intervention were designed to be delivered to teachers on-site. Although this delivery model meets the needs of some early childhood programs, on-site professional development may restrict accessibility for programs with limited personnel or financial resources or programs that are geographically remote (Dede et al., 2009). In addition, on-site delivery of professional development may be less convenient for teachers and may not be compatible with individual learning styles and pace (Ke & Xie, 2009). Finally, on-site delivery of professional development may limit the scalability of an intervention for those programs not within close proximity to the training site (Knight, 2012). Given the established efficacy of BEST in CLASS and in light of the limitations associated with the on-site delivery, we developed a web-based version of BEST in CLASS to increase accessibility, flexibility, usability, and scalability. Mirroring the on-site BEST in CLASS intervention, BEST in CLASS—Web includes the same professional development components as the on-site version. Rather than these professional development components being delivered on-site, however, all three components are delivered via the web. The teacher workshop is delivered through eight interactive online modules that teachers complete independently throughout the 14-weeks of practice-based coaching at their own pace. The practice-based coaching component is conducted synchronously over a web-based platform that allows the coach to meet with the practitioner to set weekly goals, develop action steps, provide feedback, view video recordings, access resources and supports, and self-reflect on the use of BEST in CLASS practices within the classroom. The core content provided in the online modules and virtual practice-based coaching process is the same as that in the on-site workshop and in-person coaching sessions. A more detailed description of the BEST in CLASS—Web components is provided in the Method section below.

Following development of BEST in CLASS—Web, an examination of its efficacy in relation to desired teacher and child outcomes was conducted. The extension and replication of prior BEST in CLASS research from on-site delivery to web-based delivery is timely as the need for online and accessible professional development increases (Olsen et al., 2010). Therefore, in this study, we proposed to conduct a conceptual replication (Schmidt, 2009; Travers et al., 2016) of BEST in CLASS to examine and compare the intervention effects when the professional development component was delivered over the web in comparison to on-site. Research examining the replication of BEST in CLASS delivered through a web-based format is particularly important, given the complex contexts in which interventions are delivered in early childhood classrooms (National Science Foundation and Institute of Education Sciences, 2018). For example, factors such as program characteristics, funding, compatibility with staff training, and so forth can affect the implementation process (Durlak & Dupre, 2008). In addition, replication research is important to determine if the same level of effects occur as in the prior research (Travers et al., 2016). Furthermore, the lack of empirical work examining web-based professional development interventions designed to address young children's behavior concerns in early childhood classrooms makes examining the preliminary impact of BEST in CLASS—Web an important contribution to the field.

Study Purpose, Aims, and Hypotheses

In this preliminary study, we propose to extend and replicate the findings from BEST in CLASS by testing the efficacy of a web-based version of BEST in CLASS. The overall purpose of this study was to examine the effect of BEST in CLASS—Web on child social-emotional and behavioral outcomes and compare these effects to BEST in CLASS delivered on-site. In addition, to

examine replication of previous findings, we also compared the two intervention conditions to a control group (i.e., business as usual [BAU]). We hypothesized that the treatment effect of BEST in CLASS—Web would be equivalent to that of BEST in CLASS delivered in person when compared to BAU. The following research questions were addressed in this study:

1. Is the treatment effect of BEST in CLASS—Web equivalent to that of BEST in CLASS delivered in-person when compared to a control group?
2. Is the treatment effect from the prior BEST in CLASS research replicated in the current investigation?

Method

Study Design

A small-scale multi-site RCT of BEST in CLASS—Web in comparison to the original BEST in CLASS, and a BAU condition was conducted. This trial took place over one school year across two research sites and used a pretest–posttest cluster randomized design (Spybrook & Raudenbush, 2009). Randomization occurred at the teacher level, with children nested within teachers. Teachers were randomly assigned to one of two conditions (BEST in CLASS or BEST in CLASS—Web). The participants included in the BAU condition were randomly selected from an earlier BEST in CLASS efficacy trial (see Conroy et al., 2019; Sutherland et al., 2018).

Experimental Procedures

Study staff at each research site met with interested teachers to collect informed consent. Approximately 1 month into the school year, consented teachers nominated five children in their classroom who displayed significant problem behaviors. Informed consent forms were then delivered to the caregivers of these children, and children with caregiver consent were screened for eligibility. Pretest data collection began after each classroom included one participating teacher and at least one child participant. Posttest data collection in all conditions began approximately 20-weeks later. At the completion of pretest data collection, randomization to condition occurred. Posttest data collection occurred at the completion of the intervention conditions. All study activities were approved by the Institutional Review Board (IRB) at each research site.

Setting

This study took place in three local school districts and one Head Start grantee. Early childhood classrooms were located in urban, rural, and suburban communities across two southeastern states. All participating classrooms were federally (n = 14; 73.68%) or state- (n = 5; 26.32%) funded early childhood programs serving income-eligible children from disadvantaged socioeconomic backgrounds. These classrooms were located within early childhood education centers (n = 15; 78.95%) or local elementary schools (n = 4; 21.05%). Participating classrooms averaged 16.21 (SD = 3.75) children, of which 50.97% were male and 49.03% were female.

Participants

Teachers. Teachers were eligible for participation in the study if they were the lead teacher in a classroom serving children 3 to 5 years of age and had at least one child who met criteria for

inclusion in the study. A total of 25 teachers were recruited into the study; three of these teachers withdrew consent prior to randomization, resulting in 22 teachers being assigned to either the BEST in CLASS ($n = 10$) or the BEST in CLASS—Web ($n = 12$) condition. Three teachers withdrew prior to the beginning of the intervention, resulting in 19 teachers receiving one of the two conditions. In addition, 10 teachers were randomly selected from the original BAU comparison condition. Teachers selected from the original BAU condition were located in the same geographical regions as the teachers in the current investigation. Differential attrition was 6.67% for teacher participants (see Figure 1).

Children. The Early Screening Project (ESP, Feil et al., 1995) was used to screen children for eligibility inclusion. The ESP is a multigated system used to identify children at risk for EBD. The first "gate" asks teachers to rank order the top-five children in their classroom who demonstrate the most problem behaviors, such as defiance, disruption, or aggression. The second "gate" asks teachers to complete a brief rating scale for children who were rank ordered in the top five. Children were eligible to participate if they (a) were between the ages of 3 to 5 years old and enrolled in a participating teacher's classroom; (b) were one of the two children per classroom with the most elevated risk score as ranked by the ESP; (c) were not identified as at risk for developmental delay by scoring within the normal range on the Battelle Developmental Inventory, 2nd Edition (BDI-2; Newborg, 2005); (d) were proficient in English; and (e) had caregiver consent.

Fifty-one children were consented to participate. Twelve of these children did not meet inclusion criteria, with a total of 39 child participants (referred to as focal children) allocated to the BEST in CLASS or BEST in CLASS—Web condition. In addition, five child participants did not begin intervention (see Figure 1), resulting in a total of 54 children (BEST in CLASS $n = 16$, BEST in CLASS—Web, $n = 18$, BAU $n = 20$) included in data analyses. No significant differences were observed among the three conditions on child demographic characteristics, as seen in Table 1. Differential attrition was 2.38% for child participants (see Figure 1).

Dependent Measures

Given this study was a conceptual replication of the efficacy study of BEST in CLASS, we were interested in assessing the same outcomes as the original study, and therefore, we employed the same dependent measures.

Social Skills Improvement System—Teacher Rating Scale. The Social Skills Improvement System—Teacher Rating Scale (SSIS-TRS; Gresham & Elliott, 2008) is a standardized teacher-report rating scale consisting of 76 items. The Social Skills scale and Problem Behaviors scale of the SSIS-TRS were used to measure children's performance of social skills and problem behaviors. Teachers complete the measure using the 4-point Likert-type scale (i.e., *N = Never*, S = *Seldom*, O = *Often*, and A = *Almost Always*). As reported by Sutherland et al. (2018), Cronbach's alpha for the BAU condition indicated high internal consistency (McDonald, 1999; Novick & Lewis, 1967), with .94 and .95 for the Social Skills scale at pretest and posttest, and .90 and .92 for the Problem Behaviors scale at pretest and posttest, respectively (Sutherland et al., 2018). Given the small sample size, the Cronbach alphas are not reported for the current sample.

Caregiver—Teacher Report Form. The Caregiver—Teacher Report Form (C-TRF; Achenbach & Rescorla, 2000) is a standardized teacher-report rating scale designed to measure behavioral or emotional problems in children ages 18 months to 5 years. Teachers scored each of the 99 items on a 3-point Likert-type scale (i.e., 0 = *Not True [as far as you know]*, 1 = *Somewhat or*

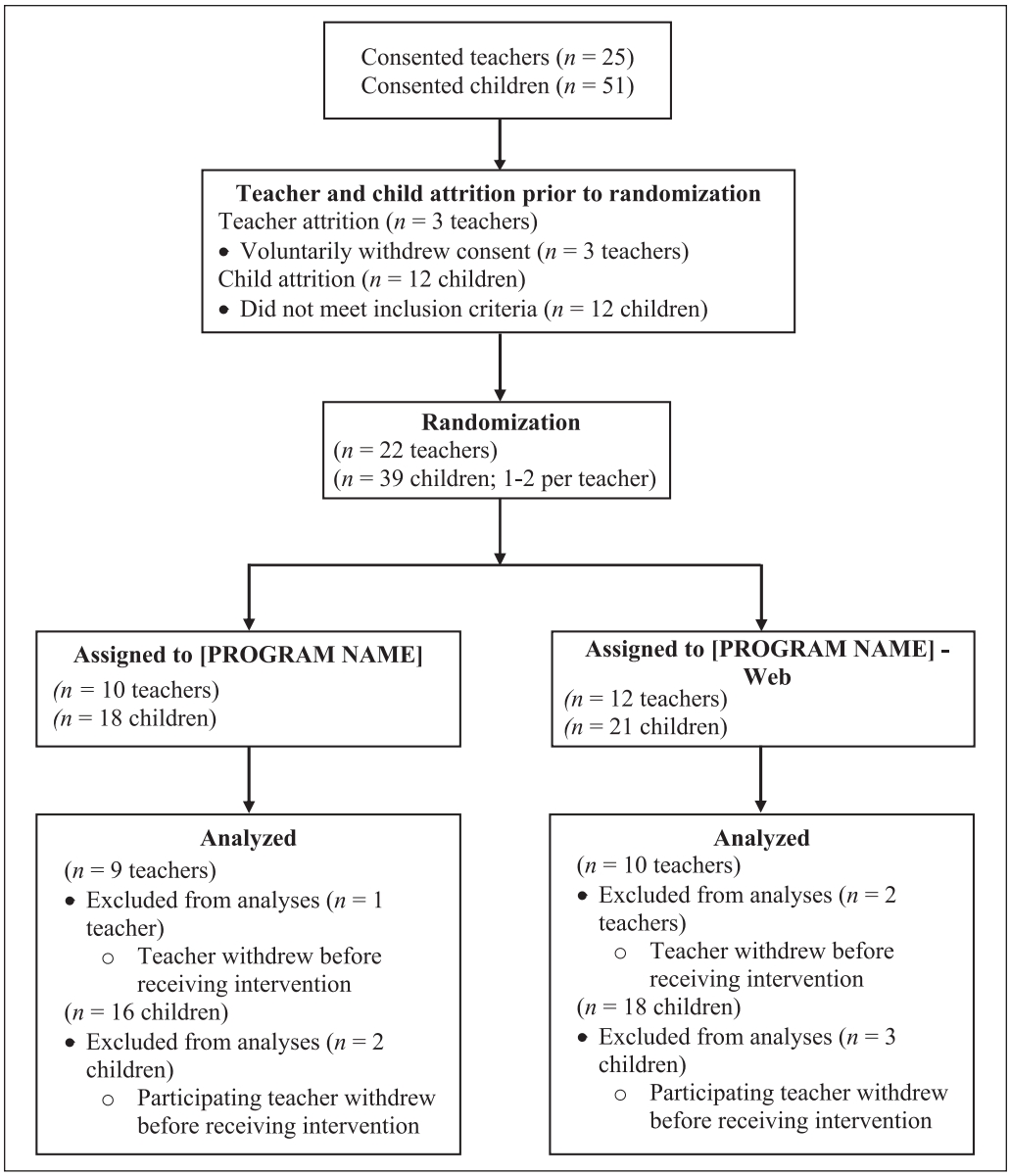

Figure 1. CONSORT flowchart of participants.

Sometimes True, and 2 = *Very True or Often True*). The C-TRF yields a raw score for each of the three scales (i.e., Externalizing Behavior, Internalizing Behavior, and Total Problems), which were converted to *T* scores for further data analyses. Similar to the SSIS, the Cronbach's alpha is not reported for the current sample but was calculated using the sample obtained from the previous BEST in CLASS study for the BAU participants. Cronbach's alpha ranged between .86 and .95 for pretest and above 0.90 for posttest (Sutherland et al., 2018).

Student Teacher Relationship Scale. The Student Teacher Relationship Scale (STRS; Pianta, 2001) is a teacher-report, 15-item rating scale designed to measure teachers' perceptions of their

Table 1. Participant Demographics.

Baseline characteristic		BEST in CLASS	BEST in CLASS – Web	BAU	Total
Teachers, *N* (%)		9 (31.03)	10 (34.48)	10 (34.48)	29
Racial background	African American	3 (33.33)	4 (40.00)	4 (40.00)	11 (37.93)
	Asian	2 (22.22)	—	—	2 (6.90)
	Hispanic	2 (22.22)	1 (10.00)	—	3 (10.34)
	Caucasian	2 (22.22)	5 (50.00)	6 (60.00)	13 (44.83)
	Other	—	—	—	—
Gender	Male	—	1 (10.00)	—	1 (3.45)
	Female	9 (100.00)	9 (90.00)	10 (100.00)	28 (96.55)
Level of education	High school	—	—	2 (20.00)	2 (6.90)
	Associate	2 (22.22)	4 (40.00)	—	6 (20.69)
	Bachelor	4 (44.44)	2 (20.00)	5 (50.00)	11 (37.93)
	Master	2 (22.22)	3 (30.00)	3 (30.00)	8 (27.59)
	Other	1 (11.11)	1 (10.00)	—	2 (6.90)
Age range	18–25	1 (11.11)	4 (40.00)	1 (10.00)	6 (20.69)
	26–35	1 (11.11)	2 (20.00)	1 (10.00)	4 (13.79)
	36–45	3 (33.33)	3 (30.00)	5 (50.00)	11 (37.93)
	46–55	—	—	2 (20.00)	2 (6.90)
	Over 55	4 (44.44)	—	1 (10.00)	5 (17.24)
	Prefer not to say	—	1 (10.00)	—	1 (3.45)
Years teaching	Mean (*SD*)	18.67 (13.00)	6.80 (6.91)	9.30 (9.91)	11.34 (11.01)
Children, *N* (%)		16 (31.37)	18 (35.29)	17 (33.33)	51
Racial background	African American	12 (75.00)	13 (72.22)	13 (76.47)	38 (74.51)
	Asian	—	—	—	—
	Hispanic	—	1 (5.56)	—	1 (1.96)
	Caucasian	3 (18.75)	3 (16.67)	2 (11.76)	8 (15.69)
	Other	1 (6.25)	1 (5.56)	2 (11.76)	4 (7.84)
Gender	Male	11 (68.75)	13 (72.22)	12 (70.00)	36 (70.37)
	Female	5 (31.25)	5 (27.78)	5 (30.00)	15 (29.63)
Age at entry	Mean (*SD*)	4.39 (0.53)	4.46 (0.43)	4.60 (0.42)	4.49 (0.46)

Note. Demographic data represent 51 children due to missing data from unreturned demographic surveys. BAU = business as usual.

relationship with a student in two domains: Closeness and Conflict. Teachers scored each of the 15 items using a 5-point Likert-type scale (1 = *Definitely does not apply*, 2 = *Not really*, 3 = *Neutral, not sure*, 4 = *Applies somewhat*, and 5 = *Definitely applies*). The scores on the items were averaged for the Closeness and Conflict domains, respectively. Cronbach's alpha (Cronbach, 1951) is not reported for participants in the current BEST in CLASS and BEST in CLASS—Web because of the small sample size; however, it was calculated with the sample obtained from the previous BEST in CLASS study. For the BAU condition, Cronbach's alpha was .82 and .81 for Closeness at pretest and posttest, and .86 and .88 for Conflict at pretest and posttest, respectively (Sutherland et al., 2018).

Teacher–Child Interaction Direct Observation System. The Teacher–Child Interaction Direct Observation System—Research Version 3.1 (TCIDOS−RV3.1; Conroy et al., 2015; Sutherland et al., 2013) is a researcher-developed, partial-interval direct observation coding system. This coding system was designed by the authors to measure teachers' implementation of targeted instructional practices, children's engagement and problem behaviors, and teacher–child interactions (i.e., positive interactions and negative interactions). This article focuses on child participant

outcomes; therefore, data are reported for (a) child engagement, (b) child problem behaviors, (c) positive teacher–child interactions, and (d) negative teacher–child interactions.

Trained observers, who were not blind to condition, conducted classroom observations with teacher–child dyads. The observation sessions were 15 min in length and included sixty 15-s intervals. In each interval, an observer determined whether a target behavior occurred at least once in the first 10-s observation period and recorded the observed behaviors using the Lily Data Collector (Tapp, 2010) in the following 5-s recording period. Data (i.e., occurrences and nonoccurrences of a target behavior in each of the 60 intervals) were transformed into the percentage of intervals where a target behavior occurred using the Multi-Option Observation System for Experimental Studies (MOOSES; Tapp et al., 1995). A secondary observer independently collected data on 30.49% of the observation sessions. Interobserver agreement was calculated using the occurrence and nonoccurrence percentage of agreement (i.e., number of intervals with agreements divided by the sum of intervals with agreements and disagreements then multiplied by 100). Interobserver agreement was calculated for each target behavior: child engagement .94 (range = .74–1.00), child problem behaviors .94 (range = .80–1.00), positive interactions .94 (range = .78–1.00), and negative interactions .94 (range = .80–1.00).

BEST in CLASS Intervention[1] and Comparison Conditions

The BEST in CLASS intervention was created as a value-added intervention to support teachers' systematic, increased, and targeted use of evidence-based practices at a Tier-2 level with children who demonstrate chronic problem behaviors (i.e., focal children). Teachers delivered increased use of the BEST in CLASS and BEST in CLASS—Web instructional practices (i.e., rules, precorrection, opportunities to respond, behavior specific praise, corrective feedback, and instructional feedback) to focal children during authentic learning activities throughout the day. Randomly selected teacher and student participants from the BAU condition were included to assess the impact of receiving BEST in CLASS or BEST in CLASS—Web and to verify the successful replication of the BEST in CLASS intervention. These three conditions are described below.

BEST in CLASS. The BEST in CLASS workshop provided teachers an overview of BEST in CLASS, a timeline for the coaching process, and a description of each practice. The workshop occurred in a single day and began with an overview of challenging behavior followed by an introduction to each of the practices included within the intervention. Video exemplars of the practices and interactive learning activities occurred throughout the workshop. During the workshop, BEST in CLASS coaches worked alongside teachers to complete a workbook developing strategies to implement each practice with high quality. Teachers also received a BEST in CLASS teacher manual during the workshop, with more in-depth explanations and examples of the evidence-based practices throughout seven modules (one module per practice). The teachers used this manual over the 14 weeks of practice-based coaching, which began approximately 1 week following training. Each week in the practice-based coaching meetings, the teacher and coach met in person and collaboratively followed a cyclical process including three steps: goal setting and action planning, focused observation, and reflection and performance feedback (for a description of practice-based coaching, see Snyder et al., 2015). During each coaching session, the coach provided supportive and constructive feedback, while the teacher self-reflected on the implementation of the practices during the previous week.

Six coaches were trained to criterion in the BEST in CLASS practice-based coaching model and assigned to work with both BEST in CLASS and BEST in CLASS—Web teachers. All

coaches were female and ranged in age between 21 and 55 years. Four of the coaches held master's degrees, one coach held an educational specialist degree, and one coach held a bachelor's degree. In the first step of coaching, the coach supported the teacher's understanding of the current BEST in CLASS practice of the week and development of an action plan, setting specific goals tailored to the individual needs of the child. Next, the coach collected data on quantity and quality of the teacher's practice use and the focal child's engagement and problem behaviors during a routine classroom activity. During the meeting, the coach encouraged the teacher to self-reflect on their use of the practices over the past week and during the time of the focused observations. The coach then reviewed the data collected from the focused observations (e.g., TCIDOS-RV3.1 [Sutherland et al., 2013; BEST in CLASS Adherence and Competence scale [BiCACS; Sutherland & McLeod, 2010; described below], anecdotal notes) with the teacher and provided feedback, relative to the goals specified by the teacher. In addition to these three key components, the coach provided strategies for communicating with the caregivers of the focal children on ways to use the practices at home. Further support was offered by the coach, such as modeling, prompting, sending emails or text reminders, and providing job-embedded materials (e.g., visual cues), any of which the teacher was free to select. This process was repeated each week until the end of the intervention. Each coach uploaded video recordings of their coaching meetings to a secure server for review by a trained observer to ensure that the coaches maintained fidelity in their delivery of coaching throughout the entire intervention.

BEST in CLASS—Web. BEST in CLASS—Web maintained the same manualized processes of the BEST in CLASS intervention; however, it operated remotely via online platforms and applications. The 1-day teacher training was replaced with online modules developed by Technology-based Behavioral Interventions Delivery System (T-BIDS; Feil et al., 2008; 2020), and the practice-based coaching was facilitated by the TORSH Talent application and occurred via video conferencing on Zoom©. Each coach delivered their BEST in CLASS—Web teacher an iPad® and necessary accessories, the teacher manual, a teacher user manual (providing step-by-step instructions for operating the online platforms, applications, and equipment), and a brief technology training. Eight BEST in CLASS modules were located in one online platform; the first module was an introduction to BEST in CLASS, and each subsequent module related to one of the BEST in CLASS practices. Modules provided comparable information to that covered in the workshop, along with video exemplars (e.g., recordings of teachers using the BEST in CLASS practices in their classrooms) and interactive knowledge checks (e.g., multiple choice, fill in the blank). One difference between the in-person and online versions is that the teacher completed one module per week compared to in-person teachers who attended a 1-day workshop. When developing the web-based format, teachers indicated that this type of delivery would be more feasible than completing all modules at a single time. Each week during the coaching session, the coach and the teacher reviewed the information in the module and the coach provided feedback on the knowledge checks.

Two introductory meetings between the teacher and coach were held before the 14 practice-based coaching sessions to build rapport and to troubleshoot potential technology issues. All meetings were conducted synchronously via Zoom, a video conference application. Coaches used the document and task management system provided by the TORSH Talent platform (2018) to facilitate the online coaching process. Teachers video recorded themselves during the pre-planned observation time through the TORSH Talent application (v.1.0.11). These videos were automatically uploaded to the TORSH Talent platform, where the coach would perform the focused observation and provide time-stamped comments directly to the video. Also stored in the

TORSH Talent platform were video exemplars and additional resources designed to support teachers' understanding and delivery of the BEST in CLASS practices.

Business-as-usual. The BAU condition was comprised of typical programing and instruction occurring in federally and state-funded early childhood programs, such as Head Start or state-funded prekindergarten. Children in these classrooms received daily instructional activities, such as circle time, center time, small group activities, and outdoor play. Children in the BAU condition did not receive the BEST in CLASS intervention.[2]

Fidelity Measures

Teachers' implementation of the BEST in CLASS practices was assessed using the BiCACS (Sutherland & McLeod, 2010; Sutherland et al., 2014). The Adherence scale of the BiCACS measures how closely teachers delivered the practices as intended (e.g., thoroughness and frequency), and the Competence scale measures the skillfulness (e.g., clarity of language and communication with the child when implementing a practice) and degree of responsiveness (e.g., appropriate timing of actions) demonstrated by teachers when delivering the BEST in CLASS practices. Both scales consist of six items corresponding to each of the six BEST in CLASS practices. In addition to the items comprising the Adherence and Competence scales, a Child Responsiveness scale consisting of two items (i.e., child responsiveness and child engagement) was also included to examine the extent to which the teacher's efforts to engage the child maintained the child's attention and participation (Sutherland et al., 2013). Trained raters observed teachers' instruction with their focal child for 10 to 15 min in the classroom and immediately completed the Adherence and Child Responsiveness scales using a 7-point Likert-type scale (1 = *Not at all*, 2 = *Limited*, 3 = *Some*, 4 = *Moderate*, 5 = *Considerably*, 6 = *Extensive*, and 7 = *Very Extensive*). After completing the Adherence scale, the raters scored only on the observed items (i.e., items scored greater than 1 on the Adherence scale) on the Competence scale using a 7-point Likert-type scale (1 = *Very Poor*, 3 = *Acceptable*, 5 = *Good*, and 7 = *Excellent*).

Ninety-nine observation sessions were conducted across pretest and posttest. The scores on the Adherence and Competence scales, and Child Responsiveness item across each research condition are displayed in Table 2. A secondary rater independently scored 24.24% of these observation sessions. The single measure intraclass correlation coefficient (ICC) was .74 and .60 for Adherence and Competence scales, respectively. ICC was .69 for Child Responsiveness.

Results

Due to the small sample size, we were unable to conduct analyses that accounted for nesting of children in classrooms. Therefore, analyses and inferences about observed differences between matched group means were conducted on the basis of repeated measures analysis of variance (ANOVA) methods (for mean comparisons between repeated assessments on the same subjects [e.g., on pretest and posttest scores]). All inferences about differences between independent group means were conducted on the basis of completely randomized ANOVA methods (for mean comparisons between assessments on participants from different groups). To determine baseline similarity between BEST in CLASS, BEST in CLASS—Web, and BAU conditions, mean differences on all study variables were evaluated at pretest and found to be nonsignificant.

To test the extent to which BEST in CLASS—Web was effective in improving child outcomes, we calculated effect sizes (Cohen's *d*) between observed means at pretest and posttest. To test if the treatment effect of BEST in CLASS—Web was equivalent to that of BEST in CLASS, we calculated effect sizes (Cohen's *d*) between observed means at posttest for BEST in CLASS—Web, BEST in CLASS, and BAU conditions. Given the current debate regarding the use of

Table 2. Teacher Implementation Fidelity.

	BEST in CLASS		BEST in CLASS—Web		BAU	
BiCACS Scales	Pretest *M* (range) *N* = 16	Posttest *M* (range) *N* = 14	Pretest *M* (range) *N* = 18	Posttest *M* (range) *N* = 15	Pretest *M* (range) *N* = 19	Posttest *M* (range) *N* = 17
Adherence	2.41 (1.25–5.75)	3.60 (1.86–6.14)	2.38 (1.00–5.44)	3.22 (1.33–6.20)	2.42 (1.11–4.79)	2.29 (1.41–4.59)
Competence	4.31 (3.50–5.10)	5.16 (4.50–6.14)	3.93 (2.50–6.09)	4.53 (3.00–5.53)	3.39 (1.00–5.12)	4.52 (4.00–5.00)
Child responsiveness	4.44	5.21	4.56	5.53	4.84	5.24

Note. Teachers' implementation fidelity as measured with the BEST in CLASS Adherence and Competence scale and Child Responsiveness was scored for each teacher–child dyad. BAU = business as usual.

significance testing versus effect size for the determination of treatment effects, we elected to report Cohen's *d* rather than statistical significance test of differences (*p*-values)[3] (Patriota, 2017). Cohen's *d* effect size was selected instead of statistical significance for the following reasons: (a) Cohen's *d* can be readily used to examine the magnitude of differences between two (or more) groups on a given variable; (b) this statistic is not affected by the number of comparisons made; and (c) Cohen's *d* is not affected by whether the compared groups or treatments are dependent or independent (Cohen, 1988; Nakagawa & Cuthill, 2007). In other words, Cohen's *d* effect size provides practical information about the magnitude of the observed differences, which is particularly important with a small sample size. Means for study outcomes variables and Cohen's *d* are presented in Table 3. Common standards for assessing effect size using Cohen's *d* are: .20 is small, .50 is moderate, and .80 is large (Cohen, 1988).

Child Problem Behavior and Social Skills Outcomes

Study findings revealed BEST in CLASS—Web showed substantial reduction from pretest to posttest for children's Problem Behavior scale scores measured by the SSIS (Gresham & Elliott, 2008) (Cohen's $d = -.54$). In addition, there was a substantial increase in children's Social Skills scale scores measured by the SSIS (Cohen's $d = 1.37$). Similar patterns of results were also observed for the children in the BEST in CLASS intervention condition (Cohen's *d's* = −.71, .38, for Problem Behavior and Social Skills, respectively). No noteworthy changes from pretest to posttest scores were observed for children in the BAU condition. Inspection of Cohen's *d* effect sizes at posttest show posttest means for Social Skills under the BEST in CLASS—Web condition were on the average much higher than those for both the BEST in CLASS (Cohen's $d = 1.16$) and the BAU conditions (Cohens $d = .85$).

In regards to the Child Behavior Checklist Teacher Report Form (C-TRF; Achenbach & Rescorla, 2000), which measured children's Externalizing, Internalizing, and Total Problems, visual inspection of the means reveal BEST in CLASS—Web showed substantial reduction from pretest to posttest in children's Externalizing, Internalizing, and Total Problems (Cohen's *d's* = −.92, −.92, −.92, respectively). Moreover, posttest means of the domain scale scores were on the average much lower than those for both the BEST in CLASS and the BAU conditions; however, no significant differences were found (see Table 3).

The STRS (Pianta, 2001) was used to measure teacher−child Closeness and Conflict. BEST in CLASS—Web increased teacher−child Closeness from pretest to posttest, (Cohen's $d = 1.00$) and showed a small reduction in teacher−child Conflict (Cohen's $d = -.26$). Results also show posttest means of teacher−child Closeness were on the average slightly higher than those for both the BEST in CLASS and the BAU conditions; however, these findings were not significant (see Table 3).

Child engagement and problem behaviors were assessed with the Teacher−Child Interaction Direct Observation System Version 3.1 (TCIDOS-RV3.1). Inspection of the means indicate that for the BEST in CLASS—Web condition there were substantial changes from pretest to posttest for child engagement (see Table 3). For child problem behavior and negative interaction, the changes were small. Table 3 displays Cohen's *d* effect sizes comparing posttest TCIDOS-RV3.1 scores of children in the BEST in CLASS—Web condition to those in the BEST in CLASS and the BAU conditions. Results show small differences on the posttest means compared to those for the BEST in CLASS and moderate to large differences compared to the scores for the BAU condition (see Table 4).

Table 3. Means and Effect Sizes for Study Variables.

Measures	BEST in CLASS			BEST in CLASS—Web			BAU		
	Pretest *M*	Posttest *M*	Cohen's *d*	Pretest *M*	Posttest *M*	Cohen's *d*	Pretest *M*	Posttest *M*	Cohen's *d*
SSIS									
Problem behaviors	129.44	117.71	−0.71*	123.50	116.47	−0.54*	115.26	117.24	0.13
Social skills	71.25	78.43	0.38	73.67	88.82	1.37*	80.21	81.18	0.08
C-TRF									
Externalizing problems	68.81	65.29	−0.27	69.22	60.35	−0.92*	63.11	62.88	−0.02
Internalizing problems	60.31	58.36	−0.15	61.78	56.06	−0.92*	56.68	58.94	0.18
Total problems	66.38	64.57	−0.13	67.33	59.53	−0.92*	61.68	62.18	0.05
STRS									
Closeness	31.06	33.86	0.45	30.39	35.65	1.00*	32.0	33.06	0.26
Conflict	22.69	18.64	−0.54	21.50	19.35	−0.26	19.00	20.71	0.24
TCIDOS									
Problem behavior	0.22	0.24	0.07	0.17	0.24	0.24	0.05	0.13	0.56
Positive interaction	0.80	0.94	0.79*	0.88	0.96	0.68*	0.89	0.88	−0.07
Negative Interaction	0.22	0.24	0.08	0.17	0.24	0.24	0.07	0.13	0.40
Engagement	0.797	0.934	0.816*	0.870	0.959	0.772*	0.911	0.883	−0.169

Note. BAU = business as usual; SSIS = Social Skills Improvement System; C-TRF = Caregiver—Teacher Report Form; STRS = Student Teacher Relationship Scale; TCIDOS = Teacher–Child Interaction Direct Observation System.
*$p < .05$.

Table 4. Summary of Effects on Posttest Scores Across Treatment Conditions.

	BEST in CLASS -Web vs BEST in CLASS	BEST in CLASS vs BAU	BEST in CLASS -Web vs BAU
Measurees	Cohen's *d*	Cohen's *d*	Cohen's *d*
SSIS			
Problem behaviors	−0.11	0.04	−0.11
Social skills	1.16*	−0.15	0.85*
C-TRF			
Externalizing problems	−0.62	0.24	−0.32
Internalizing problems	−0.38	−0.05	−0.48
Total problems	−0.63	0.09	−0.33
STRS			
Closeness	0.36	0.10	0.52
Conflict	0.09	−0.26	−0.17
TCIDOS			
Problem behavior	−0.01	0.47	0.46
Positive interaction	0.11	0.23	0.33
Negative interaction	−0.01	0.47	0.46
Engagement	0.10	0.22	0.32

Note. BAU = business as usual; SSIS = Social Skills Improvement System; C-TRF = Caregiver—Teacher Report Form; STRS = Student Teacher Relationship Scale; TCIDOS = Teacher-Child Interaction Direct Observation System.
*$p < .05$.

Discussion

The purpose of this study was to conduct a conceptual replication by extending and replicating the findings from earlier research on BEST in CLASS by testing the efficacy of a web-based version of BEST in CLASS. Specifically, we examined the effects of BEST in CLASS—Web on child social-emotional and behavioral outcomes and compared these effects to BEST in CLASS delivered on-site and to a BAU control group. Findings from this preliminary RCT suggest that BEST in CLASS—Web is a promising Tier-2 approach for supporting young children with problem behavior in early childhood classrooms. Following is a brief discussion of the results, limitations of this study, and implications for future research and practice.

In summary, findings suggest that in this study, BEST in CLASS—Web was as efficacious as the on-site version and that both intervention models resulted in reduced child problem behavior; these findings are in line with previous research on BEST in CLASS (e.g., Sutherland, et al., 2018). To illustrate, in this study, effect size estimates for the Problem Behavior scale of the SSIS were −.54 and −.71 for BEST in CLASS—Web and on-site, respectively. Similarly, effect size estimates for the Externalizing Problem scale of the C-TRF were −.92 and −.27 for the BEST in CLASS—Web and BEST in CLASS conditions, respectively. Taken in sum, both conditions resulted in reductions in problem behavior from pretest to posttest, and these findings replicate previous findings on BEST in CLASS, with effect sizes of −.42 for both the Problem Behavior scale of the SSIS and the Externalizing Problem scale of the C-TRF (Sutherland et al., 2018). While these findings were not confirmed with the direct observation data of problem behavior, large effects (.77 and .82) were noted for direct observations of child engagement for the BEST in CLASS—Web and BEST in CLASS conditions, respectively.

Relatedly, large effects were found for the Social Skills scale of the SSIS for the BEST in CLASS—Web condition (1.37), while smaller effects (.38) were found for the BEST in CLASS condition. These findings replicate the positive findings ($d = .42$) on this scale from previous research (Sutherland et al., 2018), as do the improvements in teacher reported closeness ($d =$ 1.00 and .45, respectively for Web and on-site) and reductions in conflict (d] $= -.26$ and −.54,

respectively for Web and on-site) found on the STRS. These results approximate findings from Sutherland et al. of .26 (Closeness) and −.29 (Conflict) from the STRS. Furthermore, significant increases in observed positive teacher–child interactions were found in both the BEST in CLASS—Web (.68) and BEST in CLASS (.79) conditions; no reductions in negative interactions were noted.

Comparison of conditions suggest that BEST in CLASS—Web had larger, but not significant (other than social skills), effects than the on-site condition across a number of measures. While these results should be interpreted with caution, given the small sample in this study, they do highlight the potential of the web-based version of BEST in CLASS.

Limitations

Findings from this study should be interpreted with several limitations in mind. First, this was a preliminary efficacy trial that was part of an Institute of Education Sciences funded development study and as such had a small sample of teachers and children. Given this small sample interpreting between group differences should be done with caution due to the lack of statistical power. Second, due to the small sample size we were not able to conduct analyses that accounted for nesting of children within classrooms and as a result conducted multiple comparisons, which might impact confidence in the reported results. To address this limitation, our findings focused on effect size differences of mean measures rather than statistical analysis findings. In addition, observers were not blind to condition, which may have impacted their coding of behaviors. While the treatment fidelity data indicate that the teacher training and practice-based coaching of both intervention conditions resulted in improvements from pretest to posttest for adherence, competence, and child responsiveness, there was also an increase in competence of delivery noted for the BAU condition; however, these changes are relatively small. Since BEST in CLASS is a value-added program, it is to be expected that teachers are already using many of the practices that comprise BEST in CLASS. In addition, the decrease across time for adherence in the BAU condition suggests that while teachers were using the practices with more competence when they used them, they were not using them as often or as extensively as the two BEST in CLASS conditions. Fourth, while the standardized measures used in this study have evidenced strong psychometric properties in previous research, due to the small sample, we were unable to provide internal consistency statistics for the current sample. Finally, child responsiveness also increased slightly in the BAU condition from pretest to posttest, although this increase was not as large as the increase for both the Web and on-site conditions.

Implications for Future Research and Practice

While this study was limited by a small sample, the findings are interesting and warrant further investigation. One important aspect of the study design is it allowed a direct comparison of BEST in CLASS—Web to both on-site delivery as well as BAU conditions. This is important, given the recent increase in use of web-based professional development training for teachers (Bragg et al., 2021). Future research should replicate this study with samples large enough to more confidently interpret differences between conditions and account for nesting of children within classrooms. Before investing limited resources in the delivery of a web-based professional development model, more research is needed to determine the efficacy of approaches, such as BEST in CLASS—Web. At the same time, web-based delivery does appear to be a promising approach for scaling up evidence-based programs (Knight, 2012).

In addition, future research should examine the feasibility and usability of web-based platforms for intervention delivery to maximize the effectiveness of these intervention delivery models. Finally, future work should examine further scale-up of programs such as BEST in

CLASS—Web to include program-level staff as coaches to examine sustainment of these types of programs in real-world settings.

Clearly, there is need for evidence-based professional development interventions that can be delivered effectively over the web to support teachers use of effective practices addressing the social-emotional and behavioral needs of young children in their classrooms. Professional development interventions delivered over the web can help provide support to teachers who may be located in areas that have limited access (e.g., rural areas) (Dede et al., 2009) or provide a more convenient way to provide professional development (Ke & Xie, 2009; Olsen et al., 2010). However, as early childhood programs consider using web-based professional development interventions, they should consider a number of practical implications including program characteristics (e.g., accessibility to high-speed internet), funding (e.g., technology), and compatibility with skills and time (Durlak & Dupre, 2008). For example, many early childhood programs are located in rural areas and access to high-speed internet may not be available. In addition, there are technology costs associated with delivering professional development over the web. Finally, teachers may have varying technology capabilities that may contribute to or interfere with learning over the web.

Conclusion

This study examined the effects of BEST in CLASS—Web on child outcomes in comparison to on-site and BAU conditions. Findings from this study are promising and indicate that more research is warranted to better understand how the use of web-based delivery platforms can enhance the delivery of evidence-based practices to young children with problem behavior. As we learn more about factors associated with the sustainment of evidence-based programs within implementation research (Shelton et al., 2018; Wiltsey Stirman et al., 2012), web-based delivery of evidence-based programs may be particularly important not only for supporting children who have social-emotional learning difficulties but also for providing access to evidence-based professional development to teachers across early childhood settings.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the U.S. Department of Education, Institute of Education Sciences (Grant Number R324A160158). The opinions expressed by the authors are not necessarily reflective of the position of or endorsed by the U.S. Department of Education.

ORCID iDs

Maureen A. Conroy https://orcid.org/0000-0002-0170-3299
Katerina M. Marcolouides https://orcid.org/0000-0001-8829-870X

Notes

1. The BEST in CLASS intervention manuals and materials are available upon request.
2. Note: BAU data were collected during a previous efficacy trial and collected during a different time period than the two intervention conditions. Geographical location, classroom composition, and settings were similar to the two intervention conditions.

3. Although Cohen's *d* effect size measures are highlighted and interpreted in this text, statistical significance probability (*p*) values are also noted in Table 3. Readers should use caution when interpreting the reported significance values, not only because there were multiple comparisons conducted in this study; but, also because the sample sizes were small. In addition, the utility of significance values for evaluating the noteworthiness of research is debated (Thompson, 2002).

References

Achenbach, T. M., & Rescorla, L. A. (2000). *CBCL/1, 5-5 & C-TRF/1, 5-5 profiles*. Research Center for Children, Youth, and Families.

Beyer, T., Postert, C., Muller, J. M., & Furniss, T. (2012). Prognosis and continuity of child mental health problems from preschool to primary school: Results of a four-year longitudinal study. *Child Psychiatry and Child Development*, *43*(4), 533–543. https://doi.org/10.1007/s10578-012-0282-5

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: Experiential canalization of brain and behavior. *American Psychologist*, *67*(4), 309–318. https://doi.org/10.1037/a0027493

Bragg, L. A., Walsh, C., & Heyeres, M. (2021). Successful design and delivery of online professional development for teachers: A systematic review of the literature. *Computers & Education*, *166*, Article 104158. http://www.elsevier.com/locate/compedu

Briggs-Gowan, M. J., & Carter, A. S. (2008). Social-emotional screening status in early childhood predicts elementary school outcomes. *Pediatrics*, *121*(5), 957–962. https://doi.org/10.1542/peds.2007-1948

Brown, W. H., & Conroy, M. A. (2012). Social-emotional competence in young children with developmental delays: Our reflection and vision for the future. *Journal of Early Intervention*, *33*(4), 310–320. https://doi.org/10.1177/1053815111429969

Bulotsky-Shearer, R. J., Bell, E. R., & Dominguez, 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*(6), 775–798. https://doi.org/10.1016/j.jsp.2012.08.001

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, *12*(3), 467–488. https://doi.org/10.1017/S0954579400003114

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.

Conroy, M. A., Bulotsky-Shearer, R., Morris, C., & Hetrick, A. A. (2020). Prevention and intervention in preschool and early elementary school years. In T. W. Farmer, M. A. Conroy, E. M. Z. Farmer, & K. Sutherland (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 35–50). Routledge.

Conroy, M. A. & Sutherland, K. S. (2011). *BEST in CLASS–PK intervention: Research version 2.0* [Unpublished intervention]. Anita Zucker Center of Excellence in Early Childhood Studies, University of Florida.

Conroy, M. A., Sutherland, K. S., Algina, J., Ladwig, C., Werch, B., Martinez, J., Jessee, G., & Gyure, M. (2019). Outcomes of the BEST in CLASS intervention on teachers' use of effective practices, self-efficacy, and classroom quality. *School Psychology Review*, *48*(1), 31–45. https://doi.org/10.17105/SPR-2018-0003.V48-1

Conroy, M. A., Sutherland, K. S., Wilson, R. E., Martinez, J., Whalon, K. J., & Algina, J. (2015). Measuring teacher implementation of the BEST in CLASS intervention program and corollary child outcomes. *Journal of Emotional and Behavioral Disorders*, *23*(3), 144–155. https://doi.org/10.1177/1063426614532949

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. https://doi.org/10.1007/BF02310555

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, *60*(1), 8–19. https://doi.org/10.1177/0022487108327554

Driscoll, K. C., & Pianta, R. C. (2010). Banking time in Head start: Early efficacy of an intervention designed to promote supportive teacher-child relationships. *Early Education and Development*, *21*(1), 38–64. https://doi.org/10.1080/10409280802657449

Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3-4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Feil, E. G., Baggett, K., Davis, B., Landry, S. H., Sheeber, L. B., Leve, C., & Johnson, U. (2020). Randomized control trial of an internet-based parenting intervention for mothers of infants. *Early Childhood Research Quarterly*, *50*, 36–44. https://doi.org/10.1016/j.ecresq.2018.11.003

Feil, E. G., Baggett, K. M., Davis, B., Sheeber, L., Landry, S., Carta, J. J., & Buzhardt, J. (2008). Expanding the reach of preventive interventions: Development of an Internet-based training for parents of infants. *Child Maltreatment*, *13*(4), 334–346. https://doi.org/10.1177/1077559508322446

Feil, E. G., Walker, H. M., & Severson, H. H. (1995). The Early Screening Project for young children with behavior problems. *Journal of Emotional and Behavioral Disorders*, *3*(4), 194–202. https://doi.org/10.1177/106342669500300401

Forness, S. R., Freeman, S. F. N., Paparella, T., Kauffman, J. M., & Walker, H. M. (2012). Special education implications of point and cumulative prevalence for children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *20*(1), 4–18. https://doi.org/10.1177/1063426611401624

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Teacher rating scales*. Pearson Assessments.

Hemmeter, M. L., Ostrosky, M., & Fox, L. (2006). Social and emotional foundations for early learning: A conceptual model for intervention. *School Psychology Review*, *35*(4), 583–601. https://doi.org/10.1080/02796015.2006.12087963

Hetzner, N. P., Johnson, A. D., & Brooks-Gunn, J. (2011). Poverty, effects on social and emotional development. In S. Jarvela (Ed.), *Social and emotional aspects of learning* (pp. 77–86). Elsevier.

Ke, F., & Xie, K. (2009). Toward deep learning for adult students in online courses. *The Internet and Higher Education*, *12*(3–4), 136–145. https://doi.org/10.1016/j.iheduc.2009.08.001

Knight, D. S. (2012). Assessing the cost of instructional coaching. *Journal of Education Finance*, *38*(1), 52–80. https://www.jstor.org/stable/23259121

Kupersmidt, J. B., Bryant, D., & Willoughby, M. T. (2000). Prevalence of aggressive behaviors among preschoolers in Head Start and community child care programs. *Behavioral Disorders*, *26*(1), 42–52. https://doi.org/10.1177/019874290002600105

McCabe, P. C., & Altamura, M. (2011). Empirically valid strategies to improve social and emotional competence of preschool children. *Psychology in the Schools*, *48*(5), 513–540. https://doi.org/10.1002/pits.20570

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*(4), 947–959. http://dx.doi.org/10.1037/0012-1649.43.4.947

McDonald, R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, *82*, 591–605.

National Science Foundation and Institute of Education Sciences. (2018). *Companion guidelines on replication & reproducibility in education research: A supplement to the common guidelines for education research and development*. https://ies.ed.gov/pdf/CompanionGuidelinesReplicationReproducibility.pdf

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

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*, 1–13. https://doi.org/10.1007/BF02289400

Olsen, H., Donaldson, A. J., & Hudson, S. D. (2010). Online professional development: Choices for early childhood educators. *Dimensions of Early Childhood*, *38*(1), 12–18. https://www.learntechlib.org/p/107028/

Patriota, A. G. (2017). On some assumptions of the null hypothesis statistical testing. *Educational and Psychological Measurement*, *77*(3), 507–528. https://doi.org/10.1177/0013164416667979

Pianta, R. C. (2001). *Student-Teacher Relationship Scale–short form*. Psychological Assessment Resources.

Ringeisen, H., Stambaugh, L., Bose, J., Casanueva, C., Hedden, S., Avenevoli, S., . . . West, J. (2017). Measurement of childhood serious emotional disturbance: State of the science and issues for consideration. *Journal of Emotional and Behavioral Disorders*, *25*, 195–210. https://doi.org/10.1177/1063426616675165

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). Wiley.

Schindler, H. S., Kholoptseva, J., Oh, S. S., Yoshikawa, H., Duncan, G. J., Magnuson, K. A., & Shonkoff, J. P. (2015). Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis. *Journal of School Psychology*, *53*(4), 243–263. https://doi.org/10.1016/j.jsp.2015.04.001

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108

Shelton, R. C., Cooper, B. R., & Stirman, S. W. (2018). The sustainability of evidence-based interventions and practices in public health and health care. *Annual Review of Public Health*, *39*(1), 55–76. https://www.annualreviews.org/doi/pdf/10.1146/annurev-publhealth-040617-014731

Skinner, B. F. (1953). *Science and human behavior*. Simon & Schuster Inc.

Skyhar, C. (2020). Thinking outside the box: Providing effective professional development for rural teachers. *Theory & Practice in Rural Education*, *10*(1), 42–72. https://doi.org/10.3776/tpre.2020.v10n1p42-72

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*(3), 133–143. https://doi.org/10.1177/0271121415594925

Snyder, P. A., Hemmeter, M. L., McLean, M., Sandall, S., McLaughlin, T., & Algina, J. (2018). Effects of professional development on preschool teachers' use of embedded instruction practices. *Exceptional Children*, *84*(2), 213–232. https://doi.org/10.1177/0014402917735512

Spybrook, J., & Raudenbush, S. (2009). An examination of the precision and technical accuracy of the first wave of group-randomized trials funded by the Institute of Education Sciences. *Educational Evaluation and Policy Analysis*, *31*(3), 298–318. https://doi.org/10.3102/0162373709339524

Sutherland, K. S., Conroy, M. A., Algina, J., Ladwig, C., Jessee, G., & Gyure, M. (2018). Reducing child problem behaviors: A randomized controlled trial of BEST in CLASS. *Early Childhood Research Quarterly*, *42*(1), 31–43. https://doi.org/10.1016/j.ecresq.2017.08.001

Sutherland, K. S., Conroy, M. A., Abrams, L., Vo, A., & 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. https://doi.org/10.1177/1534508412463814

Sutherland, K. S., & McLeod, B. D. (2010). *Treatment adherence and competence measure for classroom based interventions* [Unpublished instrument]. Department of Special Education and Disability Policy, Virginia Commonwealth University.

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. https://doi.org/10.1177/1063426613497258

Tapp, J. (2010). *Lily data collector* [Computer software]. Vanderbilt Kennedy Center.

Tapp, J., Wehby, J. H., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 25–31. https://doi.org/10.3758/BF03203616

Thompson, B. (2002). Statistical, practical, and clinical: How many kinds of significance do counselors need to consider? *Journal of Counseling and Development*, *80*, 64–71.

TORSH Inc. (2018a). *Today's One Room School House: TORSH Talent*. Louisiana Entertainment. https://www.torshtalent.com/

TORSH Inc. (2018b). *TORSH Talent* (Version 1.0.11).[Mobile app]. App store. https://apps.apple.com/us/app/torsh-talent/id577665140

Travers, J. C., Cook, B. G., Therrien, W. J., & Coyne, M. D. (2016). Replication research and special education. *Remedial and Special Education*, *37*(4), 195–204. https://doi.org/10.1177/0741932516648462

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. National Staff Development Council.

Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation Science*, *7*(1), Article 17. https://doi.org/10.1186/1748-5908-7-17

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, *67*(4), 272–284. https://doi.org/10.1037/a0028015