

# Preliminary Study of the Effects of BEST in CLASS–Elementary on Outcomes of Elementary Students With Problem Behavior

Journal of Positive Behavior Interventions  
2020, Vol. 22(4) 220–233  
© Hammill Institute on Disabilities 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1098300719900318  
jpb.sagepub.com  
SAGE

K. S. Sutherland, PhD<sup>1</sup> , M. A. Conroy, PhD<sup>2</sup>,  
B. D. McLeod, PhD<sup>1</sup>, K. Granger, PhD<sup>1</sup>, M. Broda, PhD<sup>1</sup> ,  
and R. Kunemund, PhD<sup>1</sup>

## Abstract

Student problem behaviors in early elementary school are associated with fewer learning opportunities, poor relationships with teachers, and academic and behavioral problems, including increased risk of emotional/behavioral disorders (EBDs). This study examined the effect of BEST in CLASS–Elementary (BEST in CLASS-E), a Tier 2 intervention delivered by teachers, on student problem behavior, academic achievement, teacher–student relationships, and classroom quality using a randomized controlled trial design. Participants were 45 kindergarten to grade 3 students identified at risk of development of EBDs and their 26 teachers from three elementary schools located in an urban school district. BEST in CLASS-E was found to decrease teacher-reported student problem behavior and increase teacher–student closeness; no effects were noted for academic achievement or overall classroom quality. Results suggest the promise of BEST in CLASS-E as a Tier 2 intervention delivered by teachers in elementary classrooms. Implications and limitations of the study are discussed.

## Keywords

challenging behavior, classroom-based studies, social relationships

Approximately 0.5% to 1% of the school-age population receives special education services for an emotional/behavioral disorder (EBD) under the Individual with Disabilities Education Act (IDEA; Forness, Freeman, et al., 2012). Yet, epidemiological studies indicate that many more children and youth are at risk of or have an EBD with prevalence rates ranging anywhere from 11.5% to 30% (e.g., Forness, Kim, et al., 2012; Ringeisen et al., 2017). Whether or not students are formally identified and found eligible under IDEA, they attend school and their chronic problem behaviors can have a negative impact on their long-term academic and behavioral success in school (Kellam et al., 1998; Myers & Pianta, 2008; Spilt, Koomen, et al., 2012). One such outcome identified in the literature is the development of negative and coercive interactions with their teachers and peers (e.g., Doumen et al., 2008).

If chronic problem behaviors are not addressed early on, students with and at risk of EBD may develop negative interaction patterns with adults in their lives, including their teachers (see Gunter & Coutinho, 1997; Patterson, 2002, for a discussion). When students and teachers engage in negative interactions, teachers often lack the training and skills to proactively address these interactions (O’Conner et al., 2011;

Sutherland et al., 2008). In fact, Reinke and colleagues (2011) reported that only 17% of elementary school teachers rated themselves as having training in practices to address chronic problem behaviors in their classrooms. Rather than using evidence-based practices that de-escalate negative interactions and teach students skills to engage positively, teachers often respond with harsh, negative feedback (McClowry et al., 2013). To escape these negative interaction patterns, teachers may also decrease students’ learning opportunities by providing fewer opportunities to respond and less attention (Farmer et al., 2016; Sutherland et al., 2008). The cumulative negative outcomes associated with sustained negative interactions and lack of learning opportunities seriously affect students’ future academic, social, and behavioral outcomes (Brock et al., 2008). Therefore, a critical need exists to

<sup>1</sup>Virginia Commonwealth University, Richmond, USA

<sup>2</sup>University of Florida, Gainesville, USA

## Corresponding Author:

K. S. Sutherland, Virginia Commonwealth University, 1015 W Main St,  
P. O. Box 842020, Richmond, VA 23284, USA.  
Email: kssuther@vcu.edu

**Action Editor:** Wendy Oakes

address chronic problem behavior and negative teacher–student interaction patterns.

## Tier 2 Interventions Targeting Students With or at Risk of EBD

Over the past 30 years, there has been an increasing emphasis on the use of positive behavioral intervention and supports (PBIS) to address problem behaviors in schools. The focus of PBIS is to provide tiered systems of supports early and systematically with an emphasis on prevention and amelioration of problem behaviors (Sugai & Horner, 2009). Within a PBIS framework, Tier 2 interventions provide additional supports to students whose behavior continues to be problematic after receiving high-quality Tier 1 supports. Typically provided to a smaller group of students in need of supports beyond those of Tier 1 or more universal strategies, a number of Tier 2 interventions have been developed and have effectively addressed the behavioral needs of many students in school settings (for a review, see Anderson & Borgmeier, 2010; Bruhn et al., 2014). BEST in CLASS–Elementary (BEST in CLASS-E; Sutherland et al., 2019) is one such Tier 2 intervention designed to address the needs of early elementary-aged students who have been identified as at risk of EBD and engage in chronic problem behaviors and negative interactions with their teachers.

### Overview of BEST in CLASS-E

Through funding from the U.S. Department of Education, Institute of Education Sciences, BEST in CLASS-E was designed to increase positive teacher–student interactions and increase learning opportunities for students in kindergarten to grade 3 who are at risk of or with EBD. Adapted from the original BEST in CLASS for delivery in early childhood classrooms (see Conroy et al., 2018), BEST in CLASS-E was founded on Sameroff’s (2009) transactional theory, which when applied to interactions between students and teachers postulates that student behavior (whether positive or negative) influences the behavior of teachers and vice versa. These transactions between teacher and student influence, and are influenced by, the broader ecology within the classroom (Bronfenbrenner, 1979). An iterative development process with teachers and families of young students with and at risk of EBD, including pilot testing in urban elementary schools, indicated both promise for the intervention and contextual fit in the schools included in the current study (see Sutherland et al., 2019, for a description of this intervention development process).

BEST in CLASS-E is comprised of three components: (a) a teacher training workshop on the BEST in CLASS-E practices; (b) a teacher resource manual, which provides additional materials and supports; and (c) 14 weeks of practice-based coaching. Through these three components,

teachers learn how to use five instructional practices (i.e., supportive relationships, rules, precorrection, opportunities to respond, praise) that, when used effectively, increase learning opportunities and positive interactions with students with and at risk of EBD in their classroom who engage in elevated rates of chronic problem behaviors. In addition, teachers learn proactive ways to engage the students’ families in their child’s education (for a description, see Conroy, McKnight, et al., 2019).

A unique aspect of BEST in CLASS-E in comparison with other Tier 2 interventions, which are implemented in small groups or individually, is that teachers learn to embed the BEST in CLASS-E practices within naturally occurring learning opportunities as they interact with these students throughout their school day. BEST in CLASS-E is considered a “value-added” intervention, in that teachers may already be using many of the BEST in CLASS-E instructional practices during learning activities in their classrooms. When implementing BEST in CLASS-E as designed, teachers are taught to increase the *quantity* and *quality of delivery* of the BEST in CLASS-E instructional practices with focal students who are engaging in sustained rates of problem behaviors. For example, most teachers have classroom rules to teach all students in the classroom the behavioral expectations (i.e., a Tier 1 intervention). However, BEST in CLASS-E provides teachers with instruction and support to help them learn how to teach focal students to learn the classroom rules, and embed the use of rules more consistently and frequently within each instructional activity in the classroom, increasing the intensity and dosage of intervention. Likewise, teachers commonly provide students opportunities to respond during instructional activities; however, BEST in CLASS-E helps teachers learn how to increase the frequency and quality of opportunities to respond with the focal students to increase their engagement in the instructional activity. As a result, students engage in fewer problem behaviors and have increased learning opportunities resulting in fewer negative interactions with their teachers.

### Overview of Current Study

Previous research on BEST in CLASS has demonstrated positive outcomes for teachers (Conroy, Sutherland, et al., 2019) and children (Conroy et al., 2018; Sutherland, Conroy, Algina, et al., 2018) in early childhood settings. To illustrate, Conroy, Sutherland, et al. (2019) found that teachers in the BEST in CLASS condition had increased self-efficacy ( $ES$  range = 0.50–0.78) and Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) domain scores ( $ES$  range = 0.47–0.65). In terms of child outcomes, Sutherland, Conroy, Algina, et al. (2018) found reductions in teacher-reported problem behavior ( $ES = -42$ ) on the Social Skills Improvement System–Rating Scale (SSIS-RS;

Gresham & Elliott, 2008), as well as improved closeness ( $ES = 0.26$ ) and reduced conflict ( $ES = -0.29$ ) on the Student–Teacher Relationship Scale (STRS; Pianta, 1993). Although positive outcomes were found in early childhood settings, an investigation of the initial efficacy of BEST in CLASS-E for elementary students with and at risk of EBD who engage in chronic problem behaviors has not been conducted. This article reports findings from an initial investigation examining the effects of BEST in CLASS-E in elementary school classrooms. Several social, emotional, and behavioral outcomes for students who are at risk of or with EBD were examined, as well as the overall classroom quality. The following research questions were addressed:

**Research Question 1:** What is the effect of BEST in CLASS-E on student problem behavior?

**Research Question 2:** What is the effect of BEST in CLASS-E on student academic achievement?

**Research Question 3:** What is the effect of BEST in CLASS-E on student–teacher relationships?

**Research Question 4:** What is the effect of BEST in CLASS-E on classroom quality?

## Method

### Setting and Participants

The researchers recruited teacher participants from three elementary schools in an urban school district in a mid-Atlantic state. The participating schools served predominantly African American students (94%, 93%, and 98%, respectively) from a low-income community (82%, 96%, and 96% free and reduced lunch, respectively). The mean number of students per school was 364 ( $SD = 73.9$ ).

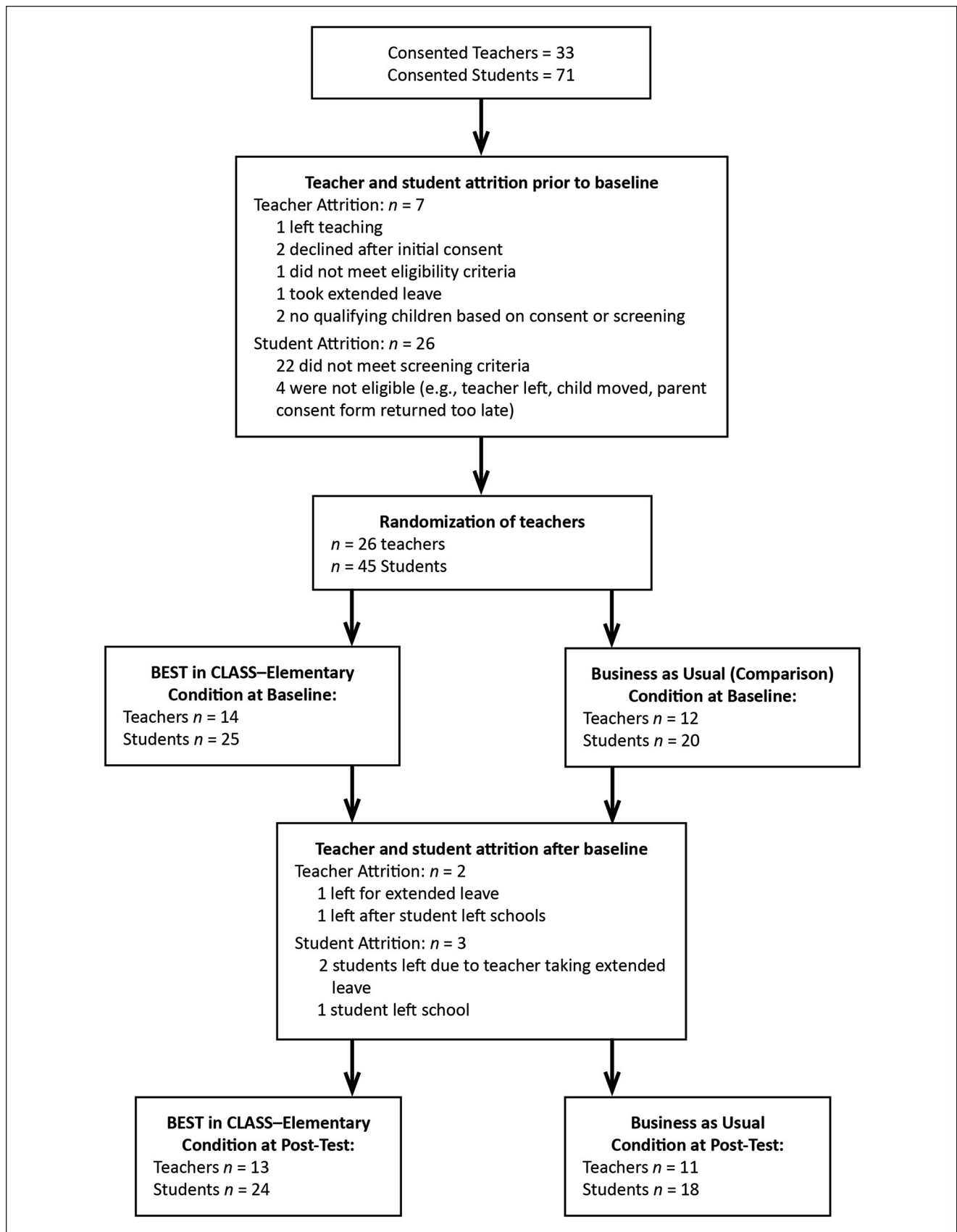
The schools indicated that no formal schoolwide positive behavior interventions and supports (SWPBIS) were in place, although evidence of informal application of SWPBIS were observed. Tier 1 supports were implemented with some variability across classrooms, and teachers engaged in classroom management practices, but at low levels of fidelity. To illustrate, the average CLASS (Pianta et al., 2008) scores for BEST in CLASS-E and business as usual (BAU) condition, respectively, for the subscales of Classroom Organization (3.91 and 3.93), Instructional Support (1.95 and 1.82), and Emotional Support (4.62 and 4.24) at pretest in the current study were considerably lower than scores found in other studies. For example, Jennings et al. (2017) noted CLASS scores of 4.86 (Classroom Organization), 2.75 (Instructional Support), and 4.92 (Emotional Support) in 224 elementary classrooms, whereas Sandilos et al. (2014) noted scores of 5.16 (Classroom Organization), 3.03 (Instructional Support), and 5.26 (Emotional Support) in 426 elementary classrooms. CLASS scores in the current study would suggest a lack of Tier 1 supports at the classroom level.

**Teachers.** A total of 26 teachers participated in the study, with 14 in the BEST in CLASS-E intervention group and 12 in BAU (see Figure 1). Teachers were eligible for inclusion if they met the following criteria: (a) taught in kindergarten to grade 3, (b) served at least one child identified as being at risk of EBDs, and (c) consented to participate. There were no significant differences between the BEST in CLASS-E intervention group and the comparison group on the teacher demographic variables in Table 1.

**Students.** A total of 45 students participated in the study, with 25 in the BEST in CLASS-E intervention group and 20 in the BAU group (see Figure 1). Students who met the following criteria were eligible for participation: (a) enrolled in a participating teacher’s classroom, (b) presence of externalizing behaviors that interfere with participation in the classroom (e.g., disruption, aggression) as indicated by the Systematic Screening for Behavior Disorders (SSBD; Walker et al., 2014), and (c) parental/guardian consent to participate. To determine eligibility for participation, in Stage 1, teachers nominated up to five students in their classroom who engaged in chronic problem behavior based on a list of example and nonexample behaviors. Caregiver consent was then obtained and systematic screening for risk of EBDs took place using the SSBD Stage 2. After screening, one to two students per classroom were selected to participate in the study, depending upon returned caregiver consents and the most elevated scores on the SSBD. All students who screened into the study met the criteria for “at risk” as defined by the normed cutoff criteria on the SSBD (see Walker et al., 2014). There were no significant differences between the intervention and comparison groups on the student demographic variables in Table 1.

### Measures

**SSBD.** The SSBD (Walker et al., 2014) is a three-stage multigate screening system designed to proactively identify students who are at risk of negative developmental outcomes associated with their internalizing and externalizing behavior patterns. The tool combines teacher ratings of the frequency and intensity of student adjustment problems in school with trained observer ratings of student’s adaptive and maladaptive behaviors. In the present study, we used the externalizing subscale in which students are categorized into risk levels associated with four scales: critical events, aggressive behavior, maladaptive behavior, and adaptive behavior. The SSBD suggests that the first two stages be completed, and allow the third stage, observation by trained observers, to be optional; we used the first two stages to identify focal students. In Stage 1, teachers rank the top five students who engage in externalizing behavior based on a list of example and nonexample behaviors. In Stage 2, students with caregiver consent are then rated by their teacher



**Figure 1.** The CONSORT flow diagram for teacher and student participation in the study.

**Table 1.** Participant Demographics by Study Group.

Participant	BAU	BEST in CLASS-E	Total
Teachers	<i>n</i> = 12	<i>n</i> = 14	26
Age range (years)			
18–25	1	3	4
26–35	5	6	11
36–45	3	2	5
46–55	3	2	5
Above 55	0	1	1
Gender			
Female	12	14	26
Hispanic/Latinx	0	1	1
Other	0	1	1
Non-Hispanic/Latinx	12	13	25
African American/Black	7	4	11
Caucasian/White	5	8	13
Other	0	1	1
Years teaching	<i>M</i> = 6.33, <i>SD</i> = 6.95	<i>M</i> = 6.07, <i>SD</i> = 7.85	<i>M</i> = 6.19, <i>SD</i> = 7.29
Education			
Bachelor's degree	8	6	14
Master's degree	4	8	12
Grade taught			
Kindergarten	2	4	6
First	3	4	7
Second	3	2	5
Third	2	2	4
Special education	2	2	4
Students	<i>n</i> = 20	<i>n</i> = 25	<i>n</i> = 45
Age	<i>M</i> = 7.48, <i>SD</i> = 1.35	<i>M</i> = 7.35, <i>SD</i> = 1.31	<i>M</i> = 7.39, <i>SD</i> = 1.31
Gender			
Male	17	20	37
Female	3	5	8
Non-Hispanic/Latinx	16	25	41
African American/Black	1	0	1
Caucasian/White	1	0	1
Other	1	0	1
Grade			
Kindergarten	4	7	11
First	5	7	12
Second	5	6	11
Third	4	5	9
Special education	2	0	2

Note. BAU = business as usual condition; BEST in CLASS-E = treatment condition; CLASS = Classroom Assessment Scoring System.

on the critical events index (CEI) and combined frequency index (CFI). Both stages exhibit strong psychometric properties. Stage 1 shows acceptable test-retest stability estimates (e.g., 69% of students ranked as the top three externalizers were also ranked in the top three at a ranking Time 2, and Spearman's rank order coefficients between Stage 1 rankings at two time points reveal a mean *r* coefficient of .76 for externalizers; Walker et al., 2014). Stage 2 also demonstrates internal consistency, with Adaptive

Behavior Scale alphas of .85 and .88, across the two ratings 1 month apart. For the Maladaptive Behavior Scale, these coefficients were .82 and .87 (Walker et al., 2014).

**SSIS-RS.** The SSIS-RS (Gresham & Elliott, 2008) is a 76-item teacher-report measure, allowing for the evaluation of social skills and problem behaviors of young students. Each item on the SSIS-RS is rated on a 4-point frequency scale, with responses ranging from 0 (*never*) to 3 (*almost*



*always*). Items are grouped into two subscales: Social Skills (e.g., completes tasks without bothering others) and Problem Behaviors (e.g., talks back to adults), with higher scores indicating more social skills or more problem behavior. For the current sample, internal consistency was acceptable with Cronbach's alpha equal to .93 and .95 for Social Skills at pretest and posttest, respectively, and .89 and .95 for Problem Behavior at pretest and posttest, respectively.

*Woodcock-Johnson-III Brief Battery (WJ-III)*. Academic achievement was measured using two subtests of the WJ-III (Letter-Word Identification and Math Applied Problems; Woodcock et al., 2007), which are ideal for measuring academic progress multiple times in 1 year. The Brief Reading and Brief Math cluster raw subscales were used (Woodcock et al., 2007), higher scores indicating more skills. The subscales of the WJ-III consistently demonstrate high reliability with internal consistency of .80 or higher.

*STRS*. Designed to measure teachers' perceptions of their relationships with students, the STRS (Pianta, 1993) is a teacher-report measure with subscales assessing domains of Closeness (the degree of warmth, positive emotions, and open communication between teacher and student) and Conflict (the degree of negative interactions and emotions involving teacher and student). Items are rated on a 5-point Likert-type scale, with responses ranging from 1 = *definitely does not apply* to 5 = *definitely applies* and scores summed for each domain, with higher scores indicating more conflict or more closeness. For the current sample, internal consistency was acceptable with Cronbach's alpha equal to .83 and .78 for Closeness at pre- and posttest, respectively, and .80 and .84 for Conflict at pre- and posttest, respectively.

*Problems Preparing Children for Academic Success (PPCAS)*. Teachers completed the PPCAS scale, which was adapted from the School Staffing Survey (National Center for Education Statistics, 1993). Teachers report on the following prompt, "How much of a problem are the factors below in preparing your children to success academically?" Seventeen factors are listed and include home/family life, parent cooperation/support, child health, inadequate nutrition, low intelligence, cultural differences, English proficiency, nonstandard English, special learning problems, behavioral problems (disruptive), inadequate supplies, students not ready academically, students have attention problems, and student tardiness/absenteeism. Teachers respond using a 4-point scale from 1 = *not a problem* to 4 = *serious problem*. Ratings were averaged across the 17 items ( $\alpha = .83$ ), with higher scores indicating higher teacher-perceived classroom-level adversity. No differences were present between BEST in CLASS-E and BAU classrooms on classroom-level adversity ( $M = 3.00$  and  $3.01$ ,  $SD = 0.61$  and  $0.64$ , respectively).

*CLASS*. The CLASS (Pianta et al., 2008) is an observational instrument designed to evaluate classroom quality. The CLASS addresses 10 separate dimensions of classroom experiences over three broad domains, including Emotional Support, Classroom Organization, and Instructional Support. Trained observers scored each of the dimensions supporting the broad domain using a 1 to 7 scale with three categories: low (1, 2), mid (3, 4, 5), and high (6, 7). A minimum of four observation cycles were completed on each classroom, and dimension scores were averaged across cycles to create broad domain scores. A higher score indicates high quality for each dimension. For the current sample, internal consistency was acceptable with Cronbach's alpha equal to .81 for Emotional Support at pretest and .81 for posttest, .93 for Classroom Organization at pretest and .93 for posttest, and .88 for Instructional Support at pretest and .85 for posttest.

Certified observers conducted observations using the CLASS at pretest and posttest. Observers participated in a 2-day training led by a certified CLASS trainer and passed the reliability test to achieve initial certification. All observers recalibrated prior to posttest data collection. Interobserver agreement data were collected on 23% of all CLASS observations, and the mean intraclass correlations (ICCs) were .81, .87, and .58 for Emotional Support, Classroom Organization, and Instructional Support, respectively.

*Treatment integrity measure*. The Treatment Integrity Instrument for Elementary School Classrooms (TIES; Sutherland et al., 2017) is an observational measure in which raters assess teachers' extensiveness (i.e., adherence; five items), quality of delivery (i.e., competence; five items), and student responsiveness (one item) using a 7-point Likert-type scale. Adherence ratings were comprised of two key components: thoroughness and frequency. Anchors on the adherence and student responsiveness items range from *not at all* to *very extensive*. Competence ratings were comprised of estimates of the skillfulness of delivery, timing, and responsiveness to a given child and situation. Anchors on the competence items range from *very poor* to *excellent*. Trained observers conducted 87 observations in BEST in CLASS-E and BAU classrooms at pretest or posttest; a secondary observer conducted reliability checks during 34 observations (39.1% of the total observations). ICCs were computed for each item on each scale. Cicchetti (1994) indicated that ICCs less than .40 reflect "poor" agreement, ICCs from .40 to .59 represent "fair" agreement, ICCs from .60 to .74 represent "good" agreement, and ICCs of .75 and higher represent "excellent" agreement. The mean ICC for the adherence scale was .82 ( $SD = 0.10$ , range = .68–.92), with all items reflecting "good" to "excellent" agreement. The mean ICC for the competence scale was .61 ( $SD = 0.11$ , range = .52–.77), with items representing "fair" to "excellent" agreement. The ICC for the student responsiveness item was .68, representing "good" agreement.

## Design and Experimental Procedures

This study was a randomized trial. Students were nested in teachers' classrooms, and teachers were randomly assigned to the BEST in CLASS-E or to a BAU comparison condition. Teachers were randomly assigned to a condition from within their school and when possible, grade level (i.e., kindergarten, grades 1, 2, 3). For example, in schools with two consented teachers per grade level, one teacher was randomly assigned to BEST in CLASS-E and one to BAU. To minimize contamination across conditions, teachers were provided information about their roles in the study and the importance of not discussing participation with any other teachers in their school.

All study activities were approved by a human participants protection board. Obtaining teacher consent and screening of student participants began approximately 1 month after the beginning of school to allow teachers to familiarize themselves with the students in their classrooms. Once teacher consent was obtained, students were screened to determine eligibility for participation and caregiver consent was received. Subsequently, pretest measures were completed. Following completion of pretest measures, randomization occurred (see Figure 1 for study flow). Teachers in the BEST in CLASS-E condition received the BEST in CLASS-E 1-day teacher training, and the following week began receiving weekly practice-based coaching (see "Method" section). In April, posttest measures were completed with all participating teachers and students.

## Treatment and Comparison Conditions

**BEST in CLASS-E.** Students in the BEST in CLASS-E condition were exposed to BAU with the addition of BEST in CLASS-E. BEST in CLASS-E is considered a Tier 2 intervention, because teachers systematically identify specific focal students in their classrooms who are at risk of EBDs, and after receiving training and practice-based coaching to increase the quality and quantity of targeted instructional practices implement these practices with these students during instructional activities throughout the school day. The intention of BEST in CLASS-E is to increase a teacher's use of specific instructional practices with focal students that facilitate positive teacher-student interactions and student engagement while decreasing the occurrence of student problem behaviors (e.g., disruption, defiance). This is accomplished through the implementation of three key components delivered to teachers: (a) the BEST in CLASS-E teacher manual, (b) the BEST in CLASS-E teacher workshop (a 1-day didactic teacher training), and (c) practice-based coaching (14 weeks of one-on-one practice-based coaching with performance feedback).

The BEST in CLASS-E teacher manual provides an overview of the intervention components and implementation

process. The manual is comprised of seven modules; five of the modules provide instructional support on the BEST in CLASS-E practices (i.e., supportive relationships, rules, pre-correction, opportunities to respond, and praise) and the final module helps the teacher learn to link these practices together. The first module during coaching focuses on supporting the teacher to partner with families of focal students, and this home-school partnership component is reviewed with teachers weekly. Each module includes a definition of each practice and steps for successful implementation with focal students. Supporting research and high-quality examples of each practice are provided. Teachers are introduced to the manual at the BEST in CLASS-E teacher workshop, but continue to use the manual throughout the 14 weeks of practice-based coaching as they master each module.

At the BEST in CLASS-E teacher workshop, led by university faculty and coaches, teachers receive detailed information about the use of each BEST in CLASS-E practice with focal students. The training includes modeling, video exemplars, and hands-on activities that are used to facilitate discussion between teachers and coaches about the specific problem behaviors of focal students in their classroom.

The BEST in CLASS-E coaching process begins the week after the teacher workshop and lasts for 14 weeks. The coaching process is cyclical, and teachers and coaches collaboratively set a new goal each week for using a BEST in CLASS-E practice. Coaches and teachers spend 2 weeks focusing their goals on the implementation of each practice as well as how that practice can be linked with other BEST in CLASS-E practices in an efficient manner. Ultimately, the coaching is intended to extend what teachers have learned from the manual and workshop to facilitate their use of the BEST in CLASS-E practices with the identified focal students in their classroom.

At each coaching meeting, a coaching plan is developed and includes specific ways the teacher will implement practices to reduce focal student's problem behaviors and increase engagement as well as how the coach will support the teacher's efforts. In addition to the development of a coaching plan, the coach also observes the teacher's implementation of the plan in the classroom with focal students during the selected activity. The coach's observation lasts approximately 15 min per teacher-student dyad during which time the coach observes and videotapes each teacher-student dyad, takes anecdotal notes, and collects observational data on teachers' use of the BEST in CLASS-E practices, teacher-student interactions, and student engagement and problem behaviors using the TIES (Sutherland et al., 2017). Following the teacher's implementation of the plan and the coach's observation, the teacher reflects and evaluates on the implementation of the plan and the coach summarizes the observation. The teacher's reflection is discussed at a subsequent coaching meeting, while the coach provides detailed performance feedback data, video examples, and anecdotal notes. Following the reflective

and performance–feedback component of the meeting, the coaching cycle begins again with goal setting and the development of an action plan. See Sutherland et al. (2015) and Sutherland, Conroy, Algina, et al. (2018) for a more detailed description of the practice-based coaching component of BEST in CLASS-E.

**Coaches.** Four coaches (75% female; 75% Caucasian, 25% African American) were part of BEST in CLASS-E. Coaches ranged in age from 30 to 36 years. All coaches held at least a bachelor’s degree, with 50% holding a master’s degree or higher, and 75% currently enrolled in a graduate program at the time of coaching. All coaches reported at least 1 year of teaching experience and 75% reported holding a teaching certificate. In addition, 50% of coaches reported at least 1 year of practice-based coaching experience.

**BAU.** Students in the comparison group experienced a BAU condition only. The BAU condition consisted of daily instructional activities typically offered by elementary teachers in classrooms. For the most part, classroom activities began with a morning meeting, which was followed by a combination of small and large-group teacher-directed activities (e.g., early literacy, mathematics). Teachers in both BAU and BEST in CLASS-E conditions received the same professional development opportunities (other than BEST in CLASS-E teachers receiving BEST in CLASS-E training and coaching), and these trainings varied somewhat across schools. In general, teachers received trainings on topics such as trauma-informed approaches, instructional strategies, and general classroom management.

### Treatment Integrity Procedures

Observational measures were used in both BEST in CLASS-E and BAU comparison conditions to assess teacher implementation of BEST in CLASS-E practices. Teacher implementation of the BEST in CLASS-E practices was assessed on three dimensions (adherence, competence, and student responsiveness) using the TIES (Sutherland et al., 2017), which was adapted from the BEST in CLASS Adherence and Competence Scale (Sutherland et al., 2014). *Adherence* refers to the extent to which the program practices were delivered as intended, whereas *competence* refers to the level of skill and degree of responsiveness demonstrated by a teacher when delivering the practices (Carroll & Nuro, 2002). *Student responsiveness* refers to the degree to which the teachers’ attempts to engage students held the interest and participation of students (Dane & Schneider, 1998; Durlak & DuPre, 2008).

**Data analyses.** To test whether BEST in CLASS-E influenced change in student behavioral and academic outcomes, teacher–student relationships, and classroom

quality, condition (scored as 0 = BAU classrooms and 1 = BEST in CLASS-E classrooms) was regressed on posttest student outcomes, teacher–student relationships, and classroom quality scores, while controlling for pretest student, student–teacher, and classroom scores (i.e., pretest scores were regressed on posttest scores). Therefore, direct effects are interpreted as the extent to which BEST in CLASS-E predicted change in student behavior, student academic achievement, student–teacher relationships, and classroom quality at posttest.

Analyses were conducted using Stata, version 15.1 (StataCorp, 2017). For the models predicting Problem Behavior, Social Skills, Reading Scores, Math Score, Conflict, and Closeness, a multilevel approach using the “mixed” command was used, with students nested in teachers. These models were estimated using restricted maximum likelihood, and a Kenward–Roger degrees of freedom correction (Kenward & Roger, 1997, 2009) was applied to account for the small sample size. This approach is in line with recent recommendations for obtaining more precise estimates from multilevel models with a small number of clusters (McNeish & Stapleton, 2016).

The models predicting Classroom Organization, Instructional Support, and Emotional Support did not converge using this approach. Therefore, we used linear regression with cluster (teacher)-robust standard errors (Huber, 1967; White, 1980). Using this specification accounts for the nonindependence of students by adjusting the standard error estimates and, therefore, removes variance due to the repetition of children across teachers (Asparouhov & Muthen, 2006). Without accounting for the nonindependence in the data, the estimated standard errors would be inflated resulting in a greater chance of committing a Type I error.

### Results

Missing data patterns revealed that two teachers were missing data on posttest study variables (see Table 2). Independent samples *t* tests and a one-way analysis of variance (ANOVA) showed that teachers for whom complete data were not available did not differ from teachers with complete data on demographics (e.g., age, years of teaching experience, race/ethnicity, gender). In addition, Little’s (1988) missing completely at random test showed that data were missing completely at random,  $\chi^2(3) = 4.13, p = .25$ . Missing data patterns also revealed that one student was missing pretest WJ-III scores, four students were each missing posttest scores on the STRS and scores on the SSIS, and five students were each missing posttest WJ-III scores. Independent samples *t* tests and a one-way ANOVA showed that students for whom complete data were not available did not differ from students with complete data on demographics (e.g., age, race/ethnicity, gender). In addition, Little’s (1988) missing completely at random test showed that data were missing completely at random,  $\chi^2(20) =$



**Table 2.** Descriptive Statistics for Study.

Condition	Instrument/subscale	Tx (n)	BAU (n)	Tx M (SD)	BAU M (SD)	Tx Min–Max	BAU Min–Max
Pretest teacher level							
CLASS							
	Classroom Organization	14	12	3.91 (0.88)	3.93 (1.15)	2.83–5.25	1.5–5.42
	Instructional Support	14	12	1.95 (0.63)	1.82 (0.56)	1.17–3.08	1.08–2.75
	Emotional Support	14	12	4.62 (0.86)	4.24 (0.79)	3.19–6.06	2.44–5.13
Pretest child level							
STRS							
	Closeness	25	20	30.84 (5.3)	30.25 (5.5)	20–39	22–40–35
	Conflict	25	20	27.04 (6.2)	26.6 (6.2)	12–36	15–36
SSIS							
	Social Skills	25	20	71.76 (10.5)	77.7 (12.9)	49–88	56–99
	Problem Behavior	25	20	131.6 (13.5)	127.85 (14.5)	106–154	99–152
	Academic Competence	25	20	86.36 (18.1)	90.35 (11.54)	63–122	67–114
WJ-III							
	Reading	26	20	25.31 (13.1)	28.19 (14.1)	4–60	12–56
	Math	26	20	13.32 (4.35)	15.85 (6.2)	4–28	6–31
Posttest teacher level							
CLASS							
	Classroom Organization	13	11	4.35 (0.85)	4.02 (1.19)	3.00–5.75	1.75–5.5
	Instructional Support	13	11	1.7 (0.51)	1.73 (0.55)	1.00–2.5	1.08–2.75
	Emotional Support	13	11	4.8 (0.81)	4.4 (1.03)	3.44–6.75	2.31–5.63
Posttest child level							
STRS							
	Closeness	25	16	35.00 (3.5)	31.12 (4.8)	28–40	23–40
	Conflict	25	16	22.72 (6.4)	24.5 (7.30)	14–35	8–34
SSIS							
	Social Skills	25	16	85.32 (13.5)	86.06 (11.9)	55–122	61–109
	Problem Behavior	25	16	118.24 (15.6)	125.4 (19.9)	91–150	92–156
	Academic Competence	25	16	90.84 (17.82)	91.25 (13.85)	65–122	65–109
WJ-III							
	Reading	24	16	27.42 (12.7)	32.56 (14.1)	9–59	14–59
	Math	24	16	14.96 (4.3)	18.63 (7.4)	8–27	10–32

Note. Tx = Treatment; BAU = business as usual group; CLASS = Classroom Assessment Scoring System (Pianta et al., 2008); STRS = Student–Teacher Relationship Scale (Pianta, 1993); SSIS = Social Skills Improvement System–Rating Scale (Gresham & Elliott, 2008); WJ-III = Woodcock–Johnson–III Test of Achievement: Brief Battery (Woodcock et al., 2007).

15.10,  $p = .77$ . The full information maximum likelihood (FIML) estimator was used to account for these missing data. This estimator retains the statistical power of the full analytic sample while minimizing bias in parameter estimates when data cannot be presumed to be missing completely at random (Enders, 2001).

The first goal was to examine the extent to which BEST in CLASS-E influenced change in student behavior (i.e., teacher-reported student problem behavior and social skills on the SSIS-RS; Gresham & Elliott, 2008) compared with BAU. Controlling for pretest scores, results revealed compared with BAU, BEST in CLASS-E did not significantly influence student social skills ( $B = 4.82, p = .16, d = 0.21$ ; see Table 3). However, results revealed, compared with BAU, BEST in CLASS-E significantly changed student problem behavior ( $B = -11.83, p < .05, d = -0.32$ ).

Next, we examined the influence of BEST in CLASS-E on student academic achievement compared with BAU (using WJ-III; Woodcock et al., 2007). Controlling for pretest scores, compared with BAU classrooms, BEST in CLASS-E did not significantly influence student math and reading scores at posttest ( $B = -0.51$  and  $-0.01, p = .56$  and  $.99, d = -0.09$  and  $0.00$ , for math and reading scores, respectively).

The third goal was to examine the influence of BEST in CLASS-E on teacher–student relationships compared with BAU (using the STRS; Pianta, 1993). Controlling for pretest scores, BEST in CLASS-E did not significantly change teacher–student conflict ( $B = -1.86, p = .23, d = -0.18$ ) in comparison with BAU classrooms. However, BEST in CLASS-E did significantly change teacher–student closeness ( $B = 3.66, p < .001, d = 0.55$ ) in comparison with BAU classrooms.

**Table 3.** Pretest Scores and Treatment Condition as Predictors of Posttest Outcomes.

Scale	Pretest score		Condition	
	B	SE	B	SE
Problem Behavior	0.64***	0.15	-11.83*	5.48
Social Skills	0.45*	0.24	4.82	3.31
Reading Scores	0.93***	0.05	-0.01	1.47
Math Scores	0.97***	0.09	-0.51	1.04
Conflict	0.55***	0.14	-1.33	1.54
Closeness	0.26**	0.09	3.66***	1.0
Classroom Organization	0.37*	0.18	0.29	0.40
Instructional Support	0.12	0.17	-0.09	0.20
Emotional Support	0.62***	0.17	0.11	0.27

Note. Condition is scored as 0 = business as usual; 1 = BEST in CLASS-E. For the models predicting Problem Behavior, Social Skills, Reading Scores, Math Score, Conflict, and Closeness, a multilevel approach was used, with students nested in teachers. These models were estimated using restricted maximum likelihood, and a Kenward–Roger degrees of freedom correction was applied to account for the small sample size. The models predicting classroom organization, instructional support, and emotional support did not converge using this approach. Therefore, we used linear regression with cluster (teacher)-robust standard errors.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

Finally, we tested the extent to which BEST in CLASS-E changed the overall classroom climate compared with BAU (using CLASS; Pianta et al., 2008). Controlling for pretest scores, BEST in CLASS-E did not significantly change Classroom Organization ( $B = 0.29$ ,  $p = .47$ ,  $d = 0.11$ ), Instructional Support ( $B = -0.09$ ,  $p = .66$ ,  $d = -0.07$ ), or Emotional Support ( $B = 0.11$ ,  $p = .69$ ,  $d = 0.06$ ) compared with BAU classrooms.

### Treatment Integrity

Adherence, competence, and student responsiveness scores increased from pretest to posttest in BEST in CLASS-E classrooms, whereas decreases in these integrity dimensions were observed in BAU classrooms from pretest to posttest (see Table 4 for descriptive statistics). Effect size estimates were  $d = 0.67$  and  $d = 1.70$ , respectively, for adherence and competence.

### Discussion

The purpose of this study was to examine the effect of BEST in CLASS-E, a Tier 2 intervention delivered by teachers, on student problem behavior, academic achievement, teacher–student relationships, and classroom quality using a randomized controlled trial design. Results suggest that the intervention had a small to moderate effect on student behavior and teacher–student relationships; however, no effects were found for academic achievement or

classroom quality. Each of these findings will be discussed, followed by implications for future research and limitations of the current study.

In terms of student problem behavior, a small effect was noted on the Problem Behavior Scale of the SSIS. These findings replicate the effects of BEST in CLASS on preschool-aged children, which found effect sizes of  $-0.42$  for problem behavior (Sutherland, Conroy, Algina, et al., 2018). Results suggest the teacher training and practice-based coaching of BEST in CLASS-E is a promising approach for reducing early elementary-age students' problem behavior; it should be pointed out that TIES data suggest that teachers in the BEST in CLASS-E condition had increases from pretest to posttest in adherence and competence of delivery of the BEST in CLASS-E practices, in comparison with the BAU condition, which likely contributed to changes in students' behavior. This is important, as previous research on BEST in CLASS has shown that although adherence is important, teacher competence of delivery mediates treatment effects (Sutherland Conroy, McLeod, et al., 2018). Future work is needed to determine whether these mechanisms operate similarly for early elementary school teachers and students.

At the same time, no effects were found on the Reading or Math scales of the WJ-III. Thus, although BEST in CLASS-E appears to have a desirable effect on student behavior, this effect may not be enough to result in short-term academic gains, as the intervention only lasts approximately 14 weeks. Other prevention work (e.g., August et al., 2002; Kellam et al., 2014) has found more distal effects for outcomes not targeted by intervention but yet associated with those targeted outcomes. For example, Kellam et al. found effects of the Good Behavior Game, a universal classroom management program delivered by teachers in first grade, on risky sexual behavior and drug abuse in young adulthood. It may be that the proximal effects of BEST in CLASS-E on student behavior contribute to more distal improvements in academic achievement via reductions in classroom behavior problems and increased engagement; however, this is an empirical question that calls for longitudinal follow-up of students.

Furthermore, given longitudinal associations in the literature between early teacher–student relationships and interactions and students' academic achievement (e.g., Pakarinen et al., 2017; Spilt, Hughes, et al., 2012), potential effects of BEST in CLASS-E from the current study on teacher-reported closeness may also potentially contribute to distal student outcomes. The effect size in the current study was 0.55, compared with a smaller effect ( $d = 0.26$ ) noted in the BEST in CLASS preschool study (Sutherland, Conroy, Algina, et al., 2018). The elementary version of BEST in CLASS does include specific training and coaching on supportive relationships (see Sutherland et al., 2019), whereas the preschool version does not specifically target

**Table 4.** TIES Adherence and Competence Descriptives at Pre- and Posttest by Study Group.

Descriptive	BEST in CLASS-E			BAU		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Adherence						
Pretest	2.88	1.62	1.04–5.11	2.63	1.62	1.25–5.15
Posttest	3.18	1.66	1.50–5.54	2.14	1.44	1.06–4.56
Competence						
Pretest	3.79	1.06	2.00–4.67	3.82	0.61	3.00–4.50
Posttest	4.78	0.31	4.50–5.25	3.56	0.97	2.00–4.60
Student responsiveness						
Pretest	4.81			4.79		
Posttest	5.08			4.45		

Note. BAU = business as usual condition; BEST in CLASS-E = treatment condition; CLASS = Classroom Assessment Scoring System.

this as one of the instructional practices. However, given the small sample size in the current study, compared with the preschool study, the magnitude of this effect is surprising. On one hand, perhaps BEST in CLASS-E has a greater impact on the teacher-reported closeness with older children via improvements in teacher–student interactions during increasingly demanding academic interactions. However, we surmise that an additional explanation may be the overall level of risk presented in the classrooms as measured by the PPCAS measure (Abry et al., 2018; National Center for Education Statistics, 1993). Specifically, teachers completed this measure as an estimate of the overall level of classroom risk (e.g., toxic stress, learning problems) faced by students in their classrooms. Ratings on this measure ( $M = 3.00$  and  $3.01$ ) were higher than previous studies using this measure ( $M = 1.78$ ; Abry et al., 2018), which is not surprising given the environmental challenges present in the community where these schools were located (i.e., high poverty, violence exposure, food desert, limited transportation). It may be that given the overall needs of a large number of students in these classrooms that the teacher attempts at increasing positive interactions and being supported to do so by the BEST in CLASS-E training and coaching led to teachers feeling closer to focal students.

At the same time, there were no significant effects noted for teacher-reported conflict or any of the dimensions of the CLASS measures, which assesses overall classroom quality. Previous research on the preschool version of BEST in CLASS (Conroy, Sutherland et al., 2019) found small effects for reductions in teacher-reported conflict ( $d = -0.29$ , compared with  $d = -0.18$  in the current study) and small to moderate effects for the three dimensions of the CLASS. It is possible the small sample size may have contributed to an increased likelihood of Type II error due to low statistical power. Alternatively, BEST in CLASS-E may not have been strong enough to overcome the existing risk as assessed by the PPCAS. Thus, although teachers reported feeling closer

to the focal students as a result of the training and coaching support provided by BEST in CLASS-E, these supports did not result in their feeling significantly less conflict toward the focal students nor were these supports sufficient to overcome high rates of classroom adversity.

### Study Limitations and Implications

Several limitations of this study should be considered when interpreting results. First, problem behavior and social skills in the current study were assessed via the SSIS, which is a teacher-report measure. Future work should also include direct observations of student behavior to strengthen study findings. Furthermore, this study, which was an underpowered promise study of an Institute of Education Sciences development project, utilized a small sample of teachers and students from a geographically unique setting; therefore, generalizing these findings to other students, teachers, or communities should be done cautiously. The small sample may limit our ability to detect group differences that may in fact be present. Future research should use samples large enough to detect group differences across diverse settings to advance our understanding of BEST in CLASS-E effects on elementary students with problem behavior who are at risk of EBD and the classrooms in which they are taught. Relatedly, the data from the PPCAS suggest that these classrooms served students who had experienced a number of adverse events. Thus, results from this study may not generalize to classrooms or schools where families and students experience fewer adverse events. Also, the lack of Tier 1 interventions in place in the schools in the current study limit our ability to assess the fit of BEST in CLASS-E within a PBIS framework delivered with high fidelity, and future work should attempt to implement BEST in CLASS-E within PBIS systems delivered with high fidelity. In addition, although the treatment integrity data from the current study replicate findings from the BEST in CLASS

preschool study, the small sample size in the current study precludes examining adherence and competence as potential mediators of treatment effects. Future research with larger samples can help examine the relationship between outcomes and treatment integrity. In addition, the low ICC noted for the Instructional Support domain of the CLASS limits the interpretability of this measure. Finally, although there were no effects found for academic outcomes in the current sample, longitudinal research is necessary to help examine potential distal effects of interventions such as BEST in CLASS-E on downstream outcomes (e.g., Kellam et al., 2014).

## Conclusion

BEST in CLASS-E appears to have promise for improving behavioral outcomes and teacher–student relationships for young elementary students with and at risk of EBD. These findings replicate those found for BEST in CLASS-E in early childhood classrooms and highlight the promise of the professional development approach—evidence-based practice elements supplemented with practice-based coaching—used in BEST in CLASS-E. At the same time, more work is needed to better understand potential long-term impacts on both teacher and student behavior, not to mention whether there are distal effects on academic achievement. Better understanding these downstream effects can help inform future work on Tier 2 interventions such as BEST in CLASS-E to maximize their impact on outcomes for this group of vulnerable students.


## 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 grants from the U.S. Department of Education’s Institute for Education Sciences (R305A150246) and Office of Special Education Programs (325H140001). 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

K. S. Sutherland,  <https://orcid.org/0000-0002-5314-090X>

M. Broda,  <https://orcid.org/0000-0002-0019-5528>

## References

- Abry, T., Granger, K. L., Bryce, C. I., Taylor, M., Swanson, J., & Bradley, R. H. (2018). First grade classroom-level adversity: Associations with teaching practices, academic skills, and executive functioning. *School Psychology Quarterly, 33*(4), 547–560. <https://doi.org/10.1037/spq0000235>
- Anderson, C. M., & Borgmeier, C. (2010). Tier II interventions within the framework of school-wide positive behavior support: Essential features for design, implementation, and maintenance. *Behavior Analysis in Practice, 3*(1), 33–45. <https://doi.org/10.1007/BF03391756>
- Asparouhov, T., & Muthen, B. (2006, August). Multilevel modeling of complex survey data. In T. Asparouhov & B. Muthen (Eds.), *Proceedings of the survey research methods section* (pp. 2718–2726). American Statistical Association.
- August, G. J., Hektner, J. M., Egan, E. A., Realmuto, G. M., & Bloomquist, M. L. (2002). The early risers longitudinal prevention trial: Examination of 3-year outcomes in aggressive children with intent-to-treat and as-intended analyses. *Psychology of Addictive Behaviors, 16*(Suppl. 4), S27–S39. <https://doi.org/10.1037/0893-164x.16.4s.s27>
- Brock, L. L., Nishida, T. K., Chiong, C., Grimm, K. J., & Rimm-Kaufman, S. E. (2008). Children’s perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the Responsive Classroom approach. *Journal of School Psychology, 46*(2), 129–149. <https://doi.org/10.1016/j.jsp.2007.02.004>
- Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge, MA: Harvard University Press.
- Bruhn, A. L., Lane, K. L., & Hirsh, S. E. (2014). A review of Tier 2 interventions conducted within multitiered models of behavioral prevention. *Journal of Emotional and Behavioral Disorders, 22*, 171–189. <https://doi.org/10.1177/1063426613476092>
- Carroll, K. M., & Nuro, K. F. (2002). One size cannot fit all: A stage model for psychotherapy manual development. *Clinical Psychology: Science and Practice, 9*(4), 396–406. <https://doi.org/10.1093/clipsy.9.4.396>
- 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. <https://doi.org/10.1037/1040-3590.6.4.284>
- Conroy, M. A., McKnight, K., & Sutherland, K. S. (2019). Partnering with families of students at risk for emotional and behavioral disorders. In L. Lo & Y. Xu (Eds.), *Family, School and Community Partnerships for Individuals with Disabilities*. (pp. 57–69) New York: Springer.
- Conroy, M. A., Sutherland, K. S., Algina, J., Ladwig, C., Werch, B., Martinez, J., . . . 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*, 31–45. <https://doi.org/10.17105/SPR-2018-0003.V48-1>
- Conroy, M. A., Sutherland, K. S., Algina, J., Werch, B., & Ladwig, C. (2018). Prevention and treatment of externalizing problem behaviors in young children: Clinical implications from a randomized controlled trial of BEST in CLASS. *AERA Open, 4*(1), 1–16. <https://doi.org/10.1177/2332858417750376>
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23–45. [https://doi.org/10.1016/S0272-7358\(97\)00043-3](https://doi.org/10.1016/S0272-7358(97)00043-3)
- Doumen, S., Verschueren, K., Buyse, E., Germeijs, V., Luyckx, K., & Soenens, B. (2008). Reciprocal relations between teacher–child conflict and aggressive behavior in kindergarten: A three-wave longitudinal study. *Journal of Clinical*



- Child & Adolescent Psychology*, 37(3), 588–599. <https://doi.org/10.1080/15374410802148079>
- 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. <https://doi.org/10.1007/s10464-008-9165-0>
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6(4), 352–370.
- Farmer, T. W., Sutherland, K. S., Talbott, E., Brooks, D. S., Norwalk, K., & Huneke, M. (2016). Special educators as intervention specialists: Dynamic systems and the complexity of intensifying intervention for students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 24, 173–186. <https://doi.org/10.1177/1063426616650166>
- Forness, S. R., Freeman, S. F., 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>
- Forness, S. R., Kim, J., & Walker, H. M. (2012). Prevalence of students with EBD: Impact on general education. *Beyond Behavior*, 21(2), 3–11.
- Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales manual*. Pearson Assessments.
- Gunter, P. L., & Coutinho, M. J. (1997). Negative reinforcement in classrooms: What we're beginning to learn. *Teacher Education and Special Education*, 20, 249–264.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221–233.
- Jennings, P. A., Brown, J. L., Frank, J. L., Doyle, S., Oh, Y., Davis, R., . . . Greenberg, M. T. (2017). Impacts of the CARE for teachers program on teachers' social and emotional competence and classroom interactions. *Journal of Educational Psychology*, 109, 1010–1028. <http://doi.org/10.1037/edu0000187>
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, 10(2), 165–185.
- Kellam, S. G., Wang, W., Mackenzie, A. C. L., Brown, C. H., Ompad, D. C., Or, F., . . . Windham, A. (2014). The impact of the good behavior game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science*, 15(Suppl. 1), S6–S18. <https://doi.org/10.1007/s11121-012-0296-z>
- Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983–997.
- Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, 53, 2583–2595.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202.
- McClowry, S. G., Rodriguez, E. T., Tamis-LeMonda, C. S., Spellmann, M. E., Carlson, A., & Snow, D. L. (2013). Teacher/student interactions and classroom behavior: The role of student temperament and gender. *Journal of Research in Childhood Education*, 27(3), 283–301. <https://doi.org/10.1080/02568543.2013.796330>
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28, 295–314.
- Myers, S. S., & Pianta, R. C. (2008). Developmental commentary: Individual and contextual influences on student–teacher relationships and children's early problem behaviors. *Journal of Clinical Child & Adolescent Psychology*, 37(3), 600–608. <https://doi.org/10.1080/15374410802148160>
- National Center for Education Statistics. (1993). *The schools and staffing survey: 1993–1994 school year*. U.S. Department of Education.
- O'Connor, E. E., Dearing, E., & Collins, B. A. (2011). Teacher–child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal*, 48, 120–162. <https://doi.org/10.3102/0002831210365008>
- Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Salminen, J., Silinskas, G., Siekkinen, M., & Nurmi, J. E. (2017). Longitudinal associations between teacher–child interactions and academic skills in elementary school. *Journal of Applied Developmental Psychology*, 52, 191–202. <https://doi.org/10.1016/j.appdev.2017.08.002>
- Patterson, G. R. (2002). Etiology and treatment of child and adolescent antisocial behavior. *The Behavior Analyst Today*, 3(2), 133–144.
- Pianta, R. C. (1993). *The student–teacher relationship scale*. University of Virginia.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system™: Manual K-3*. Paul H Brookes.
- Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, 26(1), 1–13. <https://doi.org/10.1037/a0022714>
- 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. (2009). The transactional model. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 3–21). American Psychological Association. <https://doi.org/10.1037/11877-001>
- Sandilos, L. E., & DiPerna, J. C., & The Family Life Project Key Investigators. (2014). Measuring quality in kindergarten classrooms: Structural analysis of the classroom assessment scoring system (CLASS K–3). *Early Education and Development*, 25, 894–914.
- Spilt, J. L., Hughes, J. N., Wu, J. Y., & Kwok, O. M. (2012). Dynamics of teacher–student relationships: Stability and change

- across elementary school and the influence on children's academic success. *Child Development*, 83(4), 1180–1195.
- Spilt, J. L., Koomen, H. M., Thijs, J. T., & van der Leij, A. (2012). Supporting teachers' relationships with disruptive children: The potential of relationship-focused reflection. *Attachment & Human Development*, 14(3), 305–318. <https://doi.org/10.1080/14616734.2012.672286>
- StataCorp LLC. (2017). *Stata (version 15.1)*.
- Sugai, G., & Horner, R. H. (2009). Defining and describing school-wide positive behavior support. In W. Sailor, G. Dunlap, G. Sugai, R. Horner, W. Sailor, G. Dunlap, . . . R. Horner (Eds.), *Handbook of positive behavior support* (pp. 307–326). Springer. [https://doi.org/10.1007/978-0-387-09632-2\\_13](https://doi.org/10.1007/978-0-387-09632-2_13)
- Sutherland, K. S., Conroy, M. A., Algina, J., Ladwig, C., Jessee, 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. <https://doi.org/10.1016/j.ecresq.2017.08.001>
- Sutherland, K. S., Conroy, M. A., McLeod, B. D., Algina, J., & Wu, E. (2018). Teacher competence of delivery of BEST in CLASS as a mediator of treatment effects. *School Mental Health*, 10, 214–225. <https://doi.org/10.1007/s12310-017-9224-5>
- Sutherland, K. S., Conroy, M. A., McLeod, B. D., Granger, K., Nemer, S. L., Kunemund, R. L., . . . Miles, C. (2019). Adapting an evidence-based early childhood tier 2 program for early elementary school. *Elementary School Journal*, 119, 542–561. <https://doi.org/10.1086/703103>
- Sutherland, K. S., Conroy, M. A., Vo, A., & Ladwig, C. (2015). Implementation integrity of practice-based coaching: Preliminary results from the BEST in CLASS Efficacy Trial. *School Mental Health*, 7, 21–33. <https://doi.org/10.1007/s12310-014-9134-8>
- Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *Journal of Special Education*, 41, 223–233. <https://doi.org/10.1177/0022466907310372>
- 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, 249–259. <https://doi.org/10.1177/1063426613497258>
- Sutherland, K. S., McLeod, B. D., & Kunemund, R. L. (2017). Scoring manual for the observational treatment integrity instrument for elementary school classrooms [Unpublished manuscript]. Virginia Commonwealth University.
- Walker, H. M., Severson, H. H., & Feil, G. E. (2014). *Systematic screening for behavior disorders (SSBD): Administrator's guide*. Pacific Northwest Publishing.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48, 817–838.
- Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2007). *Woodcock-Johnson III test of achievement: Form C brief battery*. Riverside.