Hammill Institute on Disabilities

Assessment for Effective Intervention
39(1) 12–23

DOI: 10.1177/1534508412463814
aei.sagepub.com
SAGE

An Initial Evaluation of the Teacher–Child Interaction Direct Observation System: Measuring Teacher–Child Interaction Behaviors in Classroom Settings

Kevin S. Sutherland, PhD[1], Maureen A. Conroy, PhD[2], Abigail Vo, PhD[1], Lisa Abrams, PhD[1], and Paula Ogston, MA[1]

Abstract

Children who engage in chronic problem behavior can present challenges for their early childhood teachers, and are more likely to develop negative relationships with these and other adults with whom they interact. Unfortunately, children who develop negative relationships with their teachers early in their school careers are more likely to have later academic and behavioral problems than children who have positive, supportive relationships. The salience of interactions between teachers and young children highlights the need for interventions that target building positive teacher–child interactions and increase the likelihood that these interactions are developmentally appropriate and supportive of children's emotional and behavioral growth, as well as assessment tools to evaluate the effectiveness of these interventions. The purpose of this article is to describe the Teacher–Child Interaction Direct Observation System (*TCIDOS*), an observation system designed to capture teacher and child behaviors during classroom interactions to evaluate the effectiveness of a classroom-based intervention targeting improvements in teacher–child interaction patterns. Following a description of our conceptual framework and the development of the *TCIDOS*, we discuss preliminary reliability findings and future research directions as well as challenges inherent to collecting direct observational data in classroom settings.

Keywords

classroom observation, teacher behavior, child behavior

The prevalence of young children at elevated risk for emotional and behavioral disorders (EBD) is a national health concern, with data indicating that approximately 12% to 25% of young children display chronic problem behaviors that affect their current and future performance in school (Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Hamre & Pianta, 2001; Loeber & Farrington, 2000; Webster-Stratton, 1997) and positive trajectory into adulthood (Dishion, French, & Patterson, 1995; Patterson, Reid, & Eddy, 2002). Although the development of EBDs can progress from these early chronic behavior problems, research also indicates that EBDs develop cumulatively and interactively with a number of other factors (Myers & Pianta, 2008). Children who are exposed to a high number and combination of risk factors, such as poverty and coercive or poor parenting practices, are more likely to demonstrate chronic behavior problems, eventually be identified with EBD, and be found eligible for special education services to address these deficits (Dunlap et al., 2006; Webster-Stratton, 1997). Although the trajectory for many of these children is troubling to say the least, Farmer, Quinn, Hussey, and Holahan's (2001) concept of correlated constraints suggests that multiple factors can also work together as a protective system and contribute to the development of prosocial and adaptive behaviors. From this perspective, a combination of protective factors tends to promote stability in each other as well as the manifestation and expression of behavior (Magnusson & Cairns, 1996). For example, one factor that might play a significant role in a protective system is an early childhood

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

Corresponding Author:
Kevin Sutherland, Department of Special Education and Disability Policy, Virginia Commonwealth University, 1015 W. Main St., Richmond, VA 23284, USA.
Email: kssuther@vcu.edu

classroom that has a positive atmosphere and high-quality interactions between teachers and children.

As more young children are attending early childhood programs (Barnett, Carolan, Fitzgerald, & Squires, 2011), one key target for intervention is coercive interactions that may develop between children at risk of EBD and their early childhood teachers. Coercive interactions occur when one member (e.g., child) in an interaction sequence demonstrates behavior that the other member (e.g., teacher) of the interaction finds aversive, resulting in responses that seek to terminate or avoid the aversive behavior (Patterson & Reid, 1970). These interaction patterns become particularly toxic when a teacher engages in behaviors that are aversive to the child to escape/avoid the child's aversive behavior (Gunter et al., 1994). Recent research indicates that interactions between teachers and children in early childhood classrooms have emerged as an important factor and are associated with later school adjustment and social, emotional, behavioral, and developmental outcomes (e.g., Hamre & Pianta, 2001; Henricsson & Rydell, 2004; Ladd & Burgess, 1999). Furthermore, children who engage in chronic problem behavior are more likely to develop negative relationships with their teachers (Ladd & Burgess, 1999). Unfortunately, children who develop such relationships with their teachers early in their school careers are more likely to have later academic and behavioral problems than students who have positive, supportive relationships with their teachers early in their school careers (Hamre & Pianta, 2001). In a recent study, Doumen and colleagues (2008) found that young children's aggressive behavior at the beginning of kindergarten was associated with increases in teacher–child conflict midyear, which subsequently was associated with increases in aggressive behavior at the end of the school year. They noted that "when children enter kindergarten, it seems to be particularly the child aggressive behavior that starts the accumulation of negative processes throughout the year" (Doumen et al., 2008, p. 596).

The development of coercive interaction patterns between children and their teachers is particularly troubling as the nature of relationships between teachers and young children tends to be stable over time (Henricsson & Rydell, 2004). In addition, research suggests that these interactions are bidirectional and transactional (Doumen et al., 2008; Sutherland & Oswald, 2005); thus, the interactions between teachers and young children with chronic problem behavior can ultimately be viewed as either a risk or protective factor. There is a need for interventions that focus on building positive teacher–child interactions to increase the likelihood that these interactions are developmentally appropriate and supportive of children's emotional and behavioral growth. To address this need, we developed a classroom-based intervention (*BEST in CLASS*; see Sutherland, Conroy, Abrams, & Vo, 2010; Vo, Sutherland, & Conroy, 2012) that, through teacher training and performance-based coaching, increases positive reciprocal teacher–child interaction behaviors. As there are few observational tools that directly measure behavioral interactions between teachers and their students, we designed the Teacher–Child Interaction Direct Observation System (*TCIDOS*) to facilitate evaluation of *BEST in CLASS*. To provide a context for this measure, we first describe the conceptual framework guiding *BEST in CLASS* as well as existing behavioral measurement tools, including limitations in these measures that led to the development of the *TCIDOS*.

BEST in CLASS

The conceptual framework guiding the *BEST in CLASS* intervention is based on three distinct, but compatible, theoretical approaches: bioecological (Bronfenbrenner, 2005), transactional (Sameroff, 2009), and behavioral (Skinner, 1953). Each theory provides a component of the *BEST in CLASS* conceptual framework, which represents the complex relationships among child, teacher, and context within early childhood classrooms serving young children at risk of EBD (see Figure 1). In our initial investigation of the *BEST in CLASS* intervention, we used the *Classroom Assessment Scoring System* (*CLASS*; Pianta, La Paro & Hamre, 2008) to measure domains at the bioecological level (i.e., classroom); the *TCIDOS* was developed from the two more discrete theories supporting our conceptual framework to measure specific teacher and child behaviors, and the interactions between the two, within the classroom ecology. As such, we next focus on describing the transactional and behavioral theories that supported the development of the *TCIDOS*.

Transactional theory. At the level of the individual teacher and child, the transactional approach (Sameroff, 2009) provides a means for clearly conceptualizing the mutually influential relationship between individuals and the ecologies in which they are embedded. This distinct developmental systems model emphasizes the bidirectional nature of the influence between individuals and their environment. It differs from other theories in the emphasis it places on the potential impact of the individual on his or her environment. For example, due to his unique biopsychosocial characteristics, Child A may elicit different responses from his environment (including other individuals) than Child B. In turn, aspects of the environment may differentially affect the behavior and development of Child A and Child B (Sameroff, 2009).

For children with a history of chronic problem behavior, preexisting negative and coercive patterns of interaction with caregivers are likely to carryover into the classroom setting and result in coercive patterns of interactions with their teachers (Gunter & Coutinho, 1997; Wahler & Dumas, 1986). Over time, children experiencing

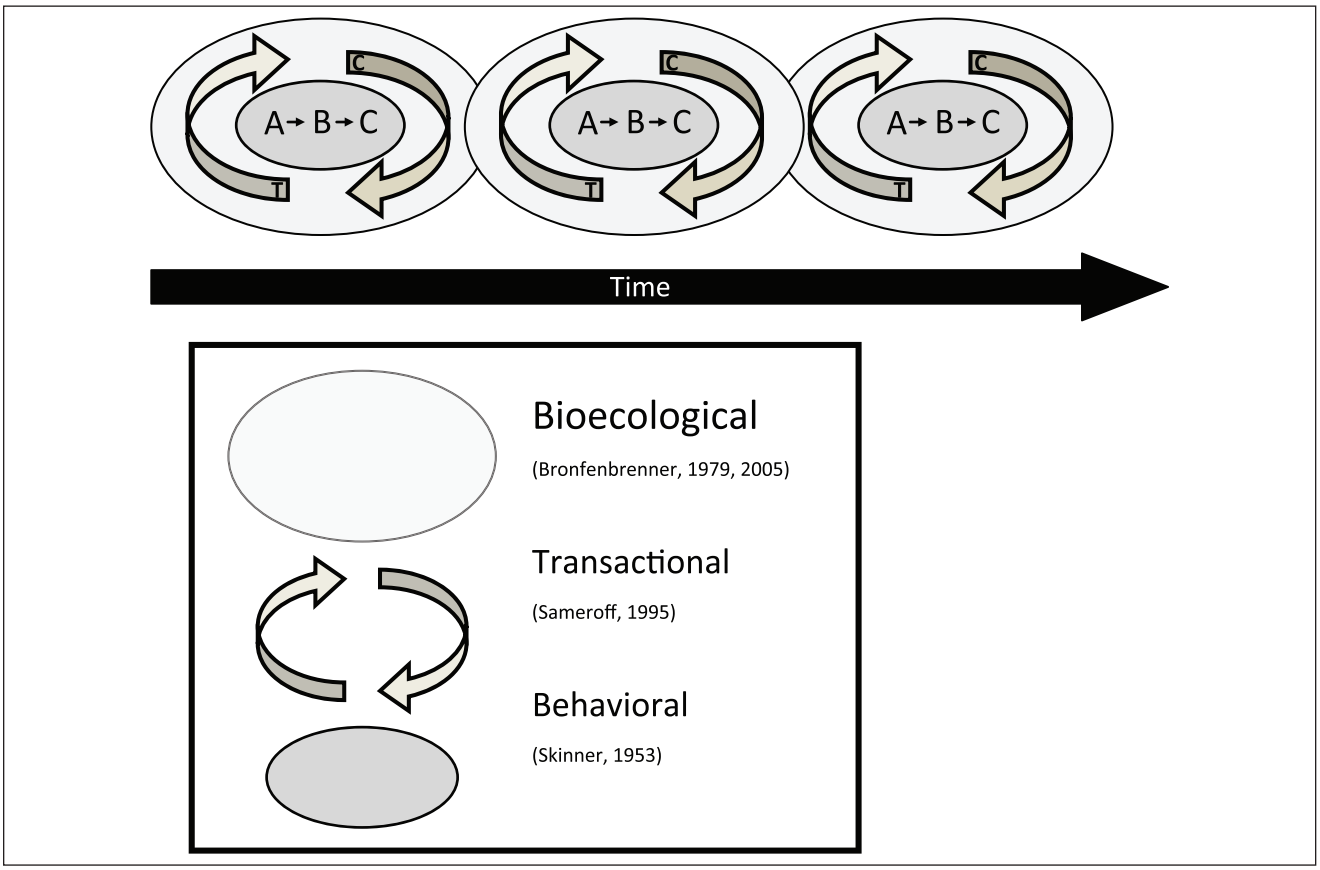

Figure 1. *BEST in CLASS* Theoretical Framework.

coercive interactions with teachers tend to receive fewer instructional opportunities and positive social interactions than their peers, which may contribute to the documented long-term adverse effects for children who exhibit chronic problem behavior at early ages (Doumen et al., 2008). *BEST in CLASS* instructs early childhood teachers to implement effective instructional practices aimed at preventing and reducing problem behavior in children who demonstrate chronic problem behaviors. At the level of the teacher–child relationship, the use of such strategies has the potential to prevent or interrupt coercive interaction patterns. We hypothesize that such changes can (a) improve the quality of teacher–child interactions; (b) increase children's proactive and adaptive behaviors, which in turn build children's capacity to function adaptively in school settings; and (c) ultimately, reduce their risk of developing EBD and associated detrimental outcomes.

Behavioral theory. Behavioral theory (Skinner, 1953), the most discrete and prescriptive of the frameworks comprising our theoretical model, aids in the conceptualization of how specific contingencies, in the form of teacher and child responses, shape behavior over time. Behavioral theory provides the basis for the specific intervention strategies included in the *BEST in CLASS* intervention model as well as the discrete teacher and child behaviors measured by the *TCIDOS*. To illustrate, teacher–child interactions include observable behavioral events that can be measured and analyzed. Accumulated evidence indicates the importance of antecedent instructional stimuli as well as reinforcement of responses as strategies for increasing desirable academic and social behavioral responses (e.g., children's correct responses and compliance; Kern & Clemens, 2007; Lewis, Hudson, Richter, & Johnson, 2004). As a result, increases in the use of these strategies can promote positive teacher–child interactions, enhance child engagement, increase learning, and decrease the occurrence of problem behaviors (e.g., Sutherland & Wehby, 2001; Werts, Wolery, & Holcombe, 1991). In turn, increases in the use of effective instructional practices can lead to increased desirable behavior by children at high risk for EBD and their peers. For example, a teacher asks a child a question (i.e., provides an opportunity to respond [OTR]), the child answers the question (e.g., correct or incorrect response), and the teacher provides feedback

(e.g., corrective feedback, instructive feedback, or behavior specific praise [BSP]). This sequence represents a three-event behavioral sequence (antecedent, behavior, consequence).

Central to behavioral theory is the understanding that behavior is functionally related to the environment in which it occurs (Baer, Wolf, & Risley, 1968). Thus, as variables in the environment are modified, the likelihood of occurrence of a target behavior is affected. As such, teachers can present instructional stimuli (antecedents) or consequent events (consequences) that can increase the likelihood that children will exhibit desirable behavior or decrease the likelihood that they will exhibit challenging behavior. These particular antecedent and consequent events, particularly as they relate to desirable student outcomes, fall into the framework of "evidence-based kernels" (Embry, 2004). Embry (2004) described an evidence-based kernel as "an irreducible unit of behavior-change technology that produces an observable, reliable result" (p. 578). He pointed out that many programs identified as evidence-based "best practices" in prevention are comprised of behavioral kernels, such as reinforcement procedures. Moreover, Embry noted that evidence-based kernels can also be combined to produce positive results and he refers to these combinations of evidence-based kernels as "behavioral vaccines."

BEST in CLASS has been designed to encompass a series of behavioral kernels, which we believe will result in a behavioral vaccine. To illustrate, we focus our intervention on those "behavioral kernels" that comprise teacher–child interactions, while recognizing the transactional nature of social interchanges and how affecting behavior and transactions (e.g., improving teacher–student interactions) can influence the broader ecology, in this case the classroom environment (see Figure 1). In this figure, the child and teacher influence each other reciprocally across time points via interactions (i.e., three-term behavioral contingencies); these interactions in turn influence and are influenced by the environment (i.e., classroom). The increasing boldness of the horizontal arrow represents the accumulating influence of the subsystems on subsequent transactions. This model serves as the foundation for the key instructional practices that comprise the model intervention, and ultimately the observable behaviors (i.e., teacher–child interactions) that are coded in *TCIDOS*.

Measurement of Behavior and Behavioral Interactions

In general, measurement issues have long plagued the field of EBD and the measurement of behavioral interactions is no exception (for a discussion, see Conroy, Stichter, Daunic, & Haydon, 2008). Currently, many of the existing tools used by researchers to examine student outcomes following an intervention only minimally address interactions between teachers and their children (e.g., *Caregiver-Teacher Report Form*; Achenbach & Rescorla, 2000; *Social Skills Improvement System*; Gresham & Elliott, 2008). In addition, the majority of the tools use indirect assessment methods (e.g., caregiver or teacher report), which are prone to bias. Direct observation methods hold several advantages over indirect reports of behavior that are often used (Conroy et al., 2008). For example, Conroy et al. (2008) point out that the use of standardized measures completed by an observer (i.e., teacher) may not provide the precision of measurement necessary to determine intervention effectiveness because the observer (in this case, the teacher) is often a part of the behavioral interaction and/or the intervention. Direct observation allows for a detailed description of behavior and the context in which it occurs, which is likely to reduce the bias that may occur from a retrospective report and increase the objectivity and accuracy of the assessment (Yoder & Symons, 2010). In addition, report measures may be subject to observer bias, in that the observer (i.e., teacher) is not independent of the context in which data are being collected (Fox & Conroy, 1995). Conroy and colleagues call for the use of multiple informant measurement (e.g., direct observation as well as more indirect measures such as teacher reports) in examining relations between teacher and child behavior in classroom-based research.

In alignment with this recommendation, several early intervention programs delivered by teachers have demonstrated efficacy in randomized controlled trials (RCTs) using outcomes measured by teacher and parent reports as well as direct observational measures (e.g., First Steps to Success, Walker et al., 2009; Second Step, Hussey & Flannery, 2007; Preschool PATHS, Domitrovich, Cortes, & Greenberg, 2007; and Incredible Years, Webster-Stratton, Reid, & Stoolmiller, 2008). Unfortunately the observational measurement systems used in these research studies have several limitations. For example, direct observation systems used in most of these studies have (a) included only a narrow scope of behaviors, (b) only a limited amount of time in which direct observational measurement occurred, and/or (c) have not included measurement across repeated time points or contexts. For example, Webster-Stratton and colleagues (2008) only conducted two 30-min observations, which included three teacher behaviors and six child behaviors when investigating the efficacy of the Incredible Years intervention. During an efficacy trial of First Steps to Success (Walker et al., 2009), researchers also only observed children's academic engagement at two time points.

Descriptive research on teacher–child interactions also has limitations. To illustrate, Henricsson and Rydell (2004) collected observational data on teacher–child interactions for 95 children during 10 to 20 five-min sessions using an observational system comprised of 31 child codes and 29 teacher codes. However, reliability data were collected on only 8% of the total observations and no reliability data

were provided for individual codes. Other studies (e.g., Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008; Doumen et al., 2008) have used teacher reports to examine the association between conflictual teacher–child relationships and key child outcomes.

The lack of ongoing observational data of teacher and child behavior across time and contexts is a significant limitation of the current research (experimental and descriptive) and may limit our ability to determine the efficacy of interventions delivered by teachers, particularly those that focus on improving ongoing interactions between teachers and children. These data would (a) enhance the interpretation of data from standardized measures of child outcomes, (b) increase the ability to determine the efficacy of interventions in RCTs, and (c) advance science in descriptive and prevention research.

We have used gaps in the existing measurement systems as well as the conceptual framework of *BEST in CLASS* to develop the *TCIDOS*. The *TCIDOS* allows us to examine (a) the occurrence of teacher and child behaviors and interactions prior to, during, and following intervention; (b) the relationship between these behaviors and child outcomes; and (c) the effect of our intervention model on teacher and child behaviors and teacher–child interactions.

Thus, a benefit of the *TCIDOS* is the ability to examine behaviors as they occur in real time across a variety of contexts and time points. This characteristic makes the *TCIDOS* particularly relevant for the study of interactions and also provides a mechanism for measuring teacher and child behaviors of interest (Bakeman, 2000). Evaluations of *BEST in CLASS* include two teacher-report measures related to child behavior (*Caregiver-Teacher Report Form*; Achenbach & Rescorla, 2000; *Social Skills Improvement System*; Gresham & Elliott, 2008) to measure child outcomes of interest. However, ultimately we are most interested in data from the *TCIDOS*, as these data provide for greater accuracy regarding the change that may occur in teachers' implementation of specific practices, children's responses to those practices, and the influence of these teacher–child behaviors and interactions following the implementation of the intervention.

The Development of the *TCIDOS*

The *TCIDOS* was created to enable the measurement of the specific teacher and child behaviors of interest related to the *BEST in CLASS* intervention within the context of instructional activities occurring in early childhood classrooms. The *TCIDOS* allows for data collection on seven effective instructional teaching practices (rules, BSP, precorrection, active supervision, OTR, teacher feedback) and five other behavioral codes of empirical and theoretical importance that measure other teacher and child behaviors (see Table 1).

The process of designing the *TCIDOS* began with a comprehensive review of the literature and identifying codes that corresponded with (a) our theoretical framework and (b) individual variables at each level of this framework. Thus, not only were we interested in measuring specific teacher behaviors representing "kernels" from *BEST in CLASS* (e.g., BSP, precorrection), but we were also interested in measuring child behaviors hypothesized to be affected by these teacher behaviors (e.g., engagement, disruptions, defiance, etc.). The process next involved a careful selection and refinement of *codes* (i.e., behavioral descriptions) that are specifically and operationally defined to reduce the potential variability that might occur as observers determine whether a particular behavior falls under a specific code (Thompson, Symons, & Felce, 2000). Throughout the process of training individuals to use this system, codes were modified to better fit the operationalization of the construct represented by each code's definition. For example, after code definitions were initially created and memorized by research assistants, 1-min video clips of child and teacher behaviors were viewed by the research team. Exemplars of codes were identified while clarification and refinement was provided when questions arose. Over time, the research team continued to refine code definitions when disagreements occurred. Eventually final definitions and exemplars (including examples and non-examples) of codes were agreed upon.

Following the creation of code definitions, we increased the validity of the data produced by the system by ensuring the system measured the intended constructs. Yoder and Symons (2010) described the validation of observational systems as "a purpose-specific, ongoing and cumulative process" (p. 184); as such, validation procedures of the *TCIDOS* are ongoing. Initial purposes of validation efforts were focused on establishing content validity. As described above, the purpose of the observation system was to capture teacher and child behaviors associated with the implementation of the *BEST in CLASS* intervention components. The codes comprising the system and the operational definitions are well aligned with the instructional practices and the related coaching teachers received. An external review of the observational system coding manual was conducted by two experts in the field of observational measurement systems and early childhood problem behavior. These reviewers provided clarification on item definitions as well as increasing the library of examples and non-examples for each code.

We next describe a preliminary investigation designed to examine the reliability of the *TCIDOS*. This study was part of a larger, nonexperimental study designed to obtain preliminary data on the promise of the *BEST in CLASS* intervention. Following a description of the coding procedures and behavioral codes, we discuss preliminary reliability of the *TCIDOS*. We then discuss limitations of the

Table 1. Direct Observation Codes and Description.

Code and Abbreviation	Definition	Description	Theory
Positive/neutral interaction (+)	Exchange in which teacher and child are exhibiting positive/neutral behavior/affect	MTS	Transactional
Negative interaction (−)	Exchange in which teacher or child are exhibiting negative behavior and/or affect	MTS	Transactional
Rule (R)	Statements that contain the word "rule"	PIR	Behavioral
Precorrection (PC)	Reminders of expectations prior to entering a situation	PIR	Behavioral
Opportunity to respond (OTR)	Questions, requests, commands, or gestures that seek a response	PIR	Behavioral
Behavior-specific praise (BSP)	Statements that indicate approval and specify the behavior being praised	PIR	Behavioral
Instructive feedback (IF)	Acknowledgement of an appropriate behavior and provision of instructional information	PIR	Behavioral
Corrective feedback (CF)	Acknowledgement of an error and provision of additional instructional information	PIR	Behavioral
Reprimand (RP)	Expressed disapproval of a behavior without provision of additional information	PIR	Behavioral
Disruption, aggression, defiance (DAD)	Verbalizations, acts, or gestures that interrupt/have the potential to interrupt instruction	PIR	Behavioral
Active supervision (AS)	Teacher is actively engaged with the child	MTS	Behavioral
Engagement (EN)	Child is participating appropriately/working on assigned/approved activity	MTS	Behavioral

Note. MTS = momentary time sampling code; PIR = partial-interval code.

TCIDOS, as well as implications for research and practice associated with collecting direct observational data in classroom settings.[1]

Method

Participants and Setting

Nine early childhood teachers and 18 focal children at risk of EBD (2 children per classroom) from two school districts and 1 teacher and 1 child from a university-based early childhood program participated in the study, for a total of 10 teachers and 19 focal children. This study took place in the southeast region of the United States. All teacher participants were female; 5 were White, 4 were African American, and 1 was Latina. Eight of the teachers held a master's degree, and the teachers' ages varied, with 2 between 18 and 25 years old, 2 between 26 and 35 years old, 1 between 36 and 45 years old, 3 between 46 and 55 years old, and 2 above 55 years of age. The teachers' years of experience teaching preschool-age children also varied, with a mean of 10.1 years (range = 3–34 years).

After informed consent was obtained for teacher participants, teachers nominated five children who demonstrated problem behavior for possible inclusion in the study. After children were nominated, informed consent from parents or guardians of these children was obtained. Selection criteria for focal children included: (a) being between 3 and 5 years old; (b) enrollment in an early childhood program; (c) being at elevated risk for EBD as indicated by the *Early Screening Project* (ESP; Walker, Severson, & Feil, 1995); (d) having average or above average cognitive/intellectual abilities as indicated by the *Battelle Developmental Inventory–Second Edition Screener* (BDI-II Screener; Newborg, 2005); and (e) demonstrating externalizing behaviors that interfered with classroom participation. After this screening procedure, the top one to two children in each classroom with the most extreme scores on the ESP were selected. Fourteen of the children were male and at the start of the study, 2 were 3 years old, 16 were 4 years old, and one was 5 years old. Fourteen children were African American, 2 children were White, 1 child was Asian/Pacific Islander, and race information was not provided for 2 children; all children (except for the child from the university-based classroom) qualified for free and reduced lunch.

Dependent Measure

The *TCIDOS* was used to examine the relationship between teacher and child behaviors across time as well as the impact of the intervention on the nature of teacher–child interactions.

TCIDOS observational categories and codes. Using the *TCIDOS*, data were collected within early childhood classroom ecologies representing transactional (e.g., teacher–child interactions) and behavioral theory (e.g., discrete child- and teacher behaviors). Data were collected using partial-interval recording on the specific instructional practices included within *BEST in CLASS* (i.e., rules, precorrection, OTR, BSP, instructive feedback, corrective feedback) as well as on children's behaviors (i.e., disruption, aggression, and defiance) and teacher reprimands (see Table 1). In addition, child engagement, teacher active supervision, positive/neutral teacher–child interactions, and negative teacher–child interactions were measured using momentary time sampling. All codes are operationally defined, which helps to eliminate variability between coders (Thompson et al., 2000).

Observer training and coding guidelines. Due to the complex nature of the behaviors and interactions examined by the *TCIDOS*, specific coding guidelines were established to facilitate standardization between coders and increase reliability (Thompson et al., 2000). Research assistants spent approximately 10 hr memorizing the initial iteration of codes. Next, discussions that occurred among the research team while watching video clips of teacher and child behavior resulted in a further refinement of codes; this process occurred over approximately 2 weeks. Next, the research assistants double-coded videotapes until 80% interobserver agreement (IOA) (A/A + D × 100) was reached across all codes for three out of five straight sessions. The average length of training was 7 weeks, across which trainees double-coded approximately 20 videos. Finally, prior to commencement of data collection, research assistants reached 80% agreement across all codes on one session coded live in a classroom. Retraining procedures for any research assistants having more than one live coding session with IOA below 80% included reviewing coding definitions, viewing and discussing practice video clips with a reliable coder, and achieving 80% IOA with a reliable coder on all codes across three out of five video sessions before returning to classrooms for live coding.

Coders observed teacher and child behaviors for 20-min observations that were divided into 10-s intervals. Coders observed teacher and child behaviors for 10 s, which was then followed by a 10-s record interval during which coders noted the occurrence or nonoccurrence of target behaviors during each interval, including the momentary time sample codes which where coded at the beginning of the 10-s record interval. This resulted in a total observational session of 20 min per child equating to 10 min of observational data. Ten-second observation intervals were selected to be able to balance having a large sample of codes to observe and having intervals long enough to capture the behaviors of interest. For example, many of the coded behaviors (e.g., teacher statements) are several seconds in length and we wanted intervals to be long enough to capture these behaviors. Ten-second coding intervals were used to give observers enough time to note the occurrence of codes on the coding sheet. Finally, 10 min of observational data provided us with a robust sample of teacher and child behavior, as Gunter and Reed (1996) suggested that 5-min vignettes of instructional sessions, as a minimal duration, can provide insight into teacher instructional behavior.

Observation procedures. Observations using the *TCIDOS* occurred in each early childhood classroom during small and large group teacher-directed instructional activities. Direct observation of each focal child occurred four times for a total of 40 min of observational data during the pretreatment phase, and two observations each week for a total of 20 min of observational data following the introduction of the teacher training and coaching for 14 weeks. We also collected 40 min of follow-up observational data per child to examine maintenance effects 1 month after the intervention had concluded. This resulted in a total of 474 observations across all phases of the study, with an average number of 24.95 (SD = 4.87) observations per child.

Observers positioned themselves in classrooms such that they could have a clear line of vision to the teacher and children, and were close enough to hear verbalizations but not so close as to serve as a distraction. Observers visited classrooms 1 to 2 times prior to observations beginning to attenuate children to their presence. Observers used a computer-based timer and one earphone during observations to prompt them to "observe" and "record" every 10 s. During IOA checks, secondary observers followed these same procedures, collecting data at the same time as the primary observer using an earphone connected to the same computer to ensure that observers were coding the same interval.

Results

IOA. IOA was assessed for the occurrence or nonoccurrence of the dependent variables represented by intervals coded for 24.7% of the observation sessions (across phases). IOA was calculated by the number of agreements divided by the number of agreements and disagreements (A/A + D) as well as by computing Cohen's kappa, which is a more conservative measure of agreement as it adjusts for chance agreement among codes (Cohen, 1960). Overall estimates of IOA ranged from 87% to 99%; the overall kappa for the measurement system was .54 (SD = 0.12). In general, kappas greater than .75 are considered excellent, between .60

Table 2. Interobserver Reliability Estimates: Percent Agreement and Kappa.

	Percent Agreement			
Code	*M*	*SD*	Kappa	*SD*
Positive/neutral interaction	95.92	5.90	.50	0.39
Negative interaction	99.71	1.25	.41	0.42
Rule	99.33	1.17	.73	0.37
Precorrection	97.76	3.27	.61	0.40
Opportunities to respond	87.32	9.66	.73	0.23
Behavior-specific praise	98.29	2.31	.61	0.38
Instructive feedback	95.94	6.31	.45	0.41
Corrective feedback	97.49	3.44	.47	0.44
Reprimands	99.65	0.69	.55	0.46
Disruption/ aggression/defiance	96.82	3.69	.37	0.33
Active supervision	97.71	4.55	.66	0.39
Engagement	94.70	4.91	.45	0.35

and .75 are considered good, between .40 and .60 are considered fair, and below .40 are considered poor (Fleiss, 1981). However, Yoder and Symons (2010) cautions that the meaning or interpretation of kappa varies depending on the base rates of observable behaviors. Behaviors with low base rates are more likely to have lower values for kappa as total chance agreement is greater in this instance. Reporting IOA and kappa values provides greater information about the reliability of the *TCIDOS* when considered according to the base rates of coded behaviors. According to Fleiss's guidelines, five individual codes can be considered good, six can be considered fair, and one poor. See Table 2 for percent agreement estimates and kappas by code.

Discussion

In this article, we described the *BEST in CLASS* intervention model, which is designed to improve teacher–child interactions in early childhood settings. We used the description of this model, and the theoretical framework that supports it, to describe the development and preliminary data of an empirically- and theoretically derived direct observational coding system, *TCIDOS*, which provides a method for collecting data on key teacher and child behaviors, as well as teacher–child interactions. The *TCIDOS* provides a means to evaluate the influence of teachers' use of effective instructional practices applied in early childhood settings on the problem behaviors of young children as well as changes in teacher–child interactions. Specifically, the *TCIDOS* allows us to examine micro-level aspects (i.e., specific teacher and child behaviors) of our intervention that may influence the occurrence, transactional nature, and association among these behaviors. Through direct and repeated measures of behaviors and interactions over time, we are able to capture the nature and patterns of the interactions between young children and their early childhood teachers. In addition, we will also be better positioned to determine whether the *BEST in CLASS* intervention model is effective in reducing young children's chronic problem behaviors and increasing their positive interactions with their teachers.

The development of *BEST in CLASS* and the *TCIDOS* has occurred through an iterative process. Throughout this process, we have been guided by the conceptual framework that supported the development of the *BEST in CLASS* intervention and ultimately informed our theory of change, as well as that of the *TCIDOS*. By grounding our intervention in theory, and then developing our observation system from the same conceptual framework, we are able to link our coded variables to our research questions in a way that is conceptually and empirically relevant. In turn, data produced by this system will be particularly beneficial to our research as it will be linked conceptually to the components of our intervention and the key child outcomes we hypothesize that our intervention will affect.

Although there are many advantages in using a complex direct observational coding system like the *TCIDOS*, there are also a number of challenges. For example, one major challenge is that data collection can be time-consuming and personnel intensive. Observers first must be trained to criterion, and there are also logistical challenges inherent in collecting observational data across time in a large number of classrooms. An additional challenge is that obtaining IOA on a large number of variables may be difficult to achieve, and if so, adds to the expense and time constraints. These issues may be related to the limitations of direct observation data collection systems used in other research that investigates teacher and child behavior change through larger RCTs (e.g., Walker et al., 2009; Webster-Stratton et al., 2008). In general, direct observation remains a challenge for the field, particularly in the face of decreasing opportunities for research funding (Mervis, 2011).

Limitations

There are several limitations of the current study that provide areas for future development work and research. Preliminary IOA estimates of the *TCIDOS* are promising but do raise issues related to reporting percent agreement versus the more conservative kappa statistic. For example, the child code DAD had low base rates of occurrence, and also had the lowest and only kappa in the poor range; however, the percent agreement estimate occurrence and nonoccurrence was 96.82. The use of the kappa statistic allowed us to

account for chance occurrence of codes, thus not overestimating agreement based on nonoccurrence (i.e., intervals where both coders agreed that a DAD did not occur, which were the large majority of coded intervals). The low kappa for DAD suggested to us that we needed to refine the definition to ensure that, when DADs did occur, observers would be more likely to accurately code this behavior. Thus, in refining the *TCIDOS* for use in our current research (i.e., an RCT efficacy study of the *BEST in CLASS*), we focused on refining the DAD code by providing more clarity for the definition in the coding manual, adding examples and nonexamples, and having more discrepancy discussions during training and weekly staff meetings.

Reactivity to videotaping or observers is always a limitation to using direct observational measurement, which can result in an artificially inflated occurrence of behaviors of interest. We attempted to address reactivity by having observers spend time in the classrooms on one to two occasions prior to beginning data collection to attenuate teachers and children to their presence; of course, this strategy adds time and expense to the data collection process but may be a necessary step to address reactivity. Finally, although the preliminary reliability data on the *TCIDOS* reported in this article are promising, the lack of validity analyses remains a limitation of the measurement system and should be addressed in future studies.

Implications for Research

Although the *TCIDOS* was developed as a measurement tool for the *BEST in CLASS* intervention, we believe that it may also be useful to other researchers who are interested in measuring teacher instruction, child behavior, and teacher–child interactions. In essence, the *TCIDOS* behavioral codes represent effective instructional practices that all teachers use to some degree and child responses that are applicable across a wide range of grade levels and classroom settings. Prior to widespread dissemination and use by other researchers, however, further research needs to be conducted to examine its accuracy in capturing and reflecting change in the constructs in which it is designed to measure. Although the *TCIDOS* uses direct observation to measure behaviors, further validation of the definitions used to describe behaviors is needed to examine the overall construct validity of the measure. In addition, more research is needed to examine the sensitivity of the measurement system, including the amount of observation time (based on overall base rates of behavior) that is needed to obtain a representative sample of behaviors. Finally, the unit of measure used for recording behavior should be examined to assure a representative sample is obtained in an efficient manner. Clarification of these issues could potentially support the use of the *TCIDOS* by other researchers doing prevention work in classrooms, addressing issues of time constraints and expense that might otherwise discourage their use of observational data collection.

Having additional observational measurement tools is an advantage for researchers working in classroom settings. As Conroy et al. (2008) pointed out, the use of multiple informant measurement (e.g., direct observation as well as more indirect measures such as teacher reports) in examining relations between teacher and child behavior in classroom-based research is desirable. Thus, in addition to the *TCIDOS*, we are using teacher report forms (*Caregiver-Teacher Report Form*; Achenbach & Rescorla, 2000; *Social Skills Improvement System*; Gresham & Elliott, 2008) as well as a more global measure of classroom climate (*CLASS*; Pianta et al., 2008) to assess child- and classroom-level variables in our RCT. In addition, we are using the *InCLASS* (Downer, Booren, Lima, Luckner, & Pianta, 2010), which is a relatively newer measure with promising psychometrics that measures teacher–child interactions. By using a variety of measures to assess child- and classroom variables, we will be in a better position to examine the validity of the *TCIDOS* in the near future, addressing a major limitation of the current study.

Finally, the initial IOA findings described in this article highlight the importance of using kappa as an estimate of agreement during training, particularly when there are low rates of behaviors. To illustrate, in training, we used 80% IOA (A/A + D × 100) of double-coded sessions as a criteria for observers to begin live coding. By using this procedure to estimate IOA, we may have overestimated agreement, particularly with low base rate codes. We are currently using a computer-based version (Tapp, 2010) of the *TCIDOS* to collect observation data in a larger efficacy study, which allows us to more closely monitor kappa estimates in a timely manner. Other researchers may want to consider using kappa as a criterion for observer checkout during training, particularly when there are low base rates of observational codes.

Implications for Practice

As behavior is context specific, the precise and accurate measurement of behaviors occurring within natural, authentic contexts is important. Through the direct observation of the behaviors we have identified and included in the *TCIDOS* coding system, we are able to obtain data on the complex interactions young, high-risk children have with their early childhood teachers. Collection of these data will help to determine whether our intervention is able to affect change in these patterns, indicating an improved behavioral trajectory. Thus, the *TCIDOS* may contribute to practice via helping us better understand how teacher behaviors influence children's behavior, and vice versa. Furthermore, given the universal nature of the teacher and child observational codes in the *TCIDOS*, other researchers working in classrooms with young children may find this system useful

in their work, leading to a better understanding of the dynamics between teachers and young children. In turn, researchers and teacher trainers may be better positioned to provide prevention programs and teacher training that are based on reliably observed teacher and child behaviors.

In this article, we have described an observational measurement system with clear theoretical and empirical links to a prevention program, *BEST in CLASS*. Although preliminary psychometric data on the system are promising, more work is necessary, particularly related to the validity of the system. That said, we view this as an important first step in the measure development process. The collection of direct observation data is resource intensive; however, these data also have the potential to enhance our ability to determine the efficacy of prevention programs in RCTs. This is a critical step, particularly as efficacious programs go to scale in effectiveness trials.

Authors' Note

The opinions expressed by the authors are not necessarily reflective of the position of or endorsed by the U.S. Department of Education.

Declaration of Conflicting Interests

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

Funding

The authors 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, Institute of Education Sciences (R324A080074, R324A110173).

Note

1. A comprehensive manual (available from the first author) was developed for the purposes of coding the behaviors and interactions of interest. The manual provides detailed instructions both for observational procedures as well as operational definitions for and exemplars of each code.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*, 91–97.

Bakeman, R. (2000). Behavioral observation and coding. In C. M. Judd (Ed.), *Handbook of research methods in social and personality psychology* (pp. 138–159). New York, NY: Cambridge University Press.

Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). The state of preschool, 2011: State preschool yearbook. *National Institute for Early Education Research*. Retrieved from http://nieer.org/sites/nieer/files/2011yearbook.pdf

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (Ed.). (2005). *Making human beings human: Bioecological perspectives on human development*. Thousand Oaks, CA: SAGE.

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology*, *40*, 415–436.

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, F. (2008). Classroom problem behavior and teacher-child relationships in kindergarten: The moderating role of classroom climate. *Journal of School Psychology*, *46*, 367–391.

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Conroy, M. A., Stichter, J. P., Daunic, A., & Haydon, T. (2008). Classroom-based research in the field of emotional and behavioral disorders: Methodological issues and future research directions. *Journal of Special Education*, *41*, 209–222.

Dishion, T. J., French, D. C., & Patterson, G. R. (1995). The development and ecology of antisocial behavior. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, Vol. 2. Risk, disorder, and adaptation* (pp. 421–471). New York, NY: John Wiley.

Domitrovich, C. E., Cortes, R. C, & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention*, *28*, 67–91.

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*, 588–599.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Education Research Quarterly*, *25*, 1–16.

Dunlap, G., Strain, P. S., Fox, L., Carta, J., Conroy, M., Smith, B. J., & Sowell, C. (2006). Prevention and intervention with young children's challenging behavior: Perspective regarding current knowledge. *Behavioral Disorders*, *32*, 29–45.

Embry, D. D. (2004). Community-based prevention using simple, low-cost, evidence-based kernels and behavior vaccines. *Journal of Community Psychology*, *32*, 575.

Farmer, T. W., Quinn, M. M., Hussey, W., & Holahan, T. (2001). The development of disruptive behavior disorders and correlated constraints: Implications for intervention. *Behavioral Disorders*, *26*, 117–130.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York, NY: Wiley.

Fox, J. J., & Conroy, M. A. (1995). Setting events and behavior problems: An interbehavioral field analysis for research and practice. *Journal of Emotional and Behavioral Disorders*, *3*, 130–140.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system—Rating scales*. Minneapolis, MN: 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.

Gunter, P. L., Denny, R. K., Shores, R. E., Reed, T. M., Jack, S. L., & Nelson, C. M. (1994). Teacher escape, avoidance, and countercontrol behaviors: Potential responses to disruptive and aggressive behaviors of students with severe behavior disorders. *Journal of Child and Family Studies*, *3*, 211–223.

Gunter, P. L., & Reed, T. M. (1996). Self-evaluation of instruction: A protocol for functional assessment of teaching behavior. *Intervention in School and Clinic*, *31*, 225–230.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationship and the trajectory of children's school outcome through eighth grade. *Child Development*, *72*, 625–638.

Henricsson, L., & Rydell, A. (2004). Elementary school children with behavior problems: Teacher-child relations and self-perception. A prospective study. *Merrill-Palmer Quarterly*, *50*, 111–138.

Hussey, D. L., & Flannery, D. J. (2007). Implementing and evaluating school-based primary prevention programs and the importance of differential effects on outcomes. *Journal of School Violence*, *6*, 117–134.

Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools*, *44*, 65–75.

Ladd, G. W., & Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development*, *70*, 919–929.

Lewis, T. J., Hudson, S., Richter, M., & Johnson, N. (2004). Scientifically supported practices in emotional and behavioral disorders: A proposed approach and brief review of current practices. *Behavioral Disorders*, *29*, 247–259.

Loeber, R., & Farrington, D. P. (2000). *Serious and violent juvenile offenders: Risk factors and successful interventions*. Thousand Oaks, CA: SAGE.

Magnusson, D., & Cairns, R. B. (1996). Developmental science: Toward a unified framework. In R. B. Cairns, G. H. Elder, Jr., & E. J. Costello (Eds.), *Developmental science* (pp. 7–31). New York, NY: Cambridge University Press.

Mervis, J. (2011). Fewer dollars, forced choices. *Science*, *334*, 750–752.

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*, 600–608.

Newborg, J. (2005). *Battelle Developmental Inventory–2nd Edition: Examiner's Manual*. Rolling Meadows, IL: Riverside Publishing.

Patterson, G. R., & Reid, J. B. (1970). Reciprocity and coercion: Two facets of social systems. In C. Neuringer & J. L. Michael (Eds.), *Behavior modification in clinical psychology* (pp. 133–177). New York, NY: Appleton.

Patterson, G. R., Reid, J. B., & Eddy, M. J. (2002). A brief history of the Oregon model. In J. B. Reid, G. R. Patterson & J. Snyder (Eds.), *Antisocial behavior in children and adolescents* (p. 321). Washington, DC: American Psychological Association.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system: Manual K-3*. Baltimore, MD: Paul H. Brookes.

Sameroff, A. J. (1995). General systems theory and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, Vol. 1. Theory and methods* (pp. 659–695). New York, NY: Wiley.

Sameroff, A. J. (2009). *The transactional model of development: How children and contexts shape each other*. Washington, DC: American Psychological Association.

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

Sutherland, K. S., Conroy, M., Abrams, L., & Vo, A. (2010). Improving interactions between teachers and young children with problem behavior: A Strengths-based approach. *Exceptionality*, *18*, 70–81.

Sutherland, K. S., & Oswald, D. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies*, *14*, 1–14.

Sutherland, K. S., & Wehby, J. H. (2001). Exploring the relation between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with emotional and behavioral disorders: A review. *Remedial and Special Education*, *22*, 113–121.

Tapp, J. (2010). *Lily data collector*. Nashville, TN: Vanderbilt Kennedy Center.

Thompson, T., Symons, F. J., & Felce, D. (2000). Principles of behavioral observation: Assumptions and strategies. In T. Thompson, D. Felce, & F. Symons (Eds.), *Behavioral observation: Technology and applications in developmental disabilities* (pp. 3–15). Baltimore, MD: Paul H. Brookes.

Vo, A., Sutherland, K. S., & Conroy, M. A. (2012). BEST in CLASS: A classroom-based model for ameliorating problem behavior in early childhood settings. *Psychology in the Schools*, *49*, 402–415.

Wahler, R., & Dumas, J. E. (1986). "A chip of the old block": Some interpersonal characteristics of coercive children across generation. In P. Strain, M. Guralnick, & H. M. Walker (Eds.), *Children's social behavioral: Development assessment and modification* (pp. 49–91). Orlando, FL: Academic Press.

Walker, H. M., Seeley, J. R., Small, J., Severson, H. H., Graham, B. A., Feil, E. G., & Forness, S. R. (2009). A randomized controlled trial of the first step to success early intervention. *Journal of Emotional and Behavioral Disorders*, *17*, 197–212.

Walker, H. M., Severson, H., & Feil, E. (1995). *Early screening project: A proven child find process. Examiner's Manual*. Longmont, CO: Sopris West.

Webster-Stratton, C. (1997). Early intervention for families of preschool children with conduct problems. In M. J. Gurlanick (Ed.), *The Effectiveness of early intervention* (pp. 429–454). Baltimore, MD: Paul H. Brookes.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry*, *49*, 471–488.

Werts, M. G., Wolery, M., & Holcombe, A. (1991). *Instructive feedback: Increasing opportunities for learning through the addition of incidental information*. Pittsburgh, PA: Allegheny-Singer Research Institute. (ERIC Document Reproduction Service No. ED407755)

Yoder, P., & Symons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer.