Journal of Early Intervention
2022, Vol. 44(4) 353–370

DOI: 10.1177/10538151211069528
journals.sagepub.com/home/jei

Comparing Coding Viewing and Recording Methods to Quantify Embedded Instruction Learning Trials

Mackenzie K. Martin[1,2], Patricia A. Snyder[1,2], Brian Reichow[1,2], and Crystal D. Bishop[1]

Abstract

The purpose of this study was to examine the comparability of counts of embedded instruction learning trials when different methods of viewing and recording direct behavioral observations were used. In 13 classrooms, while videotaping embedded instruction implementation for a larger randomized controlled efficacy trial was occurring, teachers' implementation of trials was coded in situ using pencil-and-paper methods. Videos were later coded using computer-assisted methods. Dependent-samples *t* tests, Pearson product-moment correlation coefficients, and additional score agreement calculations were conducted. Statistically significant differences were found in the estimates of trial frequency. Correlational analyses showed positive and strong relationships between the coding methods. Coding agreement was higher across the entire observation versus during 10-min continuous event blocks. In situ coding took significantly less time than video coding. Results provide empirical evidence for the advantages and disadvantages of common viewing and recording methods for quantifying behavior as part of systematic observation systems.

Keywords

direct behavior observation system, embedded instruction, observational measurement, coding viewing and recording methods

When conducting observational measurements of behavior, a variety of formats and methodologies are available for use. Although informal observations can be conducted and recorded using descriptive notes or rating scales, direct behavioral observation systems (DBOS) often are used when the research questions of interest focus on systematically quantifying counts or durations of behavior (Yoder et al., 2018). One advantage of using these systems is that observers can code a range of behaviors from context-dependent to more generalized characteristics. DBOS are systematically developed or adapted to optimize the data collected on a behavior(s) of interest based on procedural decisions about behavior sampling, participant sampling, and coding viewing and recording methods (Yoder et al., 2018). Each of these procedural decisions has strengths

[1]Anita Zucker Center for Excellence in Early Childhood Studies, University of Florida, Gainesville, USA
[2]School of Special Education, School Psychology, and Early Childhood Studies, University of Florida, Gainesville, USA

Corresponding Author:
Mackenzie K. Martin, Anita Zucker Center for Excellence in Early Childhood Studies, School of Special Education, School Psychology, and Early Childhood Studies, University of Florida, Gainesville, FL 32611, United States.
Email: mkmartin@ufl.edu

and limitations as these decisions influence data collection, data integrity, and the analyses that can be conducted with the data (Lane & Ledford, 2014). Of particular interest is the influence of viewing method (how coders observe the behavior) and recording method (how coders record their data and coding decisions). Without careful consideration of the strengths and limitations of these decisions and their interrelatedness, confounds can arise that affect the score reliability and validity of the observational data.

Methods for observational coding have progressed in part due to advancements in technology. In a 1991 summative review of technologies for collecting behavioral observation data, Farrell described recording methods ranging from the traditional in situ pencil-and-paper coding to bar-code scanning, microcomputer switches, and palm pilots (Farrell, 1991). Since then, researchers have used additional methods such as computer-assisted coding with touchscreen tablets, smart-phones, and telemetric devices with video or audio recording (Cohen & Rozenblat, 2015). As new technology emerges and applications for conducting systematic observations of behavior advance, there is a concurrent need to conduct studies that examine how decisions about behavior sampling, participant sampling, and coding viewing and recording methods affect score reliability or validity. Studies conducted on behavior and participant sampling have generally found consistent results favoring continuous methods (e.g., Lane & Ledford, 2014; Mudford et al., 2009; Powell et al., 1977; Rapp et al., 2011), although recent guidance suggests circumstances when interval sampling might be justified (cf. Yoder et al., 2018). Studies comparing how behavior is viewed (in situ vs. from video) and recorded (pencil-and-paper vs. computer-assisted) have been less frequently conducted. Recommendations related to selecting and using particular viewing and recording methods have primarily relied on scholars' experiences and anecdotal observations. From these experiences and observations, strengths and limitations of using various viewing and recording methods have been described or proposed in the literature (e.g., Suen & Ary, 2014; Wessel, 2015; Yoder et al., 2018).

Given the coronavirus disease (COVID-19) pandemic, research in this area has become perhaps even more important. Researchers had to shift their methods for data collection when they lost the ability to observe in situ. This historical event further highlighted the importance of conducting research that compares viewing and recording methods so that researchers can make data-based informed decisions about how to use resources and adapt DBOS, particularly as viewing methods change. Although conducted prior to COVID-19, the primary purpose of the present study was to examine the extent to which two combinations of viewing and recording methods produced comparable estimates of the frequency of implementation of embedded instruction learning trials. In addition, we examined interobserver agreement between learning trial data coded in situ using pencil-and-paper versus when coded from video using computer-assisted software.

Comparing Coding Viewing and Recording Methods

Behavior viewing method refers to whether the observer is conducting the observation in situ or by watching a video. The recording method refers to the way in which the observer records the decisions and codes about the behavior of interest. The focus of the present study was the combination of in situ pencil-and-paper methods compared with video computer-assisted methods.

Figure 1 shows a summary of considerations that have been described in the literature for four different combinations of behavior viewing and recording methods: (a) in situ pencil-and-paper, (b) video pencil-and-paper, (c) in situ computer-assisted, and (d) video computer-assisted. Among the considerations for selecting one of these combinations are the behavior of interest; materials required; and data collection, analysis, and use. In Figure 1, filled circles are used to indicate strengths for a combination of viewing and recording methods that have more practical utility given a particular consideration. A combination of methods that has limitations for a particular consideration and less practical utility is shown by hollow circles. Half-shaded circles are used to

Considerations	In Situ Pencil-and-Paper[a]	Video Pencil-and-Paper	In Situ Computer-Assisted	Video Computer-Assisted[a]
Behavior Sampling				
Multiple passes	○	●	○	●
Coding time	●	◑	●	◑
Stop-and-go	○	●	○	●
Physical flexibility/adaptability	●	○	●	○
Materials				
Time to access manual	◑	●	◑	●
Financial costs	◑	◑	◑	○
User-friendly	●	◑	◑	◑
Data Collection, Analysis, and Use				
Specificity of data	◑	●	◑	●
Technology failures	●	○	◑	○
Time to process	◑	○	●	○
Time to analyze	○	○	●	●
Blind interobserver agreement	○	●	○	●
Resources for other purposes	○	●	○	●

Figure 1. Considerations for using four different viewing and recording methods.
Note. ● = strength; ◑ = conditional; ○ = limitation. These considerations are tailored to direct behavioral observation systems that use continuous event sampling.
[a]Methods of focus in the present study.

reflect situations in which the utility may be dependent upon other aspects of the DBOS (i.e., conditional). For example, coding time is denoted as conditional for all video methods because coding time may depend on the degree to which a coder uses stop-and-go or multiple-pass coding. If a coder uses stop-and-go coding or multiple passes, then the coding time will be longer than that of in situ coding. If, however, the video coder uses single-pass coding and never stops the video to replay portions of the observation, coding time will likely be similar to in situ coding.

Related to the viewing methods used in the present study, in situ coding provides practical utility as it allows an observer to be present in the environment, which offers flexibility and adaptability. Alternatively, an observer using video must rely on data that have already been collected and may restrict what the observer can see or hear. A situation where video coding may be more practical would be when a researcher wants to use the video data for additional purposes, such as during a coaching session. For recording methods, an observer could use either pencil-and-paper coding or computer-assisted software based on resource availability, but computer-assisted software might be more efficient for data processing and data analyses. Alternatively, pencil-and-paper coding may be more user-friendly and would minimize the risk of data lost due to technological malfunctions.

Beyond the considerations shown in Figure 1, we identified five studies in the extant literature (Bernal et al., 1971; Gridley et al., 2018; Johnson & Bolstad, 1975; Kent et al., 1979; Tapp et al., 2006) in which data were compared when different viewing or recording methods were used. Table 1 shows information about each study for variables coded, sampling methods used, comparisons calculated, and findings reported. Across studies, mixed results were reported, often based on the statistics used to examine the comparability of methods. In some studies, correlational statistics were used to examine comparability. Findings in these studies showed associations that ranged from no correlation to strong correlations (Gridley et al., 2018; Kent et al., 1979). In other studies, researchers evaluated comparability by examining mean differences, or absolute differences, using analyses of variance (ANOVAs) or *t* tests (Gridley et al., 2018;

Table 1. Literature Comparing Direct Behavioral Observation Systems.

Author	Coding variable	Behavior sampling method	Participant sampling method	Viewing methods compared	Recording methods compared	Methods to compare findings	Results
Bernal et al. (1971)	Mother commands	Hybrid-event sampling within blocks	C	In situ vs. audiotapes	NR	Pearson *r* correlations	Strong positive correlation
Gridley et al. (2018)	Parent–child interactions	Hybrid-continuous event sampling within blocks	C	In situ vs. single pass video	NR	Spearman's correlation; *t* test	Moderate to strong correlation for 92% of codes No statistically significant difference in means for 3 of 4 composite coding categories
Johnson and Bolstad (1975)	Parent–child interactions	Continuous event sampling	C	In situ vs. audiotapes	NR	ANOVA; correlations	No significant main effects or interactions found for viewing method Moderate correlations found for two of three coding categories
Kent et al. (1979)	Child disruptive behavior	Partial interval sampling	F	In situ vs. mirror vs. television	NR	ANOVA; correlations	No significantly significant difference in means for eight of nine coding categories Moderate correlation found between methods
Tapp et al. (2006)	Caregiver–child interactions	Partial interval sampling	C	Videos	PP vs. pocket PC	Kappa for within methods	Kappa higher for pocket PC but above acceptable levels for both methods

Note. C = conspicuous; F = focal; NR = not reported; ANOVA = analysis of variance; PP = pencil-and-paper; PC = personal computer.

Johnson & Bolstad, 1975; Kent et al., 1979). In these studies, some researchers found no differences between coding methods (Johnson & Bolstad, 1975), whereas others found differences for a subset of behaviors coded (Gridley et al., 2018; Kent et al., 1979). The viewing and recording methods used in each study and additional details about each study are discussed below.

In two studies, researchers compared continuous timed-event data using two different viewing methods: in situ at home or audio recordings of interactions in the home (Bernal et al., 1971; Johnson & Bolstad, 1975). Bernal et al. examined relationships between the recorded frequency of commands given by a mother to a child when an investigator-developed audio recorder was used to capture interactions in the home compared with an observer who was physically present to observe the interactions. Coders recorded the frequency of commands by tallying occurrences within 10-min blocks. No explicit recording format was reported. The authors compared command frequencies between the audio recordings and visual observations by calculating Pearson product-moment correlations and found a statistically significant and noteworthy correlation between the methods ($r = .86, p < .01$). Johnson and Bolstad (1975) examined differences in the recorded frequency of parent and child discrete behaviors during a 45-min continuous event sampling by observers in situ and those coding from audiotape recordings. The authors did not report the recording method used. No statistically significant differences in the absolute counts were found between viewing methods. In addition, moderate correlations were found for two of the three coding categories ($r = .51$ to $.68$, $p < .05$). A disadvantage of using audiotapes, noted by both Bernal et al. (1971) and Johnson and Bolstad, was technology malfunctions. Issues with the audiotapes resulted in lost and unrecoverable data in both studies.

Kent and colleagues (1979) compared frequency data gathered from the same DBOS that measured nine child behaviors using 20-s partial interval recording and three different viewing methods (i.e., in situ, behind a one-way mirror, and via closed-circuit television). No information on the recording method was reported by authors. No statistically significant differences were found in the estimates of frequency data between the methods for eight of the nine child behavior codes. The recorded rate of child vocalizations was higher for the in situ viewing method compared to the other two viewing methods. Rank order comparisons across the coded behaviors were variable and ranged from no correlation to strong correlation ($r = .00$–$.70$). When all behaviors were collapsed together as a composite, correlation coefficients were moderate ($r = .50$–$.58$).

Only one study explicitly compared recording methods between observers (Tapp et al., 2006). Tapp and colleagues compared data collected when observers viewed videos and used pencil-and-paper recording methods to observers who viewed videos and used computer-assisted recording methods. Observers coded caregiver–child interactions in a child-care setting for 30-min, using 30-s partial interval sampling. The researchers compared the methods based on preparation time, duration of data entry, duration of kappa calculations, accuracy (kappa scores for coders within the same recording method), and overall cost. The authors found that the preparation times were comparable between the two recording methods. Kappa was high for observers coding within the same recording method ($\kappa = .90$ and $.80$, respectively, for computer-assisted and pencil-and-paper methods). The duration of data entry, kappa calculations, and overall cost indicated the computer-assisted method as the superior method because it was less costly, more reliable, and more time-efficient.

Gridley and colleagues (2018) compared continuous event sampling within six, 5-min blocks from observations where observers coded parent–child interactions for a sample of 40 dyads either in situ at home or via single-pass viewing of videotapes. Observers coded parent–child interactions on 29 discrete and non-contingent parent or child behaviors. No information about coding methods was reported. Gridley et al. found that the estimates of observed behaviors that resulted from the two viewing methods were positively correlated ($r = .57$–$.96$, $p < .001$) for all but two discrete behaviors (physical intrusions and physical negatives). When comparing associations for the four composite categories, all four composites had statistically significant and strong positive correlations between coding methods. Absolute differences were calculated for the four composite categories (i.e., positive parenting, negative parenting, child positives, and child negatives). There were statistically significant differences for only one composite category (i.e., child positive).

For each study discussed above and shown in Table 1, researchers compared viewing and recording methods for discrete behaviors that were coded separately from any other participant's behavior. Coders did not have to consider a linked set of codes that were based on individuals (i.e., practitioner and child) responding contingently to each other during an interaction, which was an issue in the present study. In the present study, we compared viewing and recording methods for a set of linked stimuli or behaviors (i.e., antecedent, behavior, consequence) coded as an event (embedded instruction learning trial) in a continuous event coding system. To inform future studies focused on quantifying embedded instruction learning trials, we were interested in determining if direct behavioral observation data collection obtained under two viewing and recording methods (i.e., in situ viewing with pencil-and-paper recording, video viewing with computer-assisted recording) yielded comparable results.

Quantifying Embedded Instruction Learning Trials Using DBOS

Instructional interactions between a teacher and child comprise elements of a three-term-contingency (i.e., antecedent, child behavior, consequence; Skinner, 1953). The combination of these teacher and child behaviors has been identified as a complete learning trial, which is defined as the teacher delivering all procedural components of a learning trial with fidelity (Snyder et al., 2017). Measuring complete learning trials is a complex process for several reasons. First, observers are required to consider a sequence of interlocking behaviors occurring between people during an interaction (e.g., teacher and child). Second, each child may be working on different skills, resulting in variability of the behavior of interest during teacher–child interactions within and across observations. Finally, the behavior of interest may be complex due to the target skill being a discrete behavior, chained behavior, or part of a response class that may vary in topography throughout the observation. Complete learning trials can be implemented as either embedded learning trials or as decontextualized instructional learning trials. Although several differences exist between embedded learning trials and decontextualized instructional trials, the most notable difference is that embedded trials are implemented in free-operant environments by familiar adults during typically occurring activities, routines, and transitions.

Researchers have used a variety of DBOS viewing and recording methods to measure the implementation of embedded learning trials. Kohler et al. (1997) had observers code teacher implementation of instructional episodes on children's target objectives during 1:1 activities, small group activities, and large group activities. Observers used the in situ viewing method and coded with continuous timed-event behavior sampling. The authors did not report the type of recording method used by observers. Horn et al. (2000) measured embedded learning opportunities by viewing videotapes of teachers' implementation and recorded learning target behaviors using a 10-s partial-interval system with multiple passes (one pass per learning target behavior). The authors did not report the recording method used. Although the viewing and recording methods for coding embedded learning trials differed across these studies, there were no comparisons of viewing or recording methods examined within the studies.

The primary focus of the present study was to evaluate the comparability of quantifying embedded instruction learning trials when observations for the same session were conducted using two different viewing and recording methods (i.e., in situ/pencil-and-paper, video/computer-assisted observational software). In addition, we were interested in examining interobserver agreement across the two methods. As part of the present study, the first author adapted a video and computer-assisted DBOS being used in a larger randomized controlled efficacy trial (Snyder et al., 2015) for use in situ with pencil-and-paper methods. Specifically, we addressed two research questions focused on whether the two methods yielded comparable results: (a) Are there statistically significant and noteworthy differences in the frequency of embedded instruction learning trials between video and computer-assisted coding methods versus in situ pencil-and-paper methods? and (b) Are there statistically significant and noteworthy associations for the frequency of embedded instruction learning trials across the two methods? To explore interobserver agreement (IOA), two questions were addressed: (a) What is IOA between the two methods when the frequency of embedded instruction learning trials is calculated for the total observation? and (b) What is IOA when it is calculated within 10-min continuous event blocks? Finally, we explored whether there were statistically significant and noteworthy differences in the time spent observing and recording data across the two methods.

Method

Context for Present Study

Embedded instruction is a multicomponent naturalistic instructional approach. In the larger embedded instruction study (Snyder et al., 2015), the researchers used a DBOS to measure the rate of teachers' implementation of embedded instruction learning trials. Observers applied the codes for this DBOS while viewing video from the classroom and used computer-assisted recording methods with multiple passes for each child's embedded instruction priority learning target. Given the time and costs associated with these viewing and recording methods, the authors of the present study were interested in exploring the comparability and feasibility of alternative viewing and recording methods to inform future decisions about data collection for quantifying embedded instruction learning trials. An initial step was to compare in situ pencil-and-paper coding methods with video viewing and computer-assisted coding methods.

In the larger efficacy study, 111 preschool teachers were recruited at two performance sites in two cohorts (Snyder et al., 2015). Within each cohort, teachers were assigned randomly within schools to one of three experimental conditions. The first experimental condition included professional development workshops and materials for teachers focused on embedded instruction, followed by on-site coaching. The second experimental condition included professional development workshops and materials, followed by self-coaching. The third condition was a business-as-usual (BAU) condition, where teachers attended the typical professional development provided by the district. Across both cohorts, teachers participated in the study for up to 2 years (i.e., an intervention year and a sustainability year). Activities for the present study occurred with the second cohort of teachers at one site during their sustainability year of the study (Snyder et al., 2015). Thirty-four teachers remained in the study for the sustainability year, and 17 were enrolled at this site for this study. In situ data were collected at this site during the final data collection time point for the larger study.

Participants

Teacher participants. To be selected for the present study, teachers had to submit at least one priority learning target for embedded instruction that was observable and measurable. Of the 17 teachers, 16 at the site for this study met this criterion. After accounting for scheduling availability and balancing observations across experimental conditions, 13 female preschool teachers for children with disabilities participated in the present study. Of the 13 teachers, five were in the on-site coaching condition, four were in the self-coaching condition, and four were in the BAU condition. Because teachers were randomly assigned to experimental conditions, they had different experiences with the professional development and embedded instruction intervention variants that were part of the larger study. Including teachers in the counterfactual conditions in the present study was important given our interest in measuring teacher's implementation of embedded instruction who may be at different levels of fidelity. The mean age for participating teachers in the present study was 48.07 years (*SD* = 8.75), and their average years of experience working with young children with disabilities was 8.23 years (*SD* = 5.71). Six teachers had a bachelor's degree and seven had a master's degree. Eight teachers served preschool-aged children in inclusive classrooms and five teachers served preschool-aged children in classrooms that enrolled only children with disabilities. Nine teachers identified their ethnicity as White, three teachers identified as Hispanic, and one teacher identified as Black.

Child participants. The present study included 30 children (21 males and 9 females) in the 13 classrooms of the participating teachers. Each teacher had up to three children from their classroom enrolled in the larger study as target children. Target children were the children with whom the teacher was focusing implementation of embedded instruction as part of the larger study. All target children in the present study had identified disabilities and were observed in the present study because their teacher submitted at least one priority learning target for them that was observable and measurable. All children had to be present on the day of data collection to participate. Two additional children had observable and measurable priority learning targets and were potential participants but were absent during data collection and could not be observed.

Observation Methods

To compare viewing and recording methods, observers in the present study used two different methods for coding the frequency of embedded instruction learning trials. The first coding method quantified embedded learning trial frequency using in situ viewing and pencil-and-paper recording. The first author observed and coded in situ, while another data collector videotaped the observation for the comparison coding method. After videotapes were collected, they were screened to ensure videos met quality standards for coding and did not need to be recollected. Videotapes that met screening criteria were used for coding using computer-assisted observational software.

Coding Variable

The coding variable of interest was the frequency of complete embedded learning trials implemented by the teacher on the target children's priority learning targets. A complete learning trial (CLT) was defined as all procedural components of an embedded learning trial being implemented correctly by the teacher (i.e., antecedent, behavior, consequence, error correction, child behavior after error correction, feedback/consequence). The rationale for comparing only CLTs only was based on utility and feasibility. Coding a sequence of interlocking behaviors between a teacher and child in a free operant environment across numerous learning targets and children can be a complex and demanding task. We selected to code CLTs because it provided data on the frequency and accuracy with which teachers delivered embedded learning trials, which was the primary variable of interest in the larger efficacy trial.

Measures

Embedded Instruction Observation System–Revised (EIOS-R). The EIOS-R (Snyder et al., 2017) was the DBOS used in the larger efficacy study to measure embedded instruction learning trials delivered by a teacher for children's priority learning targets. EIOS-R observers watched videotaped observations and used Noldus Observer XT 12.5 (Noldus Information Technology, 2015) to record the frequency of embedded instruction learning trials. EIOS-R observers used multiple pass participant sampling to code each child's priority learning target during a separate pass of the video collected in each teacher's classroom. EIOS-R observers coded and timestamped a trial occurrence using the computer-assisted observational software and then recorded whether each component of a trial (i.e., antecedent, behavior, consequence, error correction, child behavior after error correction, feedback/consequence) occurred. An antecedent was defined as an event that set the occasion for a child's priority learning target behavior to occur. The child's behavior was coded as either a correct target behavior or an incorrect behavior, meaning the child did not exhibit the target behavior. When a correct target behavior occurred, the observer recorded the occurrence of a logical and timely natural or planned consequence. After an incorrect behavior, observers recorded whether error correction procedures occurred. If the child emitted the target behavior after error correction, an observer recorded whether the teacher delivered a consequence. If a child did not perform the target behavior following error correction, the observer recorded whether the teacher provided feedback to end the trial.

At the end of each trial, observers entered a trial summary code (i.e., complete learning trial, incomplete learning trial, or uncodeable learning trial). Uncodeable trials were observable and measurable learning targets that could not be coded via video because the EIOS-R coder could not see or hear all components of the trial, preventing them from determining if the trial was complete or incomplete based on EIOS-R coding decision rules. Although all types of trials were coded using the EIOS-R, the current study only compared CLTs across the two methods.

Embedded Instruction Observation System–In Situ (EIOS-I). The EIOS-I (Martin, 2019) was an adaptation of the EIOS-R (Snyder et al., 2015) designed for the present study to conduct in situ classroom observations on teacher implementation of CLTs. The first author piloted and modified the EIOS-I before the present study began and was the EIOS-I coder in the present study. The EIOS-I coder was a graduate assistant trained on the EIOS-R by the EIOS-R developers. She was a reliable EIOS-R coder with 2 years of experience using the EIOS-R before the present study was conducted. One EIOS-I observer (first author) coded in situ using pencil-and-paper recording. She used continuous event sampling with trial frequency recorded within 10-min blocks. The decision to record continuously and use 10-min blocks was informed by previous literature (e.g., Bernal et al., 1971; Johnson & Bolstad, 1975) and pilot testing. CLT data were recorded on an EIOS-I data collection form (Figure 2) that grouped the priority learning targets that were observable and measurable for each target child and provided rows to indicate each occurrence of a CLT for each target. After observing an embedded learning trial, the EIOS-I observer determined if the trial was complete (i.e., a CLT) or incomplete. Only CLTs were recorded on the data collection form and were differentiated by two summary codes. When an observed CLT involved an antecedent, correct child behavior, and consequence, the EIOS-I observer circled "C" on the coding form. When a CLT with error correction procedures was observed, the EIOS-I observer circled "H" to indicate an additional help trial occurred.

Data Collection Procedures

Preparing for coding. Teachers submitted up to four priority learning targets for each target child in the classroom to the site coordinator for the efficacy trial before each observation. The priority learning targets specified the behaviors or skills for which teachers planned to provide embedded learning trials for each child. For the present study, we used three indicators (i.e., behavior specified, demonstration specified, and observable) from the Learning Targets Rating Scale-Research Version (Snyder et al., 2016) to determine whether the priority learning target was observable and measurable. Observable and measurable priority learning targets were targets for which the

Unique ID: _______________
Cycle: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Activity Description ______________________
Start Time_______ End Time_______

Child ID _____ Child Description

Learning Target 1:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 2:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 3:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 4:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	

Child ID _____ Child Description

Learning Target 1:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 2:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 3:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 4:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	

Child ID _____ Child Description

Learning Target 1:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 2:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 3:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	
Learning Target 4:	C H	C H	C H	C H	C H	C H	C H	C H	C H	C H	

Figure 2. Embedded Instruction Observation System–In Situ (EIOS-I) coding form.
Note. C = observed a complete learning trial that involved an antecedent, correct behavior, consequence; H = observed an additional help complete learning trial that involved error correction procedures and either feedback or a consequence.

behavior specified was observable and measurable and the ways in which the learning target behavior was to be performed were specified (e.g., using single words to make a request, identifying colors by saying the name of the color). Priority learning targets that were not used in this study were those that did not specify an observable or measurable behavior or did not describe how the child would demonstrate the behavior (e.g., identify letters without specifying if the child would point to letters vs. say letter names).

Sixty-nine percent (83 of 120) of priority learning targets submitted by the 13 teachers in the present study were observable and measurable. The median and mode number of observable and measurable priority learning targets per observation was 6. The number of submitted observable and measurable targets per teacher ranged from 3–11, with some teachers having only a single child with observable and measurable learning targets and other teachers having all three children with observable and measurable learning targets. After determining whether a priority learning target was observable and measurable, the project coordinator for the larger study created a Learning Target Consensus Form (LTCF) for all observable and measurable priority learning targets. The LTCFs provided additional information about the target behavior and reminders about decision rules, derived from the EIOS-R coding manual (Snyder et al., 2017), to use while coding using either recording method.

EIOS-I coding and materials. Before EIOS-I coding, the first author reviewed the LTCFs and added all observable and measurable priority learning target behaviors for each target child in the classroom to an EIOS-I coding form. The prepopulated EIOS-I coding forms and LTCFs were printed and brought to the observation along with a clipboard, pencil, and smartwatch with a timer. The first author coded each teacher's implementation of CLTs across all observable and measurable priority learning targets using pencil-and-paper. Continuous event behavior sampling was used for in situ coding with trial frequency recorded within 10-min blocks. In situ data were gathered by tallying frequencies within blocks to reduce errors when yoking CLTs coded using the EIOS-I to the timestamped consensus-coded CLTs obtained using the EIOS-R.

When EIOS-I coding began, the observer started a 10-min timer on a cell phone using the Multitimer App®. To ensure the same length of observation was observed across video and in situ methods, the in situ observer only coded when the videographer was filming, which involved a signaling system used by the data collectors. Every 10 min, the observer was notified that the block had ended via a vibration on a Bluetooth® paired watch and switched to a new coding sheet. When a block ended before 10 min due to the teacher or children leaving the classroom, EIOS-I coding ceased, and the observer noted the duration of the block. Given these procedural decisions, in addition to prepopulating the priority learning targets onto the recording forms, the first author recorded (a) the block number, (b) the activity in which trials occurred, (c) when each block began and ended, and (d) any CLTs that occurred during the block. Collecting this information provided additional context when aligning EIOS-I blocks to the corresponding 10-min video block.

After returning from the in situ observation, the EIOS-I observer entered EIOS-I data into an Excel® spreadsheet to transfer the CLT frequency data from paper to electronic files. All data from the in situ forms were entered into the spreadsheet, including block number, activity type, start and stop observation time, and CLT frequency data. Only two differences existed between the pencil-and-paper data collection forms and the spreadsheet data: (a) frequencies of CLTs within the block were summed when transferred to the spreadsheet, and (b) data from all blocks were recorded on one sheet rather than using a separate sheet for each block.

EIOS-R coding and EIOS-R consensus coding. For viewing and coding using the EIOS-R, two observers trained to agreement standards in the larger study independently coded video observations using the EIOS-R at least 30 days after in situ coding was completed. The 30-day minimum was specified because one observer also conducted the EIOS-I observation. The EIOS-R observers used the same LTCFs as used during the EIOS-I observation. After each EIOS-R coder completed the observation, consensus coding meetings were held.

Consensus coding for the EIOS-R was used to quantify CLTs for the present study to address potential observer bias given that it was not feasible to have a second coder present during the in situ observations. Three individuals were involved in consensus coding: the two EIOS-R coders and the project coordinator for the larger study who was involved in the development of and observer training for the EIOS-R. The project coordinator conducted the consensus meetings to ensure objective and consistent decisions were made regarding which video-coded trials were CLTs. During consensus coding meetings, all three individuals watched and discussed trials identified by either of the two EIOS-R coders. All CLTs coded by either EIOS-R coder were reviewed regardless of whether the EIOS-R coders agreed or disagreed on CLT occurrence. After watching trials and jointly identifying trial components, each trial was classified as either a consensus-coded CLT or not a CLT. Interobserver agreement was calculated between the first author's EIOS-R coding and the EIOS-R consensus-coded data for the frequency of CLTs using a point-by-point method. The average IOA between the first author's EIOS-R coder and EIOS-R consensus-coded observations was 83.40% (SD = 18%) and ranged from 53%–100%.

Data Analysis

To address the research questions focused on method comparability, we compared the recorded frequency of CLTs coded in situ using pencil-and-paper to the recorded frequency of CLTs coded using video and computer-assisted observational software. Descriptive statistics were calculated, a dependent-measures t test was conducted to compare the means from the two different viewing and recording methods (i.e., the EIOS-I and EIOS-R), and a Pearson product-moment correlation was used to examine the relationship between the two methods.

To address the research questions focused on IOA, we compared EIOS-R consensus-coded data and EIOS-I data using three different agreement calculations. Two occurrence agreement calculations and one non-occurrence agreement calculation were used. The first agreement calculation was overall proportion agreement, which examined IOA for the total observation across the two methods. Overall proportion agreement was calculated by summing the number of CLTs for each method and then dividing the smaller number of CLTs observed by the larger number of CLTs observed.

For the second agreement calculation, correspondence agreement, we determined when an EIOS-R consensus-coded CLT mapped onto an EIOS-I in situ coded CLT during a corresponding 10-min block. By recording EIOS-I data using event frequency counts within a 10-min block, we were able to examine trials coded by both methods using the same unit of time (within 10-min blocks). To calculate correspondence agreement, we matched each EIOS-I 10-min block to the corresponding 10-min video block. Occurrence agreement was examined within each 10-min block by identifying CLTs coded using each method and aligning them within the same 10-min block. The purpose of comparing agreement within 10-min blocks was that it would likely produce a more conservative agreement estimate than proportion agreement by providing information about whether the same trials were coded across methods, particularly when compared with the overall proportion agreement metric.

The third agreement statistic examined was non-occurrence agreement, defined as agreeing a CLT did not occur within a 10-min block. Agreement was 100% when a CLT was not coded for consensus-coded EIOS-R and EIOS-I within a block and 0% when a CLT within a block was coded for one method but not the other. After examining non-occurrence within the block, the data were averaged across the observation.

To address the research question about coding time, we calculated time spent coding using each method and examined differences using a dependent-samples *t* test. Time onset was when the EIOS-I coder began coding, which was marked as the start time on the first EIOS-I data collection form. Time offset was determined using the end time marked on the final EIOS-I data collection form. When the EIOS-I coder was in the classroom but not coding (because the teacher or students left the classroom), this time was included in the calculations. For EIOS-I time calculations, time spent (a) preparing for coding, (b) transferring codes to spreadsheets, and (c) conducting analyses was not included. For time spent coding using the EIOS-R, we calculated the time spent coding in addition to time spent on short breaks taken to reduce fatigue. When a teacher had only one observable and measurable priority learning target, and no breaks were taken, the time to code using video viewing and computer-assisted coding was the length of the video in minutes. If a teacher submitted multiple observable and measurable priority learning targets, coders watched the video one time for each priority learning target submitted by the teacher. For EIOS-R time calculations, the amount of time spent (a) gathering or processing videos, (b) exporting observations, and (c) conducting analyses was not included.

Results

Frequency Comparison

Figure 3 shows the frequency of trials coded by each method during each observation. Across the 13 observations, the frequency of CLTs coded in situ using the EIOS-I form ranged from 0–43, and the frequency of CLTs consensus-coded from video using the EIOS-R and observational software ranged from 0–48. Across the 13 classrooms, the mean frequency of trials coded in situ using the EIOS-I (M = 12.07; SD = 14.58) was greater than the number of consensus-coded EIOS-R trials (M = 9.38; SD = 14.52). This was a statistically significant and noteworthy difference between the two methods, $t(12) = 2.28$, $p = .04$, $d = 0.63$. A statistically significant strong positive correlation was found between the methods ($r = .96$, $p < .001$), with 92% of the variance shared between the two methods.

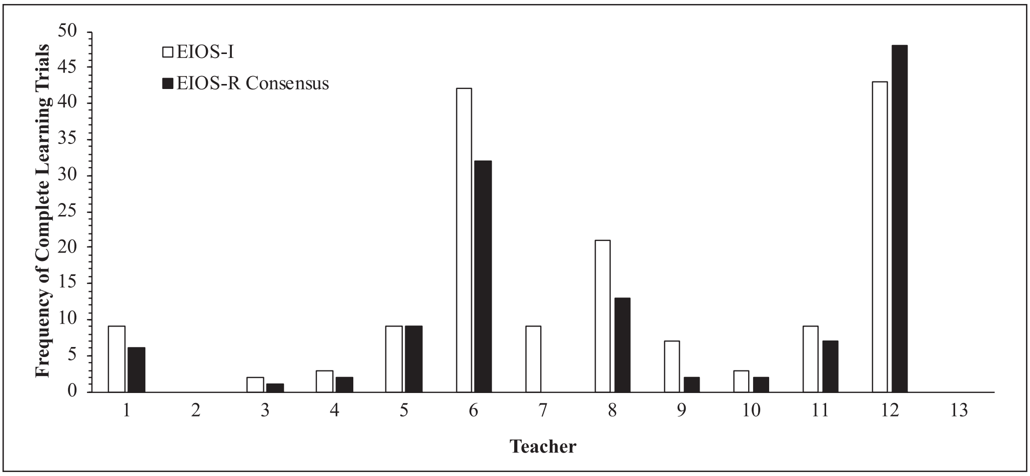

Figure 3. Complete learning trials coded by viewing and recording method.
Note. EIOS-I = Embedded Instruction Observation System–In Situ; EIOS-R Consensus = Embedded Instruction Observation System–Revised Consensus. For Teachers 2 and 13, no complete learning trials were coded using either method. For Teacher 7, no complete learning trials were coded during EIOS-R consensus coding.

Coding Method Interobserver Agreement

Overall proportion agreement between the EIOS-I and the EIOS-R consensus coding ranged from 0%–100% with a mean of 68% (*SD* = 29%). For the teacher who implemented the most trials (n = 43 using EIOS-I and n = 48 using EIOS-R), the overall proportion agreement was 90%. Five teachers implemented three or fewer trials, making a disagreement of one CLT occurrence result in low levels of agreement for three teachers. Video coding decision rules prevented consensus coding of EIOS-R trials for one teacher (Teacher 7) that were able to be coded in situ using the EIOS-I. The inability to code the consensus trials with the EIOS-R resulted in 0% agreement between the methods for this one observation.

Correspondence agreement was calculated within 10-min blocks across the 13 observations. Correspondence agreement between the two coding methods ranged from 0%–100% with a mean of 57% (*SD* = 29%). Of the 181 CLTs coded using either method, 95 trials (52%) were coded in the same 10-min block across the two methods. The most notable difference in an agreement percentage between overall proportion and 10-min block correspondence agreement was for the teacher who implemented the most trials. For this teacher, the overall proportion agreement was 90% and the 10-min block agreement was 49%.

Non-occurrence agreement ranged from 77%–100% and averaged 96% (*SD* = 6.5%) across all 13 classrooms. Except for observations for two teachers where non-occurrence agreement across the two methods was 77% and 88%, respectively, the non-occurrence agreement was above 97% for all observations. Non-occurrence agreement was lower for two observations that were of shorter duration. When disagreements occurred during these shorter observations, the average disagreement was greater due to fewer numbers of blocks.

Differences in Coding Time

On average, EIOS-I coding took 201 min (*SD* = 60 min) with the longest observation lasting 240 min and the shortest lasting 120 min. On average, the length of the video observations was 151 min (*SD* = 32 min). The shortest video was 89 min, and the longest video was 183 min. The difference between the length of EIOS-I coding and the length of video observation was due to the time the EIOS-I observer was in the classroom but not coding (e.g., the class left the room to go to lunch and recess). On average, EIOS-R coding took 304 min (SD = 139.8 min). A two-tailed dependent-samples t test showed EIOS-I coding took less time than EIOS-R coding; $t(12) = -2.80$, $p = .02$, $d = 0.77$.

Discussion

The primary purpose of the present study was to explore the extent to which in situ pencil-and paper-coding and computer-assisted video coding yielded comparable results for quantifying embedded instruction learning trials. Secondary purposes were to explore interobserver agreement data between the two methods and to compare the time spent coding using each method. We addressed these research questions by adapting a direct behavioral observation coding system that used video and computer-assisted scoring to quantify embedded instruction learning trials (i.e., EIOS-R) to a system that used in situ pencil-and-paper coding (i.e., EIOS-I). Adaptations included the viewing method (i.e., from video to in situ) and the recording method (i.e., from computer-assisted to pencil-and-paper). The first author conducted observations of 13 teachers' implementation of embedded instruction learning trials and coded in situ using the EIOS-I, while video data were collected for later consensus coding using the EIOS-R.

A statistically significant difference in the frequency of CLTs was found between the two coding methods. This finding differs from Johnson and Bolstad (1975), who found no difference in absolute counts of behaviors of interest, but replicates Gridley et al. (2018) and Kent et al. (1979). These latter authors found absolute differences in count data for select observed behaviors across coding methods. Given the t test involved a direct comparison of the number of CLTs observed using each method, it was the most conservative estimate of agreement. When examining correlations (i.e., rank order of CLT counts and intervals between these counts across the 13 observations), the magnitude of the relationship was positive and strong. Other researchers have found strong correlations when comparing counts of behavior across different viewing methods (Bernal et al., 1971; Gridley et al., 2018).

For the questions focused on interobserver agreement, CLT agreement data for EIOS-I and consensus-coded EIOS-R data were compared for the total observation versus 10-min blocks. In general, agreement across the 13 classrooms was higher when calculated for the total observation than when calculated by 10-min blocks. When examining non-occurrence across the 10-min blocks, the agreement between the two coding methods was consistently high. By coding and calculating agreement within 10-min blocks rather than across the entire observation, the authors gained the ability to calculate agreement more precisely (i.e., within 10-min blocks). In this case, it also provided evidence that comparing agreement data across an entire observation rather than shorter time blocks may inflate agreement data. Methods used in the present study offer a format other researchers might use when calculating IOA for continuous event data in the absence of timestamped data.

For the research question focused on time, observing and recording using the EIOS-R took significantly more time than the EIOS-I. This finding differs from what was reported by Tapp et al. (2006), who found a computer-assisted coding method took less time than a pencil-and-paper coding method. This difference should be considered in the context of the time tracking method used in the present study versus in Tapp et al. For EIOS-I data, time spent preparing for observations (including travel time to the location of data collection), transferring data to spreadsheets, and using statistical software were not included when calculating time in the present study. For EIOS-R data, time spent collecting videos, checking the quality of the audio and video data prior to coding, exporting computer-assisted data, and transferring data to statistical software were not included when tracking time. Tapp et al. included time calculations for preparing data for analysis, conducting data analyses, and conducting IOA analyses. Given the differences in time spent coding in the present study, researchers might consider whether the advantages of video coding (e.g., multiple passes, stop-and-go coding) offset potential limitations (e.g., time) of video coding when determining how to observe the behaviors of interest in their research (Figure 1).

Limitations

One limitation of the present study that was imposed by the in situ pencil-and-paper method, which did not permit comparisons of trial occurrence across the two methods using exact time-stamps. To address this limitation, we used continuous event counts within 10-min blocks to yoke embedded instruction trial occurrence across the two methods. Second, IOA data were not gathered in situ for EIOS-I coding, which precludes statements about the dependability of observations across in situ observers. Due to the combination of limited resources (i.e., personnel), the practicality of scheduling observations, and the feasibility of having three additional staff in preschool classrooms during data collection of the primary dependent measure of the efficacy trial, a second EIOS-I observer was not available for the purpose of gathering IOA data in situ. A decision to create and use EIOS-R consensus coded data for comparison with EIOS-I data was made to address this limitation. Using this method, IOA data between the first author's EIOS-R codes and the EIOS-R consensus coded data were acceptable (i.e., above 80%). Finally, a potential generalization limitation of the present study was the experience of the EIOS-I coder. Before the present study, the first author had 2 years of experience coding trials using the EIOS-R. The generalization of findings from the present study might be limited to trained coders with similar or greater coding experiences as the first author.

Implications

When comparing two methods for collecting direct behavioral observational data on embedded instruction CLTs, we found that the two methods of data collection resulted in a statistically significant difference in absolute count data, but noteworthy associations. The implications of these findings for research might depend on the questions of interest. In situations where the research question focuses on differences in the absolute number of CLTs implemented by practitioners within or across experimental conditions, findings from the present study suggest the two methods might produce different results. If interest is focused on examining associations within or across experimental conditions, findings from the present study suggest the two methods might produce similar results. Because our findings showed strong positive correlations (i.e., associations) between the two coding methods, researchers interested in examining associations might consider in situ coding as a more efficient method. Based on the findings in the present study, researchers should evaluate the relative strengths or limitations of pencil-and-paper in situ coding to those of computer-assisted video coding when determining which observational or recording method offers the utility for their needs.

Strengths and Limitations of Pencil-and-Paper In Situ Coding

More trials were identified using the EIOS-I than the EIOS-R, replicating previous findings that in situ observers code more occurrences of target behaviors than observers using other coding methods (Bernal et al., 1971; Johnson & Bolstad, 1975; Kent et al., 1979). Similar to what others have reported in the extant literature, in the present study, the restrictions in the visibility or audibility of the video data resulted in some trials not being coded using video methods that were coded as CLTs in situ. For example, more than 12 discrepancies within corresponding 10-min blocks were accounted for when consensus coders mapped these uncodeable EIOS-R trials onto

CLTs coded using the EIOS-I. For one teacher (Teacher 7), visibility restrictions accounted for 100% of disagreements. During the observation for Teacher 7, the ability of the in situ coder to adjust her positioning so that she was able to see the antecedent stimuli used during trials was a strength that could not be replicated when using the EIOS-R to code the video. Evidence in the present study suggests benefits for in situ coding of CLTs versus coding from videotaped observations because in situ coders can reposition during the observation to hear and see all components of a trial.

Strengths and Limitations of Computer-Assisted Video Coding

For the present study, coding CLTs in situ required the in situ coder to recall and apply coding decision rules without being able to reference the EIOS-R coding manual, which describes coding decision rules. Computer-assisted video coding permits coders to stop the video and reference coding manuals or decision rules and then resume coding. In the present study, stop-and-go coding was used during consensus coding. During EIOS-R consensus coding meetings, it was common for all three coders to watch the trial several times to determine how to code the trial. The longer time for EIOS-R coding was likely due to the strengths of video viewing methods. When the behavior of interest is more complex (e.g., chained behaviors or response class behaviors), the utility of having video data that can use multiple passes and stop-and-go coding to ensure trials are coded accurately could be more important than coding in situ in a time-efficient manner. Video coding may therefore be a more accurate way to code trials for some behaviors, especially behaviors without clear onsets or offsets. In the present study, EIOS-R video coding permitted coders to focus on one target at a time by using multiple passes. In situ coding required the coder to observe multiple trials when they occurred simultaneously for different target behaviors or different children.

Recommendations for Future Research

Given the initial findings in the present study about the comparability of various viewing and recording methods, additional studies that compare data gathered using different viewing and recording methods are still needed. Future research might explore comparing a computer-assisted recording method in situ with a computer-assisted recording method using videotaped observations. This comparison would permit exact timestamps to be applied to trials across both methods and better yoked comparisons, permitting examination of other influential factors that might affect comparability across the two methods. Studies might also be conducted to examine comparability across different types of learning trial behaviors (Wolery & Hemmeter, 2011). For example, to what degree are the methods comparable when coding discrete behaviors versus response class or chained behaviors? Studies focused on examining IOA for in situ observers who code complex behaviors, such as embedded instruction CLTs, are also needed.

Initial empirical evidence focused on the comparability of direct behavioral observation data for two different viewing and recording coding methods was provided in the present study. Studies such as this one demonstrate a need for additional research to understand further how decisions about viewing and recording methods influence data collection, data integrity, data analyses, and results.

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: The research reported here was supported, in part, by the Institute of Education Sciences, U.S. Department of Education (R324A150006) and by the Office of Special Education Programs, U.S. Department of Education (H325D150079) to the University of Florida. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, U.S. Department of Education or the Office of Special Education Programs, U.S. Department of Education.

ORCID iDs

Mackenzie K. Martin https://orcid.org/0000-0002-4589-9270
Brian Reichow https://orcid.org/0000-0001-8305-3105
Crystal D. Bishop https://orcid.org/0000-0001-8769-9085

References

Bernal, M. E., Gibson, D. M., Williams, D. E., & Pesses, D. I. (1971). A device for recording automatic audio tape recording. *Journal of Applied Behavior Analysis*, *4*(2), 151–156. https://doi.org/10.1901/jaba.1971.4-151

Cohen, E., & Rozenblat, R. (2015). Applied, behavioral, analytic and. . . technological: A current literature review on the use of technologies in behavior analysis. *European Journal of Behavior Analysis*, *16*(2), 178–187. https://doi.org/10.1080/15021149.2015.1085720

Farrell, A. D. (1991). Computer and behavioral assessment: Current applications, future possibilities and obstacles to routine use. *Behavioral Assessment*, *13*(2), 159–179.

Gridley, N., Baywater, T. J., & Hutchings, J. M. (2018). Comparing live and video observation to assess early parent–child interactions in the home. *Journal of Child and Family Studies*, *27*(6), 1818–1829. https://doi.org/10.1007/s10826-018-1039-y

Horn, E., Lieber, J., Li, S., Sandall, S., & Schwartz, I. (2000). Supporting young children's IEP goals in inclusive settings through embedded learning opportunities. *Topics in Early Childhood Special Education*, *20*(4), 208–223. https://doi.org/10.1177/027112140002000402

Johnson, S. M., & Bolstad, O. D. (1975). Reactivity to home observation: A comparison of audio recorded behavior with observers present or absent. *Journal of Applied Behavior Analysis*, *8*(2), 181–185. https://doi.org/10.1901/jaba.1975.8-181

Kent, R. N., O'Leary, K. D., Dietz, A., & Diament, C. (1979). Comparison of observational recordings in vivo, via mirror, and via television. *Journal of Applied Behavior Analysis*, *12*(4), 517–522. https://doi.org/10.1901/jaba.1979.12-517

Kohler, F. W., Strain, P. S., Hoyson, M., & Jamieson, B. (1997). Merging naturalistic teaching and peer-based strategies to address the IEP objectives of preschoolers with autism: An examination of structural and child behavior outcomes. *Focus on Autism and Other Developmental Disabilities*, *12*(4), 196–206. https://doi.org/10.1177/108835769701200402

Lane, J. D., & Ledford, J. R. (2014). Using interval-based systems to measure behavior in early childhood special education and early intervention. *Topics in Early Childhood Special Education*, *34*(2), 1–11. https://doi.org/10.1177/0271121414524063

Martin, M. (2019). *Comparing in situ and video coding to quantify embedded instruction complete learning trials* [Unpublished master's thesis]. University of Florida.

Mudford, O. C., Martin, N. T., Hui, J. K., & Taylor, S. A. (2009). Assessing observer accuracy in continuous recording of rate and duration: Three algorithms compared. *Journal of Applied Behavior Analysis*, *42*(3), 527–539. https://doi.org/10.1901/jaba.2009.42-527

Noldus Information Technology. (2015). *The Observer XT (Version 12.5)* [Computer software].

Powell, J., Martindale, D., Kulp, S., Martindale, A., & Bauman, R. (1977). Taking a closer look: Time sampling and measurement error. *Journal of Applied Behavior Analysis*, *10*(2), 325–332. https://doi.org/10.1901/jaba.1977.10-325

Rapp, J. T., Carroll, R. A., Stangeland, L., Swanson, G., & Higgins, W. J. (2011). A comparison of reliability measures for continuous and discontinuous recording methods: Inflated agreement scores with partial recording and momentary time sampling for duration events. *Behavior Modification*, *35*(4), 389–402. https://doi.org/10.1177/0145445511405512

Skinner, B. F. (1953). *Science and human behavior*. Macmillan.

Snyder, P., Algina, J., Hemmeter, M. L., & McLean, M. (2015). *Impact of professional development on preschool teachers' use of embedded-instruction practices: An efficacy trial of tools for teachers* [Abstract]. https://ies.ed.gov/funding/grantsearch/details.asp?ID=1619

Snyder, P., Bishop, C., Hemmeter, M. L., Reichow, B., McLean, M., & Crow, R., & Embedded Instruction for Early Learning Project. (2017). *EIOS: Embedded instruction for early learning observation system-revised* (Version 1.1) [Manual and training videos, Unpublished instrument]. Anita Zucker Center for Excellence in Early Childhood Studies, University of Florida, Gainesville, FL.

Snyder, P., McLaughlin, T., Sandall, S., McLean, M., Hemmeter, M. L., Crow, R., & Scott, C., & Embedded Instruction for Early Learning Project. (2016). *LTRS: Learning target rating scale* (Version 2.3) [Manual, Unpublished instrument]. Anita Zucker Center for Excellence in Early Childhood Studies, University of Florida, Gainesville, FL.

Suen, H. K., & Ary, D. (2014). *Analyzing quantitative behavioral observation data* [eBook]. Psychology Press. https://doi.org/10.4324/9781315801827

Tapp, J., Ticha, R., Kryzer, E., Gustafson, M., Gunnar, M. R., & Symons, F. J. (2006). Comparing observational software with paper and pencil for time-sampled data: A field test of Interval Manager (INTMAN). *Behavior Research Methods*, *38*, 165–169. https://doi.org/10.3758/BF03192763

Wessel, D. (2015). The potential of computer-assisted direct observation apps. *International Journal of Interactive Mobile Technologies*, *9*(1), 31–40. https://doi.org/10.3991/ijim.v9i1.4205

Wolery, M., & Hemmeter, M. L. (2011). Classroom instruction: Background assumptions, and challenges. *Journal of Early Intervention*, *33*(4), 371–380. https://doi.org/10.1177/1053815111429119

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). *Observational measurement of behavior* (2nd ed.). Springer.