Research in Autism Spectrum Disorders 7 (2013) 1026–1034

Contents lists available at SciVerse ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

A comparison of data collection techniques used with discrete trial teaching

Mitchell T. Taubman, Ronald B. Leaf, John J. McEachin, Sasha Papovich, Justin B. Leaf *

Autism Partnership, United States

ARTICLE INFO

Article history:
Received 3 March 2013
Received in revised form 5 May 2013
Accepted 6 May 2013

Keywords:
Autism
Data
Data collection
Measurement
Time sample
Trial by trial

ABSTRACT

This study was designed to examine the comparative value of three discrete trial teaching data collection techniques: Continuous recording, time sample, and estimation. The data collection was conducted by behavior interventionists while teaching children diagnosed with autism spectrum disorder skills using discrete trial teaching. Utilizing a counterbalanced design, data collection techniques were examined in regards to their accuracy, that is, their correspondence to the independent measurement of a primary observer collecting contemporaneous trial-by-trial data. Also assessed were the relative impacts of the various techniques on efficiency of therapy and rate of children's acquisition. Finally, interventionists rated their preference of and satisfaction with each of the three techniques. Continuous recording was generally the most accurate, although the other methods were accurate to a degree when used by trained recorders to justify their usage in applied settings. Estimation was the most efficient and time sample was the most preferred.

One of the cornerstones of applied behavior analysis (ABA) is the reliance on objective, cotemporaneous observational measurement (Baer, Wolf, & Risley 1968; Cooper, Heron, Heward, 2007). The use of objective measurement has enabled systematic and reliable analysis of treatment procedures. For children with autism spectrum disorder (ASD) this has yielded a wide range of treatment programs, protocols, and procedures that have been demonstrated to produce profound improvement (Lovaas, 1987; Leaf & McEachin, 1999). Data collection tools and methodologies have been refined over the years, and those in current widespread use for capturing important dimensions of behavior include continuous recording (Leaf, Sheldon, & Sherman, 2010), time-interval (Repp, Roberts, Slack, Repp, & Berkler, 1976), frequency (Kamps et al., 1992), duration (Cooper et al., 2007), and ratio (Adelinis & Hagopian, 1999) recording.

One teaching methodology that is commonly implemented with children on the autism spectrum, and that relies heavily on objective data collection, is discrete trial teaching (DTT) (Lovaas, 1987). During DTT, the target behavior or skill is broken down into small instructional episodes called trials. Each trial begins with the teacher providing a discriminative stimulus, then the learner is provided a time-limited opportunity to respond. Finally, the teacher provides feedback to the learner for his or her response (e.g., reinforcement for a correct response and corrective feedback for an incorrect response). An optional fourth step is for the teacher to prompt the learner so that he or she may display the correct response. There is a brief inter-trial interval that separates trials from each other and the length of the interval largely determines the pace of instruction. Within the inter-trial interval or at the end of the instructional session, the teacher records the learner's responses to track his or her skill acquisition. The teacher has a number of options for data recording methodology, which we will now describe.

* Corresponding author at: Justin B. Leaf, PhD, BCBA-D, 200 Marina Drive, Seal Beach, CA 90740, United States. Tel.: +1 562221 8581.
E-mail address: Jblautpar@aol.com (J.B. Leaf).

1750-9467/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.rasd.2013.05.002

One data collection methodology which is commonly utilized is continuous recording, also known as trial-by-trial data (e.g., Leaf et al., 2010). Continuous recording traditionally involves the teacher recording the learner's response immediately after each trial. Although continuous recording is often considered the most complete record of student performance, it poses several challenges. Continuous recording can interfere with the teaching process by reducing momentum as well as interfering with rapport building between teacher and student when recording data (Leaf, Taubman, & McEachin, 2008). Additionally, it can interfere with managing a learner's disruptive behaviors (Leaf et al., 2008). These disadvantages, coupled with the demands of correctly delivering DTT, often make continuous recording difficult and it may be useful to consider alternative ways for teachers to record the learner's performance.

One of these alternative approaches is time sampling, which involves recording the learner's response only during predetermined segments. A segment may be defined in terms of a number of trials or the amount of time spent teaching. For example, a teacher may record a trial or set number of trials at the beginning of the lesson, during the middle, or at the end. Alternatively, during a 5 min teaching session the teacher may decide to record the learner's response on a teaching trial after every 15 s has passed. Therefore, the teacher is required to take less data, but the intention is to obtain a representative sample of the learner's progress during teaching.

Yet another way that teachers have recorded data during DTT is estimation recording, also known as summarization or retrospective entry, at the end of a teaching session (e.g., Leaf et al., 2008). This estimation data methodology involves the teacher providing an estimate of how well the learner preformed, in a quantitative or qualitative sense, across all of the teaching trials without the need for recording data after every single teaching trial or after set periods of time.

Although there have been several studies that have compared different data methodologies when using computer models or when measuring everyday or aberrant behavior (e.g., Green, McCoy, Burns, & Smith, 1982; Harrop & Daniels, 1986; Meany-Daboul, Roscoe, Bourret, & Aheran, 2007), there are limited studies that have compared different data collection techniques during DTT. One study that did compare different data collection techniques was conducted by Cummings and Carr (2009). Cummings and Carr (2009) compared continuous recording to recording data only on the first trial when implementing DTT for six children diagnosed with ASD. The researchers evaluated the two data systems across several programs (e.g., non-vocal imitation, tacts, drawing, and receptive discrimination) across the six participants. The results of this study suggested that overall, first-trial-only data collection resulted in participants reaching mastery quicker. However, continuous recording resulted in stronger maintenance. One limitation of this study was that all researchers had extensive experience (e.g., over 10 years) with behavioral intervention; therefore, it is not known what the results would be for less experienced teachers. Furthermore, additional studies using different teachers and students are warranted prior to claiming that one measurement system is superior.

Because the literature comparing different data systems is so limited, the present study was designed to compare three data collection techniques used in support of DTT. All data, across all methods, were collected using paper and pencil. Continuous recording, time sample, and estimation data collection methods were compared in terms of their accuracy, efficiency, and observer satisfaction when used to measure student performance during DTT sessions.

1. Methods

1.1. Participants

Adult participant: The researchers evaluated the ability of three behavior therapists to implement DTT and take data across three data systems throughout the course of the study. The therapists all had at least a bachelor's degree in psychology or in a related field and had been working in the field of applied behavior analysis an average of 4 years. Therapist 1 was a 25-year old female who held a Bachelor Degree; she had 3 years of experience in the field of ABA and 36 months working for the current agency. Therapist 2 was a 26-year old male who held a Bachelor Degree; he had 4 years of experience in the field of ABA and 30 months working for the current agency. Therapist 3 was a 29-year old female who held a Masters' Degree; she had 5 years of experience in the field of ABA and 16 months working for the current agency.

Child participant: In order to assess the behavior therapists' use of the measurement systems during intervention, the researchers initially selected three children diagnosed with ASD to participate in the study. During the final phase of the study one of the participants (child #3) dropped out of the study and, therefore, the researchers replaced him with a fourth participant (child #4) who was similar in age, and functioning level, and had similar skills. The researchers selected participants with a wide age range and capabilities to include a representative sample of children who are on the autism spectrum. All children were receiving behavioral services prior to and during this study from a private agency that provides ABA intervention to children and adolescents diagnosed with ASD. Table 1 provides demographic information on the four child participants in the study.

1.2. Setting

This study took place at a private agency that provides behavioral services to children and adolescents diagnosed with ASD. The researchers implemented the study in a small therapy room measuring 2.7 m by 2.8 m. In the room there was a table and two chairs, reinforcing items, a stopwatch, clipboard, data sheet and writing materials. Two to four research sessions were conducted per child per day, two days a week. Research sessions lasted 3 min each and the number of trials per

Table 1
Child participant characteristics.

Participant	Gender	Diagnosis	Age	Vineland adaptive behavior composite score
1	Male	Autistic disorder	6 years 9 months	70
2	Male	Autistic disorder	2 years 11 months	80
3	Male	Autistic disorder	3 years 6 months	58
4	Male	Autistic disorder	3 years 6 months	60

session were varied (see below). During a research session the child participant, the adult participant, and one researcher (fourth author) who collected and scored data as the primary observer were present in the room.

1.3. Dependent variables and measurement

Accuracy of data collected: During each session of the study, across all measurement conditions (described below) the researcher independently scored the response of the child participant during each teaching trial. Contemporaneous in vivo recording was utilized so that the conditions under which the researcher collected data would generally correspond to the conditions experienced by the adult participants. On a trial-by-trial basis the researcher recorded the child's responses as either correct, incorrect, or prompted. The instruction provided to the child and the number of trials conducted per session was also recorded. During each session, the adult participant also recorded data according to the data system being evaluated (described below). Thus, the primary measure of this study was the correspondence between the researchers' data (trial by trial) record and that of the adult participants, under the three different data systems being evaluated. This method corresponds to the methodology for determining accuracy suggested by Cone (1998).

Efficiency measures: The second measure evaluated in this study was the instructional efficiency of the three data systems. To do so, each data system was compared as to its impact on the rate of the child participant's skill acquisition. That is, for each technique, across all adult participants and child participants, the number of trials per three minute session as well as the number of sessions required for mastery (80% correct responding) of the target skills were determined. Comparisons between the techniques on these efficiency measures were then made.

Satisfaction measures: A third area that was evaluated in this study was the adult participants, satisfaction with the three measurement techniques, both prior to and subsequent to the initiation of the study, as assessed by questionnaires. These questionnaires probed the adult participants' relative preference for the practicality, efficiency, benefit, usefulness and accuracy of the various data systems. That is, for each question the participant ranked the measurement techniques (first, second, and third). Further, subsequent to the study, the adult participants completed a social validity questionnaire rating each technique on such dimensions as practicality, efficiency, benefit, usefulness, and accuracy). This survey consisted of nine different questions on a five-point Likert scale.

Child participant targeted behaviors: All but one of the child participants worked on the same skill area, gestures, throughout the three rounds (see below for definition of a round) of the counter-balanced (see Section 1.7 below) design (one child worked on another target area, as noted below, during the third round). The child participants learned to either non-verbally produce gestures (e.g., palm facing out with arm extended for "stop"; shoulders raised with arms bent at the elbow and palms upturned for "I don't know") when instructed to demonstrate the gesture or to expressively identify a gesture when it was presented to him (e.g., "That means stop"). Numerous, simple gestures were identified for each child participant through consultation with treatment agency staff and these were randomly assigned to the study's various conditions. Each gesture was specifically operationalized in written form for the adult participants. Two gestures were randomly assigned to each adult participant for each child participant per round. This meant that each child participant had six total instructional targets across all three adult participants in each round. In the last round of the study, because he had learned all relevant gestures, the targets were changed for one of the highly capable child participants to simple problem solving (e.g., "What do you do when your mom is late picking you up") and "why" questions regarding associations (e.g., "Why does an umbrella go with rain"). Again, for this child participant, new targets in the third round were randomly assigned to data collection conditions and therapists.

1.4. General research session overview

Across the three conditions (described below) the adult participant implemented DTT as described by Leaf and McEachin (1999). A teaching trial in this study included: (a) the adult participant delivering a discriminative stimulus (e.g., "Show me 'stop'"); (b) a brief period of time (3–5 s) for the child participant to respond to the instruction (e.g., demonstrating the appropriate gesture); (c) the adult participant providing reinforcement (as determined by informal assessments and held constant across all conditions) or corrective feedback to the child participant contingent upon his response; (d) the adult participant providing and fading prompts as necessary to the child participant using a flexible prompt fading strategy (Soluaga, Leaf, Taubman, McEachin, & Leaf, 2008); and (e) a brief inter-trial interval (lasting approximately 5 s), during which the adult participant made necessary treatment decisions and recorded the child participant's response as dictated by the different data systems (described below). Although it is likely that each adult participant utilized slightly different styles in the delivery of the discrete trial method, the design used (see below) controlled for these potential differences and variations.

In this study, there were three distinct data collection conditions (i.e., continuous recording, time sample, and estimation data). During each research session, the adult participant worked with the child participant, providing DTT (described above) on pre-selected targets (described above) and measured the child participant's performance utilizing one of the three measurement systems, predetermined under the study's design (described below). Only one of the conditions was implemented by the adult participant during a research session and each research session lasted a total of 3 min.

Concurrent with this, a researcher (primary observer), independently and simultaneously recorded data on every teaching trial utilizing a trial-by-trial data system (as described above). The primary observer indicated when each three-minute session began and ended and insured that the indicated measurement technique (see below) was implemented. The adult participant continued to work on a specific target with a child participant across sessions until 80% correctness (as measured by data taken by the primary observer across all three conditions) was achieved within a session. Only one skill was taught per session; a condition ended for a particular child participant when mastery was achieved on both skills assigned to that condition in that round.

1.5. Staff training

In the initial part of the present study, but prior to work with children and collection of data, the researchers trained the adult participants on the three data systems (i.e., continuous recording, time-sample, and estimation data). Training consisted of both didactic training (i.e., discussions of the three measurement systems) and scoring various samples of DTT sessions from videotapes. The researchers and the adult participant scored the videotapes independently but simultaneously and the researcher considered the adult participant to have mastered a given data system once there was at least 80% agreement between the researcher and the adult participant on three consecutive segments. The researchers (one served as primary observer and another served as reliability observer, see below) also scored the videotapes independently but simultaneously and they were considered to have mastered the continuous data system once there was at least 80% agreement between them on three consecutive segments. The researcher who was the primary observer, had seven-years experience in ABA and autism (including 2 years with the agency that served as the setting for the study; responsibilities included treatment and research.). The researcher who served as the reliability observer, had 15 years of experience in ABA, 10 of which were in autism (8 of those with the agency that served as the setting of the study; responsibilities included treatment, research, supervising and mentoring).

1.6. Conditions

Continuous recording: During this condition the adult participant implemented DTT as described above. Utilizing this measurement technique, immediately upon completion of a trial, the adult participant recorded information on the child participant's performance. This occurred subsequent to each and every trial, that is, on a trial-by-trial basis. The adult participant recorded the instruction provided to the child, whether a prompt was utilized and whether the child's performance was correct or incorrect. Concurrently, the researcher scored the same information, contemporaneous to the discrete trial teaching, for all trials.

Time-sample: During this condition the adult participant implemented discrete trial teaching as described above. In the time-sample condition, prior to the initiation of the session, the primary observer randomly selected a single time point occurring during the three minutes (e.g. 90 s into the session) designating the moment at which measurement would occur. That is, during the three minute session, when that time point occurred as indicated on a timer, the adult participant would score the next full trial occurring after that time point. Only that trial would be scored as a sample of that session. The adult participant would score that trial immediately subsequent to the trial's completion, during the inter-trial interval. For the sampled interval, the adult participant recorded the instruction provided to the child participant, whether a prompt was utilized and whether the child participant's performance was correct or incorrect. For the entire duration of the session, the researcher scored the same information, contemporaneous to the discrete trial teaching, for all trials (just as was done in the continuous measurement condition). Additionally, in the time sample condition, the researcher also noted on her score sheet which trial was the sampled one.

Estimation data: During this condition the adult participant implemented DTT as described above. However, in the estimation data condition, the adult participant recorded data only at the end of the three-minute session. No data (continuous, sample or otherwise) was taken by the adult participant during the three minutes of the session. Instead, at the session's end the adult participant recorded, based on recollection, information in summary fashion. That is, at session's end the adult participant recorded an estimation (again, a retrospective summary recorded subsequent to the completion of the instruction) of how many trials occurred during the session and an estimation of the percentage of those trials that were prompted. The adult participant also recorded an estimation of the percentage of trials in which the child participant's performance was correct (independent of prompts). For example, once the three-minute session ended, the adult participant might have recorded that twelve trials occurred, that 35% of them contained prompts, and that the child participant was correct on 20% of the trials. As in the other measurement conditions, the researcher concurrently scored every trial during the session, including the instruction provided to the child, whether a prompt was utilized and whether the child's performance was correct or incorrect.

Table 2
Representation of the counter balance design.

	Therapist 1	Therapist 2	Therapist 3
Round 1	*Estimation*	*Time sample*	*Continuous*
	Child 1	Child 1	Child 1
	Child 2	Child 2	Child 2
	Child 3	Child 3	Child 3
Round 2	*Continuous*	*Estimation*	*Time sample*
	Child 1	Child 1	Child 1
	Child 2	Child 2	Child 2
	Child 3	Child 3	Child 3
Round 3	*Time sample*	*Continuous*	*Estimation*
	Child 1	Child 1	Child 1
	Child 2	Child 2	Child 2
	Child 4	Child 4	Child 4

1.7. Experimental design

The researchers implemented a counterbalanced design (Campbell & Stanley, 1963) to ensure that all adult participants would experience all data systems with all child participants and that the order in which they did so would be balanced (see Table 2). Each adult participant conducted three rounds; in each round a different measurement system was evaluated for a given therapist. The order of the rounds was randomized ahead of time and no adult participant had the same order of evaluating the measurement systems. Each child participant received intervention in each round (except for child participant 3 who did not receive intervention in the final round and child participant 4 who received intervention only in the final round.

The order of rounds for the first adult participant was the estimation data condition, followed by continuous recording and ending with time sampling. The order of rounds for the second adult participant was first the time-sample condition, followed by the estimation data condition, and lastly the continuous recording condition. For the third adult participant the order of rounds was first the continuous recording condition, then the time-sample condition, and finally the estimation data condition. A round was completed when all child participants reached mastery criterion.

1.8. Fidelity considerations

As the independent variable in this study was not a treatment intervention, typical fidelity measures were not taken (e.g., not on the DTT provided which served as a context for the study's measurement analysis). Fidelity of the observational techniques was addressed in a number of ways. Firstly, the accuracy measure is, in a sense, a fidelity measure as well. That is, accuracy is an analysis of the fidelity of a measurement technique to a standard. Conversely, and again, atypically (as noted above), part of the study was in fact concerned with whether some measurement techniques were more highly subjected to disruption of fidelity (accuracy) by their use simultaneous to the provision of instruction. The inter-observer agreement measures (see below) taken between the primary and secondary observers served as a type of fidelity measure of the primary observer's observation in the same way as the accuracy measures did for the adult participants. Further, in terms of the use of the techniques according to the provisions of the study, the simultaneous recording and running of the sessions by the primary observer during every research session meant that the procedures were used according to protocol. That is, the primary observer was always present to make sure the correct technique was used, that 3 min sessions started and ended on time, and that samples were taken when indicated. Data sheets produced in each session served as permanent product evidence of this (contact authors for example of data sheets).

1.9. Interobserver agreement

During 25% of all sessions, across all techniques, rounds, and participants, a second independent observer collected continuous, trial-by-trial data along with the primary observer. Inter-observer agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. Inter-observer agreement was 94%, ranging from 87.5% to 100%.

2. Results

2.1. Accuracy of data collection

We first evaluated the percentage of agreement between the researcher and the adult participant during sessions of discrete trial teaching across three different measurement systems. The adult participants implemented a total of 5298 teaching trials across the three measurement conditions. The overall percentage agreement between the researcher and the

adult participant for the child participants' correct responding across all research sessions was 84% for continuous recording, 72% for time-sample, and 72% for estimation data. An ANOVA was conducted and the differences between the agreement scores were found to be statistically significant ($p < .001$). It should be noted that with such large sample sizes, even small differences are likely to be statistically significant.

Second, the overall percentage agreement between the researcher and the adult participants for the total number of trials during a research session was evaluated for the continuous measurement system and estimation measurement system. For the continuous recording the average percent agreement was 93%; for the estimation data the average percent agreement was 86%. A *t*-test was conducted and the differences between the agreement scores were found to be statistically significant ($p < .05$)

Third, the overall percentage agreement between the researcher and the adult participant for the total number of prompted trials in a given research session was also evaluated for the continuous recording and estimation data. For continuous recording the average percent agreement was 83%; for the estimation data the average percent agreement was 73%. A *t*-test was conducted and the differences between the agreement scores were found to be statistically significant ($p < .03$).

A point by point analysis was done for the adult participants' percentage agreement with the accuracy data of the researcher on instructions, prompts, and correctness for both continuous and time sample (since data at the trial level was available for both). Those results were as follows: for continuous recording the percent agreement on instructions was 87%, for prompts was 83% and for child participant correct responding was 85%; for time sample the percentage agreement on instructions was 96%, for prompts was 96% and for child participant correct responding was 95%. A *t*-test was conducted and the differences between the measurement systems on agreement scores for instructions, prompts, and child participant correct responding were all found to be statistically significant ($p < .006$, $p < .001$, $p < .004$ respectively).

As a further examination of the correspondence between the data collected utilizing the three measures and the primary observer, correlational analysis were also conducted on the data. Pearson Product Moment correlations were run on the adult participant and primary observer data pairs on the percentage of correct responding by the child participants, across all sessions, for continuous, time sample and estimation methods. For continuous recording, the Pearson correlation was .86; the Pearson correlation for time-sample data was .78; the Pearson correlation for estimation data was .76 ($p < .001$ for each).

2.2. Efficiency measures

The average number of trials conducted per three-minute segment across all rounds was 8.7 for continuous recording, 9.3 for time sample, and 12.2 for estimation data. As an additional analysis of efficiency, we evaluated the total number of 3 min sessions necessary for all children participants to reach mastery criterion. The total number of sessions needed for child participants to reach mastery criterion on all skills taught across all rounds and adult participants was 156 for continuous recording, 152 for time-sample, and 142 for estimation data.

2.3. Satisfaction measure

A summary of the preference and satisfaction rankings provided by the adult participants, in regards to the measurement techniques, both before and after the study, across a number of usage relevant areas, is depicted in Fig. 1. Across the *x*-axis represents the various questions that the adult participants had to answer and along the *y*-axis is the combined ranking for all of the adult participants for each the data collection system across all of the preference and satisfaction rankings. Each panel represents a different data collection system with scores ranging from the most preferred, second most preferred, and third most preferred. Thus, we are able to evaluate change in score for each individual data collection system and across the three data collection systems. It can be seen that in post-investigation measures, continuous recording was ranked first in five areas (as compared to 4 during pre-investigation measures), time sample was ranked first two times (as compared to 1 pre-investigation) and estimation data was ranked first one time (as compared to 3 pre-investigation).

Additionally, we conducted a post-study social validity analysis with all adult participants across the 3 different data collection systems and across 9 different questions. Results of the social validity questionnaire indicate that the time sample technique received an average rating of 4.0, continuous recording received an average rating of 3.7, and estimation data received an average rating of 3.3 (see Table 3).

3. Discussion

This study investigated and compared three different data systems in regards to their accuracy, efficiency, and user preference. On most, but not all measures, the continuous recording was the most accurate, with time sampling less accurate, and estimation data, relatively speaking, the least accurate. On one of the accuracy measures, consisting of point-by-point comparison between continuous recording and time sample, the latter was more accurate. It should be noted that anecdotal report indicated that intervention requirements may have disrupted the more labor-intensive continuous data collection techniques in some cases, affecting accuracy. In measures designed to examine the efficiency of the three techniques, the estimation data technique prevailed as the most efficient. That is, more trials could be performed within a

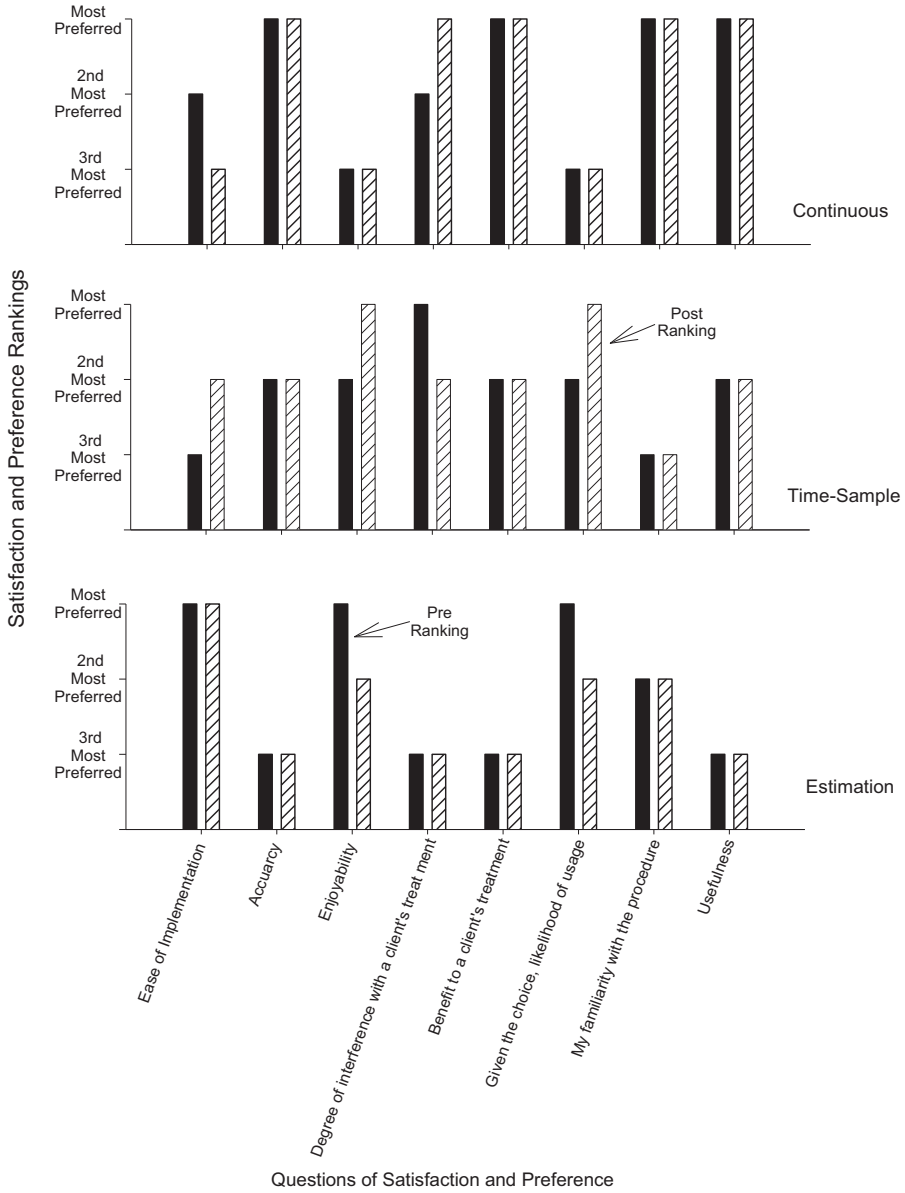

Fig. 1. Satisfaction and preference rankings for the three different data collection systems across all adult (therapist) participants.

session and mastery was achieved more rapidly. The time sample technique was shown to be relatively efficient as well, with continuous recording being least efficient.

In general, post-investigation, continuous recording was the technique most preferred by the behavior therapists (although time-sample was most preferred on an overall measure), followed by time-sample, with estimation data least preferred. It should be noted that when the adult participants were questioned post analysis, the results of the study were not known. Anecdotal inquiry pertaining to the overall scores provided indicated that the adult participants believed that, while estimation data was likely the most efficient, continuous recording was likely most accurate; in their opinion, accuracy was a more critical component for measurement than efficiency. Furthermore, the adult participants felt that time sample represented the best compromise accuracy and efficiency.

Not surprisingly, the data suggests that the various data collection procedures offer different strengths and attributes, which should influence method selection. That is, when accuracy is most critical, such as when precise data may help solve complex interventional challenges or when meticulous documentation is required, continuous recording may be the protocol of choice. Continuous recording may also be indicated when a separate data collector, distinct from the interventionists but able to collect contemporaneous data, is available; however, such arrangements are often not practical in applied treatment settings. In non-complex situations with limited resources, it may be reasonable to use estimation data

Table 3
Social validity results from adult participants.

Item	Continuous data collection	Time-sample data collection	Estimation data collection
The data collection procedure I used was easy to implement.	2.6	4.3	3.3
The data collection procedure I used was accurate.	4.3	3	3
The data collection procedure I used was enjoyable to implement.	1.6	3.6	2.3
The data collection procedure I used could provide me with helpful or useful information.	4.3	4.3	3.3
The data collection procedure I used did not interfere with treatment.	1.6	3.7	3.0
The data collection procedure I used could provide benefits to a client's treatment.	4.3	4.3	4.3
The data collection procedure I used could present disadvantages to a client's treatment.	4.0	1.7	2.7
Given the choice, I would be likely to use this procedure as part of a client's treatment.	4	4	4.7
Overall, I am satisfied with this data collection procedure.	3.7	4.0	3.3

for a portion of measurement efforts, supported by periodic continuous or contemporaneous sample data. Sample data, with such limitations as are noted elsewhere in the literature (Haynes & O'Brien, 2000; Hartmann, Barrios, & Wood, 2004; Tryon, 1998), including the need for cross-sectional representativeness and concerns with interval sampling, offers a balance of accuracy and efficiency when a variety of needs are presented.

It should be noted that a substantial amount of technique training was contained within this analysis, and proficiency in each technique, as well as its resultant accuracy, was substantially impacted by this training. This consideration may be especially applicable to estimation data collection, given its reliance on retrospective analysis. Given that each technique presents with strengths and deficiencies, advantages and limitations, a comprehensive data collection protocol that includes all elements may be most robust. For example, a data collection plan could include estimation data, taken on a regular basis with supportive continuous recording periodically sampled. Such an arrangement would furnish a core of objective, sampled data that could also serve as accuracy standards for the remainder of the data, collected through estimation methods.

While this study does provide implications for practical application, limitations do exist. The study, utilizing a counter balanced type design, was conducted with a small number of participants, children and staff, which restricts its external validity. Further, measurement techniques were largely applied on only one area of instruction. Each method may operate differently when used in other instructional contexts. Further, one subject did not complete the study and another was substituted in the final phase. Training and implementation as conducted in the present analysis may not be applicable or replicable in applied settings (for example, extensive training in estimation techniques may not be practical in some treatment venues). It was also the case that, since experienced staff was utilized, prior training in at least some of the measurement techniques had likely occurred. Finally, data collection in this study was conducted using paper and pencil methods. Data collected using electronic means (such as software or mobile apps) may produce different results.

As the independent variable in this study was the measurement technique utilized, fidelity information was furnished by the accuracy and reliability data generated, the simultaneous presence of the primary observer, and the data sheets, which provided permanent product and supportive information on protocol fidelity. However, as a further potential limitation of this study, fidelity data was not collected on the implementation of discrete trial teaching by the behavior therapists, which could be pertinent to the efficiency analysis.

Future researchers should focus on attempting to replicate the findings of the current investigation across a broader range of instructional arrangements, skill or behavioral targets, and treatment populations, and with inexperienced therapists. Studies should also examine the relative benefits of the techniques when behavioral functioning, rather than skill acquisition, is the measurement target. Further, more specific analyses of the training requirements and advantages and disadvantages of these and other measurement techniques (e.g., whole or partial interval measurement), are still necessary. Future research should additionally be conducted to examine the comparative benefits of the various methods conducted through current, electronic means.

Within a comparative analysis, the present study indicated that three data collection techniques—continuous recording, time sample, and estimation data – when utilized to measure the performance of children with ASD during discrete trial instruction, offer different advantages and disadvantages. Such findings provide implications for application by community based practitioners as well as for potential future research.

References

Adelinis, J. D., & Hagopian, L. P. (1999). The use of symmetrical do and don't requests to interrupt ongoing activities. *Journal of Applied Behavior Analysis, 32*, 519–523.

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.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally College Publishing Company.

Cone, J. D. (1998). Psychometric considerations. Concepts, contents, and methods. In S. Alan, Bellack, & Michael Herson (Eds.), *Behavior assessment. A practical handbook* (4th ed.). MA: Allyn and Bacon, Needham Heights.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Education Inc.
Cummings, A. R., & Carr, J. E. (2009). Evaluating progress in behavioral programs for children with autism spectrum disorders via continuous and discontinuous measurement. *Journal of Applied Behavior Analysis, 42*, 51–71.
Green, S. B., McCoy, J. F., Burns, K. P., & Smith, A. C. (1982). Accuracy of observational data with whole interval, partial interval, and momentary time sampling recording techniques. *Journal of Behavioral Assessment, 4*, 103–116.
Harrop, A., & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. *Journal of Applied Behavior Analysis, 19*, 73–77.
Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In Stephen, N., Elaine, M., *Comprehensive handbook of psychological assessment, behavior assessment* (3). Hoboken, NJ: John Wiley & Sons, Inc.
Haynes, S. N., & O'Brien, W. H. (2000). *Principles and practice of behavioral assessment*. New York, NY: Kluwer Academic/Plenum Publishers.
Kamps, D. M., Leonard, B. R., Vernon, S., Dugan, E. P., Delquardi, J. C., Gershon, B., Wade, L., & Folk, L. (1992). Teaching social skills to students with autism to increase peer interactions in an integrated first grade classroo. *Journal of Applied Behavior Analysis, 25*, 281–288.
Leaf, R., & McEachin, J. (1999). *A work in progress*. New York, NY: DRL Books, LLC.
Leaf, R., Taubman, M., & McEachin, J. (2008). *It's time for school!: Building quality ABA educational programs for students with autism spectrum disorders*. New York, NY: DRL Books, LLC.
Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Comparison of simultaneous prompting and no-no prompting in two choice discrimination with children with autism. *Journal of Applied Behavior Analysis, 43*, 215–228.
Lovaas, I. O. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9.
Meany-Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis, 40*, 501–514.
Repp, A. C., Roberts, D. M., Slack, D. J., Repp, C. F., & Berkler, M. S. (1976). A comparison of frequency, interval, and time-sampling methods of data collection. *Journal of Applied Behavior Analysis, 9*, 501–508.
Soluaga, D., Leaf, J. B., Taubman, M., McEachin, J., & Leaf, R. (2008). A comparison of flexible prompt fading and constant time delay for five children with autis. *Research in Autism Spectrum Disorders, 2*, 753–765.
Tryon, W. W. (1998). Behavior observation. In S. Alan, Bellack, & Michael Herson (Eds.), *Behavior assessment. A practical handbook* (4th ed.). Needham Heights, MA: Allyn and Bacon.