FURTHER INVESTIGATION OF CONTINUOUS AND DISCONTINUOUS DATA COLLECTION METHODS DURING DISCRETE TRIAL TRAINING

A Thesis Presented to the Faculty
of
California State University, Stanislaus

In Partial Fulfillment
of the Requirements for the Degree
of Master of Arts in Psychology

By
Ashley Greer
May 2020

CERTIFICATION OF APPROVAL

FURTHER INVESTIGATION OF CONTINUOUS AND DISCONTINUOUS DATA COLLECTION METHODS DURING DISCRETE TRIAL TRAINING

by
Ashley Greer

Signed Certification of Approval page is on file with the University Library

Dr. Katie Wiskow Assistant Professor of Psychology	Date
Dr. Bruce Hesse Professor of Psychology	Date
Shannon Bianchi Lecturer of Psychology	Date

DEDICATION

I dedicate this to my supportive family. My mother, my husband, and my children have supported and encouraged me throughout this journey. I am thankful to my mother, who educated me and prepared me for my future. She encouraged me to go to college and continue in graduate school. I am thankful to my husband and children for their love, understanding, and sacrifices. They all have inspired me to persevere through this graduate program and complete this research. They continue to support and encourage me every day and I am very grateful to them.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Dr. Katie Wiskow for all of her help and guidance in developing and conducting this research. She has provided her knowledge and support throughout this entire process. I would also like to acknowledge my fellow students and friends in the graduate program, who have helped me stay on track in completing this research.

TABLE OF CONTENTS

PAGE

Appendices

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

ABSTRACT

Previous research suggests that continuous data collection is the most accurate, but generally less preferred regarding efficiency and ease of implementation compared to discontinuous recording methods. Also, the effects of continuous and discontinuous data collection methods on skill acquisition and maintenance has shown mixed results. The purpose of this study was to examine differences across three discrete trial training data collection methods (continuous, first trial only, estimation) on accuracy, efficiency, and maintenance of skills across each data collection method. Behavioral technicians taught three children (ages 26-31 months at the start of the study), who are at risk for autism and received behavioral intervention services, multiple target skills (e.g., one-step directions, conditional discriminations). One third of the targets were measured using continuous data collection, one third of the targets were measured using first trial only data collection, and one third of the targets were measured using estimation. Additionally, we examined the influence of the number of targets (one or three) taught within each session on acquisition. Although continuous data collection was the most accurate, results showed that estimation data collection was close in correspondence to continuous data. Results also showed that participants maintained all mastered targets at the time of the maintenance probes regardless of data collection method.

INTRODUCTION

Several dimensions are typically present in the field of applied behavior analysis, in that our work should be applied, behavioral, analytic, technological, conceptually systematic, effective, and incorporates generality (Baer et al., 1968). Considering the analytic dimension, behavior analysts make judgements about behaviors, treatments, and programs based on analyses of data (Baer et al., 1968; Fiske & Delmolino, 2012). Data collection procedures are necessary for behavior analysts to analyze performance, demonstrate events causing or influencing behavior, and develop or modify behavioral interventions, treatment, and programs. The measurement of behavior to monitor progress is vital in the field of behavior analysis (Lerman et al., 2011). Behavior analysts rely on data to objectively guide the development of treatment, make treatment adjustments, evaluate the effectiveness of specific treatments, compare treatments, and monitor progress (Devine et al., 2011; Lerman et al., 2011). Behavior analysts rely on data collected during skill acquisition programs to determine when skills have been mastered. When teaching new skills, data collected during training is essential for evaluating the progress throughout the training phase. In addition, the data are important to determine when a specific target skill is considered learned, and the extent that the skill will maintain even when training is discontinued.

Data Collection In DTT

Discrete Trial Training (DTT) is a method often used to teach skills to children with autism. During DTT, training is separated into simple, discrete trials. A

trial, in this case, consists of presenting antecedent stimuli, providing prompts as needed, waiting for the learner's response, delivering reinforcement for correct responses, and ending with a brief pause between trials (Mayer et al., 2013). There are generally two types of data collection methods used in DTT: continuous and discontinuous. Continuous data collection refers to trial-by-trial recording of data. That is, the researcher records the participant's response after every trial. Discontinuous data collection refers to recording data on a sample of trials rather than for every trial. There are several types of continuous and discontinuous data collection methods that are used in behavior analysis and DTT.

Continuous data collection methods include recording frequency, duration, latency, and inter-response time (IRT). Frequency recording involves recording how often a behavior occurs within a specified interval. Duration recording involves recording the total time elapsed between the onset of a behavior and the offset of that behavior. Rate refers to the occurrence of a behavior over time (expressed as frequency divided by a standard of time). Latency recording involves recording the time elapsed between the onset of the antecedent stimulus and the onset of the response. Inter-response time recording involves recording the time elapsed between the offset of one response to the onset of the next response. In the context of training new skills, collecting trial-by-trial data involves recording the response (e.g., correct, incorrect, prompt level) immediately after every trial, which is often how data are collected during skill acquisition programs using DTT (Mayer et al., 2013).

Discontinuous data collection methods include momentary time sampling (MTS), whole interval recording (WIR), partial interval recording (PIR), and estimation (Green et al., 1982; Harrop & Daniels, 1986; Mayer et al., 2013). MTS consists of scoring an interval if the response occurred at the moment the interval ended, regardless if the response occurred before or after that moment. WIR consists of scoring an interval if the response occurred throughout the entire interval. If the response did not occur or stopped occurring at any point during the interval, the interval would not be scored. PIR consists of scoring an interval if the response occurred at any point during the interval for any duration of time. Estimation, also referred to as summarization, involves recording an estimate the learner's performance across all trials without recording data after any individual trial (Taubman et al., 2013). Another form of discontinuous data collection in DTT is recording the learner's response only after the first trial, without recording responding during other trials (Cummings & Carr, 2009).

There are advantages and limitations to continuous and discontinuous data collection methods. Continuous data collection, specifically trial-by-trial, is the most accurate method of data collection (Taubman et al., 2013) and can be useful when recording different types of responses during skill acquisition. However, continuous data collection methods can be difficult to use because they require more effort from the data collector, can disrupt the session and behavioral momentum, and may be impractical when managing problem behavior (Meany-Daboul et al., 2007). For example, continuous data collection requires a teacher to complete a trial, stop to

record the data for that trial, and then begin a new trial. When trials are presented in quick progression, behavioral momentum may develop, where the rate of responding of the learner increases. However, when implementers pause and collect the data after each trial behavioral momentum may be disrupted. In addition, the overall number of trials could be reduced when there are frequent pauses for recording trial data. Further, if problem behavior occurs, it may be difficult for the teacher to address the problem behavior while recording data. Therefore, it may be beneficial collect data on a sample of trials to avoid pausing to record the data.

Discontinuous data collection methods can be useful when it is impractical or inefficient to use continuous data collection methods. Discontinuous data collection methods may require less effort or vigilance compared to continuous data collection methods, are more practical to use when trying to reduce disruptions of the session and behavioral momentum, and could allow the teacher to perform more efficiently (LeBlanc et al., 2016; Taubman et al., 2013). However, discontinuous data collection methods do not provide a complete record of performance and are often less accurate. For example, data collected using WIR can underestimate performance because the response must occur for the entire interval to be scored (Fiske & Delmolino, 2012; Green, et al., 1982; Harrop & Daniels, 1986; Meany-Daboul et al., 2007). By contrast, data collected using PIR can overestimate performance because the response only needs to occur for any part of the interval; even a brief occurrence would result in being scored for that interval. MTS may underestimate or overestimate performance because of the restricted time when a response can be scored. This

method can be inaccurate and inefficient when recording low frequency behaviors or recording for only a short duration of time (Meany-Daboul et al., 2007).

It is important to consider the advantages and limitations of each data collection method to select the most appropriate and useful method in a specific situation (LeBlanc et al., 2016). Considering that there are advantages and limitations to both continuous and discontinuous data collection methods, researchers have conducted studies examining how both continuous and discontinuous data collection methods affect different aspects of data representation, interobserver agreement (IOA), skill acquisition, and preference for those implementing.

Comparison of Data Collection Methods

Several studies have compared continuous and discontinuous data collection methods using computer generated data collected through computer programming (Green et al., 1982; Harrop & Daniels, 1986; Rapp et al., 2011; Repp et al., 1976). For example, Repp et al. (1976) used electromechanical equipment to generate pseudo-behavior simulating different rates of responding and compared representation of data across both continuous and discontinuous data collection methods. These data were compared across interval recording, time sampling, and frequency recording. Results indicated that frequency recording was the most accurate and representative of responding. Interval recording and time sampling were similarly less accurate than frequency recording and were not adequately representative of responding.

Similarly, Rapp et al. (2011) used computer generated data and programming to record data and compared the effects of different recording methods on IOA scores. Experimenters calculated IOA scores for the continuous duration recording method, converted those data into a generated discontinuous data set, and calculated the IOA scores based on those discontinuous data. Results showed that IOA scores improved when using the discontinuous data set compared to the continuous data. Based on these findings, experimenters discussed the potential concern with overinflated IOA scores when using less-sensitive data collection methods. These results also highlight the importance of consistency in data collection. It is common in DTT for many different staff to work with a single child, where the experience level of each staff member varies. It would be ideal use a data collection method that is sufficiently accurate, consistent across staff members, and leads to appropriate skill mastery.

The results of Repp et al. (1976) and Rapp et al. (2011) have important implications for how we collect and analyze data in behavior programs; however, the data were generated and collected by computer programs. Therefore, several studies have extended this line of research to applied settings by recording responses of child participants and data collected by experimenters and adult participants (Chezan & Drasgow, 2017; Cummings & Carr, 2009; Giunta-Fede et al., 2016; Lerman et al., 2011; Najdowski et al., 2009; Taubman et al., 2013). Cummings and Carr (2009) compared the efficiency of continuous and discontinuous data collection methods on skill acquisition and subsequent percentage correct during maintenance probes in

children with autism spectrum disorder. Researchers analyzed the data using an adapted alternating treatments design to compare trial-by-trial data collection (continuous) and first trial only recording (discontinuous) for acquisition trials. Each session consisted of 20 trials: 10 acquisition trials and 10 maintenance trials. Researchers set the mastery criteria for each skill as two consecutive sessions at 100% correct responding. Results indicated that participants mastered skills in fewer sessions using the discontinuous measurement system than skills where trials were continuously measured. However, skills associated with continuous data collection were slightly better maintained at the 3-week follow-up probe than skills measured using discontinuous measurement.

Several researchers have conducted replications of the Cummings and Carr (2009) study (Chezan & Drasgow, 2017; Giunta-Fede et al., 2016; Lerman et al., 2011; Najdowski et al., 2009). For example, Najdowski et al. (2009) conducted a similar study, but altered the mastery criteria to be slightly less stringent: above 80% correct responding across three consecutive sessions for the continuous method condition and 100% correct responding across three consecutive sessions for the discontinuous method condition. Sessions consisted of 10 trials of the target skill. Each skill was considered mastered if the criteria were satisfied using the assigned data collection methods for that condition. The results were inconsistent with the results of the Cummings and Carr (2009) study in that the findings indicated that the participants reached mastery more quickly in the continuous recording condition. This suggests that discontinuous recording may be more conservative in regard to

determinations of skill mastery than continuous recording. However, the findings of both studies are not directly comparable because experimenters used different methods and criteria for evaluating skill maintenance. For example, Najdowski et al. (2009) evaluated maintenance after the skill was mastered once per week for three weeks, while Cummings and Carr (2009) evaluated maintenance once, three weeks after the skills were mastered. In addition, both studies included 10 target trials, but Cummings and Carr (2009) also interspersed 10 trials of mastered skills. Finally, the two studies differed in the consequences provided for correct responses during the maintenance sessions. Najdowski et al. (2009) did not provide any program consequences for correct responses, while Cummings and Carr (2009) provided praise for correct responses.

In another study, Lerman et al. (2011) compared all trial recording (continuous), first trial only recording (similar to Cummings & Carr, 2009), and first three trial recording across two session and three session mastery criteria (88% or above) of all trials with a mandatory correct response on the first trial. Mastery was determined based on the assigned data collection method (continuous or discontinuous) for that condition. Sessions were trial-based consisting of eight to nine trials (depending on the participant). Results indicated that (a) first trial only data led to premature skill mastery using the two session mastery criteria, (b) discontinuous data typically underestimated performance (regarding percentage of correct responding) compared to continuous data, and (c) continuous data were more sensitive to changes in the level of performance than discontinuous data. However,

results also indicated that first trial only recording may be useful in predicting performance on subsequent trials of the session and could, therefore, provide a rough estimate of overall performance for continuous recording.

Taubman et al. (2013) also compared continuous and discontinuous data collection methods using time-based sessions, which differs from previous studies that used trial-based sessions (Cummings & Carr, 2009; Lerman et al., 2011; Najdowski et al., 2009). The researchers compared continuous recording, time sampling, and estimation on accuracy, efficiency, and interventionist preference. The continuous recording condition consisted of the trial-by-trial recording, the time sampling condition consisted of using a predetermined time point in the session where the trial immediately following that time point was recorded, and the estimation condition consisted of recording a summary from recollection at the end of the session. Session were time-based, where each session lasted 3 min. Mastery was determined using the continuous data collection method for all target skills, regardless of which condition the skills were assigned. Results indicated that continuous recording was the most accurate, estimation was the most efficient to use, and time sampling was rated the most preferred method by the implementers. Thus, preference for data collection types may be influenced more by effort than accuracy. In fact, Kolt and Rapp (2014) implemented a concurrent-chains procedure and found that participants preferred MTS data collection over PIR because it was easier to use but was not necessarily more accurate.

Summary

Results of previous research indicate that although continuous data recording is accurate, it is generally less preferred compared to other discontinuous methods in terms of efficiency and ease of implementation. Research on discontinuous methods found inconsistent results regarding accuracy. Some findings indicate discontinuous methods underestimate performance (Lerman et al., 2011), lead to premature determination of skill mastery (Cummings & Carr, 2009), or lead to more conservative determinations of skill mastery (Najdowski et al., 2009). In addition, discontinuous methods have been shown to be more preferable to use in DTT by implementers because these methods are typically easier to use and are less disruptive to the teaching session.

Previous studies have used continuous measurement and compared results with various discontinuous methods that differed across studies. In addition, previous studies have had mixed results regarding the comparisons and effects of the different methods on skill acquisition and maintenance. One potential explanation for some of these differences may be in the number of targets per session and condition. For example, Cummings and Carr (2009) included one target per session whereas Lerman et al. (2011) included one to three targets per session. Furthermore, Lerman et al. (2011) extended previous research by collecting data across all training trials in the first-trial only condition for comparison; however, the researchers derived comparisons across data collections from extracted data. Therefore, it would be beneficial to compare methods using on-going data collection, where data collectors

are actively using each data collection methods in isolation. This will permit measurement of the data collection accuracy and data collector preference across conditions. Finally, best practice guidelines suggest introducing multiple acquisition targets simultaneously during teaching (Wunderlich et al., 2014); however, previous research only included one target per session (Cummings & Carr, 2009; Najdowski et al., 2009). Therefore, further investigations into the influence of data collection methods on children's skill mastery and maintenance, and data collector's accuracy and preference are necessary.

Purpose

The purpose of the present study was to extend the research evaluating continuous and discontinuous recording methods by comparing the accuracy of data collector's data using trial-by-trial, first trial recording, and estimation to the researcher's continuous data. We also compared trials to mastery, time to mastery, and maintenance of skill acquisition using DTT.

METHOD

Participants

Adult Participants

Researchers recruited two adult participants who were at least 18 years old from an agency that provides early intervention services to children with autism. These participants were trained in providing one-on-one behavioral intervention using applied behavior analysis and discrete trial training for children and were certified as registered behavior technicians (RBT) with master's degrees in behavior analysis. All adult participants had at least a high school education, experience working with children between 18 months and 7 years old, and experience collecting data on skill acquisition programs. Throughout the manuscript we will refer to adult participants as *behavior technicians.*

Child Participants

Four child participants, between 31 to 35 months old, were recruited from clients receiving services at an agency that provides early intervention services. Child participants had previously worked with or had established rapport with the behavior technicians. We included children who demonstrated the following prerequisite skills: attended to program materials and completed programs on the floor or at the table. Child participants may have had various diagnoses, such as autism spectrum disorder, developmental delays, or social deficits. Throughout the manuscript we will refer to child participants as *participants*

Researchers terminated sessions with two of the child participants during the data collection phase and excluded their data from analyses. We terminated sessions with one of the child participants because she frequently engaged in refusal behaviors during sessions and infrequently displayed prerequisite skills (e.g., attending to the BT and the stimuli) after the pretest. We terminated sessions with a second child participant because the family decided to postpone early intervention services during the COVID-19 pandemic; therefore, we could not continue the research sessions.

Setting and Materials

All sessions took place in the student participants' homes, where the students typically received behavioral services, where only the behavior technician and student were present. Materials for the skill acquisition programs varied depending on the program targets selected for each individual student. Materials for Julian included a table, chair, toy car, plastic bag, and toy pegs. Materials for Charlotte included a table, chair, and 2-dimensional (2-D) picture cards of various animals/insects. The specific skill for each participant varied to meet the individual teaching goals for each student (Table 1). Julian learned one-step instructions. Charlotte learned conditional discriminations of animals and insects. In addition, we used edibles to reinforce correct responses during teaching. For the data collection procedure, materials included a pencil/paper, data collection sheets (Appendix A), and video recording equipment.

Response Measurement

We measured *accuracy* of the data collection method compared to continuous data in the first trial only and estimation conditions. In the first trial only condition we compared the percentage correct of the target(s) to the average percentage of the primary data collector's continuous data. The estimation data was calculated the same as the accuracy of the behavior technician's data.

Efficiency of each data collection method was measured by the trials to mastery for each set of targets, the average duration of each session, and the total time to mastery across conditions. We measured *acquisition* of the target(s) by recording the percentage of correct responses each session. We scored a correct response when the student independently matched the sample to the corresponding comparison stimulus within 5 s of the instruction.

After completing the study, we gave a questionnaire (Appendix B) to the behavior technicians regarding the data collection methods they used during the study. Questions were rated on a five-point Likert scale and pertained to concepts, such as preference, practicality, efficiency, accuracy, and usefulness (Taubman et al., 2013).

Interobserver Agreement (IOA)

A second independent observer collected continuous, trial-by-trial data from video recordings of the sessions. These data collected by the independent observer were compared to continuous trial-by-trial data collected by the primary observer. Inter-observer agreement (IOA) was calculated by number of trials with agreement

divided by the total number of trials and multiplied by 100. We also calculated IOA on the accuracy measure for the data collection comparison. An independent observer compared the continuous data to the data collected by the behavior technician. The percentage of accuracy determined by the independent observer was compared to the percentage of accuracy determined by the primary data collector.

For Julian's 1-target set, we collected IOA for 36% of sessions in the continuous condition, with an average agreement of 96% (range, 92%-100%). We collected IOA for 43% of sessions in the first trial only condition with an average agreement of 97% (range, 92%-100%). We collected for 33% of sessions in the estimation condition, with an average agreement of 99% (range, 92%-100%). For Julian's 3-target set, we collected IOA for 30% of sessions in the continuous condition, with an average agreement of 94% (range, 83%-100%). We collected IOA for 36% of sessions in the first trial only condition, with an average agreement of 100%. We collected IOA for 38% of sessions in the estimation condition, with an average agreement of 100%.

For Charlotte's 1-target set, we collected IOA for 33% of sessions in the continuous condition, with an average agreement of 100%. We collected IOA for 40% of sessions in the first trial only condition with an average agreement of 100%. We collected IOA for 50% of sessions in the estimation condition, with an average agreement of 100%. For Charlotte's 3-target set, we collected IOA for 33% of sessions in the continuous condition, with an average agreement of 98% (range, 86%-100%). We collected IOA for 31% of sessions in the first trial only condition, with an

average agreement of 95% (range, 92%-100%). We collected IOA for 32% of sessions in the estimation condition, with an average agreement of 96% (range, 82%-100%).

Procedural Fidelity

Skill Acquisition Program

A section of the data sheet for the primary data collector included a section for procedural integrity of implementation of the skill acquisition program. The primary data collector indicated on the data sheet if each step of the procedure was followed by marking "Yes" or "No." Marking "Yes" indicated that the behavior technician completed that step accurately. "No" indicated that the behavior technician either did not complete the step or completed the step inaccurately. These procedural steps included: presenting relevant stimuli, delivering the correct instruction, waiting the 5 s for a response, using the prompting procedure correctly, and delivering the appropriate consequences.

Data Collection

A section of the data sheets also included procedural integrity of the data collection method. The primary data collector indicated on the data sheet if the behavior technician used the appropriate data collection method assigned for that session accurately for each trial. Marking "Yes" for a trial indicated that the behavior technician used the appropriate data collection method accurately for that trial. Marking "No" indicated that the behavior technician either used a data collection method other than the method assigned for that session or did not accurately use the

appropriate data collection method for that trial (e.g., collecting data after the trial during the estimation condition). In addition, we recorded the total number of trials and number of trials per target.

For Julian's 1-target and 3-target sets, procedural fidelity of the skill acquisition procedure steps (Table 2) were collected for 43% of sessions in the continuous condition, 50% of sessions in the first trial only condition, and 50% of sessions in the estimation conditions. Procedural fidelity was 100% for the first two steps of the procedure (presenting relevant stimuli and delivering the correct instruction) and fourth step (using the prompting procedure correctly) across all conditions in both sets. The average procedural fidelity for the third step (waiting 5 s for a response) was 99% (range, 92%-100%) for the continuous condition, 100% for the first trial only condition, and 100% for the estimation condition for both sets. The average procedural fidelity for the fifth step (delivering the appropriate consequences) was 99% (range, 90%-100%) for the continuous condition, 99% (range, 87%-100%) for the first trial only condition, and 100% for the estimation condition for both sets.

For Charlotte's 3-target and 1-target sets, procedural fidelity of the skill acquisition procedure steps (Table 2) were collected for 33% of sessions in each condition. The average procedural fidelity for the first three steps of the procedure (presenting relevant stimuli, delivering the correct instruction, and waiting 5 s for a response), was 100% across all conditions in both sets. The average procedural fidelity for the fourth step (using the prompting procedure correctly) was 99% (range, 92%-100%) for the continuous condition, 100% for the first trial only condition, and

100% for the estimation condition for both sets. The average procedural fidelity for the fifth step (delivering the appropriate consequences) was 100% for the continuous condition, 98% (range, 83%-100%) for the first trial only condition, and 100% for the estimation condition for both sets.

Procedural fidelity of the data collection methods (Table 2) were collected for 43% of sessions in the continuous condition, 50% of sessions in the first trial only condition, and 50% of sessions in the estimation conditions for Julian. The average procedural fidelity was 95% (range, 67%-100%) for the continuous condition, and 100% in the first trial only and estimation conditions. Procedural fidelity of the data collection methods was collected for 33% of sessions in each condition for Charlotte. The average procedural fidelity was 99% for the continuous condition, and 100% first trial only and estimation conditions for both sets.

Procedure

Training Session

We trained the behavior technicians to use each of the data collection methods during the pretest phase of the study. We provided both verbal and written instructions for collecting data using the trial-by-trial, first trial only, and estimation methods. After providing the instructions, the behavior technicians observed the researcher model collecting data using each of the methods with an adult model. After observing, the behavior technicians practiced each data collection method, where the researcher role played as the student.

The training sessions consisted of 12 practice trials for each session. The trainer scored the accuracy of using the data collection correctly for each trial but did not score the accuracy of the collected data. For example, during the trial-by-trial data collection method, the trainer scored a trial as correct if the behavior technician recorded a response for that trial immediately after the trial was completed regardless if the response the behavior technician recorded was the response that the student provided. Behavior technicians needed to reach passing criteria of at least 80% for one training session in each condition to complete the training. BT 1 and BT 2 reached mastery criteria after one training session.

Design

A multi-element design was used to compare skill acquisition, accuracy, and efficiency across data collection methods and a multiple probe design across sets was used to demonstrate experimental control of the teaching procedures.

Pretest

We conducted a pretest to assess current skills and probed for the skill acquisition targets for the students. We assessed current skills by implementing a current program for 12 trials and recording the student's attending skills (i.e., looking at the stimuli) and responding to instructions. After determining that the student can attend to stimuli and respond to instructions (at least 75%), we introduced potential targets for the skill acquisition programs to the student to determine if they can currently perform the tasks. The researcher presented each potential target one time. There were no programmed consequences for an incorrect response. Targets that the

student responds to incorrectly were randomly assigned to a condition. If the student responded correctly, the behavior technician delivered brief social praise and then repeated the trial for that target. If the student responded correctly the second trial, the target was excluded from the study. If the student responded incorrectly the second trial (50% correct responding across the two trials), the target remained included in the study. If the student was able to currently perform any of the tasks, new targets were selected and probed until 24 targets had been identified that the student could not currently perform. We taught the student participants a conditional discrimination, but the exact program and stimuli varied based on the student's current skill levels (e.g., matching categories, matching words, matching letters).

Target Assignment

Four targets total were introduced for each condition. Two sets were created for introducing targets for each student. One set consisted of nine targets, with three targets assigned to each condition. This set will be referred to as a 3-target set. Another set consisted of three targets, with one target assigned to each condition. This will be referred to as a 1-target set.

Within 3-target sets, we assigned three targets per condition. One session consisted of 12 trials and each target was presented for four trials in random order. Within 1-target sets, we assigned one target per condition. One session for these sets consisted of 12 trials of one target.

The order we introduced set types was counterbalanced across participants. Julian learned a 1-target set first and Charlotte learned a 3-target set first. When the

scores for all the conditions in the current set reached at least 50% or when at least one of the conditions reached mastery criteria, we conducted baseline sessions for the second set and then introduced teaching with the next set. Both participants exited their early intervention program and data collection was terminated at that time. Termination of data collection prohibited additional replication.

Baseline Probes

To ensure procedural fidelity (i.e., absence of prompting and reinforcement), the researcher conducted all baseline probe sessions. During baseline probes, the researcher presented each target once (3-target set) or one target three times (1-target set) for a total of three trials in each session. The researcher arranged four stimuli in an array for each trial. All targets were ordered randomly. The researcher collected data for these sessions using the trial-by-trial data collection method. The contingencies for correct and incorrect responses were the same as those in the pretest. That is, if the student responded correctly to two out of three (67%) targets during baseline, the targets were still included in the study. If the student exceeded this score during baseline sessions, the target was excluded from the study and replaced with a different target; however, this never occurred. After the baseline sessions for all target sets were completed, teaching began for the first target set. All subsequent target sets were probed again prior to teaching.

During the baseline phase, we conducted an initial multiple-stimulus without replacement (MSWO) preference assessment (DeLeon & Iwata, 1996) consisting of presenting several items (either all edible items or all toys) to the student

and allowing the student to select one during the first trial. For the next trial, the selected item from the previous trial was removed and the student was instructed to pick from the remaining items. This continued until two to three items were selected. These items were used for brief preference assessments prior to beginning each acquisition session. For Julian, edible items included fruit snacks, gushers, and Froot Loops. For Charlotte, edible items included crackers, chips, and cookies.

Acquisition Sessions

For all conditions, behavior technicians implemented the target skills with the student using DTT. Behavior technicians recorded data during all sessions. We randomized the order and sequence of conditions across sessions by randomizing the three conditions for the first three sessions, and then randomizing the conditions for the next three sessions, etc. This ensured that there were no more than two consecutive sessions of a condition. We video recorded all sessions for data collection purposes.

Behavior technicians identified reinforcers through a brief choice trial preference assessment prior to beginning the session and provided edible reinforcement immediately after each correct response. The behavior technicians presented the trial and allowed 5 s for the student to respond. If the student responded correctly, the behavior technician delivered the selected edible to the participant. After delivering the edible, the behavior technician began the next trial. If the student did not respond within 5 s or responded incorrectly, the behavior technician used a least-to-most prompting hierarchy to prompt the correct response and then initiated

the next trial. Contingent on prompted responses, the behavior technician delivered brief social praise. The prompting hierarchy was determined based on target skills for each student and were modified as needed.

Julian. Targets for Julian consisted of one-step instructions with contextual cues. For each trial within a session, behavior technician delivered the instruction (e.g., "sit down") along with the corresponding contextual cue (e.g., pointing to the chair). Julian responded correctly within 5 s (e.g., sitting down in the chair), the behavior technician delivered the edible. If Julian did not respond within 5 s or responded incorrectly, the behavior technician provided a full physical prompt. Contingent on prompted responses, the behavior technician delivered brief social praise.

Charlotte. Targets for Charlotte consisted of conditional discriminations using 2-D picture cards of animals and insects. For each trial within a session, the behavior technician arranged four cards in an array on the table in front of the Charlotte. The behavior technician delivered the instruction, "Touch [target]." If Charlotte selected the correct label card within 5 s, the behavior technician delivered the edible. If Charlotte did not respond within 5 s or responded incorrectly, then the behavior technician began the prompting hierarchy (gestural prompt, positional prompt, and full physical prompt). Contingent on prompted responses, the behavior technician delivered brief social praise. Due to lack of progress, modifications were introduced for Charlotte's research sessions. Beginning on session 16, the behavior technician delivered the instruction, "[Target]," instead of, "Touch [target]" for the

remainder of the research sessions. This modification was added to simplify the instruction and increase Charlotte's attending to the relevant feature of the instruction (target) by removing unnecessary information (Grow & LeBlanc, 2013). In addition, beginning on session 16, the behavior technician began delivering corrective feedback ("no") following incorrect responses along with the appropriate prompt. This modification was added because Charlotte had a learning history with this corrective feedback. During sessions 31-45 the behavior technician delivered the edible contingent on gestural and positional prompted responses, in addition to correct responses, to increase contact with reinforcement. After session 45, Charlotte no longer received the edible for gestural and positional prompted responses.

Mastery criteria for all target skills across all conditions was 100% of correct responding across two consecutive sessions (Cummings & Carr, 2009). Maintenance probes were conducted based on participant availability. The primary data collector conducted probes with the student two days to four weeks after targets were mastered to assess maintenance of the target skills. We considered target skills maintained if the performance of the student was 80% correct responding or higher.

Conditions

Trial-by-trial (continuous). During trial-by-trial (continuous) sessions, behavior technicians recorded the response of the student (correct or incorrect) immediately after completing every trial. For incorrect responses, the behavior technician included the type of prompt used to obtain a correct response. Data sheets included a space to record each response for all trials. During training, we instructed

the behavior technicians to record the response for every trial immediately following each trial. Behavior technicians needed to randomize the order of targets (if applicable to the current set) and write in the target next to the response so that the order of targets was also recorded.

First trial only (discontinuous). During first trial only (discontinuous) sessions, the behavior technician recorded the response of the student (correct or incorrect) for only the first trial of each target immediately after completing trial. For incorrect responses, the behavior technician included the type of prompt used to obtain a correct response. All remaining trials following the first trial for each session were not recorded by the behavior technician. Data sheets for 3-target sets included spaces to score the first three trials only, one for each target. Data sheets for 1-target sets included spaces to score the first trial only. During training, we instructed the behavior technicians to record only the first trial for each target and not record any subsequent trails. Behavior technicians needed to randomize the order of targets (if applicable to the current set) and not write down or take notes regarding the order of targets following the first trial of each target throughout the session.

Estimation (discontinuous). During estimation (discontinuous) sessions, the behavior technician did not record responses of the student for any individual trial. At the end of each session, the behavior technician estimated, based on their recollection of the session, how the student performed by providing a short summary written on the data sheet, including notes on any prompts the behavior technician recalled using during the session, and an estimate percentage of correct responses for each target and

overall. During training, we instructed the behavior technicians to not score any individual trial throughout the session. Data sheets provided a space only for the summary and estimated percentages overall. There were no spaces to record any individual trials or record the order of targets. Behavior technicians needed to randomize the order of targets (if applicable to the current set) and not write down or take notes regarding the order of targets throughout the session.

Maintenance

Two days to four weeks after each condition in a set was mastered, the researcher conducted maintenance probe sessions based on participant availability. Sessions were similar to the skill acquisition sessions, except sessions consisted of six trials. Sessions consisted of two trials of each target during maintenance sessions for 3-target sets and six trials of the target during maintenance sessions for 1-target sets.

The researcher presented each trial and allowed the student 5 s to respond. If the student responded correctly, then the primary data collector delivered brief social praise. If the student did not respond or responded incorrectly, the primary data collector initiated the next trial and did not prompt the student. The researcher recorded maintenance sessions using the trial-by-trial (continuous) data recording method.

RESULTS

Accuracy

Figures 1-2 show the difference in percentage correct for each session for the first trial only and estimation conditions collected by the behavior technicians compared to the continuous data for those sessions collected by the researcher for Julian and Charlotte, respectively. For Julian (Figure 1), the percentage correct for the first trial only condition sessions under- and overestimated performance more than the estimation condition. The percentage deviation in the first trial only condition averaged 25% (range, 7%-41%) in the 3-target set and 55% (range, 0%-83%) in the 1-target set. The percentage deviation in the estimation condition averaged 2% (range, 0%-41%) in the 3-target set and 3% (range, 0%-18%) in the 1-target set.

For Charlotte (Figure 2), the percentage correct for the first trial only condition sessions tended to under- and overestimate performance more than the estimation condition. The percentage deviation in the first trial only condition averaged 17% (range, 3%-33%) in the 3-target set and 14% (range, 10%-17%) in the 1-target set. The percentage deviation in the estimation condition averaged 6% (range, 0%-38%) in the 3-target set and 3% (range, 0%-13%) in the 1-target set.

Acquisition and Maintenance

Julian

The researcher introduced two sets of targets for Julian (Figure 3). Julian mastered the targets in the first trial only condition for the 1-target set first. Trials to mastery for this target were 84 trials. Total time to mastery for this target was 26 min.

The average session length of these sessions was 3.7 min. Julian mastered the target in the continuous condition for the 1-target set next. Trials to mastery for this target were 120 trials. Total time to mastery for this target was 52.1 min. The average session length of these sessions was 4.7 min. Finally, Julian mastered the target in the estimation condition for the 1-target set. Trials to mastery for this target were 216 trials. Total time to mastery for this target was 70.9 min. The average session length of these sessions was 3.9 min. Julian did not reach mastery criteria for any targets in any of the conditions in the 3-target set.

For the 1-target set, the researcher conducted a maintenance probe three weeks, two weeks, and three days after Julian mastered the first trial only, continuous, and estimation condition, respectively. The maintenance probe consisted of six trials, where all six trials were conducted for the single target. Julian maintained the targets in the first trial only and continuous conditions at 100% and the target in the estimation condition at 83% (Figure 3). Research sessions ended when Julian exited the early intervention program. Julian did not master any targets in any of the conditions in the 3-target set before exiting the early intervention program.

Charlotte

The researcher introduced two sets of targets for Charlotte (Figure 4). Charlotte mastered the targets in the first trial only condition for the 3-target set first. Trials to mastery for this condition was 180 trials. Time to mastery for this condition was 104.3 min. The average session length of these sessions was 6.1 min. Charlotte mastered the target in the first trial only condition for the 1-target set next. Trials to

mastery for this condition was 24 trials. Time to mastery for this condition was 5.8 min. The average session length of these sessions was 2.9 min. Finally, Charlotte mastered the target in the continuous condition for the 1-target set. Trials to mastery for this condition was 36 trials. Time to mastery for this condition was 14.6 min. The average session length of these sessions was 4.9 min.

Four weeks after Charlotte mastered the targets in the first trial only condition for the 3-target set, the researcher conducted a maintenance probe consisting of six trials, where two trials were conducted for each target. Charlotte maintained these targets at 100% (Figure 4). Four weeks after Charlotte mastered the target in the first trial only condition for the 1-target, the behavior technician conducted a maintenance probe consisting of six trials, where all six trials were conducted for the single target. Charlotte maintained this target at 100%. Three weeks after Charlotte mastered the target in the continuous condition for the 1-target set, the behavior technician conducted a maintenance probe consisting of six trials, where all six trials were conducted for the single target. Charlotte maintained this target at 100%. Research sessions ended when Charlotte exited the early intervention program. Charlotte did not master targets in the continuous and estimation conditions in the 3-target set, and did not master the target in the estimation condition in the 1-target set before exiting their early intervention program.

Questionnaire

After completing the research sessions, each behavior technician completed a questionnaire where they rated the data collection methods regarding preference,

practicality, efficiency, accuracy, and usefulness (Taubman et al., 2013). Table 3 shows the ratings from both BTs for each of the items of the questionnaire. Overall, BTs reported similar scores for all items of the questionnaire regarding the continuous data collection method including that continuous data collection was accurate, easy to implement, and provided them with useful information. BTs also reported similar scores for some items of the questionnaire for the first trial only data collection method, including that first trial only data collection was easy to implement and provided them with helpful information. However, there were some differences in BT ratings on the remaining items. For items of the questionnaire pertaining to estimation, both BTs reported that they neither agreed nor disagreed that this method presented limitations to treatment. For all other items of the questionnaire regarding estimation, each BT rated the items very differently. Each BT also rated the data collection methods in order of overall preference. BT 1 rated the estimation data collection method as the most preferred, the continuous data collection method as moderately preferred, and the first trial only data collection method as the least preferred. BT 2 rated the continuous data collection method as the most preferred, the first trial only data collection method as moderately preferred, and the estimation data collection method as the least preferred.

DISCUSSION

In the present study, we compared the accuracy and efficiency and effects on maintenance of using three data collection methods (continuous, first trial only, and estimation) and two target sets (1-target and 3-target sets) during skill acquisition. The results of the present study extend the research on data collection methods in four important ways. First, we found that the continuous data collection method was the most accurate, the estimation data collection method was only slightly less accurate compared to the continuous method, and the first trial only method was the least accurate. Previous research findings also found that continuous data collection was the most accurate (Cummings & Carr, 2009; Lerman et al., 2011; Najdowski et al., 2009; Taubman et al., 2013). However, these results are inconsistent with the findings of Taubman et al. (2013) that the estimation condition was less accurate than the first trial only condition. The results of the present study show that estimation was more accurate than first trial only data.

Second, although the discontinuous data collection methods were less accurate than continuous data collection, participants maintained the target skills when mastery was determined using the discontinuous methods for their respective conditions. These results are consistent with the findings of Cummings and Carr (2009) that 66% of the training sets were maintained three weeks following mastery regardless of the type of data collection method. However, these results are inconsistent with their finding that 44% of the trining sets mastered more quickly using the discontinuous data collection method were maintained more poorly than

those collected with continuous data (Cummings & Carr, 2009). The results of the present study showed that all target skills regardless of data collection method were maintained at least 83% for Julian and 100% for Charlotte. Although, it is possible that the differences in maintenance results could be due to the differences in when maintenance probes were conducted.

Third, both participants mastered target skills in the first trial only condition before any other conditions, which is consistent with the findings of Cummings and Carr (2009). Although it is possible that the first trial only data led to premature determination of skill mastery, similar to the findings of Cummings and Carr (2009) and Lerman et al. (2011), this did not seem to have an effect on participant maintenance of the target skills. Furthermore, both participants mastered the target in the first trial only condition in the 1-target set more quickly than the targets in the continuous and estimation conditions and all conditions in the 3-target set. These results are consistent with previous research that found targets were mastered earlier using first trial only data collection compared to target data collected using continuous data collection (Cummings & Carr, 2009). Julian also mastered target in the first trial only condition in the 1-target set in less overall time that the targets in the continuous and estimation conditions in that set. The length of time for sessions in the first trial only and estimation conditions in the 1-target set were slightly shorter on average than the continuous sessions. This could indicate that recording data after every trial could increase the length of the session. Cummings and Carr (2009) also

found that the sessions using the discontinuous data collection method were shorter than sessions using continuous data collection.

Finally, we measured BT preference for each data collection method. The results of the present study indicated that BT 1 highly preferred estimation, moderately preferred continuous, and least preferred first trial only. BT 2 highly preferred continuous, moderately preferred first trial only, and least preferred estimation. Taubman et al. (2013) also measure BT preferences for continuous and discontinuous data collection methods and found that estimation was the least preferred overall. It is possible that BT preferences of data collection methods can vary depending on personal variables and experience with using these methods.

When evaluating discontinuous measurement, it is important to compare to continuous data to determine how the data collection system impacts the efficiency of teaching and data-based decisions. In the context of skill acquisition, mastery criteria are an important consideration. In the present study, we compared the first trial only and estimation data collected by the behavior technicians to the corresponding continuous data for those sessions collected by the researcher. Results showed that overall, both first trial only and estimation data overestimated and underestimated the percentage of correct responses compared to the continuous data. This is somewhat inconsistent with the findings of Lerman et al. (2011) in that they found that discontinuous data overall underestimated responding, but did not find that it also overestimated performance, which were the results of the present study. For both participants, percentage of correct responses collected by first trial only data deviated

from the continuous data more than the percentage of correct responses collected by estimation data. Overall, the first trial only data collection method was the least accurate of the three data collection methods. It is possible that the first trial only method may overestimate and underestimate.

Because Charlotte was learning a conditional discrimination with label cards arranged in an array of four cards, there was a 25% probability she could select the correct card by chance and score 100% (overestimate). Alternatively, she could respond incorrectly on the first trial but correctly during subsequent trials and score 0% (underestimate). However, it is also possible that using this method could be more likely to underestimate in early acquisition and overestimate during sessions in which the participant engaged in more noncompliant or maladaptive behaviors. In early acquisition, it is possible that the child will make an incorrect response for the earlier trials in a session and could begin responding correctly later in the session. Using the first trial only data collection method, the correct responding later in the session are not accounted for and the percentage of correct responding would be underestimated. During sessions where the child is engaging in more maladaptive or noncompliant behaviors, it is possible that the child may initially respond correctly, but then begin engaging in maladaptive/noncompliant behaviors, resulting in incorrect responses, later in the session. The first trial only data, in these cases, would overestimate the child's performance during the session.

The percentage of correct responses collected using estimation also both overestimated and underestimated participant responding but did not deviate as much

from the continuous data as the first trial only data. There were sessions in the estimation condition where the data values were the same as the continuous data values. Overall, the estimation data collection method was more accurate than the first trial only method. The accuracy of data collected using the estimation method may depend on the experience of the behavior technician or data collector. The two behavior technicians who participated in the present study had 5-9 years of experience working in early intervention and implementing DTT. Additionally, both behavior technicians were also RBT certified and had master's degrees in behavior analysis. The length of the sessions and the number of trials may also affect the accuracy of the data collected using the estimation method. In the present study, the sessions in the estimation condition averaged 3.9 min and the behavior technicians conducted 12 trials each session. It is possible that the accuracy of the estimation data collection method may decrease as the length of the session and number of trials increase.

Cummings and Carr (2009) found that discontinuous data collection led to premature determination of skill mastery, according to their mastery criteria of 100% across two consecutive sessions. The results of the present study showed that the targets in the first trial only condition were determined to meet mastery criteria (100% across two consecutive sessions), but did not meet this criteria according to the corresponding continuous data for these targets. In other words, if we had used the continuous data to determine skill mastery, similar to Taubman et al. (2013), then these target skills would not have yet reached mastery at the time they did according

to the first trial only data, which would be consistent with the findings of Cummings and Carr (2009). However, both Julian and Charlotte maintained all target skills in all three conditions at the time of the maintenance probes. This could indicate that the type of data collection method did not impact maintenance of the target skills.

Unlike Cummings and Carr (2009), Najdowski et al. (2009) found that discontinuous data collection led to more conservative determinations of skill mastery. The results of the present study show that neither the first trial only nor the estimation targets for either participant would have reached mastery earlier using the continuous data collection method with the mastery criteria of 100% across two consecutive sessions, which are inconsistent with the findings of Najdowski et al. (2009). However, Najdowski et al. (2009) used a mastery criteria of 80% or above across three consecutive sessions to determine skill mastery, whereas we used a criteria of 100% across two consecutive sessions to determine skill mastery. If for the present study we used the mastery criteria of 80% or above for three consecutive sessions, Julian would have mastered the target in the estimation condition in the 1-target set 144 trials earlier than he did using the mastery criteria of 100% across two consecutive sessions. Julian also would have mastered the target in the continuous condition in the 1-target set 84 trials earlier, and would have mastered the targets in the continuous condition in the 3-target set (which he did not master during the present study) using the 80% or above across three consecutive sessions criteria. Julian was responding correctly at least 80% of the session for several sessions before reaching 100% across two consecutive sessions. Specifically, he was only responding

incorrectly for one or two trials for each of these sessions. We observed that for these trials he would either not respond within the 5 s following the instruction or he would respond incorrectly. Therefore, we could not determine if this was a skill or compliance issue. However, even if we had used the same mastery criteria as Najdowski et al. (2009), the results of the present study are still not consistent in their finding that the discontinuous data collection methods led to a more conservative determination of skill mastery. In the present study, we found that both participants would not have mastered any targets in any condition earlier according to the corresponding continuous data using either mastery criteria (80% or above for three consecutive sessions and 100% across two consecutive sessions). Alternatively, this could indicate that the mastery criteria of 80% or above across three consecutive sessions may be less stringent than 100% across two consecutive sessions, which could account for why the first trial only data led to more premature, rather than more conservative, determination of skill mastery for Julian's 1-target set in the present study. Fuller and Fienup (2018) found that higher mastery criterion resulted in higher maintenance responding. Furthermore, Richling et al. (2019) found that the 80% correct responding across three consecutive sessions was the most commonly used mastery criteria, but their results indicated that it may be insufficient in some cases regarding maintenance of target skills.

Lerman et al. (2011) also discussed that continuous data are more sensitive to changes in performance. The results of the present study suggest that the first trial only data may be less sensitive to changes than continuous data, but the estimation

data were close in its sensitivity to changes in performance compared to the corresponding continuous data. It could be that the estimation data were more sensitive to changes in performance than the first trial only data because the data were more accurate in that they corresponded more closely to the continuous data. The results of the present study support that the continuous data collection method is the most accurate, which is consistent with previous studies (Cummings & Carr, 2009; Lerman et al., 2011; Najdowski et al., 2009; Taubman et al., 2013). However, Taubman et al. (2013) found that first trial only data were more accurate than estimation data, while the results of the present study show that estimation data were more accurate than the first trial only data. The present study had trial-based sessions, while Taubman et al. (2013) had time-based sessions. In the present study, each session was 12 trials; therefore, the BT's could estimate how many correct responses there were out of 12 trials and the range of possible outcomes was always the same. It is possible that with time-based sessions, it can be more difficult for BT's to estimate how many correct responses there were because the number of trials each session may vary. This may have contributed to why the estimation data collection was found to be more accurate in the present study than in the study by Taubman et al. (2013).

There were some limitations within the present study that should be addressed. First, participants exited the program prior to mastering all targets; therefore, we cannot conclusively compare the time and trials to mastery between conditions within the same set. We had a limited time-frame to work with the participants due to the exit criteria for their early intervention program and COVID-

19. Participants exited the early intervention program on their third birthdays and early intervention services, along with the research sessions, were terminated on that date. Also due to the limited time-frame, we were not able to replicate the 1-target and 3-target sets to compare acquisition and mastery of teaching one target and teaching multiple targets. Future research could also further examine differences in mastery and maintenance when teaching multiple targets at one time compared to teaching a single target only. The present study found that teaching only one target resulted in quicker acquisition of initial targets. However, Wunderlich et al. (2017) found that teaching multiple exemplars concurrently resulted in quicker acquisition and a higher degree of generalization. Although the present study did not eventually mix targets from the 1-target set with additional targets, the participants of this study were very early learners with limited repertoires and learning histories. Therefore, it could be beneficial for future research to further examine differences in mastery and maintenance when teaching multiple targets at one time compared to teaching a single target only, especially for early learners with little to no experience with discrete trial training.

Additionally, research sessions were conducted intermittently and inconsistently that may have influenced acquisition. Behavior technicians were only scheduled with the participants once or twice per week, and some sessions with the behavior technicians were cancelled due to participant illness or vacation. The number of sessions the behavior technicians conducted each day also varied depending on how much time was needed for the participant's other early

intervention programming (e.g., assessments, reports, parent training). Furthermore, research suggests that distributed practice leads to better acquisition than mass practice when teaching skills (Haq et al., 2015). Future research could expand this area of research by examining the effects of data collection methods on mastery and maintenance when sessions are conducted more frequently and more consistently.

Another limitation is the varied time between maintenance probes. The researcher conducted maintenance probes based on participant availability and the time after mastery varied from three days to four weeks; however, scheduling issues and participants exiting the program limited when we were able to conduct maintenance probes. Therefore, the level of maintenance of the target skills between conditions should be viewed with caution.

Researchers could extend the research examining differences in maintenance of target skills across data collection methods by conducting several maintenance probes at different lengths of time after determining skill mastery (e.g., maintenance probes after two weeks, then four weeks, then six weeks) to examine effects on maintenance across time. Finally, researchers could examine differences in data collection methods in other behavioral areas, such as different areas of skill acquisition and maladaptive behaviors. It is possible that the best type of data collection may depend on the type of skill/behavior, the method of teaching/intervention, and length of the data collection period. For example, it is possible estimation may be more appropriate for brief sessions with a set number of trials, but may be less appropriate for longer sessions or sessions with a varying

number of trials. It is also possible that the best type of data collection for recording maladaptive behaviors may vary as well depending on the topography and frequency of the behavior and the type of intervention implemented for the behavior.

Accurate and efficient data collection during DTT is essential to monitoring program progress, determining mastery of target skills, and maintenance of those skills. The current study compared continuous and discontinuous data collection methods and found inconsistent outcomes compared to previous research. Therefore, future research should continue to compare data collection methods, with specific considerations for the teaching arrangement (e.g., number of targets, trials), mastery criteria, maintenance, and experience of the therapist.

REFERENCES

REFERENCES

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

Chezan, L. & Drasgow, E. (2017). A comparison of continuous and discontinuous data collection systems on the acquisition, generalization, and maintenance of tacting in three young children with autism spectrum disorder during discrete-trial training. *Evidence-Based Communication Assessment and Intervention, 11*(1-2), 1-8.

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.

Devine, S., Rapp, J., Testa, J., Henrickson, M., & Schnerch, G. (2011). Detecting changes in simulated events using partial-interval recording and momentary time sampling III: evaluating sensitivity as a function of session length. *Behavioral Interventions, 26*(2), 103-124.

Fiske, K. & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice, 5*(2), 77-81.

Fuller, J. & Fienup, L. (2018). A preliminary analysis of mastery criterion level: Effects on response maintenance. *Behavior Analysis in Practice, 11*(1), 1-8.

Giunta-Fede, T., Reeve, S. A., Debar, R. M., Vladescu, J. C., & Reeve, K. F. (2016). Comparing continuous and discontinuous data collection during discrete trial

teaching of tacting by children with autism. *Behavioral Interventions, 31*(4), 311-331.

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.

Grow, L. & LeBlanc, L. (2013). Teaching receptive language skills. *Behavior Analysis in Practice, 6*(1), 56-75.

Haq, S., Kodak, T., Kurtz-Nelson, E., Porritt, M., Rush, K., & Cariveau, T. (2015). Comparing the effects of massed and distributed practice on skill acquisition for children with autism. *Journal of Applied Behavior Analysis, 48*(2), 454-459.

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.

Kolt, L. & Rapp, J. (2014). Assessment of therapists' preferences for discontinuous measurement systems. *Behavioral Interventions., 29*(4), 304-314.

LeBlanc, L., Raetz, A., Sellers, P., & Carr, B. (2016). A proposed model for selecting measurement procedures for the assessment and treatment of problem behavior. *Behavior Analysis in Practice, 9*(1), 77-83

Lerman, D., Dittlinger, C., Fentress, L., & Lanagan, H. (2011). A comparison of methods for collecting data on performance during discrete trial teaching. *Behavior Analysis in Practice, 4*(1), 53-62.

Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2013). Behavior Analysis for Lasting Change, 3rd edition.

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.

Najdowski, A., Chilingaryan, V., Bergstrom, R., Granpeesheh, D., Balasanyan, S., Aguilar, B., . . . Roane, H. (2009). Comparison of data-collection methods in a behavioral intervention program for children with pervasive developmental disorders: A replication. *Journal of Applied Behavior Analysis, 42*(4), 827-832.

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 interval recording and momentary time sampling for duration events. *Behavior Modification, 35*(4), 389–402. https://doi.org/10.1177/0145445511405512

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.

Richling, S., Williams, W., & Carr, J. (2019). The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. *Journal of Applied Behavior Analysis, 52*(3), 701-717.

Taubman, M. T., Leaf, R. B., McEachin, J. J., Papovich, S., & Leaf, J. B. (2013). A comparison of data collection techniques used with discrete trial teaching. *Research in Autism Spectrum Disorders., 7*(9), 1026-1034.

Wunderlich, K., & Vollmer, T. (2017). Effects of serial and concurrent training on receptive identification tasks: A systematic replication. *Journal of Applied Behavior Analysis, 50*(3), 641-652.

Wunderlich, K. L., Vollmer, T. R., Donaldson, J. M., & Phillips, C. L. (2014). Effects of serial and concurrent training on acquisition and generalization. *Journal of Applied Behavior Analysis*, *47*(4), 723-737.

Table 1

Skill Acquisition Targets

Participant	Set	Continuous	First trial only	Estimation
Julian BT 1	1-target set	Sit down	Stand up	Clean up
	3-target set	Let's go Hug Stop	Wait Pick it up Put it down	Blow kiss Shake Roll car
Charlotte BT 1 BT 2	1-target set	Bee	Ostrich	Squirrel
	3-target set	Snail Swan Camel	Buffalo Shark Puffin	Dolphin Ferret Dragonfly

Note: Behavior technician (BT)

Table 2

Average Values for Procedural Fidelity Steps

Participant	Cond.	Stim.	Instruction	Wait 5 s	Prompt	Consequence	Data
Julian BT 1	C	100%	100%	99%	100%	99%	95%
	FT	100%	100%	100%	100%	99%	100%
	E	100%	100%	100%	100%	100%	100%
Charlotte BT 1 BT 2	C	100%	100%	100%	99%	100%	99%
	FT	100%	100%	100%	100%	98%	100%
	E	100%	100%	100%	100%	100%	100%

Note: Condition (Cond.), Stimuli (Stim.), Continuous (C), First trial only (FT), Estimation (E)

Table 3

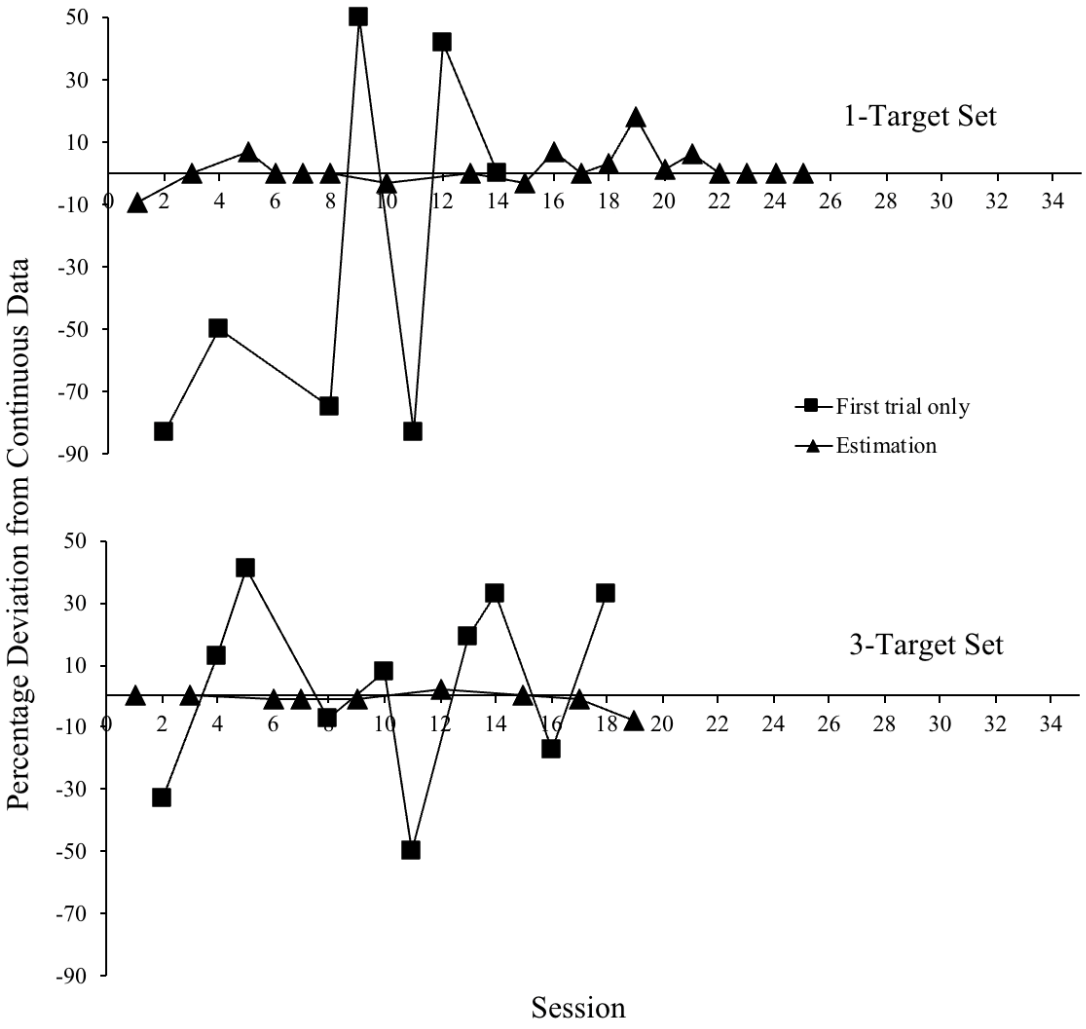
Behavior Technician Questionnaire

Questionnaire Items	BT	C	FT	E
The data collection procedure I used was easy to implement.	BT 1 BT 2	4 5	4 5	5 1
The data collection procedure I used was accurate.	BT 1 BT 2	4 5	3 5	4 2
The data collection procedure I used was enjoyable to implement.	BT 1 BT 2	4 3	4 3	4 2
The data collection procedure I used could provide me with helpful or useful information.	BT 1 BT 2	5 5	4 5	4 3
The data collection procedure I used did not interfere with treatment.	BT 1 BT 2	4 3	4 3	4 2
The data collection procedure I used provides benefits to a client's treatment.	BT 1 BT 2	4 3	4 3	4 3
The data collection procedure I used presents limitations to a client's treatment.	BT 1 BT 2	3 3	5 3	3 3
Given the choice, I would be likely to use this data collection procedure as a part of a client's treatment.	BT 1 BT 2	5 5	2 4	5 1
Overall, I am satisfied with this data collection procedure	BT 1 BT 2	5 5	2 4	5 1

Note: C = continuous data collection method; FT = first trial only data collection method; E = estimation data collection method. Likert scale ratings: 1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree.

Figure 1

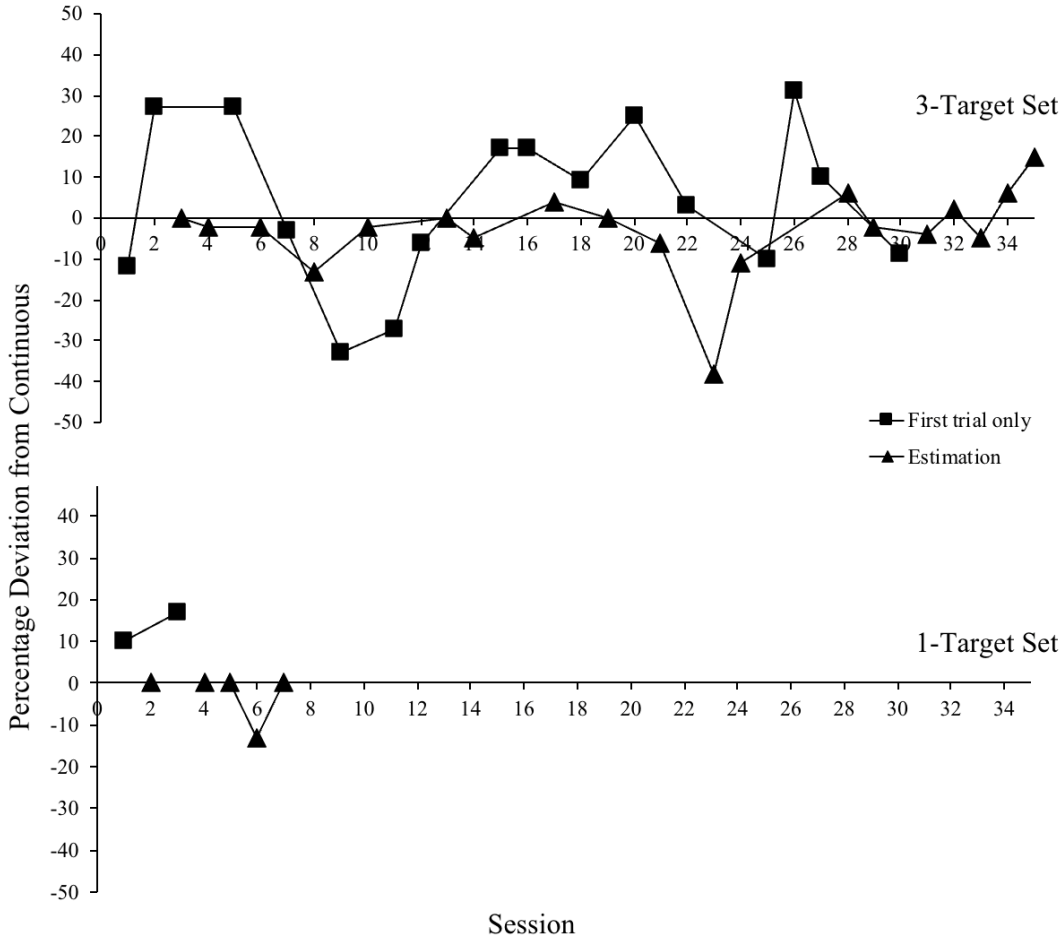
Percentage Deviation of First Trial Only and Estimation Data for Julian



Note: Percentage deviation of first trial only and estimation data for Julian's 1-target and 3-target sets collected by the BT from continuous data collected by the researcher.

Figure 2

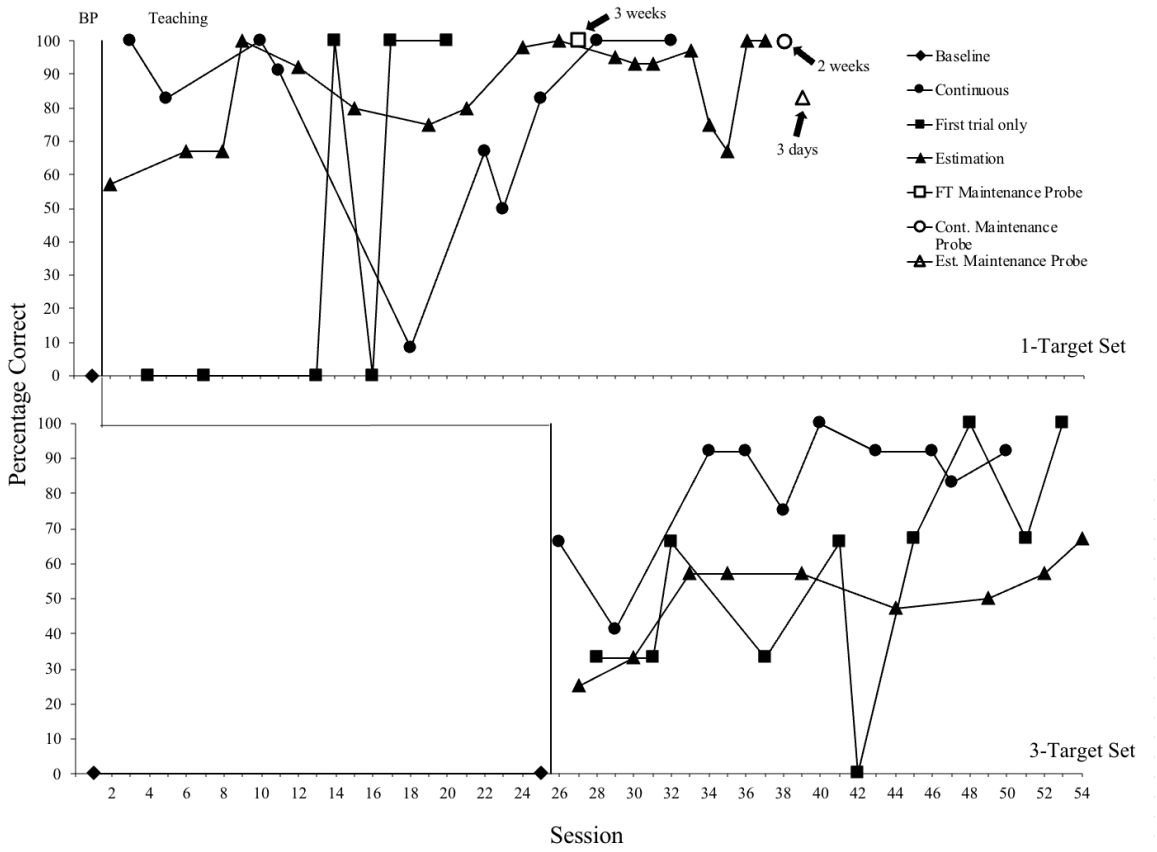
Percentage Deviation of First Trial Only and Estimation Data for Charlotte



Note: Percentage deviation of first trial only and estimation data for Charlotte's 3-target and 1-target sets collected by the BTs from continuous data collected by the researcher.

Figure 3

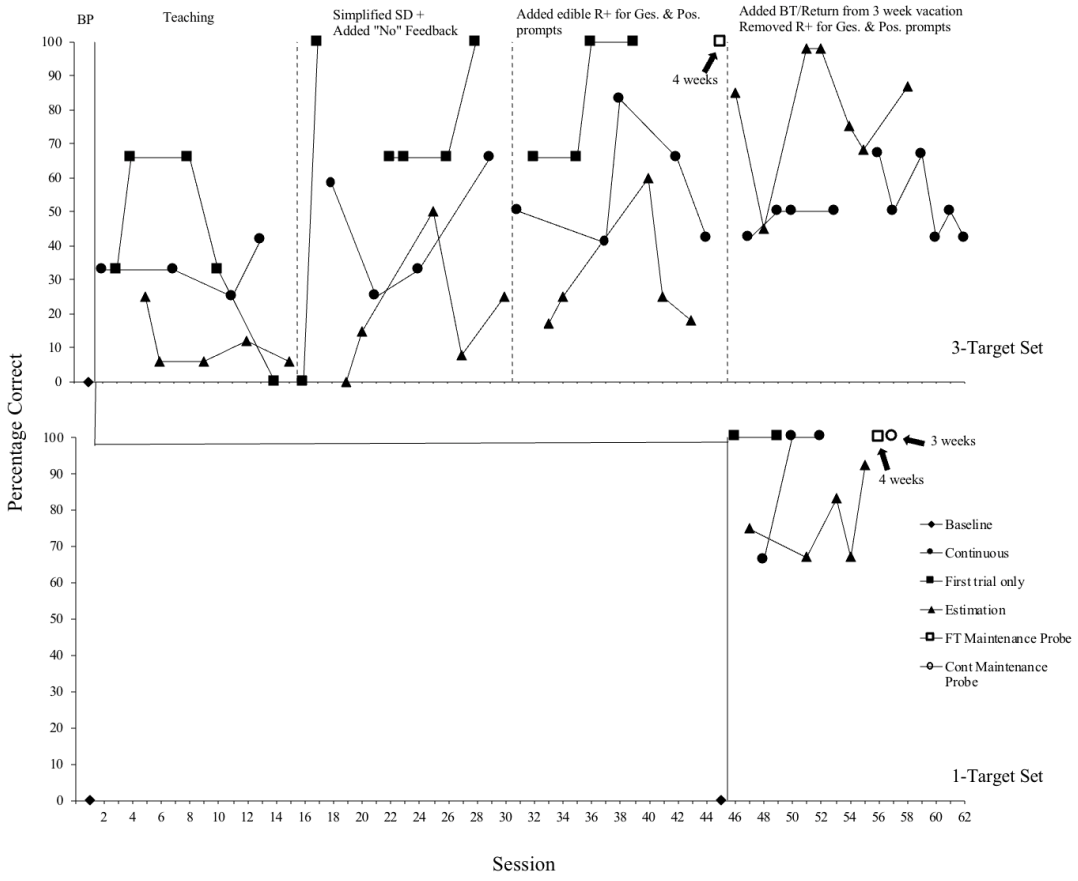
Acquisition Graphs for Julian's 1-target and 3-target Sets



Note: BP = baseline probe; FT = first trial only condition; Cont. = continuous condition; Est. = estimation condition.

Figure 4

Acquisition Graphs for Charlotte's 1-target and 3-target Sets



Note: BP = baseline probe; Ges. = gestural; Pos. = positional; FT = first trial only condition; Cont. = continuous condition.

APPENDICES

APPENDIX A

DATA SHEETS

Trial-by-trial (Continuous) Data Collection

Record data immediately after every trial.

Session duration: Start the timer immediately before you present the first trial and stop the timer immediately after you complete the last trial.

Targets: [list targets]

- For Sets A/C, conduct 4 trials of each target in random order. For Sets B/D, conduct 12 trials of the single target.

"C" = independent (without prompting) correct responses
"I" = incorrect responses.

For incorrect responses, indicate the type of prompt used to obtain the correct response under the Prompt column (e.g., gesture, model, full physical). For correct responses, write "N/A" under the Prompt column.

Set:	Condition: Trial-by-trial	Session #:	Student:	BT:	Date:

Trial	Target	Response	Prompt
1		C / I	
2		C / I	
3		C / I	
4		C / I	
5		C / I	
6		C / I	
7		C / I	
8		C / I	
9		C / I	
10		C / I	
11		C / I	
12		C / I	
TOTAL CORRECT:			
TOTAL INCORRECT:			

Session duration: _____

Notes:

First Trial Only (discontinuous) Data Collection: Sets A and C

Record data immediately after only the first trial for each target (the first three trials, with one trial per target).

Session duration: Start the timer immediately before you present the first trial and stop the timer immediately after you complete the last trial.

Targets: [list targets]

- Conduct 4 trials of each target in random order.

"C" = independent (without prompting) correct responses
"I" = incorrect responses.

For incorrect responses, indicate the type of prompt used to obtain the correct response under the Prompt column (e.g., gesture, model, full physical). For correct responses, write "N/A" under the Prompt column.

Set:	Condition: First Trial Only	Session #:	Student:	BT:	Date:

Trial	Target	Response	Prompt
1		C / I	
2		C / I	
3		C / I	
TOTAL CORRECT:			
TOTAL INCORRECT:			

Session duration: _____

Notes:

First Trial Only (discontinuous) Data Collection: Sets B and D

Record data immediately after only the first trial.

Session duration: Start the timer immediately before you present the first trial and stop the timer immediately after you complete the last trial.

Targets: [list targets]

- Conduct 12 trials of the target.

"C" = independent (without prompting) correct responses
"I" = incorrect responses.

Set:	Condition: First Trial Only	Session #:	Student:	BT:	Date:

Trial	Target	Response	Prompt
1			

TOTAL CORRECT:	
TOTAL INCORRECT:	

Session duration: _____

Notes:

Estimation (discontinuous) Data Collection: Sets A and C

Do not record data for any trial. At the end of the session, please write a short summary (based on recollection) of the student's performance during the session. Then estimate a percentage of correct responses for each target (three percentages total). Write the name of the target and then write the estimated percentage for that target. Repeat for all three targets. Then under the Overall Percentage Column, estimate an overall percentage of correct responses for the entire session (across targets).

Session duration: Start the timer immediately before you present the first trial and stop the timer immediately after you complete the last trial.

Targets: [list targets]

- Conduct 4 trials of each target in random order.

Set:	Condition: Estimation	Session #:	Student:	BT:	Date
Summary:					
Target:		Target:		Target:	
Estimated Percentage:		Estimated Percentage:		Estimated Percentage:	
Overall Session Percentage:					

Session duration: _____

Notes:

Estimation (discontinuous) Data Collection: Sets B and D

Do not record data for any trial. At the end of the session, please write a short summary (based on recollection) of the student's performance during the session. Then estimate a percentage of correct responses for the target. Write the name of the target and then write the estimated percentage.

Session duration: Start the timer immediately before you present the first trial and stop the timer immediately after you complete the last trial.

Targets: [list targets]

- Conduct 12 trials of the single target.

Set:	Condition: Estimation	Session #:	Student:	BT:	Date:
Summary:					
Target:					
Estimated Percentage:					

Session duration: _____

Notes:

Data Collection: IOA/Procedural Fidelity

Please record your name (circle either researcher or observer) and the date.

Record data immediately after every trial.

Session duration: Start the timer immediately before you present the first trial and stop the timer immediately after you complete the last trial.

Targets: [list targets]

- Conduct 4 trials of each target in random order.

"C" = independent (without prompting) correct responses
"I" = incorrect responses.

For incorrect responses, indicate the type of prompt used to obtain the correct response under the Prompt column (e.g., gesture, model, full physical). For correct responses, write "N/A" under the Prompt column

Treatment Integrity (Data Collection): If recording treatment integrity, indicate in this column if the behavior technician used the correct data collection method for each trial. Circle "Y" if data were recorded correctly for the assigned condition. Write "N" if the data were not recorded correctly for the assigned condition. Each trial under this column should indicate either "Y" or "N."

Treatment Integrity (Skill Acquisition Program): For each trial, indicate for each category if the behavior technician performed the step correctly, by circling "Y" if they performed step correctly or "N" if they did not perform step correctly. Record the total number of correct trials for each step (circled "Y") at the end of each column. Record each target and the number of trials for that target during the session (if there was only one target, write in one of these sections and leave the remaining two target sections blank). Record the total number of trials for that session.

Date: Condition: Student: Data Collector:

Session: Set: BT:

Trial	Presented Relevant Stimuli	Delivered Instruction	Target	Waited 5 s	Response	Prompt	Followed prompt procedure	Correct consequence	Correct Data Collection
1	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
2	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
3	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
4	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
5	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
6	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
7	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
8	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
9	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
10	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
11	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
12	Y / N	Y / N		Y / N	C / I		Y / N	Y / N	Y / N
Total:									

Session duration: _____

Target: _____; Number of trials: _____ Target: _____; Number of trials: _____

Target: _____; Number of trials: _____ Notes:

APPENDIX B

QUESTIONNAIRE

For the trial-by-trial (continuous) data recording method, circle the number that best indicates how much you agree or disagree with each statement.

1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree

Trial-by-trial (Continuous)	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
The data collection procedure I used was easy to implement.	1	2	3	4	5
The data collection procedure I used was accurate.	1	2	3	4	5
The data collection procedure I used was enjoyable to implement.	1	2	3	4	5
The data collection procedure I used could provide me with helpful or useful information.	1	2	3	4	5
The data collection procedure I used did not interfere with treatment.	1	2	3	4	5
The data collection procedure I used provides benefits to a client's treatment.	1	2	3	4	5
The data collection procedure I used presents limitations to a client's treatment.	1	2	3	4	5
Given the choice, I would be likely to use this data collection procedure as part of a client's treatment.	1	2	3	4	5
Overall, I am satisfied with this data collection procedure.	1	2	3	4	5

For the first trial only (discontinuous) data recording method, circle the number that best indicates how much you agree or disagree with each statement.

1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree

First trial only (Discontinuous)	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
The data collection procedure I used was easy to implement.	1	2	3	4	5
The data collection procedure I used was accurate.	1	2	3	4	5
The data collection procedure I used was enjoyable to implement.	1	2	3	4	5
The data collection procedure I used could provide me with helpful or useful information.	1	2	3	4	5
The data collection procedure I used did not interfere with treatment.	1	2	3	4	5
The data collection procedure I used provides benefits to a client's treatment.	1	2	3	4	5
The data collection procedure I used presents limitations to a client's treatment.	1	2	3	4	5
Given the choice, I would be likely to use this procedure as part of a client's treatment.	1	2	3	4	5
Overall, I am satisfied with this data collection procedure.	1	2	3	4	5

For the estimation (discontinuous) data recording method, circle the number that best indicates how much you agree or disagree with each statement.

1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree

Estimation (Discontinuous)	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
The data collection procedure I used was easy to implement.	1	2	3	4	5
The data collection procedure I used was accurate.	1	2	3	4	5
The data collection procedure I used was enjoyable to implement.	1	2	3	4	5
The data collection procedure I used could provide me with helpful or useful information.	1	2	3	4	5
The data collection procedure I used did not interfere with treatment.	1	2	3	4	5
The data collection procedure I used provides benefits to a client's treatment.	1	2	3	4	5
The data collection procedure I used presents limitations to a client's treatment.	1	2	3	4	5
Given the choice, I would be likely to use this procedure as part of a client's treatment.	1	2	3	4	5
Overall, I am satisfied with this data collection procedure.	1	2	3	4	5