# COMPARISON OF DATA-COLLECTION METHODS IN A BEHAVIORAL INTERVENTION PROGRAM FOR CHILDREN WITH PERVASIVE DEVELOPMENTAL DISORDERS: A REPLICATION

ADEL C. NAJDOWSKI, VARDUI CHILINGARYAN, RYAN BERGSTROM, DOREEN GRANPEESHEH, SUSIE BALASANYAN, BARBARA AGUILAR, AND JONATHAN TARBOX

CENTER FOR AUTISM AND RELATED DISORDERS

Cummings and Carr (2009) compared two methods of data collection in a behavioral intervention program for children with pervasive developmental disorders: collecting data on all trials versus only the first trial in a session. Results showed that basing a child's progress on first-trial data resulted in identifying mastery-level responding slightly sooner, whereas determining mastery based on all trials resulted in slightly better skill maintenance. In the current replication, no such differences in indication of mastery or maintenance were observed when data were collected on all trials or the first trial.

DESCRIPTORS: autism, data collection, discrete trials, interspersed trials, massed trials

_____

One of the defining features of applied behavior analysis is the improvement of behavior in meaningful and socially important ways (Baer, Wolf, & Risley, 1968). The primary method for assessing behavior change is through repeated data collection. Thus, behavior analysts who deliver behavioral intervention programs to children with autism have often used continuous data-collection methods in which data are collected on each learning trial. However, in an attempt to decrease time spent collecting data, clinicians have recently begun to collect data on the first trial within a block of trials (cf. Dollins & Carbone, 2003; Sundberg & Hale, 2003).

To date, one study has attempted to examine empirically the differences between all-trials and first-trial data-collection methods. Cummings and Carr (2009) collected data on either the first of 10 acquisition trials or on all 10 acquisition trials during discrete-trial instruction for 6 children with pervasive developmen-

tal disabilities. Results suggested that basing a child's progress on first-trial data led to determination that the child had met mastery levels sooner than basing progress on data that were collected on all 10 trials. By contrast, using data from all 10 trials better identified skill maintenance (i.e., retention of skill performance following initial mastery).

A potential limitation of the methods used by Cummings and Carr (2009) was that data collection did not continue for the remaining nine trials in the first-trial data-collection procedure. Thus, it was not possible to determine if an all-trials data-collection procedure would have agreed with conclusions regarding the number of sessions to mastery (within targets) using the first-trial procedure. The current study replicated procedures described by Cummings and Carr but also collected data on the remaining nine trials in the first-trial data-collection procedure to compare whether the two methods of analyzing data would yield differential identification of skill mastery and maintenance.

## METHOD

### Participants and Setting

The participants in this study were 11 children (10 boys, 1 girl) with an independent

diagnosis of autism or pervasive developmental disorder (not otherwise specified) between the ages of 3 and 11 years. Participants were receiving 5 to 38 hr ($M = 20$ hr) of one-on-one behavioral intervention and had been receiving discrete-trial instruction for a mean of 13.2 months (range, 1 to 39 months). One participant primarily communicated via a picture exchange communication system, 3 communicated with one- to two-word utterances, and 7 communicated in full sentences.

Therapists were individuals who typically implemented behavioral interventions with the participants, and who had provided such treatment to children with autism for a mean of 50.4 months (range, 17 to 144 months). All therapists were enrolled in a MA program in behavior analysis at the time of the study. Sessions were conducted in each participant's therapy room in his or her home or in the clinic.

### Data Collection and Interobserver Agreement

The principal measure of participant behavior was the correct emission of each targeted skill. The skills that were targeted for intervention depended on each child's curricular needs across three areas (i.e., language, social skills, and academic skills) and four operant classes (i.e., imitation, receptive identification, tacts, and intraverbals). (A list of specific skills targeted for all participants is available from the first author.) A *correct response* was defined as the participant emitting the target skill within 3 s of the presentation of the initial instruction; emission of a response following any additional prompting was not scored as a correct response.

Each session consisted of data collection on 10 acquisition trials. During some phases of teaching, interspersed trials of previously mastered skills were conducted, but data were not collected during these trials. Using the all-trials data-collection method, correct responses were converted to a percentage after dividing the number of acquisition trials scored as correct by the total number of acquisition trials. For the first-trial data-collection method, correct responses were converted to a percentage using the score on the first acquisition trial only (a 0 if incorrect and a 1 if correct). One data-collection method was randomly assigned to each of the participant's acquisition targets to evaluate differences in the identification of mastery across the two data-collection procedures. The mastery criterion was defined as three consecutive sessions above 80% correct responding (all-trials condition) or three consecutive sessions of 100% correct responding (first-trial condition).

Performance of the target skills yielded four dependent variables to evaluate the two data-collection procedures: (a) the number of sessions to mastery; (b) the mean percentage of correct responses during maintenance probes; (c) agreement between data-collection methods regarding identification of mastery, defined as the percentage of targets in which the two methods agreed on the number of sessions to mastery; and (d) the percentage of targets in which mastery would have been identified sooner had the alternative data-collection procedure been used, which was assessed by summing the number of target behaviors for which use of the alternative data-collection method's criterion would have resulted in a decision of mastery in fewer sessions.

Interobserver agreement was assessed by having a second observer independently collect data on 35% of sessions. Interobserver agreement scores were calculated by dividing agreements on correct responding by agreements plus disagreements and converting the resulting quotient to a percentage for each 10-trial session. Across participants, interobserver agreement for correct responses was 99% (range, 83% to 100%).

### Procedure

Each participant received discrete-trial instruction for four target skills, 1 to 5 days per week. Two target skills were from one curricular lesson (Training Set 1), and two target skills were from

a different curricular lesson (Training Set 2). At the inception of the study, one target skill from each training set was randomly paired with the all-trials condition, and the other target skill from each training set was randomly paired with the first-trial condition. Each day, one to four target skills were assessed one to three times each. For each target skill, the experiment proceeded in the following phases for 10 of the 11 participants: baseline, massed trials, interspersed trials, and maintenance (procedural details may be obtained from the first author). The massed-trials phase was omitted for 1 participant.

One baseline session was conducted per target skill, in which 10 trials of that skill were interspersed with previously mastered skills (for every acquisition trial, there were approximately one to two interspersed trials). For each trial, the instruction was vocally presented, followed by a 3-s intertrial interval. No differential consequences were arranged for the emission of a target skill, whereas correct emission of previously mastered skills resulted in brief praise. Massed-trials sessions consisted of 10 consecutive trials in which the therapist presented only the target skill, using a most-to-least intrusive prompt hierarchy (Riley & Heaton, 2000). Correct emission of the target skill following the initial instruction resulted in praise and brief (e.g., 5 s) access to stimuli that had been previously identified as preferred during brief preference assessments (DeLeon et al., 2001). Acquisition target skills progressed to the interspersed-trials phase after two to three sessions above 80% correct responding (all-trials condition) or 100% correct responding (first-trial condition). The interspersed-trials phase was conducted to ensure that participants were able to respond to acquisition stimuli in the presence of distracter stimuli (i.e., previously mastered skills). During this phase, the therapist interspersed 10 trials of the target skill with one to two trials of previously mastered skills. Correct responding during acquisition trials produced praise and brief (e.g., 5 s) access

to a preferred stimulus. This phase continued until the respective mastery criteria were achieved. Thereafter, the therapist conducted maintenance probes for each target skill once per week for 3 weeks. Each maintenance probe consisted of 10 trials of the target skill that were interspersed with previously mastered skills. The therapist provided praise for correct responses (for targeted and mastered skills); all other responses produced no differential social consequences.

*Experimental Design and Data Analysis*

An alternating treatments design was used to compare the number of sessions to mastery and maintenance for each participant's two training sets across the two data-collection procedures. However, for the purpose of evaluating differences in the four dependent measures across the first-trial and all-trials data-collection methods, data were summarized across all participants for the respective training sets. Session-by-session data are available from the first author.

## RESULTS AND DISCUSSION

Figure 1 (top) depicts the number of sessions to mastery across participants. Using the all-trials and first-trial methods, the mean number of sessions to mastery were 7.45 and 7.77, respectively, indicating little difference between the two methods. Figure 1 (middle) depicts the mean percentage of correct responding on the target skills during the maintenance probes for all participants in the all-trials and first-trial methods, yielding 95% and 97%, respectively, again indicating no differentiation between the two methods. Figure 1 (bottom) depicts the percentage of target skills for which (a) there was agreement across both methods on the number of sessions required to observed mastery level performance (39%), (b) mastery would have been identified sooner had the all-trials method been used (43%), and (c) mastery would have been identified sooner had the first-trial method been used (18%).
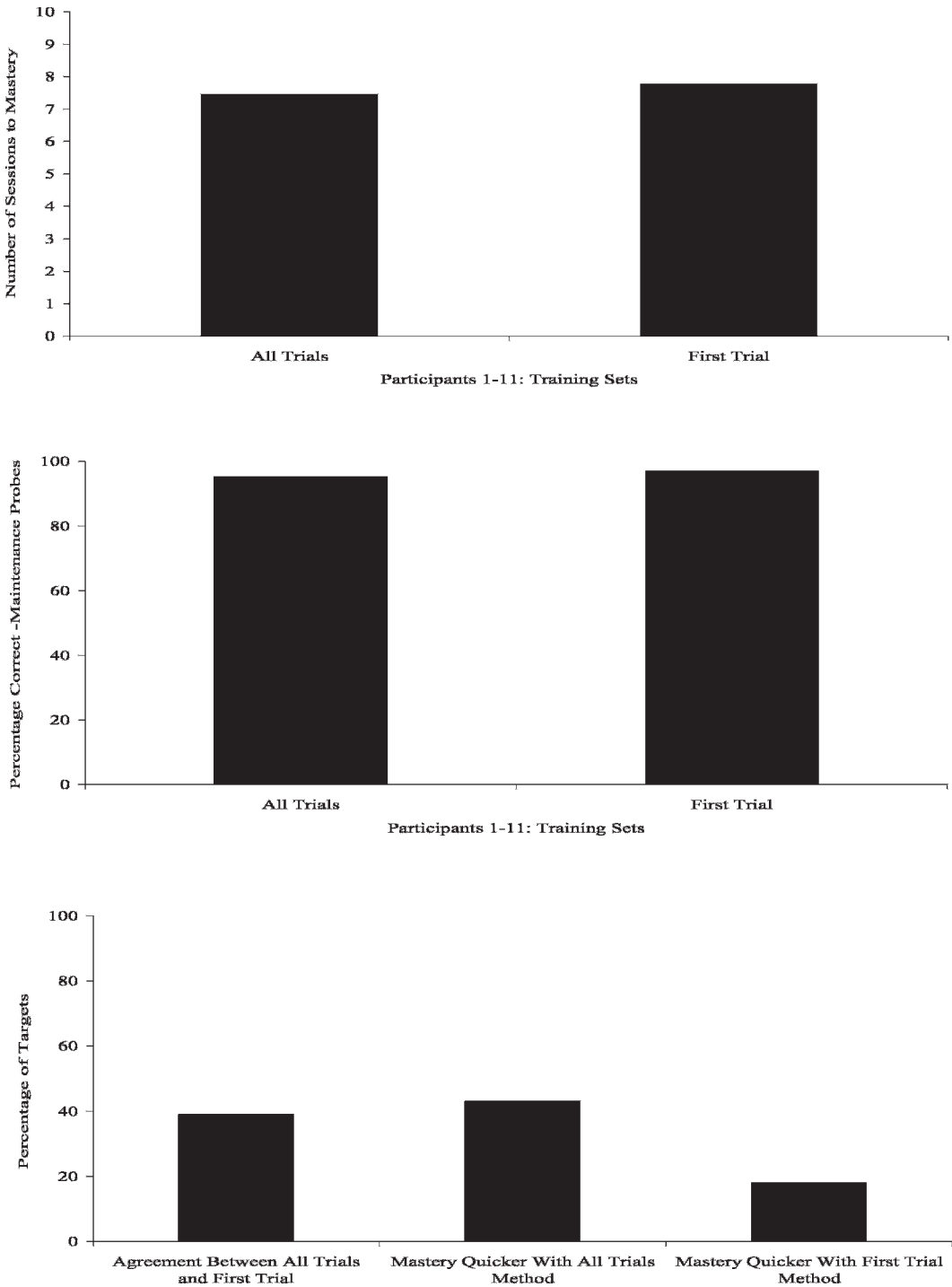
Figure 1.   Mean number of sessions to mastery (top) and mean percentage of correct responses during maintenance probes (middle) for all-trials and first-trial data-collection methods across all participants. Percentage of acquisition targets in which there was agreement between all-trials and first-trial data collection and in which mastery would have been achieved sooner had the alternative data-collection method been used (bottom).

The current results suggest that there was little difference in terms of the number of sessions in which a participant displayed mastery performance and in the percentage of correct responding during maintenance probes using either the all-trials or the first-trial data-collection method. These results differ slightly from those obtained by Cummings and Carr (2009), who noted that mastery was identified in fewer sessions using the first-trial method. However, it should be mentioned that data collection continued on the remaining nine trials during the first-trial condition in the current investigation. Thus, the current analysis also compared the two methods within each target skill. This additional analysis allowed us to see that, had the alternative data-collection method been used to identify mastery for each target, mastery would have been suggested earlier in more cases when the all-trials method was applied (43% of targets) than when the first-trial method was applied (18% of targets). Thus, the first-trial method was a slightly more conservative measure of length of time to achieve mastery-level performance.

Despite this finding, the current results should be interpreted with caution because they might have been an artifact of using a mastery criterion of greater than 80% during the all-trials condition. Given that this mastery criterion is less stringent than the 100% criterion used for the first-trial condition, it is not surprising that mastery would be suggested sooner in more cases using a lower criterion level. The above 80% criterion was used in the current investigation because it has been used to evaluate response mastery (Anderson, Taras, & Cannon, 1996). Nevertheless, future research could compare the identification of mastery-level performance using ranges of criterion levels (e.g., comparing mastery at 80%, 90%, and 100% during all trials to 100% during the first trial) to determine the impact of different criterion levels on evidence of mastery. Future research should also evaluate the extent to which data collection on all trials or only a subset of trials decreases time requirements associated with the implementation of discrete-trial instruction.

Considering the results of this study combined with the results reported by Cummings and Carr (2009), it appears that first-trial data collection might be a promising option for assessing behavior change during discrete-trial instruction. However, additional research is needed to evaluate the utility of this method. Specifically, all-trials data collection may have potential advantages in that it may help supervisors to monitor therapist behavior during sessions (i.e., procedural integrity). All-trials data collection also may help supervisors to identify errors related to prompts, presentation of stimuli, and movement through teaching phases, and to monitor how many trials are implemented.

Several limitations warrant discussion. First, no procedural integrity data were collected on implementation of discrete-trial instruction; however, the therapists in the current study were evaluated regularly for procedural integrity according to standards that have been described elsewhere (e.g., Crockett, Fleming, & Doepke, 2007; LeBlanc, Ricciardi, & Luiselli, 2005). Second, only one baseline session was conducted prior to treatment. However, all participants responded 0% to 40% correct during baseline, indicating inadequate responding. Third, the current participants may not be representative of the population of children with pervasive developmental disorders (e.g., several of the current participants had relatively well-developed verbal behavior repertoires). Thus, it is unknown if there would be a difference between data-collection methods with children who do not share these characteristics.

## REFERENCES

Anderson, S. R., Taras, M., & Cannon, B. O. (1996). Teaching new skills to children with autism. In C. Maurice, G. Green, & S. C. Luce (Eds.), *Behavioral intervention for children with autism: A manual for parents and professionals* (p. 192). Austin, TX: Pro-Ed.

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.

Crockett, J. L., Fleming, R. K., & Doepke, K. J. (2007). Parent training: Acquisition and generalization of discrete trials teaching skills with parents of children with autism. *Developmental Disabilities*, *28*(1), 23–36.

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*, 57–71.

DeLeon, I. G., Fisher, W. W., Rodriguez-Catter, V., Maglieri, K., Herman, K., & Marhefka, J. (2001). Examination of relative reinforcement effects of stimuli identified through pretreatment and daily brief preference assessments. *Journal of Applied Behavior Analysis*, *34*, 463–473.

Dollins, P., & Carbone, V. J. (2003, May). Using probe data recording methods to assess learner acquisition of skills. In V. J. Carbone (Chair), *Research related to Skinner's analysis of verbal behavior with children with autism.* Symposium conducted at the 29th annual convention of the Association for Behavior Analysis, San Francisco.

LeBlanc, M. P., Ricciardi, J. N., & Luiselli, J. K. (2005). Improving discrete trial instruction by paraprofessional staff through an abbreviated performance feedback intervention. *Education and Treatment of Children*, *28*(1), 76–82.

Riley, G. A., & Heaton, S. (2000). Guidelines for the selection of a method of fading cues. *Neuropsychological Rehabilitation*, *10*(2), 133–149.

Sundberg, M. L., & Hale, L. (2003, May). Using textual stimuli to teach vocal-intraverbal behaviors. In A. I. Petursdottir (Chair), *Methods for teaching intraverbal behavior to children.* Symposium conducted at the 29th annual convention of the Association for Behavior Analysis, San Francisco.