

Training Educators to Collect Accurate Descriptive-Assessment Data

Sacha T. Pence

Auburn University

Claire C. St. Peter

West Virginia University

Abstract

Descriptive assessments involve recording naturally occurring instances of behavior and corresponding antecedent and consequent events. Authors have argued for the use of two forms of descriptive assessment, structured and narrative Antecedent-Behavior-Consequence recording, because these methods may require little training. However, the extent to which minimal training produces accurate data with these methods is unknown. During Experiment 1, we examined teachers' accuracy when recording descriptive data from videos. Accuracy on problem behavior did not improve over time, regardless of initial exposure to structured or narrative ABC recording. Teachers preferred the structured ABC recording sheet. During Experiment 2, we provided training to participants using an automated procedure that included practice and feedback. Accurate data collection on problem behavior increased for six participants after training. Data-collection accuracy was higher for environmental events involving the presentation of stimuli (demand and attention) than the absence of stimuli (escape and low attention). Participants displayed idiosyncratic preferences for the structured or the narrative ABC recording sheet.

Keywords: ABC recording, data collection, descriptive assessment, problem behavior, teacher training

Descriptive analysis refers to the collection of data on behavior occurring in the natural environment, along with the antecedents and consequences that are associated with the behavior. Although descriptive analyses only allow for conclusions on correlations between antecedent and consequent events and behavior, and may not identify the same functions as experimental analyses (those that involve manipulation of the events preceding and following behavior;

Author Note: Sacha T. Pence is now at Drake University.

Address correspondence to: Sacha Pence, Department of Leadership and Counseling, Drake University, Des Moines, IA 50311. E-mail: sacha.pence@drake.edu

Lerman & Iwata, 1993; Pence, Roscoe, Bourret, & Ahearn, 2009; Thompson & Iwata, 2007), descriptive assessments have several advantages. Descriptive analyses allow assessment of the extent to which the environment is therapeutic (e.g., if a teacher is reliably prompting through a demand following problem behavior and that escape rarely occurs, indicating a therapeutic environment for a client with escape-maintained behavior), identification of idiosyncratic events that appear to evoke or maintain behavior (e.g., teacher helping other students serves as an evocative event and peer attention maintains behavior), and development of hypotheses regarding behavior function (Vollmer, Borrero, Wright, Van Camp, & Lalli, 2001). In addition, descriptive assessments can be used to help inform functional analysis conditions or to develop and refine operational definitions. Oliver, Pratt, and Normand (2015) surveyed behavior analysts on their use of assessment methods to determine the function of behavior. Descriptive assessments were the most commonly reported assessment with over 94% of respondents stating that they “almost always” or “always” used descriptive assessments when conducting a functional behavior assessment. Similarly, Roscoe, Phillips, Kelly, Farber, and Dube (2015) surveyed behavior analysts and school professionals and found that the majority of respondents reported using descriptive assessments (either alone or in combination with other assessment methods). Given the prevalence of descriptive assessments and the need for good behavioral observation skills, it is important to evaluate how to train practitioners to collect data on the occurrence of behavior and environmental events.

Descriptive analyses and teacher-collected data remain the most common method for identifying potential functions of behavior within school settings (Kern, Hilt, & Gresham, 2004; Weber, Killu, Derby, & Barretto, 2005). Teachers may be asked to collect descriptive-assessment data because they spend a large proportion of the school day with the child and observe ongoing problem behavior in their classrooms. Frequently, teachers are part of the functional behavior assessment (FBA) team and play a major role in behavioral assessment, including taking descriptive-analysis data. Because descriptive analyses are the cornerstone of FBA in schools, collection of accurate descriptive data is essential to ensuring the fidelity of the FBA process.

At least four factors may influence teachers’ abilities to accurately record descriptive data. First, the format in which data are collected may influence the accuracy of the data collection. At least three formats for collecting descriptive data have been reported in the literature. In more recent studies evaluating the utility of descriptive data in identifying

behavioral function, data were collected on a continuous basis using computer-based data collection programs. Although computer-assisted data collection produces detailed, often second-by-second, records of responding, it may not be feasible for classroom use because of the monetary and training resources required (Thompson & Borrero, 2011). Two paper-pencil methods have been reported in the literature; these methods have been recommended because they may not require extensive training or expertise (Thompson & Borrero, 2011). In narrative recording, the data collector writes a narrative description of the antecedents and consequences surrounding the behavior. Because the data collector describes the ongoing events, there are no constraints on what can be recorded. Because of this, narrative recording provides a considerable amount of flexibility. However, narrative recording may also lead to difficulties in pinpointing how a particular description aligns with potential functions. For example, a teacher may write “got frustrated” as the antecedent for aggression and “calmed down” as the consequence.

In structured recording, the data collector selects antecedent and consequent events from a checklist following each instance of the behavior. Structured recording requires data collectors to categorize antecedent and consequent events in a way that is likely to yield an environmental function. Additionally, structured data collection may ease mathematical analyses, such as the percentage of problematic responses that were preceded by a demand. However, structured recording limits the flexibility of the data collection and may require additional training to produce accurate data, as the same categorical event, such as an antecedent demand, can take many different forms in the environment.

Lerman, Hovanetz, Strobel, and Tetreault (2009) evaluated the accuracy with which teachers collected descriptive data using narrative and structured formats. The authors evaluated accuracy under ideal conditions: teachers collected data in a distraction-free environment in which they had no competing demands and from a videotape in which an actor engaged in instances of problem behavior that were spaced at least 1 min apart. Teachers were slightly more accurate with the structured format than the narrative format, although accuracy was relatively low regardless of the data collection type. Overall, teachers accurately recorded 64.5% of antecedents and consequences, and correctly captured all of the relevant antecedents and consequences for 27% of the instances of problem behavior. The authors suggested that these findings highlighted the need for further research in the type and extent of training necessary to produce accurate teacher-collected descriptive data.

A second factor influencing the accuracy of descriptive data are the resources available to the teacher responsible for collecting the data. Resources can take several forms, including limiting or assistance with concurrent responsibilities of the teacher during the data-collection period and the amount of training previously provided to the teacher. Accuracy may be hampered when the teacher who is recording data is also the person responsible for administering consequences for the child's behavior. Under ideal circumstances, teachers would have no other responsibilities while collecting data and would be highly trained in the data collection procedure used. In schools, however, these ideals are rarely met.

The descriptive-analysis literature is replete with examples of two highly trained observers being able to simultaneously and independently record naturally occurring responding with a high level of agreement (e.g., Lalli, Browder, Mace, & Brown, 1993; Lerman & Iwata, 1993; Pence et al., 2009; Samaha et al., 2009). However, the observers in those studies were typically undergraduate or graduate students, not teachers, and included structured practice until the observers met predetermined accuracy criteria. In contrast, the training provided to teachers is often conducted in didactic workshops that contain few practice components. Largely didactic trainings may be unlikely to foster development of data-collection skills (Crone, Hawken, & Bergstrom, 2007). Methods for improving descriptive data collected by teachers are largely missing from the literature. Ellingson, Miltenberger, Stricker, Galensky, and Garlinghouse (2000) evaluated the extent to which teachers were able to collect accurate descriptive data using a structured checklist. Although teachers were moderately to highly accurate, the type of events included on the checklist was limited and largely focused on different kinds of attention, which may have artificially inflated accuracy. In addition, teachers were able to devote time solely to data collection because paraprofessionals assumed classroom responsibilities.

A third factor influencing the accuracy of teacher-collected data is the form of the response targeted for data collection. Accurate data collection may be challenged by target behavior that occurs in bursts or at a high intensity, which may inhibit teachers' abilities to record every instance or to accurately record all antecedents and consequences. Teachers may find data collection challenging when responding occurs irregularly, such as behavior that occurs in bursts or escalates across time. For example, Rolider, Iwata, and Bullock (2012) found that highly trained undergraduates collected data less accurately when the target behavior was spaced closely together. The influence of behavioral patterning on teacher-collected data has not yet been evaluated.

Fourth, teachers' preferences for one form of data collection over another may influence the extent to which they accurately use the forms. Teachers who prefer one type of data collection to another may use that type more often and become more familiar with it, leading to improved accuracy. Lerman et al. (2009) found that teachers rated the structured format slightly higher than the narrative format, with most of the educators stating that they would prefer to use the structured format when collecting data in their classrooms. However, teachers' preferences were not strong; there was only a 2.8-point difference in the ratings on a 30-point scale. It is possible that preferences would be different if teachers used the data-collection methods under more naturalistic circumstances, or if a behavioral measure of preference was used instead of verbal report.

The purpose of the current study was to extend the findings of Lerman et al. (2009) in three ways. First, Experiment 1 evaluated potential practice effects on the accuracy with which teachers collected descriptive data from videotapes of naturally occurring caregiver-child interactions. Because the spacing of problem behavior, antecedents, and consequences was uncontrolled, the teachers' accuracy might more closely approximate typical teacher-collected data than those obtained under the ideal circumstances evaluated by Lerman and colleagues. Second, the purpose of Experiment 2 was to evaluate an automated training procedure that incorporated practice and feedback, but was easy to distribute and required few trainer resources, in an attempt to improve the accuracy with which teachers collected data. Third, across both Experiment 1 and Experiment 2 we used a behavioral-choice measure to determine whether teachers preferred narrative or structured data collection.

Experiment 1

Method

Participants and setting. Participants were eight female full-time teachers and one female clinician (Sheila) who were enrolled in a course sequence designed to prepare teachers to become Board Certified Behavior Analysts (BCBA). At the time of participation, teachers had completed two 3-credit courses on principles of behavior analysis and the application of those principles. Participants' teaching experiences and education are outlined in Table 1 (information was not available for Sheila). Six participants reported prior experience collecting narrative or structured ABC recording. Sessions were conducted

Table 1
Teaching Experience and Education for Participants in Experiment 1

Participant	Years Teaching	Master's Degree	Certifications
Carla	4.5	Special Education	Mild to Moderate Mental Impairment, Behavior Disorders, Learning Disabilities
Holly	19	Special Education	Mental Impairment, Learning Disabilities, Behavior Disorders
Jess	10	Elementary Education	Learning Disabilities, Behavior Disorders, Mental Impairment
Kari	2	None	Multicategorical Special Education, Autism
Kim	4	Elementary Education	Multicategorical Special Education
Laura	26	Special Education	Learning Disabilities, Mental Impairment
Reese	4	Special Education	Multicategorical Special Education, Autism
Shelia	—	—	—
Vickie	3	Special Education	Multicategorical Special Education

in the library at a local elementary school. The library was equipped with a projector and DVD player.

Materials. Four, 10-min video segments were selected prior to the study. Video segments were naturally occurring interactions between a mother-son dyad. Video segments were obtained from videos previously collected during a parent-training study. Parental consent was obtained to use these videos for other purposes. Nathan was a 7-year-old male who engaged in aggression, inappropriate vocals, and disruption. Aggression was defined as any instance of contact with another person in the form of hitting, slapping, kicking, biting, or hair-pulling or any instance of biting himself. Inappropriate vocals included statements of refusal using the word "No," and statements that included name-calling or the use of curse words. Disruption was defined as any instance of ripping papers, throwing objects, or slamming hands down on the table. Nathan engaged in an average of 11 (range, 6–16) instances of problem behavior during each 10-min session. Some instances of problem behavior occurred with at least 30 s elapsing before another instance of problem behavior, while other instances occurred in succession (less than 10 s between instances of

problem behavior). The videos were presented to the participants in a randomly selected order. The types of activities varied across and within the videos, but included academic work (e.g., spelling, worksheets), playing with toys, transitioning from breaks to work, and engaging in conversations. The frequency and type of antecedent and consequent events also varied across the four videos.

Antecedent events included low attention, restricted access, and demand. Low attention was defined as 3 s or more without vocal or physical interaction with another person. Restricted access was defined as the removal of a leisure item or any instance when the child was denied access to requested items or activities. Demand was defined as any vocal, model, gestural, or physical prompts to engage in a behavior or the presence of ongoing instructional activity. Consequent events included attention, access to materials, and escape. Attention was defined as vocal or physical interaction with the child, including reprimands and redirection statements. Access to materials was defined as the availability of leisure tangible items for manipulation or consumption regardless of adult permission and included if access was delivered by the adult or if the child got the item independently. Escape was defined as the removal of demand and instructional materials, or the absence of prompting if the child stopped engaging in the task for at least 3 s.

Two different ABC recording sheets were used (see Figure 1). The structured ABC sheet included an area to indicate up to three target behaviors, the general activity in place, immediate antecedents, and immediate consequences. Participants were instructed to record each instance of behavior in a separate column on the data sheet. On the narrative ABC recording sheet, participants were instructed to list each instance of problem behavior in a separate row on the data sheet. Antecedent and consequent events were recorded under the appropriate column in the same row as the corresponding instance of problem behavior.

Procedure. The course instructor gave a group lecture and assigned readings (Cipani & Schock, 2007; Lerman et al., 2009; Pence et al., 2009) on descriptive assessments prior to the first session. The lecture included an introduction to descriptive assessments, rationales for collecting these data, and a discussion of the usefulness and limitations of data collected using ABC methods. The instructor provided the participants with a copy of operational definitions for the problem behavior, common antecedent events, and common consequent events. The instructor reviewed these definitions and told the participants that the definitions applied to the structured and narrative ABC recording.

Narrative ABC Recording Sheet

Antecedent	Behavior	Consequence

Structured ABC Recording Sheet

<i>Problem Behavior 1</i>																			
<i>Problem Behavior 2</i>																			
<i>Problem Behavior 3</i>																			
General Activity																			
Leisure																			
Self-care																			
Academic (or other work task)																			
Alone																			
Antecedent																			
No attention																			
Leisure removed/denied																			
Given instruction/work																			
Consequence																			
Attention (including “no” and “stop”)																			
Redirection to another area/activity																			
Leisure item/food given																			
Work requirement terminated																			
Person moved away																			

Figure 1. Portions of the narrative ABC recording sheet (top panel) and structured ABC recording sheet (bottom panel).

The course instructor reviewed each form of data sheet. She told participants to record data on the structured ABC recording by placing a check next to the target behavior and corresponding antecedent and consequence events all within the same column. The course instructor directed participants to write other antecedent or consequent events (i.e., an event that was not prelisted) on the structured ABC recording sheet if that situation arose. For the narrative ABC recording sheet, the instructor directed the participants to list the target behavior and write a description of antecedent and consequence events within the same row. The instructor did not provide directions on the

level of detail that the participant should include in the narrative ABC recording. The instructor noted that multiple antecedent and consequent events could be recorded, regardless of the data-collection format. She answered any participant questions before participants began to practice.

We presented video segments on a large screen that participants viewed simultaneously. Half of the participants (Vickie, Carla, Sheila, Laura, and Reese) used the structured ABC recording during the first session and the other half (Jess, Kim, Kari, and Holly) used the narrative ABC recording. During the second session, teachers used the other format: Vickie, Carla, Sheila, Laura, and Reese used the narrative ABC recording and Jess, Kim, Kari, and Holly used the structured ABC recording. Participants were allowed to choose between the structured and narrative ABC recording for the third and fourth sessions.

Before each session, we instructed participants using the structured ABC recording sheet to check off the general activity, immediate (within 10 s) antecedent event(s), and immediate (within 10 s) consequent event(s) that corresponded with the occurrence of each instance of behavior. We instructed participants using the narrative ABC recording sheet to write each instance of the target behavior, and antecedent and consequence events that occurred either 10 s prior to or 10 s following each instance of behavior. After each session, the experimenter reviewed the session with the participants, including discussing any problems that arose during data collection (e.g., how to facilitate collecting data during rapid sequences of problem behavior) and answering questions (e.g., situations in which behavior should be scored or a specific antecedent or consequence event occurred). The experimenter did not review the participants' data or provide direct feedback on their individual performances.

Data analysis. The teacher-completed ABC recordings were compared to data collected by a highly trained observer who collected data using a narrative ABC recording. The data from the narrative ABC recording were used to complete a structured ABC form. The narrative or structured ABC form was used to compare the data from the corresponding teacher-completed ABC recordings. The trained observer (first author) was a doctoral student in behavior analysis who had approximately 6 years of prior experience using narrative and structured ABC recording, and a peer-reviewed publication about the use of these methods (Pence et al., 2009). The trained observer and a doctoral-level behavior analyst (the second author) independently collected data on the same videos used in the training. Interobserver agreement (IOA) was above 90% across antecedents, consequences, and target behavior for all sessions.

To calculate participant accuracy, we used total count IOA (Cooper, Heron, & Heward, 2007). Total count IOA was used because we were unable to evaluate which instances of behavior were scored or omitted by participants, precluding IOA calculations based on instance-by-instance comparisons. Total count allowed for an estimation of overall performance. We summed instances of problem behavior for each session, for each participant. This number was compared to the data collected by the trained observer. For each participant, the smaller total instances of problem behavior scored was divided by the greater total instances and then multiplied by 100. This calculation yielded a percent agreement between the participant and the trained observer across sessions.

The percent agreement for each antecedent and consequent event was also calculated. The total frequency at which a particular event was scored across all four sessions was determined (i.e., the sum of each event across all sessions). The summed frequency of each event was divided by the total instances of problem behavior scored, yielding a proportion of the event's occurrence in relation to the instances of problem behavior scored by the participant. Because we were unable to know what opportunities participants recorded or omitted, the proportional calculation allowed us to capture relative frequencies of data collection. For each participant, three proportions were obtained for antecedent events (no attention, demand, and restricted access) and three proportions were obtained for consequence events (attention, escape, and access to materials). These same proportions were calculated for the trained observer's data. The proportion obtained for each participant was compared to the proportion of the same event for the trained observer using the following calculation: the smaller proportion was divided by the larger proportion and then multiplied by 100. This provided a percent agreement between the participant and the trained observer while accounting for differences in the total instances of problem behavior recorded. For example, Shelia obtained a proportion of 0.81 for attention following problem behavior. The trained observer obtained a proportion of 0.75 for attention following problem behavior. The trained observer's proportion was divided by the Sheila's proportion ($0.75/0.81$) and then multiplied by 100 to yield a percent agreement of 92.6%.

Results and Discussion

Figure 2 depicts the percentage of agreement on the occurrence of problem behavior across sessions for all participants. Participants accurately recorded instances of problem behavior with an average of 63.9% agreement (range, 47–78.9%). Mean accuracy was higher than

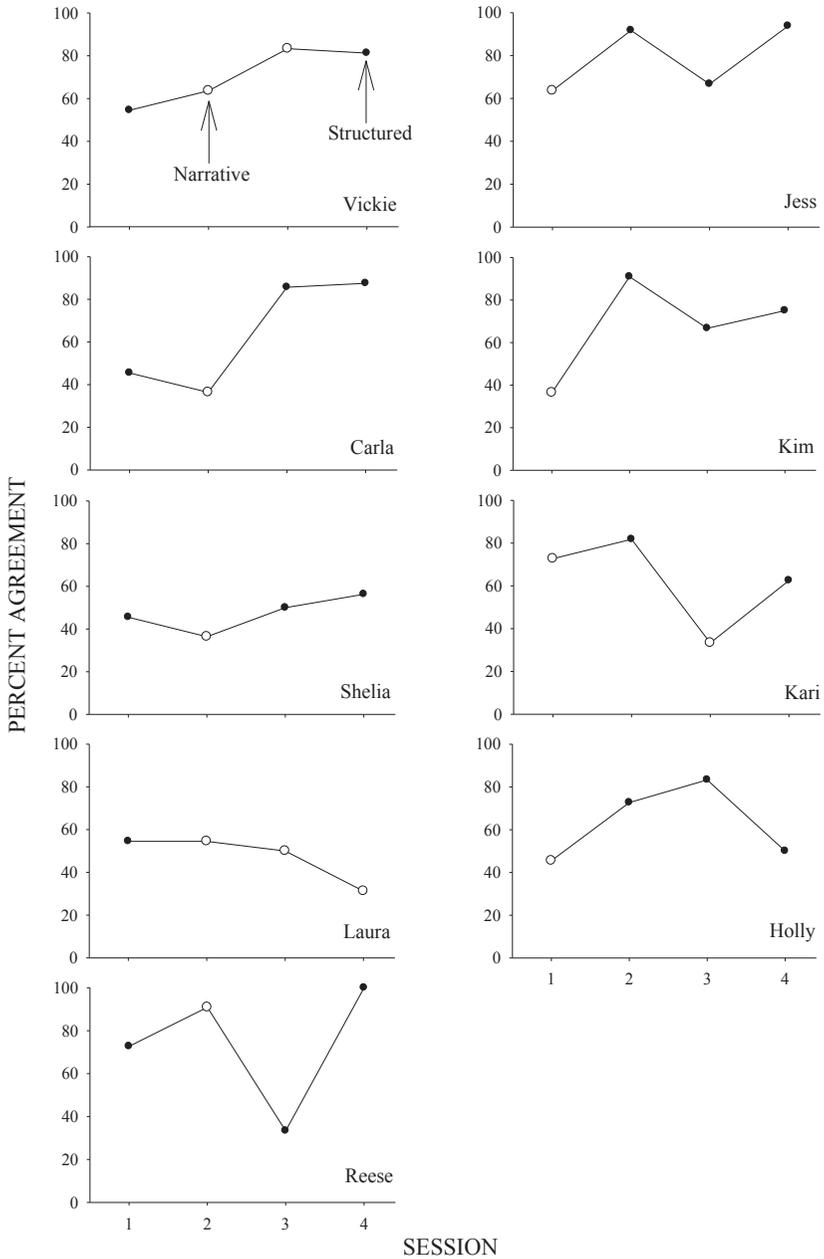


Figure 2. Percent agreement on instances of problem behavior across sessions. Sessions in which participants used the structured ABC sheet are depicted by the closed circles. Sessions in which narrative ABC recording were used are depicted by the open circles.

the levels reported by Lerman and colleagues (2009). Lerman and colleagues suggested that accuracy during narrative recording might improve after participants had experience with structured recording. Five participants (Carla, Kim, Jess, Reese, and Vickie) showed some increases in performance with repeated practice. Two of these participants were initially exposed to the narrative (right panel) and three initially with the structured (left panel) ABC recording, suggesting that the initial format with which participants collected data was unlikely to influence performance gains. The other four participants' accuracy on problem behavior did not consistently improve over time. Although mean accuracy was greater than that reported by Lerman et al. (2009), only 30.6% of all sessions across participants had greater than 80% agreement on the target behavior.

The type of data sheet selected is shown by the closed and open symbols in Figure 2. Recall that participants were assigned to a particular data sheet during the first two sessions, but could choose which data sheet to use during the third and fourth sessions. In general, participants chose structured ABC recording over narrative ABC recording. Six participants (Carla, Holly, Jess, Kim, Reese, and Shelia) selected the structured ABC recording in both sessions when provided with the opportunity to choose (refer to Figure 2). Two participants (Kari and Vickie) alternated between the structured and narrative ABC recording. One participant (Laura) selected the narrative ABC recording during both sessions.

Table 2 outlines the percent agreement between each participant and the trained observer across antecedent and consequent events. Participants accurately recorded the presentation of a demand with an average of 91.7% agreement (range, 62.9–98.7%). Restricted access was accurately recorded with an average of 40.3% agreement (range, 0–95.5%). Low attention was accurately recorded with an average of 22.4% agreement (range, 0–73.2%). Participants accurately recorded attention following instances of problem behavior with an average of 78.7% agreement (range, 51.9–97.4%). Access to materials was accurately recorded with an average of 77.8% agreement (range, 0–100%). Escape was accurately recorded with an average of 37.1% agreement (range, 6.8–81.8%). Participants were more likely to identify environment events when the event involved the presentation of stimuli (i.e., the presentation of a demand or the presentation of attention) than when the event involved the absence of stimuli (i.e. the absence of attention or the absence of prompting or demands).

The primary purpose of Experiment 1 was to evaluate if performance increased over time with practice. Overall, participants did not show consistent improvements to high levels of accuracy when

Table 2
Percent Agreement on Antecedent and Consequent Events

Participant	Antecedent Events			Consequent Events		
	No Attention	Demand	Restricted Access	Attention	Escape	Access to Materials
Vickie	70.5	93.5	71.0	55.9	61.3	100
Carla	73.2	96.4	36.6	83.3	81.8	100
Shelia	0	92.0	95.5	92.6	23.3	100
Laura	0	96.6	90.9	83.3	12.2	100
Reese	0	98.7	30.6	51.9	6.8	0
Jess	57.8	95.7	0	81.4	70.8	100
Kim	0	93.2	0	93.3	24.4	100
Kari	0	62.9	37.7	69.0	25.3	0
Holly	0	96.9	0	97.4	28.2	100

practicing collecting descriptive-assessment data in the absence of formal training. However, participants only practiced across four, 10-min videos, which might limit conclusions about sustained practice effects. It is possible that performance could continue to improve over time with additional practice. However, the training and practice provided in the current experiment exceeded that typically provided to teachers in the school district, and the teachers already had more training in basic principles and applications of behavior analysis than typical teachers. More intensive or formal training regarding descriptive data may be necessary to improve performance, but it is unlikely that school systems would have the resources for a trainer and teacher to work together until the teacher mastered data collection. One alternative to in-person training and coaching is to use an automated training that would require minimal resources. The purpose of Experiment 2 was to evaluate an automated training that would allow teachers to practice collecting descriptive-assessment data across several teaching trials while receiving automated feedback.

Experiment 2

Method

Participants and setting. Participants were another six female full-time teachers (Amy, Cali, Cam, Jan, Rya, and Tia) and two male

Master's students (Max and Tal) who were enrolled in a course sequence designed to prepare individuals to become BCBAs. As with the first experiment, participants had completed two 3-credit courses in behavior analytic principles and applications before the study, and were currently enrolled in a third course. Participants ranged in age from 22–53 years old, and had been teaching or providing services to children for 0–16 years. The teachers had Master's degrees in Special Education; the graduate students had bachelor's degrees in Nursing. Participants' teaching experience and education are outlined in Table 3. Three participants had prior experience with narrative ABC recording (Amy, Cam, and Jan), one participant had prior experience with both narrative and structured (Rya), and four participants had no prior experience with either method (Cali, Max, Tal, and Tia).

Sessions were conducted in the library of a local elementary school. The library was equipped with several computers. Participants sat at separate computer stations with headphones on during all sessions.

Materials. An automated, approximately 60-min PowerPoint presentation (available from the first author upon request) was created prior to Experiment 2. The PowerPoint presentation included an intro-

Table 3
Teaching Experience and Education for Participants in Experiment 2

Participant	Years Teaching	Master's Degree	Certifications
Amy	16	Technology Education	Behavior Disorders, Secondary Technical Education
Cali	3	Special Education	Behavior Disorders, Learning Disabilities, Mild to Moderate Mental Impairment
Cam	20	Special Education (SLD)	SLD, Mild Mental Impairment
Jan	3	Special Education (Multicategorical)	Specific Learning Disabilities
Max	0	None	None
Rya	2	Special Education	Autism, Multicategorical Special Education
Tal	0	None	None
Tia	2	Special Education (Multicategorical)	Mental Impairment, Learning Disabilities, Behavior Disorders

Note: SLD = Specific Learning Disabilities

duction to descriptive assessments, operational definitions of target behavior, operational definitions of antecedent events and consequent events, structured ABC recording data sheets, and narrative ABC recording data sheets. Participants were unable to pause the presentation, advance the slides manually, or return to previous slides. We instructed participants to listen to the voiceover audio and to read the corresponding text that appeared on each slide as the PowerPoint progressed. We provided participants with a piece of paper that included operational definitions for three topographies of problem behavior for each of the two children shown in video segments, and the definitions for antecedent and consequence events.

The PowerPoint presentation included seven video segments that were each 5 min in duration. Video segments and trials included the mother-son (Mary-Nathan) dyad from Experiment 1 and a therapist-student (Tara-Oliver) dyad. Children engaged in an average of 7 (range, 2–11) instances of problem behavior during each 5-min video. Three of these video segments were used during baseline and four were used during post-training. Additionally, 15 video trials, ranging from 20–65 s, were used during training. Antecedent and consequence events were defined as described in Experiment 1.

Three topographies of problem behavior were selected for each child. Nathan's problem behavior included aggression, inappropriate vocals, and disruption, and were defined as described in Experiment 1. Oliver was 3 years old. His problem behavior included climbing, throwing, and table slapping. Climbing was defined as any instance of standing upright on a chair or table or any instance of laying his body across a table. Throwing was defined as any instance of projecting an object towards the floor or away from him. Table slapping was defined as any instance of one or two palms coming down on the table from a distance of at least 7.6 cm that made an audible noise. Tara was provided with a script to help ensure periods of low attention and demands occurred and that attention and escape commonly followed problem behavior. However, Oliver did not follow a script and all instances of problem behavior were naturally occurring.

Participants used the structured and narrative ABC recording as described in Experiment 1. Participants were given a single data sheet that had the structured ABC recording on one side and the narrative ABC recording on the other side. Participants were instructed to collect data as described in Experiment 1. Teachers were expected to collect data on all environmental events. However, due to limited opportunities for antecedent events involving restricted access and consequence events involving access to edibles or materials, data collected on these events were not analyzed to evaluate the effectiveness

of the PowerPoint training. Participants were not permitted to ask the experimenters questions between trials or during the PowerPoint presentation.

Procedure. Prior to baseline, the course instructor assigned the article by Lerman and colleagues (2009). The instructor told participants that they would be going through an automated PowerPoint presentation that would provide an introduction and training on descriptive assessments. Participants each sat at an individual computer station with headphones on for the duration of baseline, training, and post-training sessions and followed instructions on the PowerPoint presentation. Additionally, the instructor told participants to record each instance of target behavior and the immediate antecedent and consequent event(s) that corresponded with each instance of problem behavior when video segments played. Finally, the instructor asked participants to write their names on their data sheets and told participants that all data sheets would be collected.

Baseline consisted of three 5-min video segments. We required participants to use the structured ABC sheet during the first baseline session, the narrative ABC sheet during the second baseline, and allowed them to choose a data sheet during the third baseline session. No feedback was provided to the participants during baseline.

During training, participants were exposed to 15 video trials. Participants could use the structured or narrative ABC recording sheet and were allowed to switch data sheets following any trial. Each trial was a video clip of Mary and Nathan or Tara and Oliver. We asked participants to view the video clip and record each instance of problem behavior and the corresponding antecedent and consequent events. Each trial contained one to three instances of problem behavior. Each instance of problem behavior was correlated with one or two antecedent events and one to three consequent events. The correct responses for antecedent and consequent events were displayed on the slide following each video. No behavioral criterion was established for participants to progress through the training portion or for participants to advance to post-training.

Post-training consisted of four 5-min video segments of Mary and Nathan or Tara and Oliver. Participants could use the structured or narrative ABC recording sheet and were allowed to switch data sheets before beginning a new session. Before beginning the first sessions with each dyad, the slides displayed operational definitions of problem behavior. No feedback was provided to the participants during post-training sessions.

Data analysis. We compared the teacher-completed data sheets to data collected by a highly trained observer (the first author). For

each participant, the total instances of problem behavior were summarized across all sessions for baseline and post-training. We compared the total instances of problem behavior in each video to the data collected by the trained observer using total count IOA as described in Experiment 1. We also calculated the percent agreement of each antecedent and consequent event during baseline and post-training, as described in Experiment 1. Although we did not have a criterion to progress through training, we set a standard at 80% agreement when we evaluated post-training teacher performance.

Results and Discussion

Figure 3 depicts the percent agreement on occurrences of problem behavior during baseline and post-training sessions across all participants. The percent agreement of instances of problem behavior decreased for one participant following training (Amy) and remained at levels observed during baseline for one participant (Cam). For these two participants, baseline levels were above 80% and remained above 80% during post-training sessions. For the other six participants, accuracy increased following training. Post-training percent agreement increased an average of 24.5 percentage points (range, 14.7–32.8 percentage

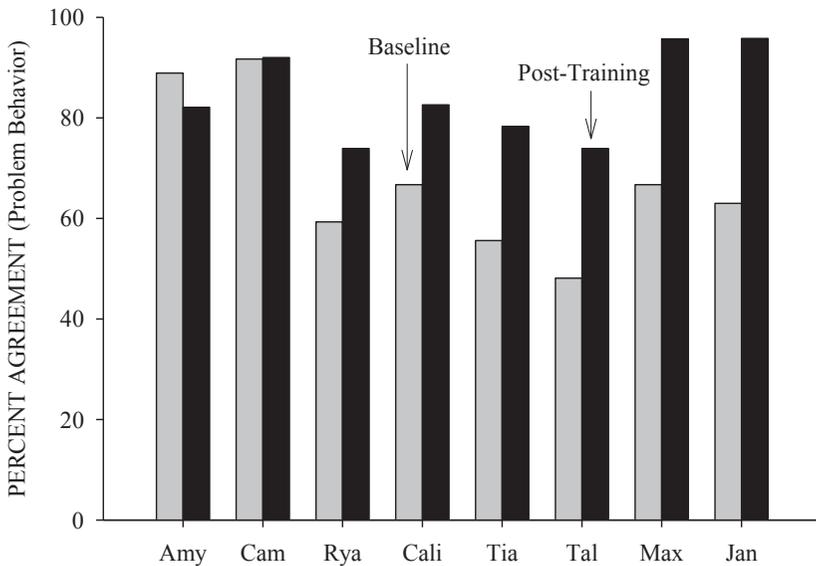


Figure 3. Percent agreement on instances of problem behavior across participants. Baseline agreements are depicted by the gray bars. Post-training agreements are depicted by the black bars.

points) over baseline levels for Rya, Cali, Tia, Tal, Max, and Jan. Although accuracy improved for six participants, only three of these participants (Cali, Max, and Jan) had accuracy levels above 80% following training.

The extent to which participants' data on occurrences of low attention preceding problem behavior agreed with a trained observers' data is shown in Figure 4 (upper panel). Following training, Tia's agreement on the occurrence of low attention decreased. Agreement on low attention increased modestly for Tal, Jan, and Max (79.3%, 31.9%, and 34.8%, respectively) and moderately for Rya, Cam, Amy, and Cali (16.9%, 30.3%, 54.3%, and 57.6%, respectively) after training. Only one participant (Cali) met mastery criteria with percent agreement above 80% post-training.

Baseline levels of percent agreement on occurrences of attention following problem behavior (lower panel, Figure 4) were above 80% for seven participants and remained above 80% following training. Max's percent agreement was 75% during baseline. Following training, Max's percent agreement increased to 91.3%.

Figure 5 (upper panel) depicts percent agreement on occurrences of demands preceding problem behavior. Levels of agreement were above 80% during baseline for five participants. Max, Jan, and Tia had moderate levels of agreement (78.6%, 60.5%, and 60%, respectively). Following training, levels of agreement were above 90% for all participants.

Percent agreements on occurrences of escape following problem behavior are depicted in Figure 5, lower panel. During baseline, low to moderate levels of agreement were observed (range, 0–75%). Following training, accuracy increased across all participants. For two participants (Cali and Amy), increases were small. Moderate to large increases in accuracy were observed for the other six participants (average, 49.4 percentage points improvement, range, 12.3–86.3 percentage points). Despite gains for participants, only three participants had greater than 80% agreements following training.

All participants had lower levels of agreement on the occurrences of environmental events that involved the absence of stimuli (low attention and escape) than environmental events that involved the presence of stimuli (attention and demand). Participants more accurately identified presentation of demands and attention than removal of those events (escape and low attention), even during baseline. Although agreement for escape and low attention increased following training, few participants met mastery criteria during post-training sessions (one participant for low attention and three participants for escape).

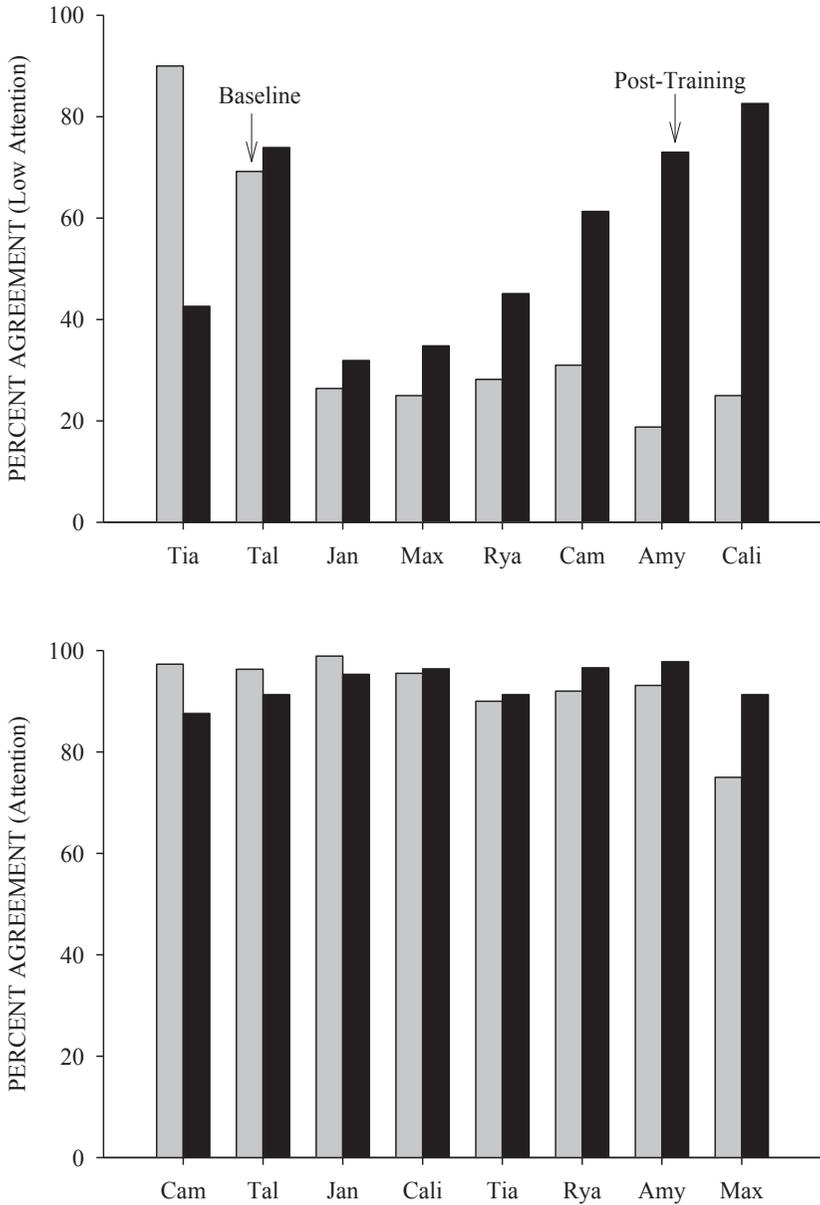


Figure 4. Percent agreement on proportion of problem behavior preceded by low attention (top panel) and following by attention (bottom panel). Baseline agreements are depicted by the gray bars. Post-training agreements are depicted by the black bars.

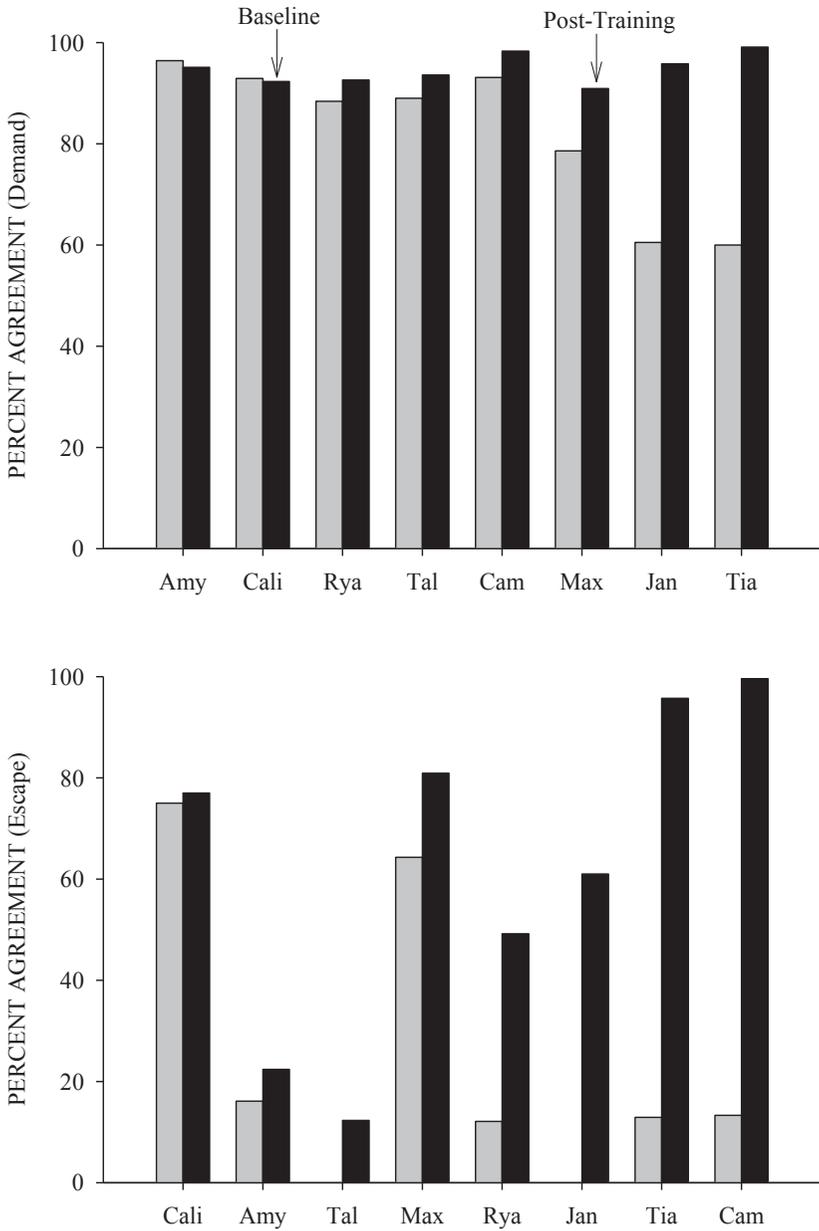


Figure 5. Percent agreement on proportion of problem behavior preceded by demand (top panel) and following by escape (bottom panel). Baseline agreements are depicted by the gray bars. Post-training agreements are depicted by the black bars.

We allowed participants to select the recording sheet to use on the third baseline session after being exposed to the structured ABC recording sheet and the narrative ABC recording sheet. Two participants (Tia and Cam) selected the narrative ABC recording sheet during the third baseline session and six participants selected the structured ABC recording sheet. Participants selected which recording sheet to use during all post-training sessions. Four participants (Tia, Rya, Jan, and Tal) used the narrative ABC recording sheet throughout post-training. Three participants (Cam, Cali, and Amy) used the structured ABC recording sheet throughout post-training. One participant (Max) used both the structured and narrative ABC recording sheets. The format of the data sheet used did not seem to affect the teachers' data-collection accuracy.

General Discussion

The current study investigated the accuracy with which teachers collected descriptive data on naturally occurring caregiver-child interactions. Although previous studies have suggested that accuracy may improve with repeated practice (Lerman et al., 2009), we did not obtain consistent evidence of practice effects in Experiment 1. In addition, teachers with more years of experience did not consistently perform better than teachers with relatively little teaching experience. For example, Laura had 26 yr of experience, but did not perform any better than our most junior teacher, Kari, who had 2 yr of experience (refer to Figure 2). We identified consistent errors made across teachers, in that all participants had more difficulty identifying the absence of events (periods of low attention and escape) than the presence of events (attention and presentation of demands). Further research should evaluate if other trainings can lead to improvements in performance during descriptive assessments. For example, perhaps competency in implementing behavioral procedures such as functional analyses, reinforcement procedures, and extinction could help improve behavioral observation skills.

Fourteen participants (seven in Experiment 1 and seven in Experiment 2) exclusively chose one type of recording method (structured or narrative) during additional sessions (Experiment 1) or during post-training sessions (Experiment 2). Of these participants, nine participants consistently selected the structured recording sheet and five selected the narrative recording sheet. Three participants used both the narrative and structured recording methods. We did not find a clear relation between the type of recording method selected and the

accuracy with which participants collected data. Yet, it remains possible that one method may result in more rapid acquisition of data-collection skills or more accurate or easily interpretable data. Future research should examine the rate at which teachers acquire data-collection skills using structured and narrative formats, as well as the extent to which teachers can be efficiently taught to interpret the data they have collected.

Lerman et al. (2009) identified a slight preference for structured recording methods based on the verbal report of their participants. Our data provide further support for the existence of teacher preferences for recording format, but we used a direct measure of behavior rather than verbal reports. Behavioral measures may provide more accurate demonstrations of preference by examining selection instead of verbal reports. However, future research should examine the correlation between selection responses and verbal reports of preference. If direct selection and verbal report are highly correlated, assessing verbal report may be a more rapid way to assess teacher preference.

One limitation of the current study was a lack of experimental design. Because the current study was a data-based case study, we cannot draw conclusions about causation, only discuss correlational observations. Experiment 1 was a descriptive evaluation of repeated practice over time and Experiment 2 used a pre- and post-training design to evaluate the effects of an automated PowerPoint training. The purpose of the current study as a case study was to evaluate a simple way to train teachers to collect descriptive-assessment data that would require minimal resources and could be immediately implemented by a school district. The training evaluated in the current study exceeds the standard didactic training that would occur during an in-service for teachers during which teachers would have minimal (if any) opportunities to practice. Despite having repeated opportunities to practice and receiving feedback during the trials within the automated training evaluated in Experiment 2, teachers were still unable to reach levels (at or above 80% agreement) demonstrating mastery of collecting descriptive-assessment data. If teachers are going to collect descriptive data during FBAs, it seems imperative that teachers receive training to a mastery criterion. Results from the current case study suggest that teachers require more systematic training to be able to collect descriptive-assessment data at mastery levels. Our findings also have implications for the use of teacher-collected descriptive data during the FBA process in that those descriptive data may not be accurate representations of the naturally occurring events to be able to make well-informed intervention decisions. Future research should use rigorous experimental designs to evaluate the necessary and suf-

ficient training components for teachers to acquire the skills necessary for accurate collection of descriptive data.

The results of the current study suggest possible directions for training descriptive-assessment data collection. Providing opportunities for teachers to practice collecting data during naturally occurring interactions did not lead to improvement in teachers' accuracy. Discrete-trial practice and feedback resulted in improvements in data-collection accuracy; however, automated practice and feedback did not increase accuracy above 80% across all participants. In addition, teachers participate in the FBA process and may be involved in the identification of target responses, development of operational definitions, and decisions regarding data-collection procedures. Based on the results of the current study, we cannot make conclusions if teacher's role in the FBA process influences his or her data-collection accuracy.

Our data suggest that teachers may require additional opportunities to practice, more explicit feedback, training to a performance criterion, and in-situ training. Future research could use the results of the current study as preliminary data to inform a training package for descriptive-analysis data collection that can be evaluated using an experimental design. In addition, future research should emphasize opportunities, practice, and feedback around detection of the absence of environmental events, including identifying periods of diverted or no attention and escape from demands. The absence of events appeared to be substantially more difficult for teachers to acquire in the current study compared to detecting the presence of events (attention, demands).

The current study focused on teacher acquisition of data-collection skills, but did not attempt to train teachers to analyze data after they had been collected. We focused on accurate data collection because having accurate data is a prerequisite for conducting a meaningful analysis. However, the extent to which teachers can rapidly acquire data-analysis skills for narrative and descriptive data remains unknown. Future studies should determine the potential impact of data-collection format on the acquisition and maintenance of data-analysis skills.

References

- Cipani, E., & Schock, K. M. (2007). *Functional behavioral assessment, diagnosis, and treatment: A complete system for education and mental health settings*. New York, NY, US: Springer Publishing Co.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ, US: Pearson Education, Inc.

- Crone, D. A., Hawken, L. S., & Bergstrom, M. K. (2007). A demonstration of training, implementing, and using functional behavioral assessment in 10 elementary and middle schools. *Journal of Positive Behavior Interventions, 9*, 15–29. doi:10.1177/10983007070090010301
- Ellingson, S. A., Miltenberger, R. G., Stricker, J., Galensky, T. L., & Garlinghouse, M. (2000). Functional assessment and intervention for challenging behaviors in the classroom by general education teachers. *Journal of Positive Behavior Interventions, 2*, 85–97. doi:10.1177/109830070000200202
- Kern, L., Hilt, A. M., & Gresham, F. (2004). An evaluation of the functional behavioral assessment process used with students with or at risk for emotional and behavioral disorders. *Education and Treatment of Children, 27*(4), 440–452.
- Lalli, J. S., Browder, D. M., Mace, F. C., & Brown, D. K. (1993). Teacher use of descriptive analysis data to implement interventions to decrease students' problem behaviors. *Journal of Applied Behavior Analysis, 26*, 227–238. doi:10.1901/jaba.1993.26–227
- Lerman, D. C., & Iwata, B. A. (1993). Descriptive and experimental analysis of variables maintaining self-injurious behavior. *Journal of Applied Behavior Analysis, 26*, 292–319. doi:10.1901/jaba.1993.26–293
- Lerman, D. C., Hovanetz, A., Strobel, M., & Tetreault, A. (2009). Accuracy of teacher-collected descriptive analysis data: A comparison of narrative and structured recording formats. *Journal of Behavioral Education, 18*, 157–172. doi:10.1007/s10864–009–9084–7
- Oliver, A. C., Pratt, L. A., & Normand, M. T. (2015). A survey of functional behavior assessment methods used by behavior analysts in practice. *Journal of Applied Behavior Analysis, 48*, 817–829. doi:10.1002/jaba.256
- Pence, S. T., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2009). Relative contributions of three descriptive methods: Implications for behavioral assessment. *Journal of Applied Behavior Analysis, 42*, 425–446. doi:10.1901/jaba.2009.42–425
- Rolider, N. U., Iwata, B. A., & Bullock, C. E. (2012). Influences of response rate and distribution on the calculation of interobserver reliability scores. *Journal of Applied Behavior Analysis, 45*, 753–762. doi:10.1901/jaba.2012.45–753
- Roscoe, E. M., Phillips, K. M., Kelly, M. A., Farber, R., & Dube, W. V. (2015). A statewide survey assessing practitioners' use and

- perceived utility of functional assessment. *Journal of Applied Behavior Analysis*, *48*, 830–844. doi:10.1002/jaba.259
- Samaha, A. L., Vollmer, T. R., Borrero, C., Sloman, K., St. Peter Pipkin, C., & Bourret, J. (2009). Analyses of response-stimulus sequences in descriptive observations. *Journal of Applied Behavior Analysis*, *42*, 447–468. doi:10.1901/jaba.2009.42-447
- Thompson, R. H., & Borrero, J. C. (2011). Direct observation. In W. W. Fisher, C. C. Piazza, H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 191–205). New York, NY: Guilford Press.
- Thompson, R. H., & Iwata, B. A. (2007). A comparison of outcomes from descriptive and functional analyses of problem behavior. *Journal of Applied Behavior Analysis*, *40*, 333–338. doi:10.1901/jaba.2007.40-333
- Weber, K. P., Killu, K., Derby, K. M., & Barretto, A. (2005). The status of functional behavior assessment (FBA): Adherence to standard practice in FBA methodology. *Psychology in the Schools*, *42*, 737–734. doi:10.1002/pits.20108
- Vollmer, T. R., Borrero, J. C., Wright, C. S., Van Camp, C., & Lalli, J. S. (2001). Identifying possible contingencies during descriptive analyses of severe behavior disorders. *Journal of Applied Behavior Analysis*, *34*, 269–287. doi:10.1901/jaba.2001.24-269

Copyright of Education & Treatment of Children is the property of West Virginia University Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.