Using modified visual-inspection criteria to interpret functional analysis outcomes

Article *in* Journal of Applied Behavior Analysis · March 2013

DOI: 10.1002/jaba.13 · Source: PubMed

5 authors, including:

H. S. Roane
State University of New York Upstate Medical University

Wayne W Fisher
Rutgers, The State University of New Jersey

Michael E Kelley
Florida Institute of Technology

Joanna Lomas Mevers
Emory University

Journal of Applied Behavior Analysis
JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2013, **46,** 130–146 NUMBER 1 (SPRING 2013)

USING MODIFIED VISUAL-INSPECTION CRITERIA TO INTERPRET FUNCTIONAL ANALYSIS OUTCOMES

HENRY S. ROANE

UPSTATE MEDICAL UNIVERSITY

WAYNE W. FISHER AND MICHAEL E. KELLEY

MUNROE-MEYER INSTITUTE AND UNIVERSITY OF NEBRASKA MEDICAL CENTER

JOANNA L. MEVERS

LOUISIANA STATE UNIVERSITY

AND

KELLY J. BOUXSEIN

NEBRASKA DEPARTMENT OF HEALTH AND HUMAN SERVICES

The development of functional analysis (FA) methodologies allows the identification of the reinforcers that maintain problem behavior and improved intervention efficacy in the form of function-based treatments. Despite the profound impact of FA on clinical practice and research, questions still remain about the methods by which clinicians and researchers interpret FA graphs. In the current study, 141 FA data sets were evaluated using the structured visual-inspection criteria developed by Hagopian et al. (1997). However, the criteria were modified for FAs of varying lengths. Interobserver agreement assessments revealed high agreement coefficients across expert judges, postdoctoral reviewers, master's-level reviewers, and postbaccalaureate reviewers. Once the validity of the modified visual-inspection procedures was established, the utility of those procedures was examined by using them to categorize the maintaining reinforcement contingency related to problem behavior for all 141 data sets and for the 101 participants who contributed to the 141 data sets.

Key words: functional analysis, visual inspection, data analysis, problem behavior

The development of the functional analysis (FA) methodology (Iwata, Dorsey, Slifer, Bauman, & Richman, 1982/1994) to assess the environmental influences on destructive problem behavior is one of the key advancements in applied behavior analysis over the last 30 years (see Hanley, Iwata, & McCord, 2003). Since its inception, FA methods have been applied to a wide range of problem behavior disorders and have repeatedly proven to be useful at identifying maintaining reinforcement contingencies. This assessment methodology has, in turn, fostered the development of effective, reinforcement-based treatments for problem behavior.

Recently, researchers have examined methods for training nonprofessionals in the implementation of FA methods. For example, Moore and Fisher (2007) trained staff in the implementation of FA methods using a combination of didactic instruction and video modeling. Similarly, Iwata et al. (2000) conducted training in FA methods with undergraduate students who had no prior history of FA implementation. In both cases, nonprofessionals acquired skills relevant to implementation of an FA in a relatively brief period of time.

We thank the reviewers who participated in this investigation and the staff who conducted the functional analyses that comprised this data set. Michael Kelley is now at the Scott Center for Autism Treatment at the Florida Institute of Technology. Joanna Mevers is now at the Marcus Autism Center.

Address correspondence to Henry S. Roane, Department. of Pediatrics, Upstate Medical University, 600 E. Genesee St., Ste. 124, Syracuse, New York 13210 (e-mail: roaneh@upstate.edu).

doi: 10.1002/jaba.13

However, during the process of conducting an FA, it is also important to master the skills required to visually inspect and interpret FA data accurately. Visual inspection of FA data is critical for identifying a functional relation between the independent variable manipulated in one or more test conditions and the occurrence of problem behavior in those test conditions. Although visual analysis generally is considered to be a conservative method for interpreting quantitative data (Baer, 1977; Parsonson & Baer, 1986), empirical studies have shown that data interpretations based on visual inspection often are subjective and vary across clinicians (DeProspero & Cohen, 1979; Jones, Weinrott, & Vaught, 1978; Kahng et al., 2010). As a result, several investigators have developed and evaluated more formal (or structured) methods for training individuals in visual inspection of single-case data (e.g., Bailey, 1984; Fisher, Kelley, & Lomas, 2003; Hagopian et al., 1997; Rojahn & Schulze, 1985).

Given that functional analyses are typically conducted in a multielement single-case experimental design, functional analyses are typically interpreted using a four-point interpretation of the data both within and across conditions (Betz & Fisher, 2011). This includes analysis of the number of data points within a specific condition or phase, variability of data points, level of data, and the direction and degree of trends. All of these elements are considered when determining if sufficient experimental control has been demonstrated in an FA.

In a noteworthy investigation on this topic, Hagopian et al. (1997) developed a set of structured criteria to facilitate consistent and accurate interpretation when conducting visual inspection of FA outcomes. First, three psychology intern raters were given data from previously conducted functional analyses. These data were corrected such that each FA test condition consisted of 10 data points. Initially, the raters had no criteria on which to base their visual inspection (with the exception of their prior graduate training in visual inspection). Following baseline data collection with the raters, an expert panel convened to obtain consensus interpretations of the FA data sets via visual inspection. Next, structured interpretation criteria were developed that produced similar interpretations of the FA data sets relative to those of the expert panel. After these visual-inspection criteria had been developed, the experimenters used didactic instruction to train the interns to use these criteria. Specifically, they taught raters to compare the number of data points in a condition that fell above and below specific criterion lines and used the expert-developed guidelines to determine if there was differentiation between each test condition and the control condition.

Prior to training, interinterpreter agreement coefficients averaged .46 among the three raters. In contrast, when the raters used the structured criteria, agreement increased to .90. In addition, agreement was .94 among members of the expert panel who applied the structured criteria to a set of FA data. These data suggested that highly structured, expert-derived visual-inspection criteria could be useful for conducting visual inspection of FA data. However, one limitation of the procedures developed by Hagopian et al. (1997) was that the criteria were developed specifically and uniquely for FA data sets that included exactly 10 data points per condition. Thus, the utility of these visual-inspection procedures for FAs of varying lengths (e.g., six sessions per condition) remains unknown.

Visual inspection of FA data permits the categorization of FA outcomes across various reinforcement contingencies (e.g., social-positive reinforcement, automatic reinforcement). Several previous investigations have categorized FA outcomes that were conducted across a relatively large number of participants. For example, Iwata et al. (1994) assessed FA outcomes for 152 individuals who exhibited severe self-injurious behavior (SIB). Iwata et al. found that differentiated FA outcomes (i.e., increased levels of SIB in one test condition relative to the control condition) were obtained in 95.4% of the analyses. In a review of published FA data sets, Hanley et al. (2003) noted that FAs identified maintaining reinforcement contingencies in 95.9% of published studies, which varied across a range of problem behaviors (e.g., SIB, aggression, property destruction). Similarly, Kurtz et al. (2003) observed differentiated FA outcomes in 87.5% of a clinical sample of young children who displayed various types of problem behavior (although differentiated outcomes were obtained in 62.1% of cases in which SIB was the behavior of interest). Likewise, Asmus et al. (2004) noted that FAs identified maintaining reinforcement contingencies in 96% of cases referred to an outpatient clinic, whereas Mueller, Nkosi, and Hine (2011) demonstrated that maintaining reinforcers were identified in 90% of school-based functional analyses. In contrast to these investigations, a recent analysis conducted by Hagopian, Rooker, Jessel, and DeLeon (2013) found that an initial standardized multielement FA effectively identified a maintaining reinforcer in 46.6% of a clinical sample. It should be noted, however, that when modifications were made to the procedures (e.g., analyses that examined the influences of idiosyncratic reinforcement contingencies) the percentage of cases for whom maintaining reinforcers were identified increased to 86.9%, a figure that is more in line with previous findings.

The purpose of the current paper was twofold. First, we attempted to modify the visual-inspection criteria described by Hagopian et al. (1997) and train reviewers in the use of these criteria for application to FAs of varying lengths. Using these procedures, the second purpose of this investigation was to determine the extent to which the categorization of FA outcomes obtained from the current clinical sample were comparable to those obtained in previous studies that presented FA outcomes across a large number of individuals.

GENERAL METHOD

Participants

Each data set was evaluated initially by four expert judges; one of the experts reviewed all of the data sets, and the other expert judges each reviewed a portion (e.g., 50%) of the data sets. Expert judges possessed a PhD in behavior analysis or a related field and had been practicing at the postdoctoral level for a minimum of 5 years. All expert judges were Board Certified Behavior Analysts (BCBA) and were currently or had previously been members of the Editorial Board for the *Journal of Applied Behavior Analysis* (*JABA*). In addition, two of the expert judges previously had held advanced editorial positions with *JABA*, and one had been a coinvestigator in the development of structured visual-inspection criteria conducted by Hagopian et al. (1997). The judges' experience in use of visual inspection for interpreting single-case design data and FA outcomes ranged from 10 to 28 years.

A separate panel that consisted of (a) two postdoctoral behavior analysts (Study 1) or (b) two master's-level trainees and three postbaccalaureate behavior therapists (Study 2) also evaluated the data sets. The postdoctoral behavior analysts had received a PhD in either school psychology with an emphasis in applied behavior analysis or in applied behavior analysis. Both postdoctoral reviewers had over 5 years of experience conducting FAs and applying visual-inspection techniques. In addition, both had been trained at programs that have long histories of producing active FA research. Finally, each of the postdoctoral reviewers had been the lead author on at least one paper published in *JABA*. One of the postdoctoral reviewers participated in this investigation following 10 months of a 12-month postdoctoral supervised practice, and the second postdoctoral reviewer participated in this investigation within the first 3 months of her 12-month postdoctoral appointment.

The master's-level reviewers each had a minimum of 2 years of graduate study in applied behavior analysis, and both were completing a field experience requirement for eventual certification as a BCBA. Both of the master's-level reviewers had completed at least 6 months of field experience prior to their participation. As part of their field experience, both had served as the lead therapist for at least two FAs. In addition, the master's-level reviewers received daily individual (10 to 15 min per meeting) and group (30 min per meeting) supervision on case-related activities that included ongoing evaluations of single-case research designs. However, with the exception of their graduate training, they had received no formal instruction in visual inspection of single-case data as part of their field experience (with the exception of their participation in this study).

The postbaccalaureate reviewers each possessed a bachelor's degree in psychology or a related field. Each postbaccalaureate reviewer was employed as a behavior therapist and had been working in an outpatient clinic that specialized in the assessment and treatment of problem behavior for a minimum of 6 months. During that time, they had participated in multiple single-case evaluations that varied relative to the needs of the clinic's clients. Thus, all postbaccalaureate reviewers had some experience in conducting FAs, although the length of experience was uncontrolled. Likewise, these reviewers attended the same daily 30-min group supervision as the master's-level reviewers, but did not participate in routine individual case supervision. Consequently, their exposure to visual inspection of single-case designs was limited to the daily discussions of client behavior during group supervision meetings.

Data Sets

The initial database for this investigation consisted of 182 multielement FAs that had been conducted at two different outpatient clinics that specialized in the assessment and treatment of severe problem behaviors. The data sets were collected from a total of 126 participants who ranged in age from 3 to 23 years old ($M = 10.8$ years old). The sample consisted primarily of males (79.4%). Diagnoses included pervasive developmental disorders (PDD; autism, Asperger's disorder, PDD-NOS), varying degrees of intellectual disabilities, psychological conditions such as attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and obsessive compulsive disorder (OCD), genetic conditions (e.g., Cornelia de Lange), traumatic brain injury, and no known diagnosis. Of the participants who comprised the sample (described below), the majority (93.1%) had been diagnosed with either a developmental disability (e.g., autism, intellectual disability), a genetic disorder that was associated with developmental disabilities (e.g., Down syndrome), or traumatic brain injury.

Thirty-two of the participants contributed multiple data sets to the original sample. For those 32 participants, the average number of FAs conducted was 2.7 (range, 2 to 5). The reasons for conducting multiple FAs for some participants varied, but usually consisted of the participant displaying multiple topographies of problem behavior that were unlikely to be in the same response class (e.g., pica and aggression), new topographies of problem behavior emerging during one FA, and modifications of previous operational definitions for targeted behaviors.

Data sets were excluded if they involved idiosyncratic test conditions (e.g., social escape, interruption of ongoing high-probability activities; 32 data sets, 17.6% of the sample), if the FA evaluated factors other than problem behavior (e.g., medication assessment, the use of protective equipment, environmental modifications; six data sets, 3.3% of the sample), or if no target behavior was observed during the FA (three data sets, 1.6% of the original sample). Once these data sets were removed, 141 data sets remained and were included for further evaluation.

The 141 data sets included FA results from 101 participants. These FAs had been conducted in a manner similar to that described by Iwata et al. (1982/1994), with modifications as described by Fisher, Piazza, and Chiang (1996). Each FA included an average of 7.1 exposures to each test condition (range, 2.8 to 20). The mean length of the multielement functional analyses was 33.8 sessions (range, 14 to 100), and each FA session was 10 min. We categorized each FA according to arbitrary lengths, including relatively short functional analyses (four or fewer exposures per condition, 27.3% of the sample); medium-length functional analyses (five to seven exposures per condition, 34.5%); long functional analyses (8 to 10 exposures per condition, 25.9%); and excessively long functional analyses (11 or more exposures per condition, 12.3%).

Of the 141 data sets, 53 included an extended analysis (e.g., pairwise comparison) based on the model suggested by Vollmer, Marcus, Ringdahl, and Roane (1995). It should be noted that data from the extended analyses were not presented to the reviewers for their initial categorization. The purpose of these comparisons was either to confirm the results of the multielement assessment (e.g., to confirm the role of two or more variables in the maintenance of problem behavior via a pairwise comparison; $n = 48$) or to evaluate the role of automatic reinforcement following an inconclusive multielement FA via a series of consecutive alone or ignore sessions ($n = 5$).

The mean number of behaviors that were targeted for reinforcement in each FA was 1.6 (range, 1 to 5). The prevalence of target behaviors across all functional analyses was as follows: aggression (e.g., hitting, kicking, or biting others, 36.4%), SIB (e.g., self-hitting, head banging, or self-biting, 27.6%), disruption or property destruction (e.g., throwing, banging on, or breaking objects, 20.9%), inappropriate vocalizations (e.g., screaming or cursing, 5.8%), dropping or elopement (e.g., lying down on the floor or ground during periods of ambulation, running away from others, or leaving an area without permission, 4.4%), stereotypy (e.g., hand flapping, object flapping, 1.8%), pica (1.3%), and masturbation (less than 1%).

Materials

For the purpose of visual inspection, each reviewer was presented with an electronic file that contained a graph of one FA in the data set and a score sheet. Each graph was created in Microsoft Excel, and one graph was presented in each Excel file. Each graph was plotted on an x–y scale, with the dependent variable plotted along the y axis and sessions plotted along the x axis. The label for each condition of the FA was provided in a figure legend on each graph. The graphs were deidentified to remove any information related to particular clients (e.g., unique client names), but each graph was given a number that corresponded to a line on the score sheet.

The score sheet also was developed in Excel and listed each graph number and columns that were labeled for each potential FA outcome. First, the reviewer was asked to type a Y (yes) or N (no) in response to whether the FA revealed a clear function. Next, in response to a Y answer, the reviewer was asked to "identify the functions by highlighting or bolding the appropriate functions." All of the potential functions (attention, escape, tangible, automatic) were listed in separate columns such that multiple functions could be identified. As in Hagopian et al. (1997), there were 12 possible categories for each FA outcome: (a) undifferentiated, (b) maintained by attention, (c) maintained by escape from demands, (d) maintained by tangible reinforcement, (e) maintained by automatic reinforcement, (f) maintained by attention and escape, (g) maintained by attention and tangible reinforcement, (h) maintained by tangible reinforcement and escape, (i) maintained by automatic reinforcement and escape, (j) maintained by automatic reinforcement and attention, (k) maintained by automatic and tangible reinforcement, and (l) maintained by attention, tangible reinforcement, and escape. Each file was protected such that it was not possible for the reviewer to modify the contents without a password (which was withheld from all reviewers, with the exception of one expert judge).

Development of Modified Visual-Inspection Criteria

The visual-inspection criteria were based on those described by Hagopian et al. (1997). Recall that the Hagopian et al. criteria were based on FA data sets that consisted of 10 points per condition, whereas the current sample consisted of FAs of varying lengths. Thus, the primary modification of the Hagopian et al. procedures involved converting the visual-inspection criteria to a percentage of data points as opposed to a specific number of data points (although the rules for low-rate behavior were not modified from those presented by Hagopian et al.). To illustrate, in Hagopian et al., upper and lower criterion lines (CL) were determined for each graph by drawing a horizontal line one standard deviation above and one standard deviation below the mean of the toy play condition. Thus, for functional analyses with 10 data points per condition, the upper CL was drawn between the second and third highest (upper CL) or lowest (lower CL) data points from the toy play condition. The resulting CLs were used to assess differentiation among conditions. In the current approach, this method was modified by developing a formula in Excel and sourcing the data from the toy play condition such that the condition mean plus the standard deviation (above or below the mean) were plotted on each graph. Thus, because the number of data points varied per FA in the current data set, the upper and lower CLs were not consistently drawn between the second and third highest (or lowest) data points (specific instructions for sourcing the CLs in Excel are available from the first author).

As in Hagopian et al. (1997), participants determined differentiation among conditions by counting the number of data points for each test condition that fell either above the upper CL or below the lower CL. However, Hagopian et al. concluded that a condition was differentiated from the toy play condition if "at least five more data points from a test condition fall above the upper CL than fall below the lower CL" (p. 324). In the current modified procedures, the rule for differentiation in a test condition was adapted such that 50% or more of the data points fell above the upper CL than fell below the lower CL. Similar modifications were made for visual-inspection criteria that related to low-magnitude effects, trends in the data, automatic reinforcement, and multiple maintaining reinforcers. The final version of the modified visual-inspection criteria is presented in the Appendix.

To illustrate, Figure 1 depicts FA Data Set 74, for which a tangible function was identified using the modified visual-inspection criteria. On this graph, the dashed horizontal line is the upper CL and the horizontal dotted line is the lower CL (which is equivalent to zero for this data set because one standard deviation below the lower mean for this data set would have resulted in a negative number; see the Appendix for considerations related to lower CL determination). The tangible condition included 11 data points. Of these, nine are above the upper CL and two are at or on the lower CL. Using the equation presented in the Appendix, the number of data points below the lower CL (2) is subtracted from the number of data points above the upper CL (9) which is then divided by the total number of data points (11). Using this formula, 64% of the data points from the tangible condition were above the upper CL, suggesting that the FA was differentiated for

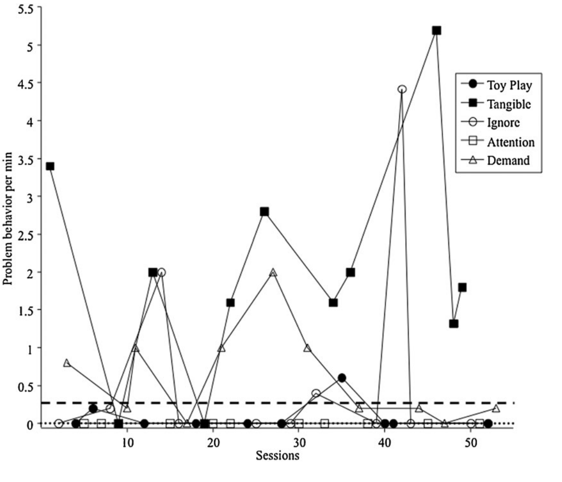

Figure 1. Example of an FA showing differentiated responding in the tangible condition. Horizontal dashed line represents the upper CL, and horizontal dotted line represents the lower CL.

this condition. In contrast, application of the modified visual-inspection criteria to the demand condition produced undifferentiated outcomes in that five data points fell above the upper CL and two fell at the lower CL. Using the formula presented in the Appendix for the 11-point demand condition, only 27% of the data points from the demand condition were above the upper CL.

Figures 2 and 3 also depict illustrative applications of the modified visual-inspection criteria. The top panel of Figure 2 shows an example of the criteria applied to the identification of an automatic reinforcement function. An automatic reinforcement function was identified if the relevant test condition (alone or ignore) was the highest overall condition, if all conditions were high and relatively stable, or if behavior tended to be higher in conditions with less external stimulation (e.g., alone, attention, tangible). As shown in Figure 2 (top), automatic reinforcement was identified because, although the ignore, tangible, and attention conditions were elevated relative to the upper CL (depicted by the dashed line and identical to the lower CL in this case given the nonoccurrence of problem behavior in the toy play condition), the ignore condition was associated with the highest overall occurrence of problem behavior.

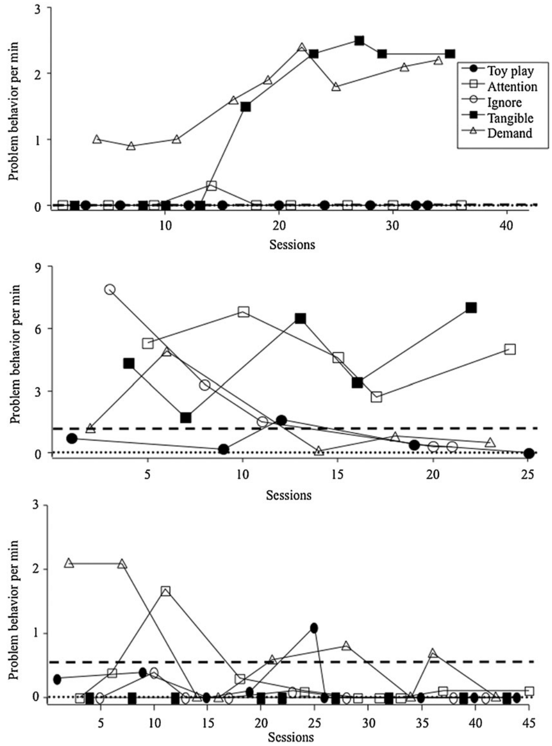

Figure 3. Examples of an FA that demonstrates the use of the upward trend criteria (top), the downward trend criteria (middle), and an undifferentiated outcome (bottom). Horizontal dashed line represents the upper CL, and horizontal dotted line represents the lower CL.

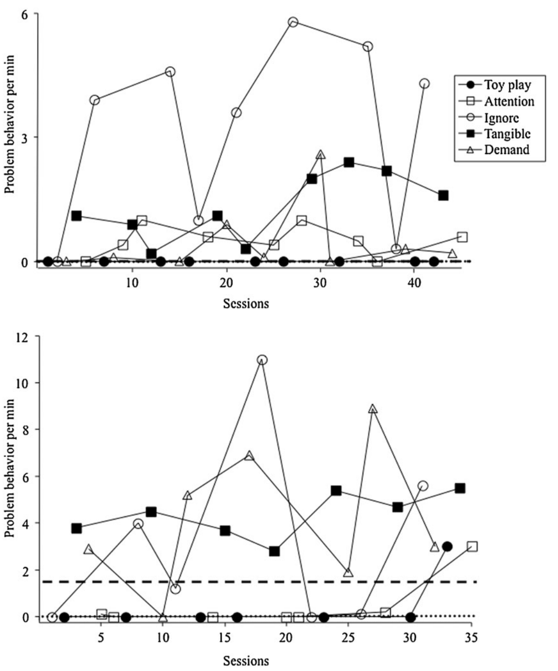

Figure 2. Examples of an FA that meets the criteria for automatic reinforcement (top) and for multiple control (bottom). Horizontal dashed line represents the upper CL, and horizontal dotted line represents the lower CL.

Figure 2 (bottom) shows an illustrative outcome of a multiply controlled FA. The general criterion for multiple control was that more than one condition met the basic criteria for differentiation (see the Appendix for additional criteria for determining multiple control). Both the tangible and demand conditions are differentiated here.

Figure 3 (top and middle) shows representative examples of cases in which additional criteria were applied for upward trends (top) and downward trends (middle). When applying the initial modified visual-inspection criteria to the data depicted in Figure 3 (top), only the demand condition is differentiated (note again in this case that the upper and lower CLs are identical due to the nonoccurrence of problem behavior in the toy play condition). However, application of the upward trend criteria further implicates the role of tangible reinforcement, such that this analysis would be scored as multiply controlled. In Figure 3 (middle), the attention, tangible, and ignore conditions each separately meet the initial visual-inspection criteria. However, the ignore condition fails to satisfy the downward trend criteria. Thus, attention and tangible functions are implicated in the results of this FA. Figure 3 (bottom) depicts an outcome in which no condition was identified as differentiated. That is, none of the established criteria were met when applied to these data.

Once the modified visual-inspection procedures were developed, interobserver agreement coefficients were calculated among all possible pairs of expert judges. Exact agreement coefficients were calculated by dividing the number of exact agreements (i.e., both judges identifying the same function from a range of potential combinations of functions) by the number of exact agreements plus disagreements. Thus, if both judges agreed on the presence of one function (e.g., attention) but disagreed on the presence of another function (e.g., one reviewer noted an escape function and the other did not), a disagreement was recorded for that data set. The average level of exact agreement among expert judges using the modified visual-inspection criteria was .91. Although this initial level of agreement was high, the expert judges subsequently met to review all points of discrepancy. Once identified, the expert judges reviewed the sources of discrepancies and agreed to any necessary modifications or clarifications of the visual-inspection criteria. Disagreements primarily occurred as a result of errors in calculation (e.g., miscounting the number of data points that fell above or on or below a CL), errors in attending to all criteria (e.g., ignoring criteria for multiple control), lack of consensus regarding the definition of *highest* for the alone or ignore condition, and lack of consensus regarding the criteria for upward trend. After each disagreement was addressed and the visual-inspection criteria were modified accordingly, average agreement increased to .99.

Study 1: Evaluation of Interobserver Agreement with and without Modified Structured Criteria

Reviewers. The reviewers for Study 1 included the two postdoctoral behavior analysts (described above). In addition, the postdoctoral reviewers' interpretations were compared to one of the expert judge's interpretations. The specific expert judge for this study had 16 years of clinical and research experience in FA methodology and related visual inspection.

Procedure. The postdoctoral reviewers were presented with two sets of graphs. One set of graphs consisted of the 141-sample FA data sets without CLs, and the participants were not given the modified visual-inspection criteria. Initially, the reviewers were instructed to apply traditional visual-inspection procedures to interpret each FA outcome. The initial set of 141 graphs were re-presented to the reviewers; however, during the second presentation the set contained predrawn CLs, and the reviewers were given the modified visual-inspection criteria. Although the reviewers had access to the modified visual-inspection criteria, they received no specific instruction in the use of those criteria. The reviewers also were given two score sheets (described above), one for each set of graphs. The FA outcomes were categorized according to the 12 possible interpretations described above.

Results and discussion. Agreement across postdoctoral reviewers was assessed using exact agreement coefficients as described above. When asked to interpret FA outcomes without the use of the modified visual-inspection criteria, the exact agreement between the two postdoctoral reviewers was .66. In contrast, when the visual-inspection criteria were employed, the exact agreement coefficient between the two postdoctoral reviewers increased to .92. Furthermore, when the postdoctoral reviewers and the expert judge used the modified visual-inspection criteria to categorize FA outcomes, the average mean exact agreement between the postdoctoral reviewers and the expert judge was .92.

FA interpretations also were examined within each postdoctoral reviewer by assessing the exact agreement with and without the modified visual-inspection criteria. For one postdoctoral reviewer, the exact agreement between interpretations with and without visual-inspection criteria was .79; for the other postdoctoral reviewer, exact agreement with and without the modified visual-inspection criteria was .66.

These results suggested that the modified visual-inspection criteria were useful in increasing agreement between and within reviewers in the interpretation of FA outcomes. These results were similar to those reported by Hagopian et al. (1997) and suggested that the modified visual-inspection criteria possess good criterion validity because the postdoctoral reviewers' interpretations shared a high level of agreement with those of the expert judge. However, these results differed from those of Hagopian et al. in that the reviewers in this study were advanced behavior analysis trainees. Thus, the extent to which the modified visual-inspection criteria would facilitate interobserver agreement among less experienced reviewers was unknown. In addition, the postdoctoral reviewers did not receive any direct instruction in the use of the modified visual-inspection procedures. It is possible that the relative expertise of the reviewers and the addition of specific instruction in the use of visual-inspection criteria could have influenced the level of interobserver agreement. Therefore, the purpose of Study 2 was twofold. First, we evaluated the use of the modified visual-inspection criteria across two groups of individuals with relatively advanced (master's-level) or relatively moderate (postbaccalaureate) previous training in FA and visual inspection. Second, we assessed levels of interobserver agreement when visual-inspection criteria were presented with and without specific instruction.

Study 2: Evaluation of Interobserver Agreement across Reviewers who Lacked Advanced Expertise in Visual Inspection

Reviewers. The reviewers for Study 2 included the two master's-level students and the three postbaccalaureate employees described above. In addition to comparing the reviewers' interobserver agreement across all possible pairs within a respective category of experience (master's or bachelor's level), we also compared each reviewer's interpretations to the interpretations of one expert judge. The expert judge for Study 2 was the same as for Study 1.

Procedure. The master's-level and postbaccalaureate reviewers were presented with the graphs that contained predrawn CLs and score sheets (described above). In addition, these reviewers also were given access to the written modified visual-inspection criteria as presented in the Appendix. The purpose of providing access to the written criteria was to assess the level of interobserver agreement when the reviewers were provided with these materials in the absence of instruction on their use. During pretraining, the reviewers had continuous access to the visual-inspection criteria while they reviewed each of the 141 sample graphs and provided their interpretation of each graph on the score sheet. Although the reviewers had access to the criteria, we did not control whether they read or applied the criteria during pretraining.

Following an initial review of the data, one of the expert judges trained the reviewers on the use of the modified visual-inspection criteria. The training procedures followed those described by

Hagopian et al. (1997) and consisted of didactic instruction, modeling, and practice with immediate feedback. The reviewers completed training after correctly applying the modified visual-inspection criteria with 100% accuracy to a random sample of five graphs from the 141-graph sample. Training across reviewers required approximately 90 min. During posttraining, the experimenters again asked the reviewers to analyze and provide their interpretation on the score sheet for each of the 141 FA graphs. The reviewers were also given continuous access to the written visual-inspection criteria during posttraining.

Results and discussion. Agreement between each master's-level and postbaccalaureate reviewer's interpretations and the interpretation of one expert judge was assessed using exact agreement coefficients as described above. During pretraining, in which the reviewers had access to the written modified visual-inspection criteria, the average agreements between the master's-level and postbaccalaureate reviewers relative to the expert judge were .73 and .80, respectively. Following training, interobserver agreement increased for both groups. Specifically, the exact agreement between the master's-level reviewers and the expert judge increased to an average of .98, whereas the average exact agreement between the postbaccalaureate reviewers and the expert judge increased to .95.

The results of Study 2 suggest that individuals without advanced experience in visual inspection of FA data could achieve fairly high levels of agreement by receiving written visual-inspection criteria alone. Interestingly, the postbaccalaureate reviewers, who had less training in visual inspection, had slightly higher initial agreement scores than the master's-level reviewers. The source of this difference is unknown. However, it is possible that the only exposure that the postbaccalaureate reviewers had to visual inspection was through their employment in the outpatient clinic, whereas the master's-level reviewers had received some training in their prior education. Thus, the individual education and employment histories could have contributed to the obtained differences.

Despite these initial differences, the results of Study 2 suggest that training in the use of these criteria produced an increase in the average level of agreement with an expert judge for both groups. Thus, these results provide additional support for the validity of the modified visual-inspection criteria. Based on the combined results of Study 1 and Study 2, the modified visual-inspection criteria appear to provide a reliable method of interpreting FAs that vary in length. In an attempt to determine the utility of the modified visual-inspection criteria, the categorization of each FA outcome was assessed across all 141 data sets and across all 101 participants who contributed to those data sets.

Study 3: Application of Visual-Inspection Criteria to Categorization of FA Outcomes

Reviewers. The four expert judges reviewed the data in Study 3.

Procedure. Each of the expert judges provided their independent interpretation of each FA using the modified visual-inspection criteria. As noted above, the initial interobserver agreement among the expert judges averaged .91. Thus, the expert judges' interpretations were not in agreement for a small proportion of graphs. To address these discrepancies, the score sheets for all expert judges were reviewed to identify all graphs that produced disagreement. As noted above, once these discrepancies were addressed, the average exact agreement coefficient across expert judges increased to .99. It should be noted that one judge did not agree to a consensus with the other three judges on two graphs; therefore, for the categorization of these two graphs, we used the interpretation given by the majority of the expert judges.

Once a consensus was achieved in the majority of cases, the FA outcomes were placed into one of six categories. These categories were similar to those used in previous research (Iwata et al., 1994) and consisted of the following: (a) social-positive reinforcement (attention), (b) social-positive reinforcement (tangible), (c) social-negative reinforcement, (d) automatic reinforcement, (e) multiple controlling variables, and (f) undifferentiated.

Results and discussion. FA outcomes were assessed in two ways. First, we examined the outcomes for each of the 141 analyses included in the sample. Results from the 141 data sets indicated that at least one function was identified for 61.7% of the analyses (see Table 1). This outcome is similar to the results reported by Kurtz et al. (2003) for the subset of participants who displayed SIB (62.1%). Similarly, the current results for socially mediated negative reinforcement (7.2%) were low, as was the case in the Kurtz et al. sample for both SIB (3.4%) and all forms of problem behavior (4.2%) relative to those reported previously (e.g., 38.1% of the Iwata et al., 1994, sample). In addition, the overall prevalence of socially mediated positive reinforcement in the current analyses (16.3%) was much lower than that reported by Kurtz et al. (37.9% for SIB only; 62.5% for all other topographies of problem behavior) and Iwata et al. (26.3% of cases). Likewise, the current percentage of differentiated FA outcomes (61.7%) was much lower than that reported by Kurtz et al. for all topographies of problem behavior (87.5%) and by others (e.g., 95.4% of the Iwata et al., 1994, sample).

Table 1
Summary of FA Results for 141 Data Sets

Behavioral function	Number of analyses	% of analyses
Social positive (attention)	8	5.7
Social positive (tangible)	15	10.6
Social negative	10	7.1
Automatic	28	19.9
Multiple control	26	18.4
Undifferentiated	54	38.3

Given these discrepancies, the sample also was examined on a case-by-case basis for each of the 101 participants who contributed to the 141 data sets. The purpose of this analysis was to determine the percentage of participants for whom a differentiated FA was eventually obtained. To pursue this, we summarized the interpretations of the expert judges for the deidentified data for all 141 analyses on one score sheet. We then decoded the data by comparing each graph number to the original participant data such that the score sheet could be sorted by participant name, which permitted an examination of FA outcomes for all 101 participants who contributed at least one data set.

Using the interpretations conducted by the expert judges, of the 101 initial multielement FAs conducted with the 101 participants, 67 (66.3%) produced clear, differentiated results, thus identifying one or more functions of problem behavior. Of the remaining 34 participants whose initial multielement FAs did not identify one or more clear functions, there were six cases in which experimenters conducted an additional FA for the same topography or topographies of problem behavior (e.g., one participant was exposed to two FAs to identify the function of SIB). When we applied the modified visual-inspection criteria to these subsequent FAs, we obtained a differentiated FA outcome for four of these six participants. Also, the initial undifferentiated FAs were followed by an extended analysis based on Vollmer et al. (1995) for four participants. A differentiated outcome was observed based on responding in the extended analyses for all four of the participants. It should be noted, however, that the extended analyses were not evaluated using the modified structured criteria because there was no comparison control condition to determine relevant CLs. Thus, operant functions of problem behavior were identified for 75 (74.3%) of the 101 participants who contributed at least one FA data set. The results of the case-by-case interpretations of FA outcomes are presented in Table 2. Overall, the

Table 2
Summary of FA Results for the 101 Participants Who Contributed Data Sets

Behavioral function	Number of participants	% of participants
Social positive (attention)	6	5.9
Social positive (tangible)	14	13.9
Social negative	10	9.9
Automatic	23	22.8
Multiple control	22	21.8
Undifferentiated	26	25.7

findings of Study 3 suggested that the sequential application of multielement FAs and extended analyses produced differentiated outcomes for a similar number of participants as those obtained in previous studies.

GENERAL DISCUSSION

Over the past 30 years, FA has emerged as the best practice for identifying the variables that maintain problem behavior. Accurate visual inspection of FA data is critical to categorizing the function of problem behavior such that a treatment can be developed that addresses the pertinent reinforcement contingencies. Although visual inspection has been a standard practice in applied behavior analysis since its inception (see Baer, Wolf, & Risley, 1968), the extent to which individuals agree on outcomes using visual inspection has been shown to vary across reviewers (DeProspero & Cohen, 1979; Jones et al., 1978; Kahng et al., 2010). In an attempt to address this concern as it relates specifically to FA outcomes, Hagopian et al. (1997) developed a set of structured criteria to be applied to FA data sets. Although these criteria produced a high level of agreement among senior and novice reviewers, the generality of those criteria was potentially limited by their application to data sets of a fixed number of data points. Thus, one purpose of the current study was to evaluate modifications to these previously validated criteria. We accomplished this by adapting the structured criteria such that they could be applied to FAs of varying lengths. In addition, we also sought to apply the modified visual-inspection criteria to a set of 141 data sets and 101 participants to examine the clinical utility of the criteria as they relate to the categorization of FA outcomes.

The current criteria were modified to examine differential responding in a percentage of data points as opposed to a specific number of data points as described by Hagopian et al. (1997). Once the modified criteria had been developed, a panel of four expert judges applied these criteria to the sample of 141 data sets. Although the initial average agreement coefficient was high, a number of common discrepancies were identified. Once clarified, the resulting criteria yielded a mean agreement coefficient of .99.

Study 1 and Study 2 were designed to assess the reliability of visual inspection when reviewers either did or did not have access to the modified structured criteria (Study 1) and when reviewers received direct instruction on the use of the modified criteria (Study 2). The results of Study 1 suggested that the interpretation of FA outcomes was less reliable in the absence of the modified structured criteria, whereas agreement increased when the criteria were used. These results, in conjunction with previous research (e.g., Fisher et al., 2003; Hagopian et al., 1997), suggest that the development and use of objective visual-inspection criteria can assist in the interpretation of single-case data. Similarly, the results of Study 2 suggested that direct instruction in the use of structured criteria improved agreement among reviewers, independent of their previous training in the interpretation of single-case data. In addition, the combined results of Studies 1 and 2 suggested that the modified visual-inspection criteria had reasonable validity in that the use of the criteria resulted in high levels of agreement between reviewers of various levels of expertise and an expert judge.

Study 3 evaluated the utility of the modified visual-inspection criteria in the categorization of FA outcomes. In this study, the interpretations of four expert judges, all of whom employed the modified structured criteria, were used to identify whether each of the data sets yielded differentiated patterns of responding across test conditions. The results of Study 3 revealed that a differentiated multielement FA outcome was obtained in 61.7% of 141 analyses reviewed and for 73.3% of the 101 participants who contributed data to the sample. We obtained lower percentages of differentiated FA outcomes than did previous investigations (e.g., Asmus et al., 2004; Hanley et al., 2003; Iwata et al., 1994).

There are several potential reasons for these disparate findings. First, it is possible that different methods used to inspect and interpret the FAs could have produced different outcomes. Prior to the visual-inspection procedure developed by Hagopian et al. (1997), no formal method of analyzing FA outcomes had been described in the literature. It is possible that less formal methods of visual inspection tend to be less conservative than methods that involve the application of structured visual-inspection criteria. This possibility is admittedly speculative at this point, and comparisons of the results of informal and structured visual inspection of FAs should be the focus of future research.

A second issue for the current divergent results related to differentiated FA outcomes could be that the current data sets as well as those presented by Hagopian et al. (2013) were drawn from clinical (or referred) samples, whereas previous data sets (e.g., Hanley et al., 2003) were drawn from published research. All other things being equal, FAs with clear outcomes are more likely to be published than those with more ambiguous results. In addition, in both the current and Hagopian et al. samples, the referral population varied more with respect to function level than the population included in Iwata et al. (1994). Also, both the current and Hagopian et al. samples primarily included individuals who were at risk of placement in a more restrictive setting and for whom outpatient services were ineffective at identifying the function of problem behavior. Thus, these participants may have had, on average, more severe and complex problem behavior than those included in previously published samples, particularly given the relatively larger proportion of participants diagnosed with PDD.

Finally, various procedural issues also could have accounted for the results of Study 3. For example, multiple topographies of behavior were reinforced in some of the FAs in the current sample (1.6 responses per FA). Derby et al. (1994, 2000) found that graphing FA results across multiple topographies of problem behavior could obscure results for individual topographies. It is possible that particular topographies occurred in particular test conditions (e.g., SIB occurred in the attention condition and aggression occurred in the escape condition, leading to results being interpreted as undifferentiated; Smith, Iwata, Vollmer, & Zarcone, 1993) or that different topographies of problem behavior occurred in a hierarchical order (e.g., Lalli, Mace, Wohn, & Livezey, 1995; Richman, Wacker, Asmus, Casey, & Andelman, 1999).

One implication of Study 3 is derived from the different outcomes that were obtained from the two data-analysis methods. There was an 11.6% increase in differentiated FA outcomes when the data were analyzed by participants who contributed at least one FA. A large part of this difference was due to the eight participants for whom a subsequent FA or extended analysis produced differentiation (recall that the modified structured criteria were not applied to extended analyses). Thus, these results are similar to previous investigations (e.g., Vollmer et al., 1995) in suggesting that additional analyses will increase the likelihood of differentiated FA outcomes.

It is possible that some individuals for whom undifferentiated FAs were obtained displayed problem behavior that was sensitive to idiosyncratic reinforcement contingencies (e.g., Bowman, Fisher, Thompson, & Piazza, 1997; Carr, Yarbrough, & Langdon, 1997; Van Camp et al., 2000). The current approach was limited to the interpretation of FAs that consisted of the conditions described by Iwata et al. (1982/1994). Presumably, the current visual-inspection procedures could be applied easily to any multielement assessment that consists of an appropriate control condition on which to base the determination of upper and lower CLs. This line of research would, in turn, permit a standard approach to the categorization of modified FA outcomes.

These results should be interpreted relative to several limitations of the current research. First, no baseline interpretations were collected for the reviewers in Study 2. Recall that these reviewers had access to the written criteria during pretraining and posttraining, such that the only difference between these two conditions was their receipt of instruction in the use of the modified visual-inspection criteria. Thus, the contribution of written instructions alone (relative to having no criteria, as was the case in Hagopian et al., 1997) was not evaluated among these reviewers. In contrast, Study 1 did include baseline data collection on the interpretations of reviewers who did not have access to the modified visual-inspection criteria. However, the reviewers in Study 1 were at the postdoctoral level and both had considerable prior experience in the visual inspection of FA data. Also, the instruction used in Study 2 consisted of multiple components (i.e., didactic instruction, modeling, and practice to criterion levels). Thus, the independent contribution of each component to the acquisition of visual-inspection skills remains unknown. The current sample consisted of functional analyses that varied in length (range, 2.8 to 20 exposures per condition). It is possible that the length of the FA affected reliability scores. Despite this variation, a review of the data revealed no significant correlations between the length of the FAs and agreement level among raters. Finally, reviewers in Studies 1 and 2 received repeated exposure to each of the 141 data sets. In both studies, the second exposure to the data sets was correlated with the administration of the relevant independent variable. Thus, the effect of exposure alone on agreement scores was not controlled.

As noted by Hagopian et al. (1997), the use of structured visual-inspection criteria does not account for clinical judgments related to the intensity of behavior, within-session patterns of responding, and multiple-treatment interference. Therefore, visual-inspection criteria should not be used as an exclusive procedure for interpreting FA outcomes. However, results of this study and others (Fisher et al., 2003; Hagopian et al.) suggest that training in the use of visual-inspection aids is an effective method for increasing agreement among reviewers of single-case data. The use of structured criteria may offer additional benefits related to the etiology of problem behavior and program evaluation in that applying a standard practice across multiple research and clinical sites would help to ensure universal procedures for data interpretation in applied behavior analysis.

REFERENCES

Asmus, J. M., Ringdahl, J. E., Sellers, J. A., Call, N. A., Andelman, M. S., & Wacker, D. P. (2004). Use of a short-term inpatient model to evaluate aberrant behavior: Outcome data summaries from 1996 to 2001. *Journal of Applied Behavior Analysis*, *37*, 283–304. doi: 10.1901/jaba.2004.37-283

Baer, D. M. (1977). "Perhaps it would be better not to know everything." *Journal of Applied Behavior Analysis*, *10*, 167–172. doi: 10.1901/jaba.1977.10-167

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.

Bailey, D. B. (1984) Effects of lines of progress and semilogarithmic charts on ratings of charted data. *Journal of Applied Behavior Analysis*, *17*, 359–365. doi: 10.1901/jaba.1984.17-359

Betz, A. M., & Fisher, W. W. (2011). Functional analysis: History and methods. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 206–225). New York, NY: Guilford.

Bowman, L. G., Fisher, W. W., Thompson, R. H., & Piazza, C. C. (1997). On the relation of mands and the function of destructive behavior. *Journal of Applied Behavior Analysis*, *30*, 251–265. doi: 10.1901/jaba.1997.30-251

Carr, E. G., Yarbrough, S. C., & Langdon, N. A. (1997). Effects of idiosyncratic stimulus variables on functional analysis outcomes. *Journal of Applied Behavior Analysis*, *30*, 673–686. doi: 10.1901/jaba.1997.30-673

Derby, K. M., Hagopian, L., Fisher, W. W., Richman, D., Augustine, M., Fahs, A., & Thompson, R. (2000). Functional analysis of aberrant behavior through measurement of separate response topographies. *Journal of Applied Behavior Analysis*, *33*, 113–117. doi: 10.1901/jaba.2000.33-113

Derby, K. M., Wacker, D. P., Peck, S., Sasso, G., DeRaad, A., Berg, W., … Ulrich, S. (1994). Functional analysis of separate topographies of aberrant behavior. *Journal of Applied Behavior Analysis*, *27*, 267–278. doi: 10.1901/jaba.1994.27-267

DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, *12*, 573–579. doi: 10.1901/jaba.1979.12-573

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, *36*, 387–406. doi: 10.1901/jaba.2003.36-387

Fisher, W. W., Piazza, C. C., & Chiang, C. L. (1996). Effects of equal and unequal reinforcer duration during functional analysis. *Journal of Applied Behavior Analysis*, *29*, 117–120. doi: 10.1901/jaba.1996.29-117

Hagopian, L. P., Fisher, W. W., Thompson, R. H., Owen-De Schryver, J., Iwata, B. A., & Wacker, D. P. (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis*, *30*, 313–326. doi: 10.1901/jaba.1997.30-313

Hagopian, L. P., Rooker, G. W., Jessel, J., & DeLeon, I. G. (2013). Initial functional analysis outcomes and modifications in pursuit of differentiation: A summary of 176 inpatient cases. *Journal of Applied Behavior Analysis*, *46*, 000–000.

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis*, *36*, 147–185. doi: 10.1901/jaba.2003.36-147

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, *27*, 197–209. doi: 10.1901/jaba.1994.27-197 (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*, 3–20, 1982)

Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., … Willis, K. D. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis*, *27*, 215–240. doi: 10.1901/jaba.1994.27-215

Iwata, B. A., Wallace, M. D., Kahng, S., Lindberg, J. S., Roscoe, E. M., Conners, J., … Worsdell, A. S. (2000). Skill acquisition in the implementation of functional analysis methodology. *Journal of Applied Behavior Analysis*, *33*, 181–194. doi: 10.1901/jaba.2000.33-181

Jones, R. R., Weinrott, M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis*, *11*, 277–283.

Kahng, S., Chung, K., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, *43*, 35–45. doi: 10.1901/jaba.2010.43-35

Kurtz, P. F., Chin, M. D., Huete, J. M., Tarbox, R. S. F., O'Connor, J. T., Paclawskyj, T. R., & Rush, K. S. (2003). Functional analysis and treatment of self-injurious behavior in young children: A summary of 30 cases. *Journal of Applied Behavior Analysis*, *36*, 205–219. doi: 10.1901/jaba.2003.36-205

Lalli, J. S., Mace, F. C., Wohn, T., & Livezey, K. (1995). Identification and modification of a response-class hierarchy. *Journal of Applied Behavior Analysis*, *28*, 551–559. doi: 10.1901/jaba.1995.28-551

Moore, J. W., & Fisher, W. W. (2007). The effects of videotape modeling on staff acquisition of functional analysis methodology. *Journal of Applied Behavior Analysis*, *40*, 197–202. doi: 10.1901/jaba.2007.24-06

Mueller, M. M., Nkosi, A., & Hine, J. F. (2011). Functional analysis in public schools: A summary of 90 functional analyses. *Journal of Applied Behavior Analysis*, *44*, 807–818. doi: 10.1901/jaba.2011.44-807

Parsonson, B. S., & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 157–186). New York, NY: Plenum.

Richman, D. M., Wacker, D. P., Asmus, J. M., Casey, S. D., & Andelman, M. (1999). Further analysis of problem behavior in response class hierarchies. *Journal of Applied Behavior Analysis*, *32*, 269–283. doi: 10.1901/jaba.1999.32-269

Rojahn, J., & Schulze, H. H. (1985). The linear regression line as a judgmental aid in visual analysis of serially dependent A-B time-series data. *Journal of Psychopathology and Behavioral Assessment*, *7*, 191–206. doi: 10.1007/BF00960752

Smith, R. G., Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1993). Experimental analysis and treatment of multiply controlled self-injury. *Journal of Applied Behavior Analysis*, *26*, 183–196. doi: 10.1901/jaba.1993.26-183

Van Camp, C. M., Lerman, D. C., Kelley, M. E., Roane, H. S., Contrucci, S. A., & Vorndran, C. M. (2000). Further analysis of idiosyncratic antecedent influences during the assessment and treatment of problem behavior. *Journal of Applied Behavior Analysis*, *33*, 207–221. doi: 10.1901/jaba.2000.33-207

Vollmer, T. R., Marcus, B. A., Ringdahl, J. E., & Roane, H. S. (1995). Progressing from brief assessments to extended experimental analyses in the evaluation of aberrant behavior. *Journal of Applied Behavior Analysis*, *28*, 561–576. doi: 10.1901/jaba.1995.28-561

Received December 22, 2011
Final acceptance August 24, 2012
Action Editor, Joel Ringdahl

APPENDIX

Modified Structured Criteria for Visual Inspection of Multielement FA Data for Analyses of Varying Lengths

Placing the Criterion Lines

The upper and lower criterion lines (CL) are placed one standard deviation above and below the mean of all points in the toy play condition (or at zero if this results in a negative number or the sum of the toy play sessions is zero). If all points in the toy play condition are zero, then both criterion lines are placed at zero.

Assessing the Percentage of Data Points Above and Below the CLs

Examine responding in each of the conditions in comparison to responding in the toy play condition. A condition is differentiated if 50% or more of the data points from a test condition fall above the upper CL than fall below the lower CL. Make sure the criterion for trends does not apply before excluding a function.

1. If the lower CL is zero, count each zero point as below the lower CL.

2. If both the lower and upper CLs are zero, count each zero point in a condition as below the lower CL.

3. To calculate the percentage of data points, subtract the number of data points below the lower CL from the number of data points above the upper CL, divide by the total number of data points in that condition, and convert the result to a percentage.

Rules for Automatic Reinforcement

Score an FA as automatic only if one of the following occurs (if one of the following does not apply, the FA cannot be scored as an automatic function only, but may be scored as multiply maintained according to the criteria below).

1. Alone is the highest condition and is significantly higher than toy play. Define *highest* by calculating the mean of each condition.

2. The rates of behavior tend to be higher in conditions with less external stimulation (i.e., alone, social attention, and tangible) than in conditions with higher external stimulation (i.e., demand and toy play).

3. All conditions are high and relatively stable with no overall trends.

Rules for Trends

Downward trends. At least 40% of the data points above the upper CL must occur in the second half of the assessment. If not, there is a downward trend and the condition is not differentiated (e.g., if there are three data points above the upper CL, two of these must occur in the second half of the assessment overall; otherwise it is considered a downward trend). Exception: For demand and tangible conditions, do not apply the differentiation rule for downward trends if there is a decreasing trend to an efficient rate of responding (e.g., if items are provided for 30 s contingent on behavior, an efficient rate of responding would be two responses per minute).

Upward trends. If 50% of all of the points fall above the upper CL in the last half of the assessment, and the last data point is above the upper CL, this condition is differentiated. For example, if there are eight data points and the first four are below the lower CL and the last four are above the upper CL, this does not meet the initial criterion for differentiation; however, because 50% of all the points are above the upper CL during the last half of the assessment, this is counted as differentiated.

Overall trends. If there is a trend in the toy play condition and most of the other conditions, any condition that is consistently higher than toy play over the course of the assessment meets criterion for differentiation.

Rules for Low-Rate Behavior

When most of the data points are at near-zero levels, the condition in which the highest rate of behavior occurs is considered differentiated (i.e., more than 50% of the high-rate sessions occur in one condition and more than 50% of the total number of behaviors in the higher rate sessions occur in that same condition). However, at least one of the high points must occur in the last half of the FA.

Rules for Low Magnitude of Effects

If a condition meets criteria for differentiation but more than 10% of the points are above the upper CL by only a small amount, raise the upper CL for that condition by 20%. Then analyze this condition using the criteria stated above with the new upper CL. Exception: If there are at least 50% of data points above the upper CL in the last half of the FA, do not adjust the upper CL and apply the criteria for an upward trend instead.

Rules for Multiple Maintaining Variables

1. If more than one condition meets the criteria for differentiation, score the analysis as multiply maintained.

2. If one of the differentiated conditions is alone, use the following criteria:

(a) If the highest condition is alone, score the FA only as automatic (see above).

(b) If there are four differentiated conditions (i.e., all test conditions are high but the toy play condition is still lower) and alone is one of the conditions but not the highest, do not score the FA as automatic. Instead score the FA as the other three differentiated conditions.

(c) If there are three differentiated conditions and alone is one of the conditions, but it is not the highest, do not score the FA as automatic. Instead, score the FA as the other two differentiated conditions.

(d) If there are two differentiated conditions and alone is one of the conditions but it is the lower of the two, score the FA as both automatic and the other condition.