RESEARCH ARTICLE

A Preliminary Evaluation of Interrater Reliability and Concurrent Validity of Open-Ended Indirect Assessment

Valdeep Saini¹ · Kelly Ubdegrove² · Shany Biran² · Rebecca Duncan³

Published online: 28 May 2019

© Association for Behavior Analysis International 2019

Abstract

Indirect assessments are a commonly used component of functional behavior assessment by behavior analysts in practice who work with individuals with severe behavior disorders. Although used frequently, closed-ended indirect assessments have repeatedly been shown to have low to moderate interrater reliability and poor concurrent validity with functional analysis. Recently, the use of open-ended interviews has become more commonly adopted in applied clinical practice, despite no studies evaluating the psychometric properties of such assessments. In the present study, we evaluated the interrater reliability and concurrent validity of an open-ended functional assessment interview. We compared the results of two open-ended indirect assessments conducted with a common caregiver and subsequently conducted functional analyses in an attempt to validate hypotheses generated from the interviews. Interrater agreement for the open-ended interviews was higher than previous research on closed-ended interviews (75%); however, concurrent validity with functional analysis was relatively poor (50%). We discuss these findings in the context of assessment and treatment for severe behavior disorders, as well as best practice methods during functional behavior assessment.

Keywords Functional analysis · Functional assessment · Indirect assessment · Reliability · Validity

Research Highlights

- This study is valuable for behavior analysts in practice who routinely conduct functional behavior assessments or work with individuals with disabilities who engage in severe problem behavior.
- Functional analysis is the gold-standard method for identifying the environmental conditions that maintain severe problem behavior.
- Recently, the use of open-ended indirect assessments to inform functional analysis design has received greater attention in the extant literature on severe behavior disorders.
- This study examined the interrater reliability and concurrent validity of open-ended indirect assessments with functional analysis.
- Interrater reliability and concurrent validity for open-ended interviews were moderate and poor, respectively.
- This study highlights the value and importance of conducting functional analyses as opposed to using less reliable and valid methods, such as indirect assessments, to inform behavioral interventions.

✉ Valdeep Saini
vsaini@brocku.ca

¹ Department of Applied Disability Studies, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON, Canada

² Surrey Place Centre, Toronto, Canada

³ ABA of Illinois, Darien, IL, USA

Functional behavior assessment (FBA) is the cornerstone method used by behavior analysts in practice for identifying the environmental variables responsible for the maintenance of severe problem behavior in individuals with autism spectrum disorder and related disabilities (Horner, Carr, Strain, Todd, & Reed, 2002; Matson & Nebel-Schwalm, 2007; Reid & Nelson, 2002; Sugai, Lewis-Palmer, & Hagan-Burke, 2000). FBA is regarded as an information-gathering process about antecedent and consequence events that correlate with incidents of severe problem behavior in an attempt to determine the function, or reason why the behavior is occurring (Gresham, Watson, & Skinner, 2001). Treatments based on a comprehensive FBA have been demonstrated to be associated with large reductions in problem behavior (Carr et al., 1999; Ellingson, Miltenberger, Stricker, Galensky, & Garlinghouse, 2000).

Functional analysis is perhaps the most resource-intensive form of FBA, but this assessment method also provides the most reliable, valid, and useful information when developing interventions for problem behavior (Saini, Fisher, Betz, & Piazza, *in press*). During a functional analysis, environmental events hypothesized to maintain problem behavior are systematically presented and withdrawn under highly controlled

conditions until the function is identified. A number of large-scale reviews and meta-analyses have directly compared treatments for problem behavior that were and were not based on a functional analysis, and have demonstrated more efficacious reductions in problem behavior for those treatments that were based on a pretreatment functional analysis (Campbell, 2003; Didden, Duker, & Korzilius, 1997; Heyvaert, Saenen, Campbell, Maes, & Onghena, 2014; Iwata et al., 1994; Kuhn, DeLeon, Fisher, & Wilke, 1999). As a result, functional analysis is considered to be the hallmark approach when conducting an FBA (Beavers, Iwata, & Lerman, 2013; Hanley, Iwata, & McCord, 2003; Iwata & Dozier, 2008).

Despite the widespread effectiveness of functional analysis, and recommendations from the applied literature for behavior analysts to conduct functional analyses when developing interventions for problem behavior, the majority of behavior analysts in practice do not conduct functional analyses and instead rely on other FBA methods, such as indirect or descriptive assessments (Oliver, Pratt, & Normand, 2015; Roscoe, Phillips, Kelly, Farber, & Dube, 2015). Indirect assessments are used to survey the frequency or severity of problem behavior under different environmental situations and can consist of interviews, rating scales, or questionnaires (Kelley, LaRue, Roane, & Gadaire, 2011). Structured interviews may be a useful preliminary method for gathering information about the conditions under which problem behavior occurs from individuals with whom the client most often interacts. Indirect assessments such as interviews are the most time- and cost-efficient form of FBA; however, interviews may provide less, or even erroneous, information about the functions of problem behavior relative to functional analysis (Kelley et al., 2011).

Most indirect assessments suffer from poor interrater reliability, which is defined as the extent to which two independent raters (or interviewers) produce the same rating or results regarding the conditions under which problem behavior occurs after the interview is administered (Barton-Arwood, Wehby, Gunter, & Lane, 2003; Iwata, DeLeon, & Roscoe, 2013; Sigafoos, Kerr, & Roberts, 1994; Zarcone, Rodgers, Iwata, Rourke, & Dorsey, 1991). For example, Zarcone et al. (1991) evaluated interrater agreement of the Motivation Assessment Scale (MAS) by administering the interview to teachers, teachers' aids, and therapists working with individuals who engaged in self-injurious behavior. They obtained only 29% agreement between raters when comparing reinforcement categories that were hypothesized to maintain the client's problem behavior. More recently, Iwata et al. (2013) evaluated the interrater reliability of the Functional Analysis Screening Tool (FAST) based on interviews administered to caregivers, relatives, teachers, teachers' aides, and direct-care staff working with individuals with disabilities and obtained a mean reliability of 64.8%.

Concurrent validity is traditionally evaluated by comparing two methods that purport to assess the same construct. In the case of FBA, concurrent validity is assessed by comparing the results of an indirect assessment to the results of a functional analysis. Because functional analysis is the only method that experimentally evaluates and confirms the maintaining reinforcers for problem behavior during an FBA, it is used as the criterion measure that indirect assessments are compared against. Along with poor interrater reliability, indirect assessments such as interviews have the additional shortcoming of weak concurrent validity with functional analysis (Crawford, Brockel, Schauss, & Miltenberger, 1992; Dracobly, Dozier, Briggs, & Juanico, 2018; Iwata et al., 2013; Paclawskyj, Matson, Rush, Smalls, & Vollmer, 2001; Newcomer & Lewis, 2004). For example, Iwata et al. (2013) obtained moderate levels of concurrent validity between the FAST and analogue functional analysis, even in cases where interrater reliability was relatively high. As a result of both low interrater reliability scores and poor concurrent validity between indirect assessments and functional analysis, some researchers have cautioned against using indirect assessments as the sole method to inform intervention (Fisher, Piazza, Bowman, & Amari, 1996; Green & Striefel, 1988; Lennox & Miltenberger, 1989; Umbreit, 1996).

Although the literature on the reliability and validity of indirect assessments is compelling, one collective limitation of these findings is that previous researchers have primarily examined the reliability and validity of closed-ended interviews and rating scales, such as the MAS or FAST. Closed-ended indirect assessments provide the informant with predetermined response options from which they must choose, limiting the informant's ability to precisely describe the conditions under which problem behavior occurs. For example, the FAST is a 16-item questionnaire about possible functions of problem behavior wherein the informant is restricted to three response options: "yes," "no," or "not applicable." One limitation of using closed-ended interviews during FBA, which might contribute to low reliability and validity scores, is that closed-ended interviews do not allow the interviewer to probe or ask clarifying questions about contextual factors, or allow the informant to provide qualitatively rich details that may be relevant to understanding the function of problem behavior. Given that FBA is an information-gathering process about environmental variables that might influence the occurrence or maintenance of problem behavior, some have questioned the utility of closed-ended interviews in practice (Fryling & Baires, 2016).

Recently, an increased focus has been placed on using open-ended interviews to guide decision-making when identifying, designing, and determining the test conditions of a functional analysis (Beaulieu, Van Nostrand, Williams, & Herscovitch, 2018; Hanley, Jin, Vanselow, & Hanratty, 2014; Herman, Healy, & Lydon, 2018; Jessel, Hanley, & Ghaemmaghami,

2016; Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018; Slaton, Hanley, & Raftery, 2017). Specifically, the open-ended functional assessment interview presented by Hanley (2012) has been relied on as the primary indirect assessment that inspires research on interview-informed functional analyses. Open-ended interviews have a number of advantages over closed-ended interviews including allowing the behavior analyst to (a) derive information about contextual factors influencing problem behavior, (b) ask follow-up or probing questions relevant to what the informant has described, and (c) build positive rapport with the informant (Fryling & Baires, 2016). Although research in this area has been promising, the interrater reliability and concurrent validity of the open-ended interview described by Hanley (2012) is currently unknown and thus should be evaluated formally.

The purpose of this study was to evaluate the interrater reliability and concurrent validity of the open-ended functional assessment interview described by Hanley (2012). We assessed interrater reliability by conducting two independent open-ended interviews with a single informant and comparing the exact agreement of those results between raters (i.e., the extent to which the two raters developed the same prediction regarding the maintaining reinforcers for problem behavior). We then used the results of each open-ended interview to conduct a functional analysis with the individual for whom the interview was based on. We assessed concurrent validity by comparing functional analysis outcomes with the predictions derived from the open-ended interviews. Both interrater reliability and concurrent validity were calculated using percentage agreement analyses.

Method

Participants and Setting

We used a consecutive-case series analysis wherein the first four children diagnosed with comorbid intellectual disability and autism spectrum disorder who were referred for the assessment and treatment of severe problem behavior were selected to participate. Sally, a 15-year-old girl, engaged in self-injury, aggression, and property destruction and communicated using two- to four-word sentences. Sally was also diagnosed with a seizure disorder. David, a 17-year-old boy, engaged in self-injury and property destruction and communicated using augmentative and alternative communication systems (e.g., Proloquo2Go, picture exchange). Alex, a 15-year-old boy, engaged in aggression and property destruction and communicated using unintelligible vocal approximations, pointing to objects and pictures, and gesturing. Alex was also diagnosed with impulse control disorder. Eric, an 11-year-old boy, engaged in self-injury and communicated using one- to

two-word vocal approximations, as well as augmentative and alternative communication systems.

Sally's, David's, and Alex's functional analysis sessions were conducted in a 6.5 m × 10 m empty room equipped with a one-way observation panel. Eric's sessions were conducted in a 3 m × 3–3.7 m empty and enclosed space within a larger special education classroom. Eric was unable to look outside of the space and unable to easily leave the space without the assistance of a therapist. Only relevant assessment materials were included during each participant's functional analysis (e.g., preferred items, task demands). These items were informed by open-ended interviews except when functional analysis results were inconclusive, in which case the materials used were based on observations of the participant in the natural environment.

At the conclusion of this study, function-based interventions were designed for all participants to reduce severe problem behavior.

Response Definitions, Measurement, and Interobserver Agreement

For Sally, David, and Alex, trained observers collected data on laptop computers behind the one-way observation panel. For Eric, data were collected by observers in the same room in which experimental sessions were conducted. Sessions were 5 min for Sally, David, and Alex and 10 min for Eric.

Response topographies and definitions were based on informant report during the open-ended interviews of each participant. The primary dependent variable was combined inappropriate behavior, which included self-injury, aggression, and property destruction. We defined *self-injury* (David, Sally, Eric) as any instances of self-biting, body slamming, self-hitting, self-scratching, pulling one's own hair, and head banging. We defined *aggression* (Sally, Alex) as any instances or attempts of hitting, kicking, pushing, pinching, scratching, grabbing, or throwing objects at the therapist. We defined *property destruction* (David, Sally, Alex) as hitting, kicking, or throwing objects; swiping materials off of furniture; and turning over furniture. Due to the severity of Eric's self-injury, we also included negative vocalizations as part of the response class for his assessment and treatment, as negative vocalizations appeared to reliably precede more severe behavior (however, we did not conduct a formal precursor analysis). *Negative vocalizations* were defined as any instances of screaming, yelling, or crying at a level above typical conversational volume for 3 s or longer.

Interobserver agreement for combined inappropriate behavior was obtained by having two observers simultaneously yet independently collect data during functional analysis sessions.

Sessions were divided into 10-s intervals, and an agreement was recorded for each interval in which both observers

measured the same number of each type of response. We summed the number of agreement intervals and then divided the number of agreement intervals by the total number of intervals within the session. Each quotient was then converted to a percentage. Interobserver agreement was collected for 36%, 45%, 66%, and 52% of sessions for Sally, David, Alex, and Eric, respectively. Agreement averaged 99% (range 98%–99%) for Sally, 91% (range 85%–97%) for David, 90% (range 80%–98%) for Alex, and 90% (range 77%–100%) for Eric.

Open-Ended Interview

The open-ended functional assessment interview described by Hanley (2012) was used for all participants. Only the questions listed in the interview were asked (i.e., no follow-up questions were probed). Two raters independently interviewed a common caregiver (Sally, Alex, Eric) or primary/lead therapist (David) about the topographies and situations in which problem behavior was observed to occur in the natural environment. Educational background of informants was unknown. Raters were either Board Certified Behavior Analysts or Board Certified Assistant Behavior Analysts, and all had experience conducting indirect assessments as part of an FBA. Independent interviews were conducted either over the phone or in person and occurred within 5 days of each other. Interviews took between 40 and 60 min to complete.

Following each independent interview, each rater independently developed a prediction regarding the function of problem behavior, which was to be tested in a subsequent functional analysis. If multiple predictions were generated (i.e., more than one potential function of problem behavior), raters aggregated potential functions into a single test condition if the results of the interview suggested that those reinforcers co-occurred in the natural environment. If potential reinforcers did not appear to co-occur in the natural environment, the rater selected the most probable function based on the interview's results.

Reliability Analysis

Interrater reliability was calculated through a dichotomous yes/no by comparing the two predictions developed by the raters. Agreement was scored "yes" when the two raters developed the same prediction and scored "no" (i.e., a disagreement) when the two raters developed different predictions. An agreement was only scored "yes" if the two raters developed exactly the same hypothesis. For example, if the first rater hypothesized "escape" as a putative reinforcer for problem behavior and the second rater hypothesized "escape plus tangible," this was not considered an agreement despite the fact there was partial overlap in the two raters' hypotheses (i.e.,

escape). We calculated reliability for the interpretation of the interview only. We used this method because the nature of the open-ended interview did not allow for a precise item-by-item analysis. This is unlike reliability analyses using closed-ended interviews, which have a predefined set of responses and corresponding scale for interpretation (e.g., Iwata et al., 2013).

Functional Analysis

We evaluated the concurrent validity between each rater's open-ended interview results with an analogue functional analysis. If there was exact interrater agreement between the two raters (i.e., both raters developed the same prediction regarding the potential function of problem behavior), only one functional analysis was conducted. If the results of each rater's interview differed, we tested each potential function in separate functional analyses. For all participants, only the potential function(s) based on the results of the interview were tested and done so using a pairwise design, except where noted. Using the logic described by Jessel et al. (2016), we used a matched control condition during each functional analysis, except where noted. No systematic observations of the participants were made between the end of the interview and the start of the functional analysis. That is, the hypotheses derived from each interview were not informed by any descriptive assessments or observations, and the design of the functional analysis was informed solely by the interview results.

If the results of the interview-informed functional analysis were inconclusive or undifferentiated, we conducted subsequent analyses to determine the function of problem behavior. Subsequent functional analyses were based on informal observations of each participant in the natural environment in a manner similar to that described by Hanley et al. (2014).

Sally In the attention plus edible condition, Sally interacted with her father for 30 s prior to the start of the session via prerecorded videos on an iPhone. The videos included a fictitious conversation where Sally's father made statements about what she was eating, her clothing, how she was behaving, what she was doing at school, or the weather. Based on informal observations, it appeared that Sally could not discriminate between live and prerecorded videos of her father. Prerecorded videos of her father appeared equally as valuable (i.e., reinforcing) as interacting with him in person (however, this was not evaluated formally). In addition to the video interaction, she was given a high-preferred edible item to consume during the interaction (i.e., bites of cheese). The session began when the therapist told Sally it was time to say goodbye to her father, turned the video off, and simultaneously removed the cheese. The therapist did not interact with Sally during the session except to inform her that the conversation with her father was ending. The therapist withdrew attention and pretended to be "busy." Contingent on the occurrence of problem behavior, the therapist began delivering bites of

cheese to Sally and also replayed the video of her father for 20 s. In the control condition, Sally had continuous access to her father's attention via prerecorded videos on an iPhone and bites of cheese were delivered on a fixed-time (FT) 10-s schedule. Cheese was delivered in this manner during the control condition (instead of continuously) to prevent Sally from consuming all of the cheese at once.

The results of the open-ended interview with Sally's father implicated a synthesis of positive reinforcers (i.e., attention plus edibles); however, the interactive and independent effects of these reinforcers could not be determined based on the initial functional analysis. Therefore, we conducted subsequent analyses to determine the nature of independent and interactive effects (Fisher, Greer, Romani, Zangrillo, & Owen, 2016).

In the attention condition, Sally interacted with her father for 30 s prior to the start of the session via prerecorded videos on an iPhone. The session began when the therapist told Sally it was time to say goodbye to her father, turned the video off, and pretended to be "busy." The therapist did not interact with Sally during the session except to inform her that the conversation with her father was ending. Contingent on the occurrence of problem behavior, the therapist replayed the video of her father for 20 s. In the *control* condition, Sally had continuous access to her father's attention via prerecorded videos on an iPhone.

In the edible condition, the therapist provided access to cheese for 30 s prior to the start of the session. The session began when the therapist removed the cheese. Contingent on the occurrence of problem behavior, the therapist provided access to bites of cheese for 20 s. In the control condition, the therapist delivered cheese to Sally on an FT 10-s schedule.

David In the alone condition, David was by himself in an empty session room to test for problem behavior maintained by automatic reinforcement (Querim et al., 2013). There were no materials or items in the room for him to engage with.

Due to variability in responding during the alone condition, we tested potential social reinforcers for problem behavior in a multielement design. Another pattern of responding consistent with automatic reinforcement is high and undifferentiated responding across conditions in a multielement design when testing for socially mediated reinforcers (Hagopian, Rooker, & Zarcone, 2015). In the attention condition, the therapist provided attention for 30 s prior to the start of the session. The session began when the therapist withdrew attention and pretended to be "busy." Contingent on the occurrence of problem behavior, the therapist provided attention for 20 s in the form of sympathy or statements of concern (e.g., "Don't do that. You'll hurt yourself."). In the escape condition, the session began when the therapist started delivering academic instructions (e.g., matching, insert puzzle) with the use of a least-to-most (i.e., verbal, model, physical) prompting procedure. Contingent on the occurrence of problem behavior,

the therapist removed all instructions and task materials and provided a 20-s break from instructions. In the control condition, David had continuous access to therapist attention and the therapist did not deliver instructions.

Alex In the attention plus edible condition, the therapist provided attention for 30 s prior to the start of the session. Alex also had free access to a highly preferred edible item (i.e., rice). The session began when the therapist withdrew attention, pretended to be "busy," and removed the rice. Contingent on the occurrence of problem behavior, the therapist provided attention for 20 s in the form of calming and soothing statements (e.g., "It's OK. You don't have to do that."). The therapist simultaneously provided bites of rice with attention (e.g., "You can eat this to calm down."). In the control condition, Alex had continuous access to therapist attention and bites of rice were delivered on an FT 10-s schedule. Rice was delivered in this manner during the control condition (instead of continuously) to prevent Alex from consuming all of the rice at once.

Based on informal observations of Alex in the natural environment, problem behavior appeared to occur highest in the presence of specific task demands. In the escape condition, the session began when the therapist started delivering athletic and fine-motor instructions (e.g., passing a ball, beading strings) with the use of a least-to-most (i.e., verbal, model, physical) prompting procedure. Contingent on the occurrence of problem behavior, the therapist removed all instructions and task materials and provided a 20-s break from instructions. In place of a matched control condition, we conducted a toy-play control condition because Alex was observed to engage in low levels of disruptive behavior in environments correlated with task demands (e.g., school) even when demands themselves were not presented. In the toy-play condition, Alex had continuous access to therapist attention and high-preferred tangible items (i.e., a pull scooter he could ride on), and the therapist did not deliver instructions.

Eric In the escape from noise condition, the session began when the therapist began playing nonpreferred sounds from an iPad close to Eric, including the sounds of dogs barking, babies crying, and a crowded mall. Contingent on the occurrence of problem behavior, the therapist paused the iPad's sounds for 20 s. The therapist did not otherwise interact with Eric. In the control condition, no sounds were played from the iPad and the therapist ignored Eric.

In the escape to tangible plus attention condition, the therapist provided access to an iPad for 30 s prior to the start of the session and concurrently delivered attention. The session began when the therapist removed the iPad and began delivering academic instructions (e.g., puzzle, matching) with the use of a least-to-most (i.e., verbal, model, physical) prompting procedure. Contingent on the occurrence of problem behavior, the therapist removed all

instructions and task materials and provided a 30-s break from instructions, while redelivering the iPad and delivering attention in the form of distraction statements (e.g., “You don’t have to hurt yourself. You can play with this.”). In the control condition, Eric had continuous access to his iPad and therapist attention and the therapist did not deliver instructions.

Based on informal observations of Eric in the natural environment, problem behavior appeared to occur at the highest rate in the presence of specific academic demands more than others, primarily when those demands followed the removal of a preferred tangible item. In the escape to tangible (modified tasks) condition, the therapist provided access to an iPad and rolling office chair for 30 s prior to the start of the session. The session began when the therapist removed the iPad and chair and began delivering modified academic instructions based on observations of Eric’s problem behavior (e.g., worksheets, spelling, tracing). Instructions were delivered with the use of a least-to-most (i.e., verbal, model, physical) prompting procedure. Contingent on the occurrence of problem behavior, the therapist removed all instructions and task materials and provided a 30-s break from instructions while Eric regained access to the iPad and rolling chair. In the control condition, Eric had continuous access to his iPad and rolling chair, and the therapist did not deliver instructions.

Concurrent Validity Analysis

Concurrent validity was calculated through a dichotomous yes/no by comparing the results of each individual interview with the outcome of the functional analysis. Agreement was scored “yes” when the comparative hypothesis and functional analysis results were the same. Similar to the reliability analysis, concurrent validity was only scored “yes” if the results of the interview were exactly the same as the results of the functional analysis. A disagreement was scored “no” when the comparative hypothesis and functional analysis results were different. Because each interviewer developed a separate and independent prediction about the potential function of problem behavior, there were two comparisons for each functional analysis. Therefore, for each participant, concurrent validity was scored using 2 as the denominator.

Results

Table 1 displays the results of each rater’s interview, as well as functional analysis outcome. For Sally, separate interviews were conducted with Sally’s father. In both interviews, Sally’s father reported that he observed problem behavior at the highest level when his attention was withdrawn during mealtime situations. He suggested that the attention of others was not as valuable (e.g., mother, school support staff) and that he frequently had to video chat with Sally throughout the day while she was at school, often commenting on what she was eating, if she had earned her reinforcer for good behavior (i.e., cheese), and what she would be having for lunch. Sally’s father also made an effort to have lunch with her at school at least once a week, and she would frequently ask school support staff when her father would be arriving. Both raters independently predicted that problem behavior was maintained by access to Sally’s father’s attention plus edible items. Figure 1 (top-left panel) displays the results of Sally’s functional analysis. Elevated and increasing levels of combined inappropriate behavior were observed during the attention plus edible condition ($M = 0.6$ responses per minute; range 0.2–1.2), and zero levels were observed during the control condition. The outcome of the functional analysis supported the prediction made from each open-ended interview. When we tested for the independent effect of Sally’s father’s attention as a reinforcer for combined inappropriate behavior, we observed zero rates of responding in both the attention and control conditions. When we tested for the independent effect of access to edible items as a reinforcer for combined inappropriate behavior, we observed near-zero levels of responding in the edible condition ($M = 0.1$ responses per minute; range 0–0.7) and zero levels of responding in the control condition.

David’s independent interviews of his problem behavior were conducted with his lead therapist in a school-based program for children with autism spectrum disorder. In both interviews, David’s therapist reported that self-injury and property destruction were topographically similar (i.e., open-handed hits to the body during self-injury and open-handed hits to objects during property destruction) and

Table 1 Results of open-ended interviews and analogue functional analysis

Participant	Open-Ended Interview 1	Open-Ended Interview 2	Functional Analysis
Sally	Attention plus edible	Attention plus edible	Attention plus edible
David	Automatic	Automatic	Automatic
Alex	Attention plus edible	Attention plus edible	Escape
Eric	Escape from noise	Escape to tangibles plus attention	Escape to tangible (modified tasks)

occurred throughout the day. David's therapist reported that these behaviors could occur several hundreds of times per day, even when David was alone. David engaged in these behaviors across a number of social situations and times of the day and in the presence of different individuals. Given that David's therapist reported to both raters that the target behaviors occurred without any clear antecedent event and persisted in the absence of social consequences, both raters independently predicted that problem behavior was maintained by automatic reinforcement. Figure 1 (top-right panel) displays the results of David's functional analysis. Responding during the alone condition was high and variable ($M = 13.9$ responses per minute; range 0.0–34.6). When we tested for possible social reinforcers for problem behavior, we observed high and undifferentiated responding between the attention ($M = 13.5$ responses per minute; range 11.4–15.6), escape (8.4 responses per minute; range 5.6–9.8), and control (12.8 responses per minute; range 8.8–19.4) conditions. Collectively, the results of the alone condition combined with the results of the undifferentiated multielement analysis of social functions support the prediction made from each open-ended interview (i.e., combined inappropriate behavior maintained by automatic reinforcement).

Open-ended interviews about Alex's problem behavior were conducted with his mother. In both interviews, she suggested that she could not identify a clear antecedent

event that evoked problem behavior but only that it occurred "randomly" or "throughout the day." Despite being unable to identify clear environmental events correlated with instances of problem behavior, she suggested that only when she gave him his favorite food (i.e., rice) and attempted to calm him down using soothing statements and gestures of comfort would problem behavior temporarily decrease. Both raters independently predicted that problem behavior was maintained by access to adult attention plus highly preferred edible items. Figure 1 (bottom-left panel) displays the results of Alex's functional analysis. We obtained undifferentiated results when testing the prediction of the open-ended interviews, with high and variable levels of responding occurring in both the attention plus edible ($M = 2.1$ responses per minute; range 0.2–2.2) and control ($M = 3.2$ responses per minute; range 1.2–4.8) conditions. Anecdotal observations of Alex at school indicated that problematic behavior occurred at the highest rate in the presence of classroom task demands that involved gross- or fine-motor movements and sometimes when task materials were placed near him (despite no instruction to manipulate or engage with task materials). We therefore tested a negative reinforcement (escape) contingency in a subsequent functional analysis. In the escape condition ($M = 2.3$ responses per minute; range 1.6–2.8),

Fig. 1 Combined inappropriate behavior per minute across functional analyses conducted for Sally, David, Alex, and Eric

levels of combined inappropriate behavior were high and stable and occurred at zero levels during the toy-play condition. The outcome of the functional analysis did not support the prediction made from each open-ended interview.

Eric's independent interviews of his problem behavior were conducted with his father. In the first interview, Eric's father indicated that self-injury and correlated precursor behavior occurred in the presence of specific auditory stimuli, including barking dogs, crying babies, and when Eric was at a crowded mall. He further reported that to calm Eric down and stop large episodes of tantrums, he needed to remove Eric from the situation where the stimuli were present (e.g., leaving the mall), and that this strategy was typically effective. He reported other situations in which problem behavior occurred, such as when Eric wanted or did not want an adult's attention, but suggested that this occurred less frequently. As a result, the first rater predicted that problem behavior was maintained by escape from aversive auditory stimulation (i.e., "noise"). In the second interview, Eric's father reported that problem behavior occurred across a number of social situations, including when he asked him to complete a chore at home (e.g., putting on shoes), when attention was diverted (e.g., when Eric's father was speaking with a houseguest), and when a high-preferred tangible item was withheld (e.g., iPad). He also reported that these events would co-occur in the natural environment (e.g., Eric engaged in problem behavior to avoid tasks and gain access to tangible items; Eric engaged in problem behavior to gain access to tangible items and his father's attention). The second rater predicted that Eric's problem behavior was maintained by escape from task demands to gain access to preferred tangible items and adult attention. Figure 1 (bottom-right panel) displays the results of Eric's functional analysis. During the initial pairwise comparison, which tested the first rater's prediction, we observed zero levels of responding in both the escape from noise condition and the control condition. In the subsequent functional analysis, which tested the second rater's prediction, we continued to observe zero levels of problem behavior, with no responding in both the escape to tangible plus attention condition and the control condition. Anecdotal observations of Eric in the natural environment revealed that certain task demands (i.e., worksheets, spelling, tracing) evoked problem behavior at a higher rate than others and that these demands were different from those reported by Eric's father during the second open-ended interview. In addition, it appeared that Eric engaged in problem behavior to regain access to specific tangible items that were often removed at the onset of instruction delivery and were not previously reported by Eric's caregiver (e.g., rolling chair). We tested whether the removal of novel tangible items and presentation of novel task materials were evocative in a subsequent functional analysis.

We observed high and somewhat variable levels of responding in the escape to tangible (modified tasks) condition ($M = 2.4$ responses per minute; range 0.4–4.0) and near-zero levels of responding in the control condition ($M = 0.1$ responses per minute; range 0.0–0.4). Therefore, the outcome of the functional analysis did not support the prediction made from either open-ended interview.

Exact agreement between raters conducting independent open-ended interviews was 75% with correspondence in predictions for Sally, David, and Alex. There was disagreement between open-ended interview interpretations for Eric. Concurrent validity between each interview prediction and outcome of each participant's functional analysis was 50%, as there was correspondence between the open-ended interviews and functional analyses for Sally and David. The predictions made by the open-ended interview were not supported by the outcome of the functional analysis.

Discussion

We evaluated the interrater reliability and concurrent validity of an open-ended functional assessment interview and obtained results consistent with previous research in this area on indirect assessments. Mean interrater reliability scores of the interview-derived hypothesis regarding potential functions of problem behavior was 75%, which represents an improvement on interrater reliability scores compared to closed-ended interviews, but still falls below the 80% criterion that is often considered the minimum score for acceptable reliability. Despite higher reliability scores relative to other, closed-ended, indirect assessments, concurrent validity was poor (50%) and less than that obtained by Iwata et al. (2013), who compared the concurrent validity of the FAST with functional analysis (63.8%).

There may be several reasons why a higher overall mean score for interrater reliability was obtained in the present study compared to previous research evaluating interrater reliability of closed-ended interviews. First, the open-ended interview allowed the informant to describe qualitatively important information about the conditions and events that surround occurrences of problem behavior that cannot be easily gleaned from closed-ended interviews, and that this information could be explained in detail to each of the raters (e.g., Sally's father describing how he frequently called her through video chat). Fryling and Baires (2016) note that most closed-ended indirect assessments overlook potentially important context-specific factors that could affect the occurrence and maintenance of problem behavior. Second, the open-ended interview allowed the rater to gather a wide range of information about historical and current environmental influences on problem behavior, which may lead to a more informed hypothesis about the function of problem behavior. Whereas closed-ended

interviews are restricted by the number of questions and response options, there are no restrictions on the amount of information that can be obtained from an open-ended interview.

Results of the present study suggest that although open-ended interviews may be more reliable at guiding the behavior analyst in the development of an individualized functional analysis, incorporating idiosyncratic details relevant to the client's problem behavior, these hypotheses are not considerably more valid than those obtained from closed-ended interviews. The validity of an indirect assessment might depend on the accuracy of the informant in identifying and reporting events that genuinely do co-occur with problem behavior. Results of Alex's concurrent validity analysis highlight this point, wherein Alex's mother provided the two raters with the same information about the conditions under which problem behavior occurred, but that information did not appear to coincide with the results of Alex's functional analysis (i.e., the perceived reinforcers described by Alex's mother were not putative reinforcers for problem behavior). Dracobly et al. (2018) found that indirect assessments conducted with expert informants (i.e., Board Certified Behavior Analysts trained in the design and implementation of FBAs and function-based treatments) had better concurrent validity with functional analyses than indirect assessments conducted with caregivers (see also Lewis, Mitchell, Harvey, Green, & McKenzie, 2015). This suggests that validity may be a product of how accurately and reliably an informant can identify environmental events that co-occur with problem behavior.

Difficulty in obtaining high concurrent validity between indirect assessments and functional analysis is further highlighted through the functional analysis outcome obtained with Eric. His father suggested to the second rater that the presentation of certain task demands and withdrawal of certain tangible items evoked the occurrence of problem behavior; however, when we tested a possible escape plus tangible plus attention function using the materials and items described by Eric's father, we did not observe any occurrences of problem behavior (Figure 1, bottom-right panel, second phase). However, when we modified the task and tangible materials to those that appeared to evoke problem behavior during informal observations of Eric, we were able to obtain clear and differentiated results in his final functional analysis (Figure 1, bottom-right panel, last phase). These results suggest that although caregivers might be accurate at describing general categories of reinforcement that may contribute to problem behavior, supplemental assessments might be necessary to specifically identify which and what kinds of stimuli should be incorporated into a functional analysis (e.g., Roscoe, Rooker, Pence, Longworth, & Zarcone, 2009).

Although the results of this study are consistent with previous research on indirect assessments, this study is

not without limitations. First, we did not ask probing questions for further clarification during the open-ended interview and adhered closely to the questions provided by Hanley (2012). Although this method might be inconsistent with how open-ended interviews are used in applied practice, we did this in an attempt to control for differences in interviewing skills between the two raters. That is, we attempted to eliminate the possibility that some raters were asking more or better follow-up questions than any other rater. Second, our study included a small sample size and utilized only one measure of reliability and validity. However, we used a controlled consecutive-case series analysis wherein we enrolled the first four participants who were referred for assessment and treatment of problem behavior instead of using retrospective data analysis, which has been done in other studies examining reliability and validity of indirect assessments (e.g., Iwata et al., 2013). Third, our final outcome agreement measures were used in light of the limitations of scoring item-by-item agreement of open-ended interviews; however, it is possible that with a larger sample size, this study could be replicated using a greater number of agreement measures, such as those suggested by Saini, Fisher, and Retzlaff (2018), to further validate the present findings. Fourth, we measured combined inappropriate behavior as our primary dependent variable, which may have inadvertently increased the likelihood of identifying multiply controlled problem behavior (Beavers & Iwata, 2011; see also Bell & Fahmie, 2018). Last, although we make reference to differences between reliability and validity of open- versus closed-ended interviews, we did not directly compare these two types of indirect assessments in the current investigation.

It is important to note that the function of problem behavior may vary across environments and that for some participants we conducted observations of their problem behavior in a context different from the one reported during the indirect assessment. For example, for Alex, the open-ended interview was conducted with his mother, who likely reported on the occurrence of problem behavior in the home setting. When results of the functional analysis were not consistent with the raters' hypothesis, a subsequent functional analysis was conducted based on direct observations of Alex at school, which suggested a possible negative reinforcement function and was then confirmed in the subsequent functional analysis. Future researchers might address this limitation by interviewing caregivers relevant to the context in which the problem behavior will be treated (e.g., teachers at school versus parents).

Unlike in Sally's case, we did not conduct an analysis of interactive and independent effects of potential reinforcers for problem behavior for Eric. That is, based on observations of Eric, we aggregated positive and negative

reinforcement contingencies during the functional analysis and therefore cannot be certain that such an aggregation was necessary to evoke problem behavior (i.e., it is unclear if problem behavior was maintained by single, multiple, or combined control). It is important to note that aggregating potential reinforcers in a functional analysis test condition could erroneously lead to false positive identification (Fisher et al., 2016).

Despite the consistent body of literature demonstrating the low to moderate reliability and validity of indirect assessments, they are considered a necessary component of the common FBA process and should continue to be improved on in future research. Moreover, indirect assessments are cost and time efficient, are used frequently by clinical behavior analysts, and represent the first step in a logical progression during FBA that proceeds from the least to the most intrusive assessment method (Fryling & Baires, 2016; Kelley et al., 2011). However, the use of successive indirect assessments in an attempt to determine a plausible hypothesis for the functions of problem behavior may ultimately be more time-consuming than simply conducting a single functional analysis. For example, indirect assessments are likely to be inaccurate in identifying functions of problem behavior, and if a clinician spends considerable amounts of time basing a treatment on those indirect assessments, he or she may unnecessarily exert more effort than basing the treatment on a single functional analysis that produces conclusive results.

Interestingly, a considerably large body of research exists examining the utility of closed-ended interviews, and relatively few studies have examined the usefulness of open-ended interviews. Open-ended interviews have the advantage of greater interrater reliability, which may contribute more readily to a hypothesis-driven functional analysis compared to closed-ended interviews, which are constrained by the general reinforcement categories offered by the interview scale. However, the extent to which open- or closed-ended interviews provide greater concurrent validity with a functional analysis should be examined within subject in future research.

The inclusion of open-ended interviews as a basis to inform functional analysis design has been growing in recent years (e.g., Jessel et al., 2016, 2018). This may be justified in some respects, as open-ended interviews appear to be superior to closed-ended interviews in terms of interrater reliability. However, in addition to poor concurrent validity with functional analysis, which should serve as a caution for practicing behavior analysts, open-ended interviews are limited by (a) taking longer to administer than closed-ended interviews (e.g., the interviews in the present study took up to 60 min to complete), (b) the informant providing information not relevant to the function of problem behavior (e.g., Alex, Eric), and (c) potentially

requiring increased expertise to interpret results compared to closed-ended interviews (e.g., raters in the present study had to develop an independent hypothesis about the function of problem behavior, as there was no scale to rely on). This latter limitation suggests that formal training on the use, application, and interpretation of open-ended interviews may be necessary to accurately glean information from the indirect assessment (cf. Iwata, Wong, Riordan, Dorsey, & Lau, 1982).

In accordance with the recommendations made by others, we suggest that increased emphasis be placed on increasing the efficiency and practicality of functional analysis procedures for clinical behavior analysts working in the area of assessment and treatment of severe behavior disorders. It is likely that the reliance on indirect assessments as a core component of FBA will continue among behavior analysts in practice until current methodological limitations of functional analysis are overcome (see Roscoe et al., 2015, for a list of common reasons behavior analysts in practice do not conduct functional analyses). Most importantly, we strongly caution practicing behavior analysts against relying heavily on an open-ended interview for informing treatments for problem behavior, as the results of the present study provide preliminary yet further evidence of the moderate reliability and poor validity of indirect assessments.

Compliance with Ethical Standards

Ethical Approval This study was approved by the Research Ethics Board at Surrey Place Centre.

Informed Consent Informed consent for assessment, treatment, and publication of findings was provided in writing by the substitute decision maker.

References

- Barton-Arwood, S. M., Wehby, J. H., Gunter, P. L., & Lane, K. L. (2003). Functional behavior assessment rating scales: Intrarater reliability with students with emotional or behavioral disorders. *Behavioral Disorders, 28*, 386–400. <https://doi.org/10.1177/019874290302800402>
- Beaulieu, L., Van Nostrand, M. E., Williams, A. L., & Herscovitch, B. (2018). Incorporating interview-informed functional analyses into practice. *Behavior Analysis in Practice, 11*, 385–389. <https://doi.org/10.1007/s40617-018-0247-7>.
- Beavers, G. A., & Iwata, B. A. (2011). Prevalence of multiply controlled problem behavior. *Journal of Applied Behavior Analysis, 44*, 593–597. <https://doi.org/10.1901/jaba.2011.44-593>.
- Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 46*, 1–21. <https://doi.org/10.1002/jaba.30>.
- Bell, M. C., & Fahmie, T. A. (2018). Functional analysis screening for multiple topographies of problem behavior. *Journal of Applied Behavior Analysis, 51*, 528–537. <https://doi.org/10.1002/jaba.462>.

- Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, 24, 120–138. [https://doi.org/10.1016/S0891-4222\(03\)00014-3](https://doi.org/10.1016/S0891-4222(03)00014-3).
- Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J. G., McLaughlin, D. M., McAtee, M. L., et al. (1999). *Positive behavior support for people with developmental disabilities: A research synthesis*. Washington, DC: American Association on Mental Retardation.
- Crawford, J., Brockel, B., Schauss, S., & Miltenberger, R. G. (1992). A comparison of methods for the functional assessment of stereotypic behavior. *Journal of the Association for Persons With Severe Handicaps*, 17, 77–86. <https://doi.org/10.1177/154079699201700202>.
- Didden, R., Duker, P. C., & Korzilius, H. (1997). Meta-analytic study on treatment effectiveness for problem behaviors with individuals who have mental retardation. *American Journal on Mental Retardation*, 101, 387–399.
- Dracoby, J. D., Dozier, C. L., Briggs, A. M., & Juanico, J. F. (2018). Reliability and validity of indirect assessment outcomes: Experts versus caregivers. *Learning and Motivation*, 62, 77–90. <https://doi.org/10.1016/j.lmot.2017.02.007>.
- 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 classroom teachers. *Journal of Positive Behavior Interventions*, 2, 85–97. <https://doi.org/10.1177/109830070000200202>.
- Fisher, W. W., Greer, B. D., Romani, P. W., Zangrillo, A. N., & Owen, T. M. (2016). Comparisons of synthesized and individual reinforcement contingencies during functional analysis. *Journal of Applied Behavior Analysis*, 49, 596–616. <https://doi.org/10.1002/jaba.314>.
- Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment to enhance reinforcer identification. *American Journal on Mental Retardation*, 101, 15–25.
- Fryling, M. J., & Baires, N. A. (2016). The practical importance of the distinction between open and closed-ended indirect assessments. *Behavior Analysis in Practice*, 9, 146–151. <https://doi.org/10.1007/s40617-016-0115-2>.
- Green, G., & Striefel, S. (1988). Response restriction and substitution with autistic children. *Journal of the Experimental Analysis of Behavior*, 50, 21–32. <https://doi.org/10.1901/jeab.1988.50-21>.
- Gresham, F. M., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review*, 30, 156–172.
- Hagopian, L. P., Rooker, G. W., & Zarcone, J. R. (2015). Delineating subtypes of self-injurious behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, 48, 523–543. <https://doi.org/10.1002/jaba.236>.
- Hanley, G. P. (2012). Functional assessment of problem behavior: Dispelling myths, overcoming implementation obstacles, and developing new lore. *Behavior Analysis in Practice*, 5, 54–72. <https://doi.org/10.1007/BF03391818>.
- 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. <https://doi.org/10.1901/jaba.2003.36-147>.
- Hanley, G. P., Jin, C. S., Vanselow, N. R., & Hanratty, L. A. (2014). Producing meaningful improvements in problem behavior of children with autism via synthesized analyses and treatments. *Journal of Applied Behavior Analysis*, 47, 16–36. <https://doi.org/10.1002/jaba.106>.
- Herman, C., Healy, O., & Lydon, S. (2018). An interview-informed synthesized contingency analysis to inform the treatment of challenging behavior in a young child with autism. *Developmental Neurorehabilitation*, 21, 202–207. <https://doi.org/10.1177/019874290302800402>.
- Heyvaert, M., Saenen, L., Campbell, J. M., Maes, B., & Onghena, P. (2014). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: An updated quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, 35, 2463–2476. <https://doi.org/10.1016/j.ridd.2014.06.017>.
- Horner, R. H., Carr, E. G., Strain, P. S., Todd, A. W., & Reed, H. K. (2002). Problem behavior interventions for young children with autism: A research synthesis. *Journal of Autism and Developmental Disorders*, 32, 423–446. <https://doi.org/10.1023/A:1020593922901>.
- Iwata, B. A., DeLeon, I. G., & Roscoe, E. M. (2013). Reliability and validity of the functional analysis screening tool. *Journal of Applied Behavior Analysis*, 46, 271–284. <https://doi.org/10.1002/jaba.31>.
- Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice*, 1, 3–9. <https://doi.org/10.1007/BF03391714>.
- Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., et al. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis*, 27, 215–240. <https://doi.org/10.1901/jaba.1994.27-215>.
- Iwata, B. A., Wong, S. E., Riordan, M. M., Dorsey, M. F., & Lau, M. M. (1982). Assessment and training of clinical interviewing skills: Analogue analysis and field replication. *Journal of Applied Behavior Analysis*, 15, 191–203. <https://doi.org/10.1901/jaba.1982.15-191>.
- Jessel, J., Hanley, G. P., & Ghaemmaghami, M. (2016). Interview-informed synthesized contingency analyses: Thirty replications and reanalysis. *Journal of Applied Behavior Analysis*, 49, 576–595. <https://doi.org/10.1002/jaba.316>.
- Jessel, J., Ingvarsson, E. T., Metras, R., Kirk, H., & Whipple, R. (2018). Achieving socially significant reductions in problem behavior following the interview-informed synthesized contingency analysis: A summary of 25 outpatient applications. *Journal of Applied Behavior Analysis*, 51, 130–157. <https://doi.org/10.1002/jaba.436>.
- Kelley, M. E., LaRue, R. H., Roane, H. S., & Gadaire, D. (2011). Indirect behavioral assessments: Interviews and rating scales. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 182–190). New York, NY: Guilford Publications.
- Kuhn, D. E., DeLeon, I. G., Fisher, W. W., & Wilke, A. E. (1999). Clarifying an ambiguous functional analysis with matched and mismatched extinction procedures. *Journal of Applied Behavior Analysis*, 32, 99–102. <https://doi.org/10.1901/jaba.1999.32-99>.
- Lennox, D. B., & Miltenberger, R. G. (1989). Conducting a functional assessment of problem behavior in applied settings. *Journal of the Association for Persons With Severe Handicaps*, 14, 304–311. <https://doi.org/10.1177/154079698901400409>.
- Lewis, T. J., Mitchell, B. S., Harvey, K., Green, A., & McKenzie, J. (2015). A comparison of functional behavioral assessment and functional analysis methodology among students with mild disabilities. *Behavioral Disorders*, 41, 5–20. <https://doi.org/10.17988/0198-7429-41.1.5>.
- Matson, J. L., & Nebel-Schwalm, M. (2007). Assessing challenging behaviors in children with autism spectrum disorders: A review. *Research in Developmental Disabilities*, 28, 567–579. <https://doi.org/10.1016/j.ridd.2006.08.001>.
- Newcomer, L. L., & Lewis, T. J. (2004). Functional behavioral assessment: An investigation of assessment reliability and effectiveness of function-based interventions. *Journal of Emotional and Behavioral Disorders*, 12, 168–181. <https://doi.org/10.1177/10634266040120030401>.
- Oliver, A. C., Pratt, L. A., & Normand, M. P. (2015). A survey of functional behavior assessment methods used by behavior analysts in practice. *Journal of Applied Behavior Analysis*, 48, 817–829. <https://doi.org/10.1002/jaba.256>.
- Paclawskyj, T. R., Matson, J. L., Rush, K. S., Smalls, Y., & Vollmer, T. R. (2001). Assessment of the convergent validity of the Questions About Behavioral Function Scale with analogue functional analysis and the Motivation Assessment Scale. *Journal of Intellectual*

- Disability Research*, 45(6), 484–494. <https://doi.org/10.1046/j.1365-2788.2001.00364.x>.
- Querim, A. C., Iwata, B. A., Roscoe, E. M., Schlichenmeyer, K. J., Ortega, J. V., & Hurl, K. E. (2013). Functional analysis screening for problem behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, 46, 47–60. <https://doi.org/10.1002/jaba.26>.
- Reid, R., & Nelson, J. R. (2002). The utility, acceptability, and practicality of functional behavioral assessment for students with high-incidence problem behaviors. *Remedial and Special Education*, 23, 15–23. <https://doi.org/10.1177/074193250202300103>.
- 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. <https://doi.org/10.1002/jaba.259>.
- Roscoe, E. M., Rooker, G. W., Pence, S. T., Longworth, L. J., & Zarcone, J. (2009). Assessing the utility of a demand assessment for functional analysis. *Journal of Applied Behavior Analysis*, 42, 819–825. <https://doi.org/10.1901/jaba.2009.42-819>.
- Saini, V., Fisher, W. W., & Betz, A. M. (in press). Functional analysis: History and methods. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (2nd ed., pp.). New York, NY: Guilford.
- Saini, V., Fisher, W. W., & Retzlaff, B. J. (2018). Predictive validity and efficiency of ongoing visual-inspection criteria for interpreting functional analyses. *Journal of Applied Behavior Analysis*, 51, 303–320. <https://doi.org/10.1002/jaba.450>.
- Sigafoos, J., Kerr, M., & Roberts, D. (1994). Interrater reliability of the Motivation Assessment Scale: Failure to replicate with aggressive behavior. *Research in Developmental Disabilities*, 15, 333–342. [https://doi.org/10.1016/0891-4222\(94\)90020-5](https://doi.org/10.1016/0891-4222(94)90020-5).
- Slaton, J. D., Hanley, G. P., & Raftery, K. J. (2017). Interview-informed functional analyses: A comparison of synthesized and isolated components. *Journal of Applied Behavior Analysis*, 50, 252–277. <https://doi.org/10.1002/jaba.384>.
- Sugai, G., Lewis-Palmer, T., & Hagan-Burke, S. (2000). Overview of the functional behavioral assessment process. *Exceptionality*, 8, 149–160. https://doi.org/10.1207/S15327035EX0803_2.
- Umbreit, J. (1996). Functional analysis of disruptive behavior in an inclusive classroom. *Journal of Early Intervention*, 20, 18–29. <https://doi.org/10.1177/105381519602000104>.
- Zarcone, J. R., Rodgers, T. A., Iwata, B. A., Rourke, D. A., & Dorsey, M. F. (1991). Reliability analysis of the Motivation Assessment Scale: A failure to replicate. *Research in Developmental Disabilities*, 12, 349–360. [https://doi.org/10.1016/0891-4222\(91\)90031-M](https://doi.org/10.1016/0891-4222(91)90031-M).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.