Behavior Analysis in Practice
https://doi.org/10.1007/s40617-021-00665-6

RESEARCH ARTICLE

On the Reliability and Treatment Utility of the Practical Functional Assessment Process

Adithyan Rajaraman[1,2] · Gregory P. Hanley[1] · Holly C. Gover[1] · Kelsey W. Ruppel[1] · Robin K. Landa[1]

Accepted: 25 October 2021

Abstract

Saini et al. (2019) urged caution with respect to the use of practical functional assessment (PFA) procedures to inform behavioral treatment when they found that responses to an open-ended caregiver interview were only somewhat reliable and showed moderate to weak correspondence with analog functional analyses. Because the practitioner's goal in conducting any functional assessment process is to inform the successful treatment of problem behavior, we replicated and extended Saini et al. by (a) evaluating the reliability of hypotheses gleaned from two independent PFA processes for each of four children, (b) conducting treatment informed by a randomly assigned PFA, and (c) determining the extent to which potentially different levels of reliability impacted the treatment utility of the PFA process. Results indicated that the reliability of the PFA process varied depending on the stringency with which it was evaluated. However, treatments developed from randomly determined PFA processes produced efficacious outcomes on problem behavior and targeted social skills that transferred to the context designed from the other PFA process in all evaluations, suggesting that the PFA has strong treatment utility despite parts of the process having ambiguous levels of reliability. We discuss implications for practitioners tasked with treating severe problem behavior.

Keywords Open-ended indirect assessment · Practical functional assessment process · Reliability · Severe problem behavior · Skill-based treatment · Treatment utility

Functional assessments for severe problem behavior have received a great deal of attention in behavioral research (Hagopian, Dozier, et al., 2013a; Hanley et al., 2003; S. S. Johnston & O'Neill, 2001). Research is meant to inform practice, and researchers over the past several decades have endeavored to evaluate and refine functional assessments in service of ultimately helping practitioners create a socially meaningful resolution of problem behavior. Perhaps unsurprisingly, over 50 years of research has led to the development of many types of functional assessment, which has set the occasion for multiple comparative evaluations to determine which assessment type is most beneficial (e.g., Alter et al., 2008; English & Anderson, 2006; Fisher et al., 2016; Iwata et al., 2013; Paclawskyj et al., 2001; Saini et al., 2019). Although well intentioned, these comparisons sometimes result in conflicting conclusions and advice from researchers regarding which assessments to use in practice. Practitioners tasked with addressing severe problem behavior may find this troubling given the volume of information to sift through in order to find the most appropriate assessment for their circumstances.

An example of conflicting recommendations from researchers pertains to the practical functional assessment (PFA) process that was originally described by Hanley et al. (2014) in which an open-ended interview with caregivers (i.e., an indirect assessment), along with a brief direct observation, informed the design of a functional analysis of problem behavior. In Hanley et al. (2014), the results of the PFA process directly informed the contingencies manipulated in a skill-based treatment that successfully eliminated problem behavior in ecologically relevant contexts for three children. During the interview, caregivers nominated concerning topographies of their child's problem behavior and reported on the environmental events that often preceded

✉ Adithyan Rajaraman
arajaraman@umbc.edu

[1] Department of Psychology, Western New England University, Springfield, MA, USA

[2] Department of Psychology, University of Maryland, Baltimore County, 1000 Hilltop Cir, Baltimore, MD 21250, USA

Published online: 09 February 2022

and followed problem behavior. The analysts who conducted the interview then considered information about the possibly relevant events and hypothesized a synthesized reinforcement contingency to evaluate in a functional analysis. This type of functional analysis was termed an *interview-informed* synthesized contingency analysis (IISCA; Jessel et al., 2016)[1] because the variables tested were primarily those reported by caregivers in the interview and because interviewing is an ongoing element of the iterative PFA process (i.e., caregivers are often present during the analysis and are sometimes asked for more specific information if an analysis is not immediately successful in functionally controlling problem behavior).

A growing body of research has demonstrated the PFA process to directly contribute to socially meaningful improvements with respect to problem behavior (Beaulieu et al., 2018; Ferguson et al., 2020; Hanley et al., 2014; Herman et al., 2018; Jessel et al., 2019; Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018a; Jessel, Ingvarsson, Metras, Whipple, et al., 2018b; Rajaraman et al., 2021; Rose & Beaulieu, 2019; Santiago et al., 2016; Strand & Eldevik, 2018; Taylor et al., 2018). In spite of these findings, practitioners have been warned against its use as a pretreatment assessment due to its reliance on a subjective, open-ended indirect assessment of unknown psychometric properties such as reliability and validity (Cooper et al., 2019; Fisher et al., 2016; Greer et al., 2020; Saini et al., 2019; Tiger & Effertz, 2021).

The psychometric construct of *reliability* generally refers to the consistency of a measurement system or assessment (Kazdin, 2011). According to Kazdin (2011), there are various ways to assess the reliability of a particular assessment tool (e.g., test–retest, internal consistency, interrater), each of which speaks to different ways in which the assessment may or may not produce consistent findings. With respect to indirect functional assessments for problem behavior, *interrater* reliability has most commonly been evaluated and refers to the degree to which two individuals agree on hypotheses regarding the variables influencing problem behavior (e.g., Alter et al., 2008; Barton-Atwood et al., 2003; Fisher et al., 2016; Iwata et al., 2013; Newton & Sturmey, 1991; Saini et al., 2019; Sigafoos et al., 1994; Zarcone et al., 1991). It has been argued that this type of reliability is a necessary precondition for an assessment's *validity*—a construct that generally describes the accuracy with which an assessment identifies the variables influencing behavior (Cooper et al., 2019; Hanley, 2010, 2012; Iwata & Dozier, 2008; Johnston et al., 2019). The issue of unreliability reduces to the question of how one could know which assessment results are accurate if they stand in disagreement.

Psychometric constructs such as reliability and validity initially gained importance in behavioral research evaluating traditional psychological assessments. Traditional psychological assessments are typically designed to capture behavior patterns indicative of hypothetical constructs presumed to (a) exist *within the individual* and (b) be responsible for the observed behavior (Shapiro & Kratochwill, 2000; Shriver et al., 2001). The overall goal of traditional psychological assessment is to gather structural information about certain patterns of behavior such that one can infer the existence of such a hypothetical construct (e.g., emotional dysregulation), and because the constructs cannot be observed or measured, the obtained data are compared to a normative sample to suggest whether or not the individual exhibits behavior consistent with others said to also possess the construct. Because of the nature of inferences made in traditional psychological assessments, reliability has indeed been a critical precondition for validation in this domain. Divergent findings from these assessments may draw into question whether or not the assessment tool itself can accurately describe a particular internal construct.

In contrast to traditional psychological assessment, the goal of any functional assessment is to identify the environmental variables responsible for the problem behavior of an individual. Functional assessments are predicated on the behavior-analytic assumptions that (a) environmental contingencies are the primary determinants of behavior and (b) each individual has a unique learning history and therefore has unique interactions with the various environments they encounter (Carr, 1977; Cooper et al., 2019). Because each individual's problem behavior is assumed to be a product of their specific learning history, and because suspected variables are environmental and therefore observable, measurable, and manipulable, functional assessment data need not be compared to a normative sample. As such, reliability may not be the most appropriate construct in validating a particular functional assessment approach. In fact, reliability across raters should be relatively unlikely to be achieved given that each rater is an independent element of the individual's environment and may themselves represent different components of the contingencies influencing problem behavior. Divergent findings from a functional assessment, then, may instead represent different environmental contexts or contingencies that may each functionally contribute to problem behavior. This means that a functional assessment of questionable reliability may still be useful for practitioners

[1] Jessel et al. (2016) invoked the term *IISCA* in reference only to the functional analysis component of the multifaceted functional assessment process; however, the term has been somewhat misconstrued to refer to the entire assessment process in everyday parlance. To keep consistent with usage in Jessel et al. (2016) and promote clarity in term usage, we use the term *IISCA* to refer to the functional analysis, and we use *PFA* or *PFA process* to refer to the entire functional assessment process that includes both the open-ended interview and the IISCA.

if it captures one of possibly many contingencies influencing problem behavior.

A third construct worth mentioning is *treatment utility*, which describes the extent to which an assessment contributes to a positive treatment outcome (English & Anderson, 2006; Hayes et al., 1987; Iversen, 2013; Slaton et al., 2017). Several researchers have suggested that treatment utility may be a more appropriate means of functional assessment validation than an evaluation of reliability in the absence of treatment (Evans & Nelson, 1977; Ghaemmaghami, Hanley, Jin, & Vanselow, 2016b; Gresham & Lambros, 1998; Gresham et al., 2001; Hayes et al., 1987; Iversen, 2013; Newcomer & Lewis, 2004; Shapiro & Kratochwill, 2000; Shriver et al., 2001; Silva, 1993; Slaton et al., 2017; Sturmey, 1994). For example, Shriver et al. (2001) argued that "an [assessment] method may have demonstrated reliability and accuracy, but if the data do not contribute to developing effective treatment then the assessment method is not considered useful" (p. 189). Indeed, English and Anderson (2006) found that analog functional analyses, which are known to be reliable and valid methods of assessment, did not consistently have strong treatment utility. These authors, along with Hayes et al. (1987), recommended an evaluation of the direct and general effects of an intervention informed by a functional assessment as a more appropriate marker of its utility (Durand & Carr, 1991).

The relationship between reliability and treatment utility is poorly understood, in large part because the majority of studies in which researchers concluded that indirect assessments were not useful for informing intervention—based on insufficient reliability—did not involve evaluation of any treatment informed by the assessments they dismissed (e.g., Alter et al., 2008; Barton-Atwood et al., 2003; Fisher et al., 2016; Iwata et al., 2013; Saini et al., 2019). This gap in the literature leaves us with little understanding of how the reliability of assessment tools relates to the fundamental practical goal of successfully treating problem behavior. Iversen (2013) summarized this sentiment cogently, commenting that

> functional assessment has appeared to have taken on a life of its own, separate from the goal of providing impetus for intervention, with publications presenting results from assessment alone without actual follow-up intervention. . . . However, *reliability of assessment outcome cannot be fully evaluated without a direct link between the inferred causes of problem behavior in assessment and the outcome of subsequent intervention using manipulations of these inferred causes* [emphasis added] for the same client. (p. 21)

A study published by Saini et al. (2019) provides an example, relevant to the PFA process, of the concern expressed by Iversen (2013). Saini et al. conducted a preliminary evaluation of the reliability and validity of the open-ended interview included in the PFA process. In this study, analysts compared their hypotheses of behavioral functions based on two independent, open-ended interviews with a common caregiver, and additionally compared those hypotheses against analog functional analyses. These authors found that, across four consecutively enrolled participants, the open-ended interview was only somewhat reliable at identifying categorical functions of behavior and resulted in poor correspondence when compared against functional analysis outcomes. Without evaluating treatment, Saini et al. concluded that the open-ended interview was limited in its utility, ultimately cautioning practitioners against using the PFA process to inform behavioral treatment.

Saini et al. (2019) suggested that treatments informed by the PFA process may be ineffective *because* they are informed by an assessment of questionable reliability. However, given that the vast majority of studies dedicated to evaluating the reliability of indirect assessments did so in the absence of a treatment evaluation—including Saini et al.—we counter that the relation between treatment utility and the more traditional concept of reliability warrants further investigation. As Iversen (2013) implied, in order to discount an assessment approach, its inability to contribute to a positive treatment outcome should be demonstrated first, which requires an evaluation of a treatment informed by the assessment findings.

We therefore attempted to evaluate both the reliability and the treatment utility of the PFA process in the present study. We asked the following questions: How reliable is the process by which an analysis is designed from the open-ended interview? If unreliable, to what extent does the variability imposed by the assessment process impact its treatment utility? By conducting two independent PFA processes across four consecutively enrolled children who engaged in problem behavior, this study sought to provide multiple measures of the reliability of the PFA process. By evaluating whether or not the effects of the treatment developed from one of the PFA processes generalized to the context informed by the other PFA process, this study sought to assess the treatment utility of the PFA process.

Method

Two teams independently conducted separate PFA processes on the problem behavior of the same child until (a) differentiated responding across test and control conditions was observed in each IISCA and (b) the teams felt confident enough in the findings of their IISCA such that they *could* develop and evaluate a skill-based treatment replicating procedures outlined in Hanley et al. (2014). Conducting two independent assessment processes allowed for an evaluation of the reliability of the PFA process (Saini et al., 2019).

The results from *one* of those processes were then randomly chosen to inform the design of a skill-based treatment; we refer to this as the *treatment context*, which describes (a) a location in which (b) a particular synthesized reinforcement contingency was manipulated by (c) a particular analyst. Skill-based treatment commenced, and once a terminal performance criterion was met (e.g., elimination of problem behavior and consistent emission of target skills during periods of nonreinforcement), we conducted posttests in both the treatment context and the context established from the other PFA process that was not selected to inform treatment. We refer to this other context as the *generality context*—describing a different location, analyst, and synthesized contingency—because treatment was never conducted in this space following the differentiated IISCA. Data from the posttests in each context were then compared to data from the test sessions of the respective IISCAs (i.e., baseline) to evaluate the direct efficacy and generality of a treatment informed by one randomly chosen PFA. This comparison provided data from which to interpret the PFA process's treatment utility in relation to its reliability. That is, if the independent PFA processes yielded unreliable findings, but a randomly chosen PFA contributed to positive, generalizable treatment effects, it would suggest that the unreliability of the PFA process did not preclude its utility in informing effective treatment.

Participants and Settings

The first six children referred to our university-based outpatient clinic for the assessment and treatment of problem behavior were slated to participate in this study following its approval from the institutional review board. Any children who had a differentiated IISCA with an analyst were included. Children who required parent implementation of contingencies in the IISCA to achieve differentiation were excluded from the study because of the increased variability that would be imposed on the treatment process and our interpretation of treatment utility if the parent required additional training on procedures. Two of the original six children were excluded and transferred to a parent-driven assessment and treatment study for this reason. The other four participated in the current study. This study therefore meets the requirements of a consecutive controlled case-series design (Hagopian, 2020; Hagopian, Rooker, et al., 2013b; Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018a; Slaton et al., 2017). All participants were children living with two parents, referred to the clinic by their pediatrician's office due to dangerous, intractable, and worsening problem behavior in their home (all children) and school (Jeffrey), and all had a history of physical aggression toward other children.

Jeffrey[2] was a White 9-year-old boy with attention-deficit/hyperactivity disorder and generalized anxiety disorder who communicated vocally and fluently. Brandon was a White 3-year-old boy with no formal psychiatric diagnosis who communicated vocally using short, broken sentences. Henry was an Arab American 5-year-old boy with autism who communicated vocally in one- to three-word utterances. Sophie was a White 4-year-old girl with no formal psychiatric diagnosis who communicated vocally and fluently.

All sessions were conducted in one of two small rooms (4 m × 3 m) or a conference room (8 m × 4 m) at the university clinic, each equipped with a video camera, a one-way observation mirror, two child-sized tables, two to three chairs, and play and academic materials as nominated in each child's caregiver interview. Sessions were 5 min in duration throughout the IISCAs and functional communication training (FCT) phases. Thereafter, session duration was determined by the time required to complete five evocative trials (i.e., five presentations of the putative establishing operation [EO]). A trial was considered complete when reinforcement was earned or 30 min had elapsed (the latter never occurred). Session duration therefore varied between three and 35 min in phases following FCT.

All sessions were conducted by Board Certified Behavior Analysts, licensed to practice in Massachusetts, who were doctoral candidates pursuing degrees in behavior analysis. Parents were required to be at the clinic during all sessions and either watched from behind an observation mirror or watched and participated in the session.

Response Definitions and Measurement

Trained observers used data collection software on laptop computers to record target responses and relevant environmental events during each session. We measured problem behavior, simple and complex functional communication responses (FCRs), tolerance responses, and contextually appropriate behavior (CAB).

For all children, target problem behavior included aggression, disruption, and other problem behavior, some of which were nondangerous but reported to co-occur with more severe topographies. For Brandon, target problem behavior also included self-injurious behavior (SIB). Tables 1, 2, 3 and 4 provide specific descriptions of topographies within these response categories, including responses that were nominated in the initial interview, as well as those that parents confirmed to be relevant during the PFA process.

The programmed simple FCR, which was scored as independent if no analyst prompt preceded it by 5 s, regardless of

[2] Data from one of Jeffrey's functional analyses, as well as his skill-based treatment, were published in Rajaraman et al. (2021).

Table 1 Two functional analysis descriptions from independently conducted practical functional assessment processes for Jeffrey and their agreement

Event type	Agreement level	Functional analysis designed from interview with mother and father (treatment context)	Functional analysis designed from interview with father (generality context)	% agreement
Establishing operations	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	Directed to terminate iPad, table games, *Gameboy* via *least-to-most physical prompting* Directed to *complete writing tasks*, *correct written mistakes* via vocal and gestural prompting Compliments and *apologies* withheld; criticisms of *thumb sucking delivered* Mands ignored	Directed to terminate iPad, table games, *Beyblades*, *Uno cards* via *vocal and gestural prompting* Directed to *play game differently*, *talk about past misbehavior* via vocal and gestural prompting Compliments withheld; criticisms of *game play delivered* Mands ignored	57
Responses	Categorical	Aggression, disruption, other	Aggression, disruption, other	100
	Specific	Headbutting, punching, kicking Disruptive vocals, sighing loudly, *banging surfaces*, *throwing items* Elopement, *head in hands*	Headbutting, punching, kicking Disruptive vocals, sighing loudly, *stomping* Elopement, *climbing*, *scowling*	50
Reinforcers	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	All demands terminated iPad, table games, *Gameboy* delivered Compliments and *apologies* delivered; analyst in close proximity to respond to bids Mands reinforced	All demands terminated iPad, table games, *Beyblades*, *Uno cards* delivered Compliments delivered; analyst in close proximity to respond to bids Mands reinforced	79

The items in agreement are in regular typeface, and items in disagreement are italicized

vocal cadence, was "My way, please" for Brandon, Jeffrey, and Henry and "Can we please play my way?" for Sophie. The programmed complex FCR for all children required them to obtain a listener vocally, wait for adult acknowledgment, and emit a mand while making eye contact. Jeffrey's and Henry's complex FCR was "Excuse me. . . . May I have my way, please?" Brandon's complex FCR was "Excuse me. . . . My way, please?" Sophie's complex FCR was "Hey, [adult's name]. . . . Can we please play my way?" The programmed tolerance response was "I'm cool with that" or "That's cool with me" for Jeffrey, "That's cool" for Brandon, and "OK" for Sophie and Henry. Independent complex FCRs and tolerance responses were scored only when the phrases were uttered with an appropriate vocal cadence and volume and were 5-s removed from an analyst prompt.

Data were also collected on the duration and number of CAB expectations programmed by analysts during delays to reinforcement and the percentage of those expectations for which children emitted CAB (i.e., cooperation with adult instruction). CAB expectations were those that were presented by the analyst upon the termination of reinforcement (in the IISCA only) and during the delay to reinforcement following a denial cue. Any instruction posed by the analyst counted as a CAB expectation (e.g., discrete academic demands, instructions to play by oneself with less preferred toys, instructions to clean up). CAB engagement, then, was specific to the expectation in place but was scored only if emitted in the absence of problem behavior or noncompliance lasting longer than 10 s (e.g., completing the academic task without problem behavior, playing with less preferred toys without problem behavior or any off-task behavior for longer than 10 s).

Counts of problem behavior, FCRs, and tolerance responses were converted to rates for all evaluations. The percentage engagement in CAB was calculated by dividing the number of independent CAB observed by the number of CAB expectations in place per session and multiplying that quotient by 100. Duration data were collected for periods of reinforcement (i.e., any interval in which all the reinforcers identified to be relevant in the PFA were available to the child) and the total session. Because we tracked when reinforcement was programmed during the IISCA, we were able to determine and report whether target behaviors occurred during reinforcement or EO periods.

Interobserver agreement was evaluated by having a trained second observer simultaneously but independently collect data on all target behavior and CAB expectations during at least 22% of randomly determined sessions within each phase—including sessions within each condition of the IISCAs—for all children (range 22%–67%). Agreement was

Table 2 Two functional analysis descriptions from independently conducted practical functional assessment processes for Brandon and their agreement

Event type	Agreement level	Functional analysis designed from interview with mother (treatment context)	Functional analysis designed from interview with mother and father (generality context)	% agreement
Establishing operations	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	Car toys, *bouncy balls*, *animal* toys removed Directed to play via *least-to-most prompting* with *iPad*, *blocks or directed to clean up toys* Analyst's *and mother's* attention diverted to *removed item* Mands ignored	Car toys removed Directed to play via *vocal prompting only* with *puzzles*, *flash cards*, *beads* Analyst's attention diverted to *iPad* Mands ignored	60
Responses	Categorical	SIB, aggression, disruption, other	SIB, aggression, disruption, other	100
	Specific	Slapping own face Hitting, spitting on others Screaming, swearing, whining Sticking tongue out, *crossing arms and pouting*	Slapping own face Hitting, spitting on others Screaming, swearing, whining, *throwing items* Sticking tongue out	81
Reinforcers	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	All demands terminated Car toys, *bouncy balls*, *animal* toys delivered Attention delivered; analyst *and mother stay* in close proximity to respond to social bids Mands reinforced by analyst *and mother*	All demands terminated Car toys delivered Attention delivered; analyst stays in close proximity to respond to bids Mands reinforced by analyst	71

The items in agreement are in regular typeface, and items in disagreement are italicized. *SIB* self-injurious behavior

calculated by partitioning each session into 10-s intervals and dividing the number of measures in agreement per interval by the number of disagreements plus agreements per interval and multiplying the quotient by 100. If both observers scored 0 for any measure in a given interval, 100% agreement was attributed. For all dependent measures, mean interobserver agreement was 98% (range 83%–100%) for Jeffrey, 97% (range 82%–100%) for Brandon, 97% (range 80%–100%) for Henry, and 97% (range 79%–100%) for Sophie.

Procedural integrity was evaluated by having a trained observer score the extent to which procedures were implemented as prescribed across each phase of the study. Procedural integrity was coded in approximately 46% (range 30%–73%) of randomly determined sessions across analyses, treatments, and posttests for each child. Items were scored as correct only if they were implemented correctly across the entire session (i.e., an error on a single trial within a session would result in that item being scored as incorrect). The percentage of correctly implemented components in the functional analysis and reversal phases (see Appendix 1), as well as the treatment and posttest phases (see Appendix 2), was calculated by dividing the number of correctly implemented procedural components in a session by the total number of components and multiplying by 100. If a checklist item did not apply to procedures for a given treatment phase (e.g., if a child did not make an unreasonable request during reinforcement; delivering prompts during posttest evaluations), it was scored as "not applicable" and the item was omitted from the calculation for that session. Across all evaluations, the mean procedural integrity for functional analysis and reversal sessions was 99% (range 98%–100%). The mean procedural integrity for skill-based treatment was 99% (range 99%–100%). Procedural integrity during posttests was 100%.

Experimental Design

We used a multielement design to compare test and control conditions in the IISCAs. In treatment, we used a reversal design to demonstrate control over problem behavior and the simple FCR for Brandon, Henry, and Sophie. In addition, for all children, we systematically and progressively changed the criterion of the response(s) required to produce reinforcement across treatment phases. Control over behavior by the synthesized reinforcement contingency identified in the IISCA was demonstrated when rates of problem behavior and alternative responses changed in predictable directions with each change in contingency.

Table 3 Two functional analysis descriptions from independently conducted practical functional assessment processes for Henry and their agreement

Event type	Agreement level	Functional analysis designed from interview with mother and father (treatment context)	Functional analysis designed from interview with father (generality context)	% agreement
Establishing operations	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	Dinosaur and animal toys, *balls*, *blocks*, *pretend-play toys* removed by analyst physically Directed to *do seated work alone* via *vocal prompting* Father's attention withheld; *analyst's attention limited to prompting of demands* Mands *denied by analyst* and ignored by father	Dinosaurs and animal toys, *iPad*, *car toys*, *Hungry Hungry Hippos* removed by analyst physically *(unless directed to clean up toys)* Directed to *put on red jacket*, *brush teeth*, *clean up toys* via *least-to-most physical prompting* Father's attention withheld; *analyst's attention limited to prompting of demands* Mands *ignored by analyst* and father	45
Responses	Categorical	Aggression, disruption	Aggression, disruption, *other*	66
	Specific	Hitting, kicking, *poking eyes* Throwing, stomping, screaming, whining, *ripping items*, *growling*	Hitting, kicking, *scratching*, *biting* Throwing, stomping, screaming, whining, *spitting on items*, *coloring surfaces* *Elopement*, *covering ears*	30
Reinforcers	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	All directives terminated Dinosaur and animal toys, *balls*, *blocks*, *pretend-play toys* delivered Attention delivered by analyst and father Mands reinforced by analyst and father	All directives terminated Dinosaurs and animal toys, *iPad*, *car toys*, *Hungry Hungry Hippos* delivered Attention delivered by analyst and father, *including bids to play chase* Mands reinforced by analyst and father	75

The items in agreement are in regular typeface, and items in disagreement are italicized

The experimental question regarding the treatment utility of the PFA process was evaluated using a pretest/posttest design with random assignment of treatment. Data from the IISCA test sessions served as the pretest against which posttest performance was compared. Treatment utility of the PFA was considered evident if the effects of the skill-based treatment on problem behavior and target skills in the treatment context were also observed in the treatment- and generality-context posttests.

There were two points of randomization determined by coin flips. The first coin flip randomly determined which team would interview first, which served to minimize potential interviewer effects across children. The second coin flip randomly determined the treatment context, as well as the default generality context. The random assignment of the treatment context served to prevent biases with respect to analyst speculation as to which context might positively moderate a transfer of effects. Treatment was therefore assigned randomly to a context, regardless of which IISCA process was conducted first.

Prior to the posttest evaluations, sessions were conducted in the treatment context in which an additional analyst (i.e., one who did not serve as the primary analyst in either IISCA context for that child) implemented skill-based treatment. Child performance requirements were identical to the CAB chaining sessions conducted immediately prior. These sessions were conducted to ensure that any deviations from predicted effects in the generality-context posttest were a function of the differences in the contextual features of the synthesized contingency and not merely the child's reactivity to a somewhat novel analyst. These sessions were alternated with standard CAB chaining sessions (i.e., with the primary treatment analyst) until stable responding was observed.

PFA Process

Conducting Two Blinded Assessment Processes

For Jeffrey, Brandon, and Henry, the treatment context was randomly determined to be designed by the team that interviewed first. For Sophie, the treatment context was randomly determined to be designed by the team that conducted the second interview (i.e., the first interview informed her generality context). For clarity, the order of events will be

Table 4 Two functional analysis descriptions from independently conducted practical functional assessment processes for Sophie and their agreement

Event type	Agreement level	Functional analysis designed from interview with mother (treatment context)	Functional analysis designed from interview with mother (generality context)	% agreement
Establishing operations	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	Barbies, baby dolls, *car* toys *sometimes* removed by *CA* Directed to play alone with *whatever toys are not being used* or *to play interactively with current toys differently* via vocal prompting Analyst attention *withheld*; *CA attention given while prompting different play* Mands ignored by analyst *and CA*	Barbies, baby dolls, *blocks*, *stickers*, *stuffed animal* toys *always* removed by *analyst* Directed to play alone with *puzzles* and *coloring books* via vocal prompting Analyst attention *diverted to removed item* Mands ignored by analyst	30
Responses	Categorical	Aggression, disruption, other	Aggression, disruption, other	100
	Specific	Hitting Arguing, *whining* Withholding items	Hitting, *pinching*, *mock hitting* Arguing, *stomping*, *screaming* Withholding items, *hiding*	36
Reinforcers	Categorical	Escape, tangibles, attention, mand compliance	Escape, tangibles, attention, mand compliance	100
	Specific	All directives terminated Barbies, baby dolls, and *car* toys delivered Attention delivered by analyst *as she mediates conflict between Sophie and CA*. Bids for attention reinforced Mands reinforced by analyst *and CA*	All directives terminated Barbies, baby dolls, *blocks*, *stickers*, *stuffed animal* toys delivered Attention delivered by analyst *with ongoing commentary and compliments*; bids for attention reinforced Mands reinforced by analyst	62

The items in agreement are in regular typeface, and items in disagreement are italicized. CA = confederate adult

described in accordance with the process experienced by Jeffrey, Brandon, and Henry, in which the first interview informed the treatment context.

The order of events during the PFA process was as follows: The first treatment team conducted a caregiver interview on the first visit to the clinic. The second treatment team conducted the second interview with the same caregiver on the second visit.[3] On a subsequent visit, the first treatment team conducted an IISCA informed by their interview in the treatment context. Then, the second treatment team conducted an IISCA in the generality context. As mentioned previously, the generality context was distinguished from the treatment context with respect to the room, the analyst(s), the potential materials included, and the potential synthesized contingency analyzed.

Teams were blinded to one another's procedures and outcomes while conducting the PFA. If the first treatment team was conducting an interview or analysis, members of the second treatment team were asked to not be present at the clinic for the duration of the family's visit (and vice versa). Furthermore, all treatment teams were asked to not share their interview or analysis results with the other team or the clinic director (i.e., Dr. Gregory Hanley) until both PFAs were completed and a treatment context had been assigned. Parents were asked to not share any information about one team's process with the other team.

PFA Procedures

The PFA process emulated procedures described in Slaton et al. (2017), which differs from Hanley et al. (2014) in one important way. Whereas Hanley et al. (2014) and other studies evaluating the PFA process described a structured observation that occurred after the interview and helped inform the design of the IISCA (e.g., Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018a; Santiago et al., 2016), Slaton et al. and the present study omitted the observation portion of the PFA. Thus, the PFA process in this study consisted of a caregiver interview and an IISCA.

[3] For Jeffrey, Brandon, and Henry, one parent participated in both interviews, but a second parent joined to participate in one of the interviews. We followed parent preference regarding involvement in the PFA process with the minimum requirement being that at least one parent needed to be present during all sessions. We chose this approach, despite the additional variance it likely imposed on interview reliability, to help make parents feel comfortable, and because it was more likely to negatively impact rather than inflate reliability, thereby providing a potentially more difficult test of treatment utility.

Analysts received training in conducting the PFA process through prior research and clinical experience as doctoral students under the second author's mentorship. From the analyst's perspective, the goal of the interview was to obtain the information necessary about all potentially ecologically relevant EOs, target problem behaviors, and reinforcers such that an analysis could be conducted wherein (a) a synthesized EO rapidly and reliably evokes (b) any topography of problem behavior reported to co-occur with the most concerning forms, which immediately cease upon the contingent delivery of (c) synthesized reinforcers. Analysts were instructed to ask the questions as outlined in Hanley (2012) while focusing on identifying (a) dangerous and associated nondangerous topographies of problem behavior, (b) all of the EOs suspected to evoke problem behavior, and (c) all of the possible reinforcers suspected to maintain problem behavior. Follow-up questions (not necessarily on the interview form in Hanley, 2012) were therefore presented for clarification as needed.

IISCAs were conducted based on hypotheses gleaned from interviews. IISCAs consisted of two conditions: a test condition in which the suspected synthesized contingency was present, and a control condition in which the synthesized contingency was absent. During the control condition, the analyst provided the child with continuous access to all of the suspected reinforcers (e.g., toys, a tablet, adult availability and compliance, and the absence of demands). The analyst did not present any of the EOs suspected to evoke problem behavior and did not provide any differential consequences if problem behavior occurred. In other words, the child experienced noncontingent reinforcement for the duration of the control condition. Each analysis began with a control condition to verify that problem behavior did not occur under the conditions designed to eliminate all relevant EOs.

The test session began with the child experiencing reinforcement. However, within the first minute of the test session, the analyst interrupted reinforcement by presenting the suspected synthesized EO (e.g., removal of toys and attention while presenting academic demands). Upon the first instance of any problem behavior or associated less dangerous behavior, the analyst terminated any expectations and immediately delivered the suspected positive reinforcers for approximately 20 s to 90 s. The process of periodically interrupting reinforcement with the imposition of EOs continued for the duration of the 5-min session. If problem behavior was not evoked immediately during the EO presentation, the analyst would progress the EO by adding more components (e.g., by escalating expectations to cooperate with an adult instruction). Similar to the procedures in the control condition, the analyst did not provide any differential consequences if problem behavior occurred during the reinforcement periods within the test session.

Certain tactics were included in the procedures to promote safety within the analyses. Importantly, all members of the treatment team were instructed to terminate the IISCA session if problem behavior escalated to an unsafe point. Furthermore, although the initial interviews provided descriptive classes of responses to be reinforced in the IISCA, all analysts were poised to reinforce *any* seemingly problematic response in the analysis, regardless of whether or not it was included in the descriptive class. If this occurred, a member of the treatment team immediately consulted with the observing parent to ask if the response usually preceded or co-occurred with severe problem behavior. This occurred in at least one IISCA for each child, and parents confirmed the relevance of the unreported topography in every instance. Receiving verbal confirmation from parents was an important reason why caregivers attended all IISCA sessions. Reinforcing associated nondangerous topographies has been shown to prevent escalation to more severe problem behavior (thereby promoting safe analyses), and these responses have been demonstrated to share response-class membership in virtually every study in which they have been evaluated (e.g., Borrero & Borrero, 2008; Herscovitch et al., 2009; Schmidt et al., 2020; Smith & Churchill, 2002; Warner et al., 2020)

Analyses were implemented by each team's primary analyst, with the exception of Sophie's treatment-context IISCA, in which the team designed a context that involved having a second analyst act as Sophie's play partner (i.e., a confederate). This modification was included because the interview suggested that any play period with Sophie's older sister or with other peers was most likely to involve problem behavior.

For all children, any lack of cooperation with CAB expectations presented by the analyst in the test sessions that did not co-occur with problem behavior was met with demand escalation that was individualized within each IISCA (see Tables 1, 2, 3 and 4 for descriptions of how demands were presented and escalated in each respective IISCA). Test and control sessions were alternated until (a) functional control over problem behavior by the synthesized contingency was demonstrated and (b) the treatment team and observing caregiver felt satisfied with the information gleaned from the process and were poised to design a skill-based treatment.

Reliability Analysis

Following the completion of the PFA process, a reliability analysis was conducted for each child with the information used in the design of each successful analysis. In other words, we evaluated the reliability of the hypothesized contingencies gleaned by analysts from the open-ended interviews. The event types reported on were the EOs programmed during the test condition, the responses included in the target contingency class, and the reinforcers provided contingent on problem behavior in the test condition and available freely in the control condition. Information regarding the EOs, responses, and reinforcers was examined through two different levels of stringency—categorical and specific—to allow for the analysis of reliability across levels.

We chose to evaluate interassessor reliability of the design of the analysis as informed by the interview because this represented the hypotheses gleaned from the interview regarding controlling variables that were verified in analysis and that could be subjected to evaluations of reliability (Saini et al., 2019, used the same tactic to examine reliability). Although evaluations of reliability in closed-ended assessments are often conducted across raters because the number of possible responses is fixed and agreement can therefore be coded as a binary measure, open-ended assessments can generate an infinite number of responses for which agreement cannot easily be calculated. Therefore, we followed the precedent set by studies that have evaluated the reliability of open-ended indirect assessments and calculated agreement based on the hypotheses extracted from the interview responses (Arndorfer et al., 1994; Fisher et al., 2016; Kinch et al., 2001; Reed et al., 1997; Saini et al., 2019; Yarbrough & Carr, 2000).

Agreement for the categorical levels of analysis within event types was calculated by measuring the partial agreement of each identified category in the analysis. For example, in categorical EOs, agreement was calculated by dividing the number of categories found to be in agreement across analyses by the total number of categories identified, and multiplying the quotient by 100 (see Table 1 for an example).

Agreement for the specific level of event types was calculated by measuring the partial agreement of each specified action and item/activity in the analysis. An example can be seen in Brandon's reliability analysis (see Table 2). Within specific EOs, four specific actions were programmed by the adult in each analysis for Brandon. With respect to tangible items in the IISCA, the analyst in the treatment context removed car toys, bouncy balls, and animal toys, whereas the analyst in the generality context removed car toys only. We measured partial agreement of this action by noting agreement between "car toys" and the action of physically "removing" items (as opposed to "asking" Brandon to relinquish them) but disagreement with respect to "bouncy balls" and "animal toys." With respect to tangible items, then, there were two agreements and two disagreements, yielding a partial agreement score of .5 for that line. We calculated similar agreement for all actions taken in each EO, dividing the number of agreements by the total number of agreements and disagreements and multiplying the quotient by 100 to generate agreement values. This calculation was applied to the specifics of all reported EOs, responses, and reinforcers across all PFA processes.

Skill-Based Treatment Procedures

Procedures for skill-based treatment closely emulated those outlined in Hanley et al. (2014). Across all treatment phases, the synthesized reinforcement contingency identified in the treatment context IISCA was arranged to teach and maintain skills, and all problem behavior was placed on extinction. We used most-to-least prompting to teach target skills, fading prompts as independent skills emerged. Criteria to progress across treatment phases (i.e., from simple FCT, to complex FCT, to tolerance response training, to CAB chaining) were two consecutive sessions with zero problem behavior during EO intervals and consistent, independent emission of the skill(s) targeted in that phase (e.g., emitting a complex FCR following each EO presentation during complex FCT). If mastered skills were not emitted independently during the EO, an additional prompt (e.g., "You can say 'Can I please have my way?'") was delivered approximately every 30 s.

Simple FCT

Sessions in this phase looked similar to the test sessions in the corresponding IISCA context, except that a simple FCR was required to produce reinforcement, whereas problem behavior was placed on extinction. Each treatment session began with the child experiencing synthesized reinforcement. Within the 1st minute of each session, the analyst interrupted reinforcement by presenting the synthesized EO and used most-to-least prompting and differential reinforcement to teach the simple FCR. If the child emitted the simple FCR, it was reinforced immediately for approximately 20 to 90 s, at which point the EO was re-presented.

Contingency Reversal

Brandon, Henry, and Sophie experienced a brief contingency reversal. Procedures were identical to the test sessions of the IISCA, wherein problem behavior was reinforced and all other responses were placed on extinction during the programmed EO.

Complex FCT

This phase consisted of a similar procedure to that of the simple FCT, except that both problem behavior and the simple FCR were placed on extinction, and instead the complex FCR was required to produce the synthesized reinforcers.

Table 5 Terminal Contextually Appropriate Behavior (CAB) expectations in the treatment and generality-context posttests

Child	Average amount or duration of terminal CAB expectations per session	
	Treatment context	Generality context
Jeffrey	5 min of writing paragraphs on an analyst-directed topic, with proofreading and editing, while the analyst diverted attention to a phone	5 min of adult-directed play with preferred or less preferred games, while the analyst gave demands to comment and reflect on past misbehavior
Brandon	17 demands (about 2 min) to clean up toys or engage with less preferred items while the analyst diverted attention to preferred toys or Brandon's mother	2 min of independent engagement with less preferred items while the analyst diverted attention to a phone
Henry	33 demands to read tasks or play independently with less preferred items	15 demands to put on a jacket, practice brushing teeth, or clean up toys
Sophie	21 demands (about 3 min) to tolerate and cooperate with a play partner (i.e., a confederate adult) while the partner picked what to play and how to play it	2 min of independent engagement with less preferred items while the analyst diverted attention to preferred toys

Tolerance Response Training

Procedures in this phase were similar to complex FCT, except that during 60% of randomly determined trials, the analyst did not immediately reinforce the complex FCR and instead required a tolerance response. During these trials, the analyst replied to the complex FCR with a denial cue (e.g., "No, you cannot have your way not now.") and subsequently prompted and differentially reinforced a tolerance response (e.g., "That's cool with me.").

CAB Chaining

This phase involved gradually introducing CAB expectations (i.e., demands that evoked problem behavior in the IISCA; e.g., instructions to clean up or complete homework) following the emission of a tolerance response on some trials. In CAB chaining, the complex FCR and tolerance response were still each reinforced immediately on one to five trials (20% each), but CAB expectations of varying length and difficulty were programmed in the remaining three of five trials (60%), all of which were randomly programmed. Engagement in CAB was required to produce reinforcement on trials in which CAB expectations were in place. Thus, an intermittent and unpredictable schedule of reinforcement was used to maintain target skills. We used within- and across-session shaping to achieve terminal goals with respect to CAB expectations (similar to procedures for *contingency-based progressive delay* in Ghaemmaghami, Hanley, & Jessel, 2016a). Context-specific terminal CAB expectations were agreed on by the child's caregiver(s) and each respective treatment team, in both the treatment and generality context, prior to initiating treatment (see Table 5 for a description of specific terminal CAB expectations for each child).

For all children, the performance criterion to move to the posttest evaluation was two sessions in the CAB chaining phase with terminal expectations in place and (a) zero problem behavior during EO intervals, (b) consistent emission of the complex FCR and tolerance response (without requiring any prompts), (c) experience in the EO interval for approximately 50% of each session, and (d) at least 90% total CAB engagement.

Posttest Evaluation

Treatment-Context Posttest

The treatment-context posttest sessions were almost identical to the final CAB chaining sessions, in which each analyst programmed the terminal range of CAB expectations during delays to reinforcement on 60% of trials, while reinforcing the complex FCR and tolerance response in the remaining trials. The only procedural difference was that no prompts were programmed, regardless of whether or not acquired skills were emitted. This posttest represented the final day of skill-based treatment in the treatment context. Posttest data for this context consisted of the average performance in the final two sessions.

Generality-Context Posttest

The generality-context posttest involved having children return to the context in which they participated in the other IISCA (i.e., the one not chosen to inform treatment), which was in a different room, with a different analyst, and a potentially different synthesized contingency from the treatment context. However, instead of an IISCA test session with reinforcement arranged for problem behavior, procedures emulated the treatment-context posttest, which were almost identical to the end of the CAB chaining phase but without prompts to emit target skills. The analyst in the generality-context posttest programmed the synthesized

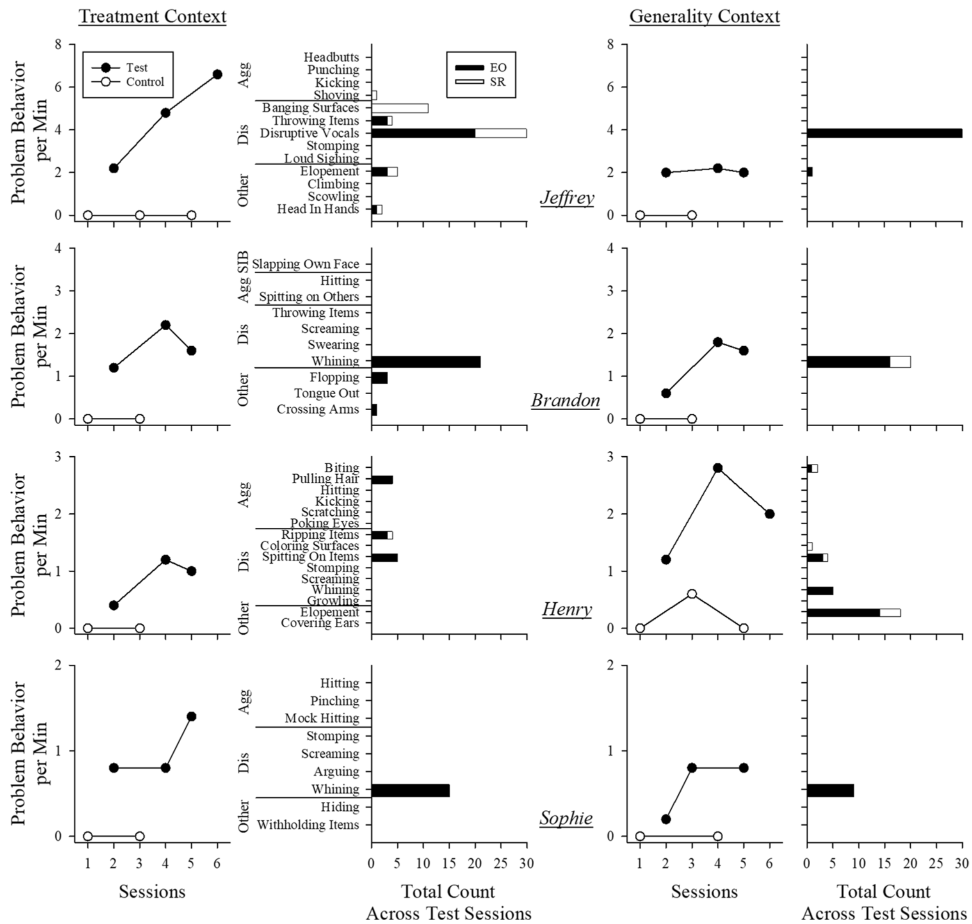

Fig. 1 Results of two independent interview-informed synthesized contingency analyses conducted with the participants. *Note.* Jeffrey's results are shown in the top panel, Brandon's in the second panel, Henry's in the third panel, and Sophie's in the bottom panel. The first and third columns display problem behavior aggregated into a single measure per session. The second and fourth columns display counts of occurrences of individual topographies of problem behavior (*y*-axis labels). Responses that occurred during the establishing operation (EO) are represented by black bars, whereas responses that occurred during reinforcement are represented by white bars. Agg = aggression; Dis = disruption; SIB = self-injurious behavior; EO = establishing operation; SR = reinforcement

contingency identified in their IISCA, including all relevant materials, individuals, and interactions, and differentially reinforced either a complex FCR, a tolerance response, or CAB on a similarly unpredictable, intermittent schedule (see the right column of Table 5 for a description of the terminal CAB expectations programmed in this posttest). All problem behavior was placed on extinction. In addition, no presession instructions were delivered to children regarding what was going to happen during the generality posttest. This was to help establish as stringent a test of the transfer of effects to the generality context as possible. This posttest occurred on the child's next visit to the clinic following the treatment-context posttest. Members of the team that conducted skill-based treatment were absent from the clinic. The generality of effects was evaluated across three sessions for Jeffrey, Brandon, and Henry and in two sessions for Sophie. Posttest data for this context consisted of the average performance in the final two sessions. Throughout the posttest process, and similar to the PFA process, treatment team members and parents were instructed to not share any information regarding one team's posttest with the other team or with the clinic director.

Results

PFA Process

Figure 1 shows the results of both IISCAs for Jeffrey, Brandon, Henry, and Sophie. All eight functional analyses demonstrated behavioral sensitivity to a synthesized reinforcement contingency of escape to tangibles, attention, and mand compliance. In all eight analyses, elevated rates of problem behavior were observed almost exclusively in test conditions. Jessel et al. (2020) summarized and validated several ways to categorize and interpret the level of functional control achieved in IISCAs. These included, but were not limited to, an examination of the percentage of nonoverlapping data (PND) between test and control data paths, and novel, multilevel structured criteria (described as an extension of the structured criteria developed by Hagopian et al., 1997). The multilevel structured criteria categorized an analysis as having *strong* control if there was no overlap in the data across conditions and if no problem behavior was observed in the control condition. Given that there was no overlap between test and control data paths (i.e., 100% PND) across all eight IISCAs, and that there was no problem behavior observed in the control conditions of seven of the eight IISCAs, we consider these results to show strong control over problem behavior (because some problem behavior was observed in one control session for Henry, his generality-context IISCA would be categorized as having demonstrated *moderate* control based on the criteria outlined in Jessel et al., 2020).

In addition, we measured whether problem behavior occurred during EO or reinforcement periods within the test sessions of each IISCA. The second and fourth columns in Fig. 1 display not only the number of occurrences of each topography of problem behavior (the absence of a bar denotes that the response was not observed during the analysis) but also the proportion of those responses that occurred in the EO relative to reinforcement. Strong control is also evident by the fact that problem behavior occurred exclusively or primarily during the EOs of the test conditions for seven of the eight analyses. In some cases (e.g., Jeffrey's treatment-context IISCA), problem behavior that was evoked by the EO persisted into the subsequent reinforcement period, which may obscure the interpretation of control. However, it is important to note that on the few occasions in which problem behavior carried over briefly into reinforcement periods in test conditions, we replicated zero levels of problem behavior in subsequent control conditions, suggesting strong control over behavior by the contingency (Hagopian et al., 1997; Jessel et al., 2020).

The specific topographies of problem behavior listed for each child in the second and fourth column of Fig. 1 include both responses that were nominated in the initial interview and responses that were initially unreported but emerged and were asserted by parents to be associated with problem behavior. Across analyses for each child, the percentage of topographies for which there was consistency with respect to occurrence and nonoccurrence was 69%, 80%, 60%, and 100% for Jeffrey, Brandon, Henry, and Sophie, respectively. All topographies of problem behavior denoted in Fig. 1 were thus targeted for reduction in treatment.

Reliability Analysis

Tables 1, 2, 3 and 4 describe the qualitative details along with the agreement between the two IISCAs informed by independent PFA processes for Jeffrey, Brandon, Henry, and Sophie, respectively. The information presented in Tables 1, 2, 3 and 4 represent the contingencies that resulted in differentiated IISCAs (i.e., those that could have been randomly chosen for treatment development).

Six of the eight analysis descriptions represent the initial design by the treatment team, which yielded differentiation in the IISCA. Two of the eight analyses required redesign following initially undifferentiated IISCAs. For example, Brandon's analysis in the treatment context was redesigned to include his mother's attention in the reinforcement interval and because an iPad that was initially reported to be a viable substitute reinforcer for his mother's phone was observed to evoke problem behavior in the initial analysis. It was subsequently removed from the array of suspected tangible reinforcers and instead programmed as an independent leisure task in the EO. Henry's analysis in the treatment context was redesigned to include his father's attention in the reinforcement interval. The analyst also observed that Henry would cooperate with adult instruction when physically guided by the analyst, and Henry's father noted that independent task demands appeared more evocative in the home; thus, Henry was instead cued to read or play with less preferred items independently during the EO of the differentiated analysis.

Figure 2 summarizes the agreement data for the differentiated analysis designs for Jeffrey, Brandon, Henry, and Sophie. At the categorical level of analysis, agreement was 100% across EOs and reinforcers programmed in the IISCAs for all children (see the top panel of Fig. 2). All problem behavior was sensitive to contingency categories of escape, attention, tangibles, and mand compliance. For Jeffrey, Brandon, and Sophie, agreement was 100% for categorical responses as well, whereas for Henry, categorical responses were 66% in agreement. Taken together, the average agreement across all categorical event types and children was 97%.

At the specific level of analysis, agreement was lower and more variable across children and event types. Across EOs, responses, and reinforcers, respectively, agreement was 57%, 50%, and 79% for Jeffrey; 60%, 81%, and 71% for Brandon; 45%, 30%, and 75% for Henry; and 30%, 36%, and 62% for Sophie. Some representative examples of qualitative differences across analysis designs include the EO for escape programmed across Jeffrey's analyses (e.g., escape from written work vs. escape from playing an adult-directed game; Table 1), the types of aggression targeted in Henry's analyses (e.g., eye poking vs. scratching and biting; Table 3), and the nature of the attention delivered during Brandon's reinforcement periods (e.g., from two adults vs. one adult; Table 2). Taken together, the average agreement across all specific event types and children was 57%.

In summary, the analysis design process predicated on an open-ended interview resulted in high agreement at the categorical level of analysis and lower, more variable agreement at the specific level of analysis. Although there is no established, universal criterion with which to consider an indirect assessment reliable, Iwata et al. (2013) suggested 80% as a reasonable marker due to its typical consideration as acceptable with respect to direct-observation measures. If this precedent is taken, one might consider the PFAs conducted in the current study to be reliable at the categorical level but unreliable at the specific level. In other words, the reliability of the PFA process depends on the level of stringency with which it is measured.

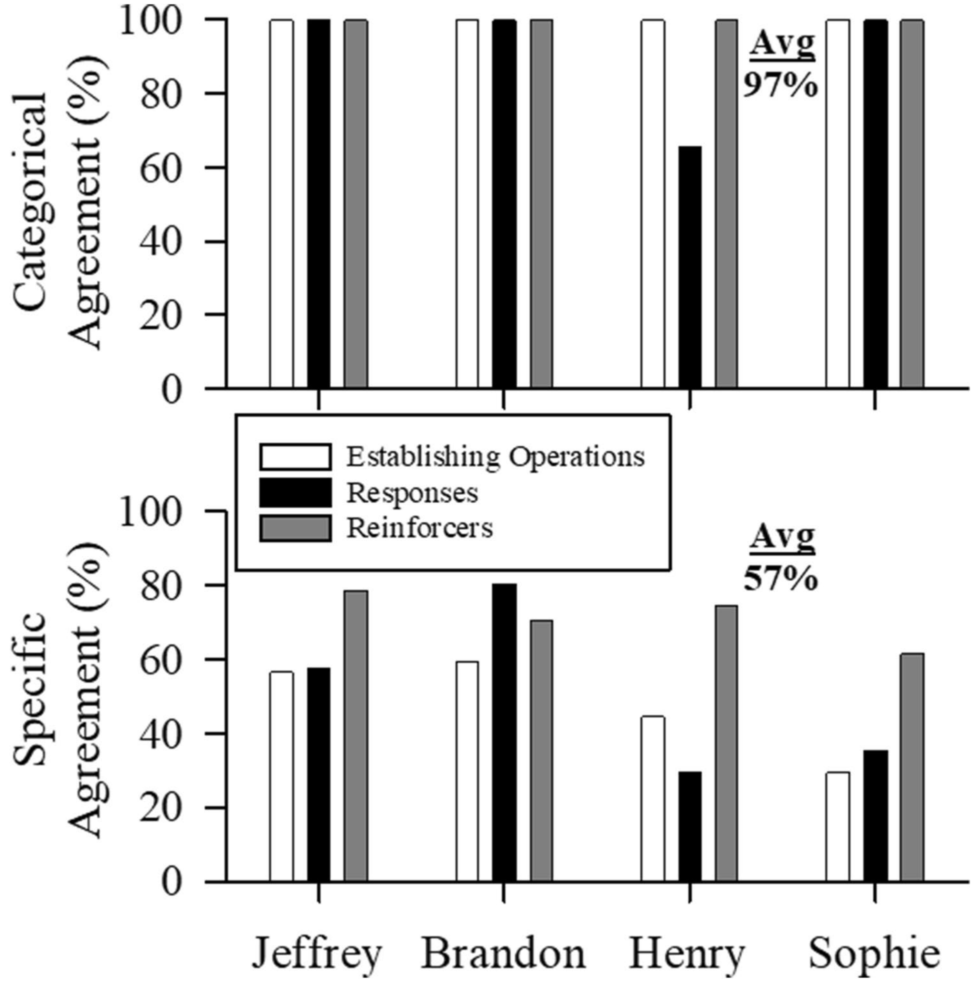

Fig. 2 Agreement percentages of the event types in two independent analysis designs informed from practical functional assessment processes for Jeffrey, Brandon, Henry, and Sophie. *Note.* The top panel shows categorical agreement percentages; the bottom panel shows specific agreement percentages. Avg = average percentage of agreement across all four children

Treatment-Context Evaluation and Posttest

The skill-based treatment procedures were successful in eliminating problem behavior and developing simple and complex FCRs, tolerance responses, and CABs during delays to reinforcement for all children in the treatment context. Figure 3, which depicts pretest, treatment, and posttest data for Brandon, serves as a representative example of the skill-based treatment process (a summary of outcomes for all participants is presented in Fig. 4; skill-based treatment process graphs are not reported because Jeffrey's data were published in Rajaraman et al., 2021). Functional control by the synthesized contingency over Brandon's problem behavior and replacement responses was demonstrated with a reversal and when targeted social skills emerged only when included in the synthesized reinforcement contingency. Figure 3 also depicts the duration that Brandon spent in EO and reinforcement periods across sessions (bottom panel), demonstrating that the problem behavior remained low compared to baseline levels despite increasing delays to reinforcement. Brandon's CABs, which were shaped during delays, were reinforced exclusively with the events identified in the PFA process (see Table 5 for a description of responses targeted during the delay and expected of Brandon in the transfer context posttest). Data for all four PFA and skill-based treatment evaluations emulated those that were published in Hanley et al. (2014) and Santiago et al. (2016).

Whereas Fig. 3 depicts the comprehensive process experienced by Brandon, including both posttests, Fig. 4 presents a summary of pretest (i.e., baseline) and posttest data for all participants. The data to the left of the dashed line in Fig. 4 represent the baseline and posttest performance in the treatment context for all children, an evaluation of the direct effects of a skill-based treatment informed by a randomly chosen PFA process.

Jeffrey displayed an average of 5.5 problematic responses per minute, 0 complex FCRs and tolerance responses, and an average of 10% CAB engagement in baseline. In the treatment-context posttest, Jeffrey exhibited 0 problem behavior, an average of 0.2 complex FCRs and tolerance responses per minute, and 100% CAB engagement during delays to reinforcement.

Brandon displayed an average of 1.9 problematic responses per minute, 0 complex FCRs and tolerance

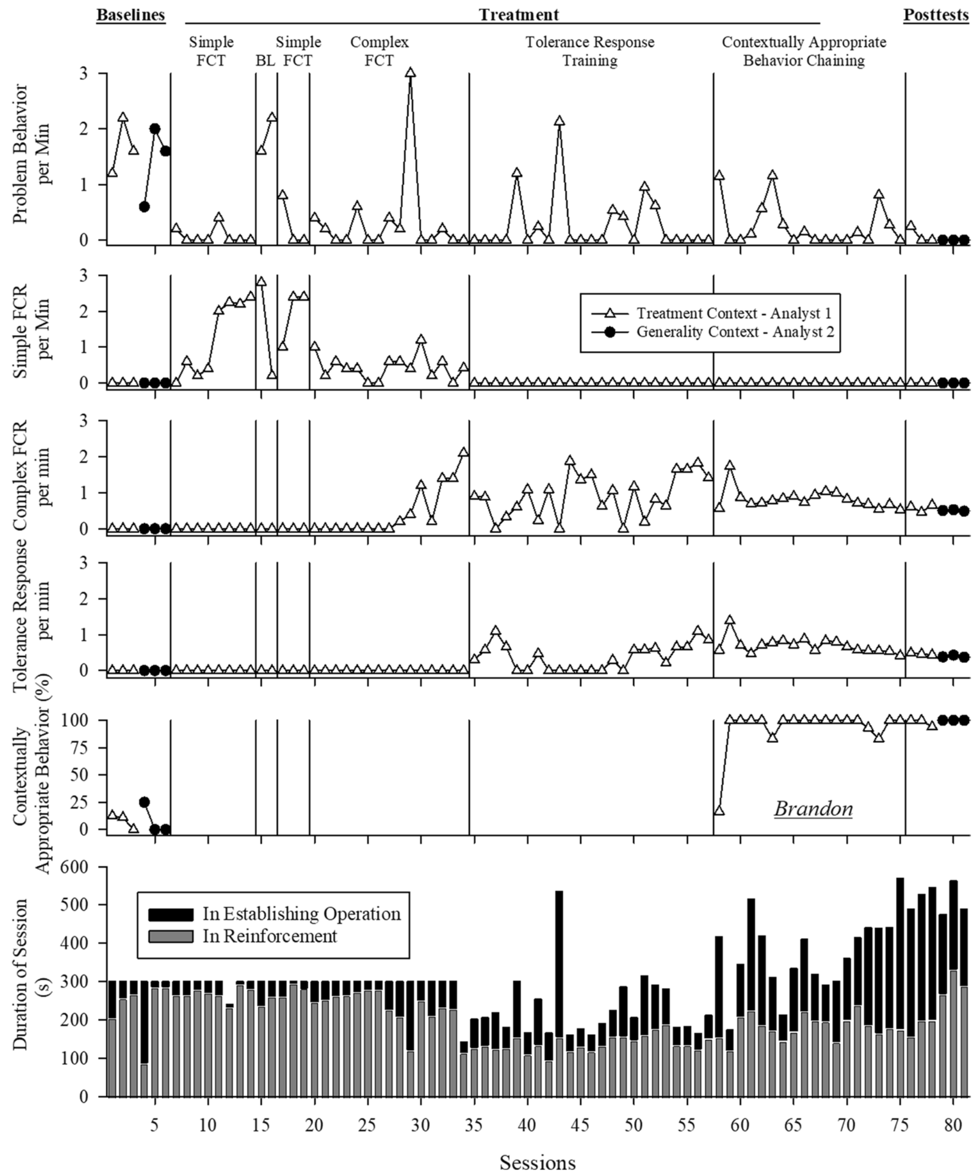

Fig. 3 Skill-based treatment evaluation and Practical Functional Assessment (PFA) Posttests for Brandon. *Note.* For Brandon, the initial PFA context was randomly assigned for treatment. FCT = functional communication training; FCR = functional communication response; BL = baseline

responses, and an average of 6% CAB engagement in baseline. In the treatment-context posttest, Brandon exhibited 0 problem behavior; an average of 0.6 and 0.5 complex FCRs and tolerance responses per minute, respectively; and 97% CAB engagement during delays to reinforcement.

Henry displayed an average of 1.1 problematic responses per minute, 0 complex FCRs and tolerance responses, and an average of 0% CAB engagement in baseline. In the treatment-context posttest, Henry exhibited 0 problem behavior; an average of 0.6 and 0.5 complex FCRs and tolerance responses per minute, respectively; and 100% CAB engagement during delays to reinforcement.

Sophie displayed an average of 1 problematic response per minute, 0 complex FCRs and tolerance responses, and an average of 35% CAB engagement in baseline. In the treatment-context posttest, Sophie exhibited 0 problem behavior, an average of 0.3 complex FCRs and tolerance responses per minute, and 97% CAB engagement during delays to reinforcement.

In summary, skill-based treatment applied in a randomly assigned IISCA context was successful in eliminating problem behavior and increasing target social skills for all children to an extent consistent with other treatment-oriented studies relying on the PFA process (e.g., Hanley et al., 2014; Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018a; Rose & Beaulieu, 2019; Santiago et al., 2016).

Almost no variability was imposed on treatment measures when an additional analyst conducted two sessions in the training context in the absence of the primary analyst for each child. Across four children, differences between the means in the treatment-context

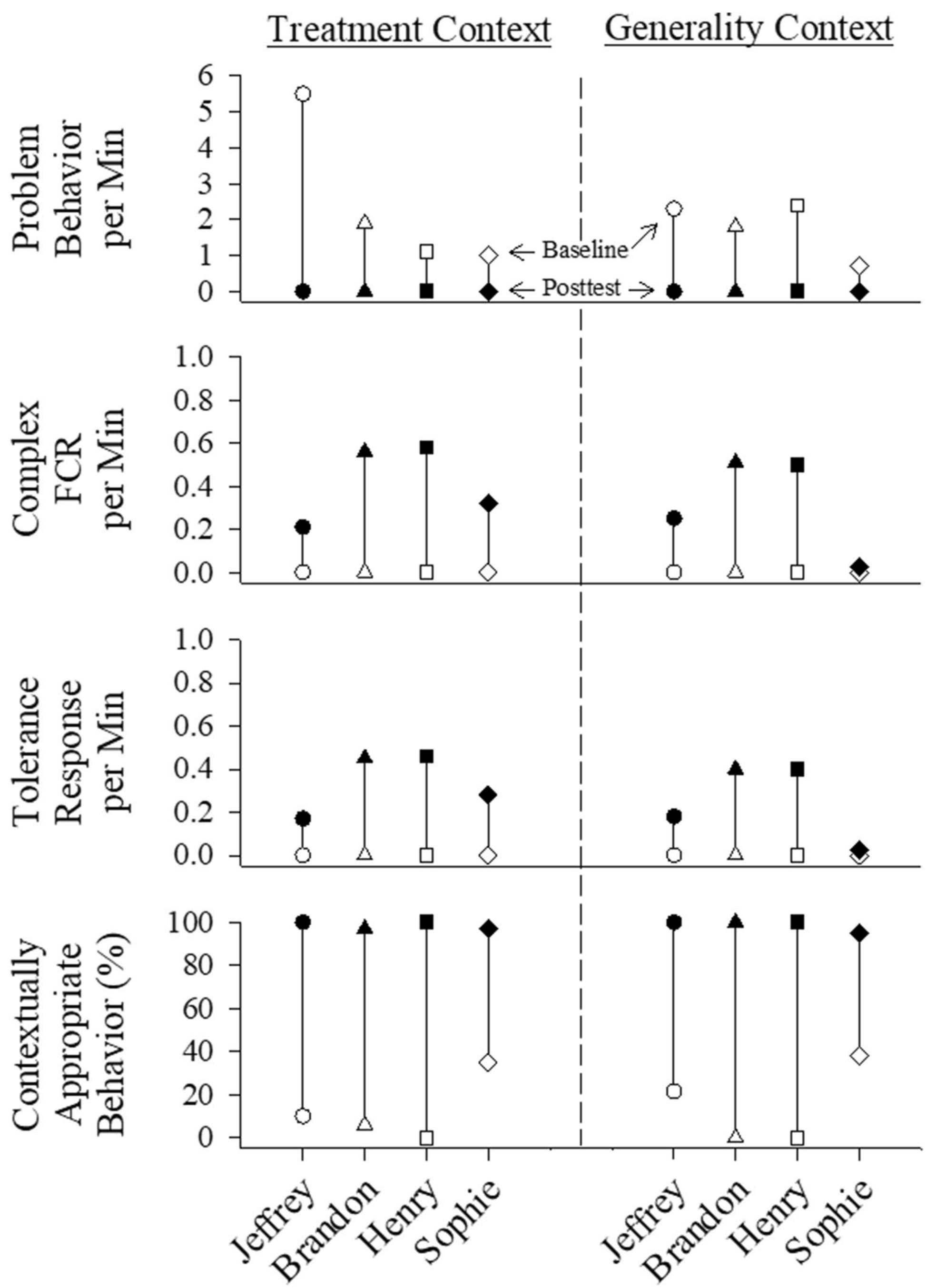

Fig. 4 Treatment and generality evaluation summary data. *Note.* Each data point represents the average of the final two sessions of each respective phase. FCR = functional communication response

posttest and the means of the two sessions conducted by the additional analyst were 0 for problem behavior per minute, 0.1 for complex FCRs per minute, 0.1 for tolerance responses per minute, and 2% for CAB. These data show that the effects of treatment generalized and maintained when implemented in the same context by a novel person. These data minimize the threat that any variability observed in the transfer test was a function of introducing a novel person, which could have otherwise complicated conclusions regarding the utility of the assessment process.

Generality-Context Posttest

The data to the right of the dashed line in Fig. 4 represent the baseline and posttest in the generality context for all children. Fourteen of the 16 effects of skill-based treatment observed in the treatment contexts transferred to the contexts designed from the other PFA process in which no teaching of skills occurred.

Jeffrey displayed an average of 2.3 problematic responses per minute, 0 complex FCRs and tolerance responses, and an average of 21% CAB engagement in baseline. In the generality-context posttest, Jeffrey exhibited 0 problem behavior; an average of 0.3 and 0.2 complex FCRs and tolerance responses per minute, respectively; and 100% CAB engagement during delays to reinforcement.

Brandon displayed an average of 1.8 problematic responses per minute, 0 complex FCRs and tolerance responses, and an average of 0% CAB engagement in baseline. In the generality-context posttest, Brandon exhibited 0 problem behavior; an average of 0.5 and 0.4 complex FCRs and tolerance responses per minute, respectively; and 100% CAB engagement during delays to reinforcement.

Henry displayed an average of 2.4 problematic responses per minute, 0 complex FCRs and tolerance responses, and an average of 0% CAB engagement in baseline. In the generality-context posttest, Henry exhibited 0 problem behavior; an average of 0.5 and 0.4 complex FCRs and tolerance responses per minute, respectively; and 100% CAB engagement during delays to reinforcement.

Sophie displayed an average of 0.7 problematic responses per minute, 0 complex FCRs and tolerance responses, and an average of 38% CAB engagement in baseline. In the generality-context posttest, Sophie exhibited 0 problem behavior, almost 0 complex FCRs and tolerance responses per minute, and 95% CAB engagement during delays to reinforcement.

Problem behavior that was initially observed at strength in the IISCAs did not occur in the transfer tests for any child, despite no experience in this context other than immediate reinforcement of problem behavior. With the exception of the complex FCR and tolerance response for Sophie, all targeted social skills were observed in the transfer test at levels similar to those observed in the treatment context. These data show that the effects of skill-based treatment on problem behavior and targeted social skills generalized to an untrained context and maintained despite similarly lengthy delays to reinforcement and in the presence of somewhat novel EOs that were demonstrated to be similarly evocative in baseline. These data demonstrate the treatment utility of the PFA process.

Discussion

In analyses informed by independent caregiver interviews, functional relations were demonstrated for four consecutive children's problem behavior across two categorically similar, but specifically unique, synthesized reinforcement contingencies. We found that the PFA process was reliable at the categorical level but showed questionable reliability at the specific level. Furthermore, the variability imposed by the different manners in which analyses were designed did not have any negative impact on the utility of the PFA process, as treatments designed from one analysis were similarly efficacious in the context associated with the other analysis. These findings suggest that the PFA process, which relies on a subjective, open-ended interview and an interview-informed analysis, has sufficient utility in assessing and treating problem behavior.

Evidence of the treatment utility of the PFA was demonstrated in four consecutive applications when (a) the skill-based treatment that was developed from the findings of a randomly determined PFA process produced robust effects on problem behavior and targeted social skills and when (b) those effects generalized to a different baseline of problem behavior established from an independent assessment process. Predicted treatment effects (e.g., low rates of problem behavior and the consistent emission of target skills) in both the treatment and generality-context posttests provided evidence for the treatment utility of the assessment process. Whereas the former demonstrated that a randomly assigned PFA process identified a functionally relevant synthesized reinforcement contingency that was directly manipulated across conditions to produce efficacious treatment outcomes, the latter suggested that any variability across baselines (i.e., any differences in context indicative of questionable reliability) did not negatively impact the likelihood of a successful treatment outcome. In addition, whereas the former represents a systematic replication of the positive outcomes of the PFA and skill-based treatment process (e.g., Beaulieu et al., 2018; Hanley et al., 2014; Herman et al., 2018; Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018a; Jessel, Ingvarsson, Metras, Whipple, et al., 2018b; Rajaraman et al., 2021; Santiago et al., 2016; Taylor et al., 2018), the latter extends the skill-based treatment literature.

The findings of the current study add to and extend Saini et al. (2019). Whereas Saini et al. found that the hypotheses gleaned from open-ended interviews were only somewhat reliable (75%) at the categorical level, the present results found almost perfect (97%) reliability when evaluated with the same level of stringency. However, the present study's findings add a level of nuance to the interpretations made by Saini et al. in that reliability was weaker and more variable (57%) when examined through a closer lens (see Fig. 2). Given the different findings with respect to reliability, the results of the current study could be interpreted in two ways. If reliability is considered at the categorical level, the PFA may be considered to possess strong treatment utility and stand as a reliable process. By contrast, if reliability is considered exclusively at the specific level, the PFA may be considered to possess strong treatment utility despite being a somewhat unreliable process. The important point for practitioners who have an opportunity to address severe problem behavior is that treatment utility of the PFA process remains independent of one's preferred interpretation of the reliability findings.

Another contribution to the literature on PFA and skill-based treatment pertains to the generality of effects observed in the current study. The existence of a baseline in the randomly assigned generality context allowed for an evaluation of generality not yet described in the skill-based treatment literature in that the effects observed in the posttest could reasonably be attributed to the treatment experienced in the other context (Durand & Carr, 1991).

The findings of the current study notably provide a preliminary understanding of the relation between the reliability of the PFA and its treatment utility—namely, that stringent reliability was not necessary to develop a useful treatment from the PFA process. These data importantly provide evidence in support of the suggestion, put forth by multiple researchers, that treatment utility may be a more appropriate means of functional assessment validation than an evaluation of reliability in the absence of treatment (Evans & Nelson, 1977; Ghaemmaghami, Hanley, Jin, & Vanselow, 2016b; Gresham & Lambros, 1998; Gresham et al., 2001; Hayes et al., 1987; Iversen, 2013; Nelson & Hayes, 1979; Newcomer & Lewis, 2004; Shapiro & Kratochwill; 2000; Shriver et al., 2001; Silva, 1993; Slaton et al., 2017; Sturmey, 1994).

We argue that emphasizing treatment utility in the validation of functional assessments could help mitigate confusion when sorting through the literature to identify optimal assessments to employ in practice. Furthermore, we submit that the commitment to determining *personalized* and *synthesized* reinforcement contingencies identified from the PFA process and verified to be functionally relevant in the IISCA was a critical aspect influencing treatment utility, and that unreliability may be a somewhat irrelevant construct in this idiographic and self-correcting process. In other words, because the PFA process is aimed at achieving some understanding of the reinforcement history of an individual's problem behavior, an emphasis on identifying personally relevant controlling variables is likely to influence treatment utility, whereas an emphasis on achieving a somewhat arbitrary level of reliability may not. Finally, we argue that the skill-based approach likely contributed to the treatment efficacy and generality demonstrated in the present study. Teaching skills, such as mands for reinforcers shown to influence problem behavior and extended chains of cooperative behavior for producing similar reinforcement, may have positively moderated the transfer of effects to the untrained context, as adaptive skills have the capacity to enable performances beyond the context in which they are taught (e.g., Durand & Carr, 1992; Hernandez et al., 2007) and may thus insulate the process from potential errors imposed by the subjective starting point of the PFA.

There are many aspects of the current study that provide opportunities for future investigation. First, given that practitioners are ultimately interested in making socially meaningful improvements with respect to their clients' problem behavior, it is important to point out that the current study was conducted by well-trained analysts in a university setting. A future study could use methods similar to those described in the current study to examine the treatment utility of the PFA in contexts relevant to the client (e.g., school, home) when the process is implemented by relevant caregivers (e.g., parents, teachers). Second, if positive treatment outcomes are interpreted with respect to the high degree of reliability at the categorical level, it could be inferred that the specifics of the analysis design did not matter as long as there was generic agreement across applications. This notion appears to be indirectly supported by the fact that all four children demonstrated problem behavior sensitive to a categorically similar synthesized reinforcement contingency (i.e., escape to tangibles, attention, and mand compliance), suggesting that a generic programming of these four common contingencies may have been sufficient for a successful treatment outcome (Tiger & Effertz, 2021). This is probably not the case with respect to the current study's findings because this interpretation ignores the fact that both synthesized contingencies for each participant were indeed individualized from interviews. It is therefore not possible to rule out the influence of the specific, personalized information obtained from the interview on the outcomes observed. Future research could again use similar methods to compare outcomes of treatments informed by analyses that incorporate personalized contingencies to those of treatments informed by analyses that incorporate categorically similar synthesized contingencies not informed by an interview. The other context could include either arbitrarily selected variables (clinic materials and typical interactions) or variables reported to be irrelevant to problem behavior, thereby clarifying the role of personalization from interviews. Finally, there are many factors that may have facilitated generalized performance that could be evaluated more closely in future research. For example, the programming of varied EOs in treatment may have facilitated generalized performances (akin to the "train loosely" tactic conveyed by Stokes & Baer, 1977). The role of unprogrammed verbal mediation, given the vocal-verbal repertoire of the children in this study, could have also positively moderated the transfer of effects into novel contexts (akin to the mediated generalization tactic conveyed by Stokes & Baer, 1977).

The treatment utility of the PFA process was supported in the current study, but its utility has been evident via other types of evaluations. First, the treatment utility of the PFA process was initially made apparent when it yielded socially validated outcomes with respect to severe problem behavior (Hanley et al., 2014; Santiago et al., 2016). Second, socially validated outcomes from PFA-informed treatments were replicated by several groups of authors across varied contexts, providing further evidence of its utility (Beaulieu et al., 2018; Ferguson et al., 2020; Herman et al., 2018; Jessel et al., 2019; Jessel, Ingvarsson, Metras, Kirk,

& Whipple, 2018a; Jessel, Ingvarsson, Metras, Whipple, et al., 2018b; Rose & Beaulieu, 2019; Strand & Eldevik, 2018; Taylor et al., 2018). Third, Slaton et al. (2017) demonstrated treatment utility in an evaluation comparing interventions developed from a PFA and a standard functional analysis (i.e., procedures developed by Iwata et al., 1982/1994) for four participants. Slaton et al. found that treatments informed by PFAs resulted in the elimination of problem behavior and the acquisition of an FCR in 100% of applications, whereas treatments informed by standard analyses only did so in 50% of applications. Fourth, Ghaemmaghami, Hanley, Jin, and Vanselow (2016b) provided evidence of treatment utility when they found that problem behavior that was shown to be sensitive to a synthesized contingency of escape to tangibles and attention in an IISCA was only eliminated in treatment when FCRs were acquired for all three reinforcers. Finally, Jessel, Ingvarsson, Metras, Kirk, and Whipple (2018a) reported effective treatment outcomes in 25 of 25 consecutive applications of skill-based treatment developed from PFAs. The current study extends the notion of strong treatment utility by replicating positive direct effects in four consecutive cases with treatments informed by PFAs and by showing generalized effects in different evocative contexts designed from independent PFA processes. Collectively, these six types of evidence for the treatment utility of the PFA process show that the subjectivity of an open-ended interview may not limit the utility of a functional assessment process that includes an analysis to verify possible controlling variables and a treatment that develops important skills.

Appendix 1

Procedural Integrity Checklist for Functional Analysis and Reversal Sessions (Adapted From Whelan et al., 2021)

Control condition	
	Begins analysis with the control condition
	Provides continuous access to synthesized reinforcers (e.g., follows child's lead) during control
	Refrains from implementing any potential establishing operations for problem behavior during control
	Ignores problem behavior if it occurs during control; continues reinforcement
Test/reversal condition	
	Initiates synthesized establishing operation upon the start of the test/reversal condition (i.e., within the first minute of the start of the session)
	Progressively initiates more components of the establishing operation if behavior is not evoked (e.g., signals the transition with position and words, removes engaging materials, presents work, escalates prompts, etc.)
	Reinforces the first instance of problem behavior immediately (even if it is not a target response) with all hypothesized reinforcers
	Allows access to reinforcers for 20–90 s
	Refrains from implementing any establishing operations for problem behavior during reinforcement intervals
	Progressively reimplements synthesized establishing operations following reinforcement intervals
	Provides salient transitions between reinforcement and establishing-operation intervals during test (e.g., body positioning, tone of voice, presentation of materials)

Appendix 2

Procedural Integrity Checklist for Skill-Based Treatment and Posttest Sessions

Reinforcement (i.e., child-led time)	
	Makes available many of the child's preferred items/activities
	Remains available to and engaged with the child (close in proximity, not distracted, and providing attention in the child's preferred manner)
	Honors all reasonable requests for items, attention, or doing things a particular way
	Programs "child-led" time for an appropriate amount of time (at least 20 s; should not feel unnaturally short or long)

	Denies and redirects to the items that are available if the child makes an unreasonable request
	Refrains from placing any demands, including instructions and questions
	Refrains from correcting the child (including providing feedback on past problem behavior) or the way they are engaging with an item/activity
	Refrains from manipulating the child's toys, unless following their lead
	Refrains from reacting in any obvious way to any inappropriate behavior; does not attempt to redirect the child following inappropriate behavior
Establishing operation (i.e., adult-led time)	
	Makes clear that the adult is directing activity by delivering an instruction as child-led time is terminated
	Delivers clear, concise instructions to the child
	Uses the three-step prompting method when delivering instruction: Tell them what to do (wait 3 s), show them what to do (wait 3 s), and help them do it.
	Only allows access to materials relevant to what the child is expected to do
	Only provides attention relevant to what the child is expected to do (prompting within the three-step method and praise for cooperation)
	Does not negotiate, argue, rationalize, or cajole
	Does not comply with the child's attempts to lead instruction (e.g., "I want to clean up before I sit at the table.")
	Does not react in any obvious way to any inappropriate behavior; proceeds with three-step prompting
	Does not change the demand contingent on problem behavior
Transition from child-led to adult-led time	
	Moves from adult-led time to child-led time only following one of these three skills: functional communication, delay/denial toleration, or compliance with the instruction/expectation following denial
	Reinforces functional communication and toleration responses *some* of the time (at least one time per session)
	Reinforces *very small chains* of cooperation following a denial at least once per session (e.g., three compliances or less; e.g., "Go take a seat.")
	Prompts communication or toleration skills if they are not occurring (i.e., if they are simply cooperating with all instructions); waits approximately 30 s between prompts
	Does not foreshadow which skills will be reinforced or how many demands will need to be completed prior to earning child-led time (i.e., keep it unpredictable)

Author Note Adithyan Rajaraman https://orcid.org/0000-0002-3671-8327 is now at the Department of Psychology, University of Maryland, Baltimore County. Kelsey W. Ruppel https://orcid.org/0000-0002-6800-8485 is now at FTF Behavioral Consulting, Worcester, MA, USA. Holly C. Gover is now at Ivymount School, Rockville, MD, USA. Robin K. Landa is now at the May Institute, Randolph, MA, USA.

This article was prepared in partial fulfillment of a PhD in behavior analysis by Adithyan Rajaraman.

We wish to thank Rachel H. Thompson, Eileen M. Roscoe, and Jason C. Bourret for their feedback on earlier versions of this article, and Jake Amatruda, Claudia Cwalina, Rachel Diverdi, and Betelehem Getahun for their assistance with data collection.

Declarations

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the institutional review board and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study does not contain any experimentation with animals performed by any authors.

Conflict of interest The authors declare that they have no conflicts of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Alter, P. J., Conroy, M. A., Mancil, G. R., & Haydon, T. (2008). A comparison of functional behavior assessment methodologies with young children: Descriptive methods and functional analysis. *Journal of Behavioral Education, 17*(2), 200–219. https://doi.org/10.1007/s10864-008-9064-3

Arndorfer, R. E., Miltenberger, R. G., Woster, S. H., Rortvedt, A. K., & Gaffaney, T. (1994). Home-based descriptive and experimental analysis of problem behaviors in children. *Topics in Early Childhood Special Education, 14*(1), 64–87. https://doi.org/10.1177/027112149401400108

Barton-Atwood, S. M., Wehby, J. H., Gunter, P. L., & Lanke, K. (2003). Functional behavior assessment rating scales: Intrarater reliability with students with emotional or behavioral disorders. *Behavioral Disorders, 28*(4), 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*(4), 385–389. https://doi.org/10.1007/s40617-018-0247-7

Borrero, C. S., & Borrero, J. C. (2008). Descriptive and experimental analyses of potential precursors to problem behavior. *Journal of Applied Behavior Analysis, 41*(1), 83–96. https://doi.org/10.1901/jaba.2008.41-83

Carr, E. G. (1977). The motivation of self-injurious behavior: A review of some hypotheses. *Psychological Bulletin, 84*(4), 800–816. https://doi.org/10.1037/0033-2909.84.4.800

Cooper, J. O., Heron, T. E., & Heward, W. L. (2019). *Applied behavior analysis* (3rd ed.). Pearson Education.

Durand, V. M., & Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. *Journal of Applied Behavior Analysis, 24*(2), 251–264. https://doi.org/10.1901/jaba.1991.24-251

Durand, V. M., & Carr, E. G. (1992). An analysis of maintenance following functional communication training. *Journal of Applied Behavior Analysis, 25*(4), 777–794. https://doi.org/10.1901/jaba.1992.25-777

English, C. L., & Anderson, C. M. (2006). Evaluation of the treatment utility of the analog functional analysis and the structured descriptive assessment. *Journal of Positive Behavior Interventions, 8*(4), 212–229. https://doi.org/10.1177/10983007060080040401

Evans, I. M., & Nelson, R. O. (1977). Assessment of child behavior problems. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment*. New York: Wiley.

Ferguson, J. L., Leaf, J. A., Cihon, J. H., Milne, C. M., Leaf, J. B., McEachin, J., & Leaf, R. (2020). Practical functional assessment: A case study replication and extension with a child diagnosed with autism spectrum disorder. *Education and Treatment of Children, 43*, 171–185. https://doi.org/10.1007/s43494-020-00015-1

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

Ghaemmaghami, M., Hanley, G. P., & Jessel, J. (2016a). Contingencies promote delay tolerance. *Journal of Applied Behavior Analysis, 49*(3), 548–575. https://doi.org/10.1002/jaba.333

Ghaemmaghami, M., Hanley, G. P., Jin, S. C., & Vanselow, N. R. (2016b). Affirming control by multiple reinforcers via progressive treatment analysis. *Behavioral Interventions, 31*(1), 70–86. https://doi.org/10.1002/bin.1425

Greer, B. D., Mitteer, D. R., Briggs, A. M., Fisher, W. W., & Sodawasser, A. J. (2020). Comparisons of standardized and interview-informed synthesized reinforcement contingencies relative to functional analysis. *Journal of Applied Behavior Analysis, 53*(1), 82–101. https://doi.org/10.1002/jaba.601

Gresham, F. M., & Lambros, K. M. (1998). Behavioral and functional assessment. In T. S. Watson & F. M. Gresham (Eds.), *Handbook of child behavior therapy* (pp. 3–22). Springer. https://doi.org/10.1007/978-1-4615-5323-6_1

Gresham, F. M., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review, 30*, 156–172. https://doi.org/10.1080/02796015.2001.12086106

Hagopian, L. P. (2020). The consecutive controlled case series: Design, data-analytics, and reporting methods supporting the study of generality. *Journal of Applied Behavior Analysis, 53*(2), 596–619. https://doi.org/10.1002/jaba.691

Hagopian, L. P., Fisher, W. W., Thompson, R. H., Owen-DeSchryver, 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*(2), 313–326. https://doi.org/10.1901/jaba.1997.30-313

Hagopian, L. P., Dozier, C. L., Rooker, G. W., & Jones, B. A. (2013a). Assessment and treatment of severe problem behavior. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 2. Translating principles into practice* (pp. 353–386). American Psychological Association. https://doi.org/10.1037/13938-014

Hagopian, L. P., Rooker, G. W., Jessel, J., & DeLeon, I. G. (2013b). Initial functional analysis outcomes and modifications in pursuit of differentiation: A summary of 176 inpatient cases. *Journal of Applied Behavior Analysis, 46*(1), 88–100. https://doi.org/10.1002/jaba.25

Hanley, G. P. (2010). Prevention and treatment of severe problem behavior. In E. Mayville & J. Mulick (Eds.), *Behavioral foundations of effective autism treatment* (pp. 233–256). Sloan.

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*(2), 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*(1), 16–36. https://doi.org/10.1002/jaba.106

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*(11), 963–974. https://doi.org/10.1037/0003-066X.42.11.963

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*(3), 202–207. https://doi.org/10.1080/17518423.2018.1437839

Hernandez, E., Hanley, G. P., Ingvarsson, E. T., & Tiger, J. H. (2007). A preliminary evaluation of the emergence of novel mand forms. *Journal of Applied Behavior Analysis, 40*(1), 137–156. https://doi.org/10.1901/jaba.2007.96-05

Herscovitch, B., Roscoe, E. M., Libby, M. E., Bourret, J. C., & Ahearn, W. H. (2009). A procedure for identifying precursors to problem behavior. *Journal of Applied Behavior Analysis, 42*(3), 697–702. https://doi.org/10.1901/jaba.2009.42-697

Iversen, I. H. (2013). Single-case research methods: An overview. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 2. Translating principles into practice* (pp. 3–32). American Psychological Association. https://doi.org/10.1037/13937-001

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., 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*(2), 197–209. https://doi.org/10.1901/jaba.1994.27-197

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*(1), 271–284. https://doi.org/10.1002/jaba.31

Jessel, J., Hanley, G. P., & Ghaemmaghami, M. (2016). Interview-informed synthesized contingency analyses: Thirty replications and reanalysis. *Journal of Applied Behavior Analysis, 49*(3), 576–595. https://doi.org/10.1002/jaba.316

Jessel, J., Ingvarsson, E. T., Metras, R., Kirk, H., & Whipple, R. (2018a). 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*(1), 130–157. https://doi.org/10.1002/jaba.436

Jessel, J., Ingvarsson, E. T., Metras, R., Whipple, R., Kirk, H., & Solsbery, L. (2018b). Treatment of elopement following a latency-based interview-informed, synthesized contingency analysis. *Behavioral Interventions, 33*(1), 271–283. https://doi.org/10.1002/bin.1525

Jessel, J., Hanley, G. P., Ghaemmaghami, M., & Metras, R. (2019). An evaluation of the single-session interview-informed synthesized contingency analysis. *Behavioral Interventions, 34*(1), 62–78. https://doi.org/10.1002/bin.1650

Jessel, J., Metras, R., Hanley, G. P., Jessel, C., & Ingvarsson, E. T. (2020). Evaluating the boundaries of analytic efficiency and control: A consecutive controlled case series of 26 functional analyses. *Journal of Applied Behavior Analysis, 53*(1), 25–43. https://doi.org/10.1002/jaba.544

Johnston, S. S., & O'Neill, R. E. (2001). Searching for effectiveness and efficiency in conducting functional assessments: A review and proposed process for teachers and other practitioners. *Focus on Autism and Other Developmental Disabilities, 16*(4), 205–214. https://doi.org/10.1177/108835760101600402

Johnston, J. M., Pennypacker, H. S., & Green, G. (2019). *Strategies and tactics of behavioral research and practice* (4th ed.). Routledge. https://doi.org/10.4324/9781315537085

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

Kinch, C., Lewis-Palmer, T., Hagan-Burke, S., & Sugai, G. (2001). A comparison of teacher and student functional behavior assessment interview information from low-risk and high-risk classrooms. *Education & Treatment of Children, 24*(4), 480–494 http://www.jstor.org/stable/42900504

Nelson, R. O., & Hayes, S. C. (1979). The nature of behavioral assessment: A commentary. *Journal of Applied Behavior Analysis, 12*(4), 491–500. https://doi.org/10.1901/jaba.1979.12-491

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*(3), 168–181. https://doi.org/10.1177/10634266040120030401

Newton, J. T., & Sturmey, P. (1991). The Motivation Assessment Scale: Inter-rater reliability and internal consistency in a British sample. *Journal of Intellectual Disability Research, 35*(5), 472–474. https://doi.org/10.1111/j.1365-2788.1991.tb00429.x

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

Rajaraman, A., Hanley, G. P., Gover, H. C., Staubitz, J. L., Staubitz, J. E., Simcoe, K. M., & Metras, R. (2021). Minimizing escalation by treating dangerous problem behavior within an enhanced choice model. *Behavior Analysis in Practice. Advance online publication.* https://doi.org/10.1007/s40617-020-00548-2

Reed, H., Thomas, E., Sprague, J. R., & Horner, R. H. (1997). The student guided functional assessment interview: An analysis of student and teacher agreement. *Journal of Behavioral Education, 7*, 33–49. https://doi.org/10.1023/A:1022837319739

Rose, J. C., & Beaulieu, L. (2019). Assessing the generality and durability of interview-informed functional analyses and treatment. *Journal of Applied Behavior Analysis, 52*(1), 271–285. https://doi.org/10.1002/jaba.504

Saini, V., Ubdegrove, K., Biran, S., & Duncan, R. (2019). A preliminary evaluation of interrater reliability and concurrent validity of open-ended indirect assessment. *Behavior Analysis in Practice, 13*, 114–125. https://doi.org/10.1007/s40617-019-00364-3

Santiago, J. L., Hanley, G. P., Moore, K., & Jin, C. S. (2016). The generality of interview-informed functional analyses: Systematic replications in school and home. *Journal of Autism and Developmental Disorders, 46*(3), 797–811. https://doi.org/10.1007/s10803-015-2617-0

Schmidt, J. D., Kranak, M. P., Goetzel, A. L., Kaur, J., & Rooker, G. W. (2020). A clinical demonstration of correlational and experimental analyses of precursor behavior. *Behavior Analysis in Practice*. Advance online publication. https://doi.org/10.1007/s40617-020-00452-9

Shapiro, E. S., & Kratochwill, T. R. (Eds.). (2000). *Behavioral assessment in schools: Theory, research, and clinical foundations*. Guilford Press.

Shriver, M. D., Anderson, C. M., & Proctor, B. (2001). Evaluating the validity of functional behavior assessment. *School Psychology Review, 30*, 180–192. https://doi.org/10.1080/02796015.2001.12086108

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*(5), 333–342. https://doi.org/10.1016/0891-4222(94)90020-5

Silva, F. (1993). Treatment utility: A reappraisal. *European Journal of Psychological Assessment, 9*(3), 222–226.

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

Smith, R. G., & Churchill, R. M. (2002). Identification of environmental determinants of behavior disorders through functional analysis of precursor behaviors. *Journal of Applied Behavior Analysis, 35*(2), 125–136. https://doi.org/10.1901/jaba.2002.35-125

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349–367. https://doi.org/10.1901/jaba.1977.10-349

Strand, R. C., & Eldevik, S. (2018). Improvements in problem behavior in a child with autism spectrum diagnosis through synthesized analysis and treatment: A replication in an EIBI home program. *Behavioral Interventions, 33*(1), 102–111. https://doi.org/10.1002/bin.1505

Sturmey, P. (1994). Assessing the functions of aberrant behaviors: A review of psychometric instruments. *Journal of Autism and Developmental Disorders, 24*, 293–304. https://doi.org/10.1007/BF02172228

Taylor, S. A., Phillips, K. J., & Gertzog, M. G. (2018). Use of synthesized analysis and informed treatment to promote school reintegration. *Behavioral Interventions, 33*, 364–379. https://doi.org/10.1002/bin.1640

Tiger, J. H., & Effertz, H. M. (2021). On the validity of data produced by isolated and synthesized contingencies during the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 54*(3), 853–876. https://doi.org/10.1002/jaba.792

Warner, C. A., Hanley, G. P., Landa, R. K., Ruppel, K. W., Rajaraman, A., Ghaemmaghami, M., Slaton, J. D., & Gover, H. C. (2020). Toward accurate inferences of response class membership. *Journal of Applied Behavior Analysis, 53*(1), 331–354. https://doi.org/10.1002/jaba.598

Whelan, C. J., Hanley, G. P., Landa, R., Sullivan, E., LaCroix, K., & Metras, R. (2021). Randomized controlled trial of seminarbased training on accurate and general implementation of practical functional assessments. *Journal of Applied Behavior Analysis, 54*(4), 1437–1455. https://doi.org/10.1002/jaba.845

Yarbrough, S. C., & Carr, E. G. (2000). Some relationships between informant assessment and functional analysis of problem behavior. *American Journal on Mental Retardation, 105*(2), 130–151. https://doi.org/10.1352/0895-8017(2000)105<0130:SRBIAA>2.0.CO;2

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*(4), 349–360. https://doi.org/10.1016/0891-4222(91)90031-M