Partially Automated Training for Implementing, Summarizing, and Interpreting Trial-Based Functional Analyses

Cassandra M. Standish¹ · Joseph M. Lambert¹ · Bailey A. Copeland¹ · Kathryn M. Bailey¹ · Ipshita Banerjee¹ · Mallory E. Lamers¹

Accepted: 7 September 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Trial-based functional analysis (TBFA) is an accurate and ecologically valid assessment of challenging behavior. Further, there is evidence to suggest that individuals with minimal exposure to behavior analytic assessment methodology (e.g., parents, teachers) can quickly be trained to conduct TBFA in naturalistic settings (e.g., schools, homes). Notwithstanding, the response effort associated with training development can be prohibitive and may preclude incorporation of TBFA into practice. To address this, we developed a partially automated training package, intended to increase the methodology's accessibility. Using a multiple-probe across skills design, we assessed the degree to which the package increased caregiver accuracy in (a) implementing TBFA, (b) interpreting TBFA outcomes, and (c) managing TBFA data. Six caregivers completed this study and all demonstrated proficiency following training, first during structured roleplays and again during assessment of their child's actual challenging behavior.

Keywords Behavioral skills training · Trial-based functional analysis · Parent training · Procedural fidelity · Implementation fidelity

We thank Amanda Sandstrom and Esther Kwan for their contributions to this study's execution. No outside funding was secured for this project.

✉ Joseph M. Lambert
joseph.m.lambert@vanderbilt.edu

¹ Department of Special Education, Vanderbilt University, PMB 230, Appleton Place, Nashville, TN 37203, USA

Introduction

Challenging behavior (e.g., aggression, self-injury, property destruction) is prevalent in individuals with intellectual and developmental disabilities (Bowring et al., 2019; McClintock et al., 2003) and is a substantial source of stress to caregivers and other invested stakeholders (e.g., Doubet & Ostrovosky, 2015). Although research has demonstrated that function-based interventions can be more effective than non-function-based interventions at addressing challenging behavior (e.g., Jeong & Copeland, 2020), the generality of such interventions can be limited unless steps to promote maintenance, and generalization are taken (e.g., Fuhrman et al., 2021; Lindberg et al., 2003). One way to achieve generality is by employing collaborative service-delivery models which ensure behavior supports are extended to home environments (e.g., Harvey et al., 2013), and which promote caregivers as primary implementers of relevant assessment and intervention procedures (e.g., Fettig & Barton, 2014; Gerow et al., 2018).

However, the success of such collaborations depends on their ability to consistently and successfully identify the function(s) of challenging behavior. Despite consensus on the value and importance of functional analysis (FA) to the assessment and treatment of challenging behavior (e.g., Hanley, 2012; Iwata & Dozier, 2008; Lloyd et al., 2021; Mace, 1994; Oliver et al., 2015; Roscoe et al., 2015), individuals who receive formal behavior interventions are unlikely to benefit from it because few practitioners qualified to conduct the analysis actually do so (Lloyd et al., 2021; Oliver et al., 2015; Roscoe et al., 2015). Reasons for lack of adoption vary but common concerns include ecological validity (e.g., Conroy et al., 1996) and resource constraints (Lloyd et al., 2021; Oliver et al., 2015). Additionally, the most commonly studied FA methodologies are session-based (Beavers et al., 2013; Hanley et al., 2003). That is, behavior-sampling procedures during FA often terminate based on the passage of time (e.g., 15 min), not performance. Thus, it is possible to see hundreds of instances of challenging behavior during a single analysis (Thomason-Sassi et al., 2011). This fact can raise concerns with risk of injury, property destruction, and/or general environmental disruption.

For decades, researchers have adapted FA methodology to improve its accessibility while maintaining its methodological integrity (Hanley, 2012; Iwata & Dozier, 2008). One such adaptation is the trial-based FA (Bloom et al., 2011; Rispoli et al., 2014; Sigafoos & Sagers, 1995). Trial-based FAs (TBFA) demonstrate functional relations through brief trials embedded into daily activities. Trials test hypotheses informed by interviews and observations and are typically divided into two segments (i.e., a control and a test).

Trial-based FAs have multiple practical advantages, including that trials: (a) are brief, (b) can be conducted when antecedents naturally occur, and (c) terminate following challenging behavior. Although simple, these adaptations address many concerns levied against session-based FAs. Specifically, the TBFA has increased social and ecological validity due to the shift from analog (i.e., contrived) settings to natural ones, as well as the shift from massed-trial (i.e., repeated re-presentation of relevant establishing operations in quick succession)

to distributed-trial (i.e., presentation of relevant establishing operations at the times, and in the locations, that they typically occur) formats.

A growing body of research demonstrates naturalistic implementers can be trained to conduct TBFA with high fidelity. In fact, TBFA have been implemented by professionals with little experience in systematic assessment and intervention procedures; including group-home staff (Lambert et al., 2013), special education teachers (Bloom et al., 2013; Kunnavatana et al., 2013a, 2013b; Lambert et al., 2012; Rispoli et al., 2015), and classroom staff (Kodak et al., 2013; Lloyd et al., 2015). Notwithstanding, training materials used in published research protocols are not widely available for general practitioner use. Thus, although practitioners know it is possible to train naturalistic implementers to fidelity, they may not be equipped to replicate. As a result, the potential impact of relevant research is diminished. To address this, Lambert et al. (2014) created a partially automated PowerPoint presentation (with voiceover narration and video models) to deliver the didactic and modeling portions of a TBFA training based on content which has facilitated favorable outcomes in previous studies (e.g., Bloom et al., 2013; Kunnavatana et al., 2013a, 2013b; Lambert et al., 2012, 2013, 2014, 2017; LeJeune et al., 2019).

Although Lambert et al. (2014) demonstrated that exposure to the narrated PowerPoint, alone, was sufficient to improve fidelity to TBFA procedures, mastery across participants was inconsistent. Their recommendation was that future research deliver the training's content using an empirically supported training framework appropriate for adult learners known as behavior skills training (BST; Miltenberger, 2012). Specifically, they recommended trainers not only expose trainees to PowerPoint content (i.e., instructions, video models), but to also provide them with opportunities to practice each condition of the analysis and receive corrective feedback until achieving empirically verifiable mastery criteria (e.g., Drifke, et al., 2017; Hahs & Jarynowski, 2019; Miltenberger, 2012; Parsons et al., 2012, 2013).

Since the publication of Lambert et al. (2014), ongoing complications associated with COVID-19 and the global pandemic, chronic nationwide shortages of Board Certified Behavior Analysts (BCBA; i.e., specialists fully qualified to address challenging behavior; Behavior Analyst Certification Board [BACB], 2020), and disproportionate availability of BCBAs in urban areas (e.g., Mello et al., 2016; Murphy & Ruble, 2012) have all forced practicing BCBAs to be innovative as they attempt to maximize their influence and impact. In this regard, an additional benefit presented by the Lambert et al. protocol is that it affords automation of critical but time-consuming aspects of training (e.g., the delivery of standardized content) which can be completed by caregivers in the absence of trainers, at convenient times and locations. In this same vein, when caregivers receive instruction about *why* they perform the tasks they perform, and how to interpret the results of their efforts, they can become more autonomous in their execution of TBFA. For example, if asked to complete a TBFA between weekly consultations with a BCBA, caregivers trained to interpret TBFA results can independently determine when a sufficient amount of data have been collected to confirm or negate hypothesized functions; thus expediting the assessment process and minimizing unnecessary exposure to countertherapeutic contingencies when a pre-determined number of trials are prescribed without consideration for emergent trends (e.g., Lambert et al., 2012). Such an approach

would also allow synchronous training minutes to be more efficiently committed to structured roleplays, performance feedback, problem solving, and other activities for which a BCBA presence is more clearly essential.

Thus, the purpose of this proof-of-concept project was twofold. First, to enhance caregiver autonomy during the TBFA process, we updated and expanded the content of the training presented in Lambert et al. (2014) to include instruction on TBFA execution, data analysis and function identification, and data management [content available in Standish et al., (2020)]. Then, using pre/post-mastery probes (i.e., structured role plays during which feedback was not delivered), and fidelity assessments during TBFAs of actual challenging behavior, we assessed the degree to which the updated training content, delivered through the BST framework, established mastery of desired targets in naturalistic implementers (i.e., caregivers of children with challenging behavior) with no formal training in behavior analysis. Through these efforts, we sought to address the following research question: to what extent will caregivers implement, interpret, and graph TBFA data to high fidelity following exposure to a partially automated training package with standardized content?

Method

Participants and Settings

We recruited six caregiver-child dyads to participate in this study (see Tables 1, 2). Five caregivers were parents, however, one (Iman) was the child's grandmother. While no caregivers reported experience delivering behavior analytic services prior to the study, educational backgrounds and income brackets varied across caregivers. Children were included if they were three years or older, had an intellectual or developmental disability, and engaged in challenging behavior. Any adult who participated in the home care of eligible children could volunteer for study participation. The first six caregiver-child dyads to meet these criteria and expressed interest in being involved in the study were recruited. Of the six participating caregivers, one (Jay) dropped out of the study prior to generalization for personal reasons.

During training, facilitators (i.e., graduate research assistants) observed caregivers consume automated training-module content and facilitated structured roleplays

Table 1 Caregiver information

Participant	Child	Age	Gender	Relationship	Race/ethnicity	Education	Income
Jay	Bob	34	Male	Father	Black	High school	< \$25,000
Iman	David	57	Female	Grandmother	White	college	\$35,000–49,999
Nora	Nick	33	Female	Mother	White	College	\$75,000–99,999
Kristin	Dax	30	Female	Mother	Black	College	NR
Goldie	Kurt	31	Female	Mother	Hispanic	High school	\$35,000–49,999
Tina	Amy	41	Female	Mother	Black	College	\$50,000–74,999

NR Not reported

Table 2 Child participant information

Participant	Age	Gender	Race	Diagnosis	Behavior	Communication
Bob	9	Male	Black	ASD	Agg, SIB	1- to 3-word utterances
David	5	Male	White	ASD; ADHD	Agg, PD, SIB	Full sentences
Nick	5	Male	White	ASD; ADHD	Agg, PD	Full sentences
Dax	5	Male	Other	ASD	Agg, PD	Full sentences
Kurt	6	Male	White	ASD	Agg, PD, SIB	1- to 3-word utterances
Amy	9	Female	Black	ASD	Agg, PD, SIB	Full sentences

ASD Autism spectrum disorder, ADHD attention deficit hyperactivity disorder, Agg aggression, PD property disruption, SIB self-injurious behavior

in common living areas when children were not present. Although modules could technically be completed without formal oversight (a noted advantage of the format), facilitators observed module completion in this study to ensure consistent exposure to content across participants and preserve the integrity of our analysis. During generalization, caregivers conducted TBFA's of their child's challenging behavior in the child's home setting, at times when challenging behavior typically occurred (e.g., during morning routines).

Experimental Design

Using a concurrent multiple-probe across skills design (Gast et al., 2018), we assessed the degree to which a BST framework, in which partially automated modules replaced in-vivo supports for the didactic instruction and modeling phases of training, could facilitate caregiver acquisition of competencies relevant to effective behavioral consultations which incorporate TBFA. Specifically, TBFA implementation (Tier 1), data interpretation (Tier 2), and data-management (Tier 3) skills were assessed for each caregiver in the study.

Response Measurement

The primary dependent variable was caregiver fidelity to established procedures for executing a series (i.e., attention, tangible, escape) of TBFA trials (Tier 1), interpreting TBFA graphs (Tier 2), and managing raw TBFA data (Tier 3).

Specifically, in Tier 1, trained observers assessed participant adherence to TBFA procedures using fidelity checklists whose content is displayed in Appendix A of Lambert et al. (2014). Observers scored a *yes* when a procedure outlined by the checklist was executed as intended, a *no* if it was not, and an *N/A* if there was no opportunity to display the skill. We calculated Tier 1 fidelity scores by dividing *yes* by the sum of *yes* and *no* for each series of trials (i.e., attention, tangible, escape) and multiplying by 100.

In Tier 2, we assessed participant adherence to visual analysis criteria (described below) using worksheets which contained three graphical displays of hypothetical (pre/post-mastery probes, self-evaluation probes) or actual (generalization probes)

TBFA outcome data. Graphed data depicted latencies to challenging behavior from trial-segment onset. The number of trials displayed in each graph of each worksheet ranged from 3 to 20 and either did or did not confirm a functional relation. Functional relations were established using ongoing visual inspection criteria appropriate for trial-based assessment formats; modeled after the work published by Roane et al. (2013) and Saini et al. (2018), and validated by Standish et al. (in press).

On each worksheet, participants assessed functional relations by circling *yes*, *no*, or *unknown* in a multiple-choice test displayed to the right of each graph. Responses were scored as *correct* when answers reflected accurate assessment and *incorrect* when they did not. Correct responses were then divided by the sum of correct and incorrect responses and multiplied by 100 to produce a single score for each assessment.

In Tier 3, we assessed participant adherence to a data-management procedure (described below) by evaluating permanent products on paper-and-pencil data sheets, and in an electronic data summary spreadsheet. On a point-by-point basis, we assessed (a) accurate identification of trials not conducted to fidelity (randomly interspersed with trials conducted to fidelity on paper-and-pencil data sheets), (b) accurate transfer of data from trials conducted to fidelity from paper-and-pencil data sheets to appropriate cells of an Excel® spreadsheet, and (c) accurate assessment of whether accumulated data justified condition termination (i.e., either because a function had been identified or because the maximum number of trials [i.e., 20] had been reached). We scored *correct* when permanent products corresponded with our master key, *incorrect* when they did not, and multiplied by 100 to produce a single score for each assessment.

Interobserver Agreement

We calculated point-by-point interobserver agreement (IOA) between trained and independent observers across 89.3% of all trials conducted, across all phases of this study. Specifically, we scored an *agreement* when observer assessments of caregiver performance corresponded and a *disagreement* when they did not. We then divided *agreements* by the sum of *agreements* and *disagreements* and multiplied by 100 to generate IOA scores for each assessment. All IOA scores fell above 95% (see Table 3).

Implementation Fidelity

Using checklists and across all phases of training, we evaluated facilitator (graduate student) adherence to training procedures. Specifically, we evaluated whether (a) correct instructions were provided, (b) correct materials were present, (c) technology worked as intended, (d) trainings were completed in their entirety, (e) caregiver responses to imbedded multiple-choice questions were reviewed by facilitators, (f) supporting materials (e.g., caregiver marks on raw data sheets) were reviewed, and (g) role playing procedures were implemented correctly.

Table 3 Interobserver agreement

Participant	Mean IOA (% of trials assessed)		
	Execution	Interpretation	Management
Jay	97.86 (71.43%)	100% (100%)	100% (100%)
Iman	98.36% (81.25%)	100% (100%)	100% (100%)
Nora	96.63% (80.00%)	100% (100%)	100% (100%)
Kristin	98.29% (57.89%)	100% (100%)	100% (100%)
Goldie	99.11% (62.50%)	100% (100%)	100% (100%)
Tina	98.89% (100%)	100% (100%)	100% (100%)

IOA Interobserver agreement

When facilitators executed a procedure as intended, the relevant checklist item was scored as *correct*. Otherwise, it was scored as *incorrect*. We calculated fidelity scores for each assessment by dividing the number of procedures implemented correctly by the sum of the procedures implemented correctly and incorrectly, and then multiplying by 100. Across all tiers and participants, implementation fidelity scores remained at 100%.

Procedural Fidelity

Likewise, during 100% of pre/post-mastery and self-evaluation probes, we used checklists to evaluate whether facilitators (a) correctly delivered relevant instructions, (b) ensured relevant materials were always present, (c) adhered to actor scripts during structured roleplays (Tier 1 only), and (d) delivered feedback according to study phase. Procedural fidelity was scored and calculated in the same manner as implementation fidelity. With the exception of Iman and Nora during Tier 1 (for whom mean fidelity scores were both 99.4%), procedural fidelity scores were 100% across all tiers and for all participants.

Procedures

Prior to initiating study procedures, we conducted an open-ended functional assessment interview (Hanley, 2012) with each caregiver to gather information about (a) response topographies of challenging behavior, (b) challenging behavior's potential controlling variables, and (c) preferred items. This information was used to design ecologically valid experimental trials during the generalization phase of this study. For example, we opted not to include ignore trials due to a lack of circumstantial support for automatic functions derived from interview results. Using confederates (Tier 1) and hypothetical data sets (Tiers 2 and 3), we then collected pre-mastery (i.e., baseline) probes.

Prior to training, we conducted pre-mastery probes of each relevant skill. Following training, we exposed participants to three conditions: post-mastery probes, self-evaluation, and generalization. Each time a participant demonstrated mastery of content targeted by a training (described below), we conducted additional

pre-mastery probes for untrained skills. The purpose of post-mastery probes was to assess the degree to which skill mastery persisted across time in the context in which it was established (i.e., structured roleplays). The purpose of self-evaluation was to assess the degree to which participants could discriminate their own errors (if they made them), successfully self-correct them (by reconducting trials identified as having errors), and only analyse data from trials implemented to fidelity. This was done because naturalistic implementers must often assess the validity of the trials which they conduct without experimenter oversight (cf. Bloom et al., 2013; LeJeune et al., 2019). Thus, we wanted to assess the degree to which their judgments aligned with our own before asking them to conduct TBFA of actual challenging behavior. During generalization, we assessed the degree to which mastered skills generalized to the actual assessment and treatment of challenging behavior.

Pre-Mastery Probes

On the day of each assessment, facilitators provided caregivers with all necessary materials (e.g., data sheets, preferred items) and instructed them to demonstrate each relevant skill (e.g., “Run a tangible trial”) during structured roleplays (Tier 1) or structured practice opportunities (Tiers 2, 3). During Tier 1, facilitators instructed caregivers to tell them when to begin and end each assessment (e.g., “tell me when to start our roleplay and then tell me when you think you have conducted a complete trial and the roleplay should stop”).

During each probe facilitators followed scripts, did not deliver corrective feedback, did not respond to questions, and did not indicate when an assessment should be terminated. To avoid frustration and maintain caregiver buy-in for study procedures, caregivers were given the option to “pass” if they did not think they could demonstrate a skill. If participants passed, we immediately moved to the next probe, and all missed opportunities were scored as incorrect.

During Tier 1 mastery probes, a confederate provided participants with structured roleplay opportunities in clusters of three (i.e., attention, tangible, escape). At the onset of each opportunity, the facilitator delivered a brief instruction (e.g., “when you’re ready, let’s run a tangible trial”). When caregivers indicated they were ready, confederates followed one of multiple scripts drafted for each TBFA condition.

In each script, up to five distinct confederate behaviors were possible: (a) target challenging behavior (i.e., self-injurious behavior), (b) non-target challenging behavior (i.e., disruption), (c) appropriate communication, (d) compliance (escape condition only), and (e) appropriate play. In all scripts, distractor behaviors were each emitted once before the first instance of target challenging behavior. To maintain unpredictability, scripts randomized the order in which distractor behaviors (i.e., disruption, communication, compliance, and play) occurred. To avoid inadvertently cueing participants about mistakes, scripts extended well beyond what would be needed to execute a trial to fidelity. That is, challenging behavior was scripted to occur before two min passed during each trial-segment. However, if caregivers did not react to challenging behavior, confederates continued to act either until the caregiver told them the roleplay was finished, or until 4 min passed.

During Tier 2, facilitators presented participants with a pen (or pencil) and one graph-interpretation worksheet at a time. Each worksheet (consisting of 3 distinct line graphs which plotted response latencies from trial-segment onset; see Standish et al., 2021) represented one assessment opportunity (i.e., trial). Each worksheet (generated by research assistants) contained at least one graph displaying a functional relation between challenging behavior and the independent variable and at least one graph in which available data did not support a functional relation. The specific graphs displayed on each worksheet varied from trial to trial, with the positions of graphs with and without functional relations randomized across trials (i.e., top, middle, bottom). To identify functional relations, we used data-interpretation criteria validated by Standish et al. (in press) which stipulates functions are confirmed when (a) at least three valid demonstrations of effect (i.e., challenging behavior is observed in a test segment and not its contiguously conducted control) are present and (b) at least 50% of all trials conducted contain valid demonstrations of effect. To prevent indefinite execution of ineffective conditions, we imposed a 20-trial ceiling (i.e., the largest number of trials conducted in published TBFA literature).

During Tier 3, facilitators presented participants with a laptop computer, an electronic data summary file (i.e., an Excel spreadsheet), and a raw data sheet containing 31 lines of hypothetical data. Each raw data sheet contained data from 20 correctly executed trials and 11 incorrectly executed trials. The content and position of correct and incorrect trials were randomized across data sheets. Caregivers were required to (a) mark out any and all trials not conducted to fidelity, (b) enter only data (e.g., trial number, latency to challenging behavior, etc.) for trials conducted to fidelity, and (c) determine when to terminate the TBFA based on the data that they had entered thus far. As in Tier 2, all graphs displayed latency to problem behavior in the test and control segment, and thus were presented as line graphs. All interactions with, and data transfers from, one raw data sheet represented a single assessment opportunity.

Training Content and Format

Training materials are now published on a website (i.e., Standish et al., 2020) and included a training summary sheet (i.e., concise rules and reminders of training content), a workbook for participants to document their responses to questions posed during partially automated PowerPoint presentations, scripts for structured roleplays (Tier 1), worksheets with graphs of hypothetical data (Tier 2), raw data sheets with hypothetical data and a pre-formatted Excel file for managing and graphing (Tier 3), and a laptop computer with partially automated PowerPoint presentations (all tiers).

All training procedures conformed to the BST framework during which instruction and modeling was delivered through partially automated PowerPoint presentations. These presentations included voiceover narrations which covered (a) background information pertinent to relevant concepts, (b) definitions and procedures, (c) examples and near non-examples (e.g., delivery of a demand in an attention control segment), and (d) opportunities to respond (i.e., multiple-choice quiz questions). The PowerPoints were fully automated except for slides in which opportunities to respond were presented. On these slides, the training stopped indefinitely and only re-initiated after participants made a valid response to the question posed by

clicking their mouse in any of a few pre-determined areas (functionally advancing the presentation to its next slide and re-initiating the automated audio files and their associated timers). Opportunities to respond covered various aspects of training content, including appropriate control- and test-segment behavior, data-management procedures, and data-interpretation procedures.

Upon completion of each PowerPoint presentation, roleplays (Tier 1) and practice opportunities (Tier 2 and Tier 3) were delivered by facilitators until caregivers demonstrated mastery of each of the targeted skills. Performance feedback was delivered after every trial. On average, trainings lasted 25.5 min (range: 20–32 min); not including participant-response times. Between one and four trainings were conducted in a single home visit. Thus, including time for roleplaying and study-specific procedures (e.g., pre-mastery probes of performance for untrained tiers), home visits could have been as short as 30 min or as long as 3 h.

Tier 1 training (i.e., TBFA implementation) was divided into four phases and included an introduction to TBFA methodology and instruction on each of three tests for social functions of challenging behavior (i.e., attention, tangible, escape). The TBFA procedures highlighted in these trainings were based on recommendations made by Bloom et al. (2011) and entailed two-segment experimental trials (i.e., control, test) for which transitions occurred either based on the occurrence of challenging behavior, or the passage of time (i.e., 2 min). Tier 1 roleplays were considered mastered once the participant roleplayed the scenario (i.e., attention, tangible, escape) to 100% fidelity.

Tier 2 content focused on interpretation of TBFA data and provided opportunities for caregivers to apply interpretative rules to a variety of hypothetical data sets of varying complexity. Tier 2 mastery entailed perfect correspondence in data interpretation between caregivers and researchers across two consecutive assessment opportunities.

Tier 3 content was designed to teach participants to: (a) manage data for invalid trials (i.e., trials not conducted to fidelity), (b) transfer data from valid trials from paper-and-pencil data sheets to appropriate cells of a pre-formatted Microsoft Excel® spreadsheet that automatically graphed data as they were entered, and (c) indicate on the spreadsheet when each TBFA condition should have been terminated, according to available data. To facilitate data interpretation, the spreadsheet also visually distinguished components of a valid demonstration of effect by coloring relevant cells either red (no demonstration of effect) or green (valid demonstrations of effect; see Fig. 1).

Post-Mastery Probes

Post-mastery probes were identical to pre-mastery probes.

Self-Evaluation Probes

Self-evaluation probes were only conducted for Tier 1 performance (i.e., TBFA execution) and were similar to Tier 1 pre/post-mastery probes except that at the end of each probe, caregivers were instructed to indicate whether they thought they had

Fig. 1 Excel file with hypothetical data

executed the skill correctly or incorrectly. When they identified an incorrectly conducted trial (regardless of the actual fidelity of the trial), they were given another opportunity to conduct the trial. This process continued, with no feedback from the facilitator, until the caregiver reported conducting a perfect trial. At that point, if the facilitator agreed, praise was delivered and the next probe initiated. If the facilitator disagreed, they discussed the noted error and gave the caregiver another opportunity to conduct the trial. This continued until consensus between facilitator and caregiver was achieved.

Generalization

For Tier 1, caregivers used the specific routines, antecedents, and consequences identified during initial interviews to complete 20 trials of each social condition (i.e., attention, tangible, escape) of a TBFA of their child's actual challenging behavior. Once more, TBFA procedures followed the recommendations of Bloom et al. (2011). Specifically, each trial began with a control segment in which programmed consequences were abolished (e.g., continuous attention during an attention control segment), followed by a test segment in which programmed consequences were established (e.g., denied access to attention during an attention test segment). Transitions from control to test occurred following the first instance of challenging behavior, or after two min elapsed (whichever came first). During test segments, programmed reinforcers (e.g., statements of concern during an attention trial) followed challenging behavior and trials terminated either after reinforcer delivery, or after 2-min elapsed (which came first). To assess generalization of Tier 1 skills, facilitators evaluated caregiver fidelity to programmed TBFA procedures for three series of social conditions (i.e., attention, tangible, escape). As in the self-evaluation phase, caregivers self-reported their own fidelity of implementation and re-conducted any trials in which they stated they incorrectly implement the trial.

To assess generalization of Tier 2 skills, facilitators graphed caregiver-obtained child-outcome data and presented it to caregivers using the same worksheet format

described above (i.e., a graduate research assistant transferred obtained data to three graphs [attention, tangible, escape] on a single, vertically displayed, worksheet with spaces to identify functional relations). Because each data sheet consisted of three graphs, and there were only three conditions evaluated, there was only one generalization data point for Tier 2.

To assess generalization of Tier 3 skills, caregivers were asked to transfer raw data from their own paper-and-pencil data sheets to the electronic data summary file and indicate when they *would have* ended the TBFA condition, if they were following rules established in the training. To generate comparison data for this project's companion piece (i.e., Standish et al., in press), participants were asked to conduct 20 trials regardless of emergent child-behavior patterns.

Safety Precautions

Because antecedents and consequences to challenging behavior were individualized for each participant and were designed to mirror commonly occurring events, we did not anticipate caregivers would interact with intensities or topographies that they were not already accustomed to seeing. Notwithstanding, we walked each through a cost-benefit analysis prior to study onset and obtained written acknowledgment of associated risks during the consenting process. Further, we instructed each to immediately abandon research protocol and to engage in business-as-usual de-escalation procedures if they ever encountered an intensity or topography of challenging behavior that they were uncomfortable seeing, or that they felt posed a risk to them, their child, or their property. Following TBFA completion, we added details to this guidance, indicating that abolishing operations of confirmed functions (i.e., caregiver performance during control segments of relevant TBFA trials) were likely to enhance de-escalation efforts. Importantly, no caregiver ever reported abandoning protocol or needing to attempt de-escalation.

Construct and Social Validity

To assess the degree to which demonstrations of training content mastery corresponded to accurate and valid assessments of actual challenging behavior (i.e., construct-validity; Yoder et al., 2018), we evaluated whether TBFA's conducted during generalization (a) produced interpretable child-outcome data, and (b) informed the design of effective function-based interventions (executed by caregivers after completing similarly formatted, partially automated, intervention-training modules [i.e., Bailey et al., 2020a, 2020b]). Additionally, we asked each caregiver to complete a social validity questionnaire adapted from the Treatment Evaluation Inventory Short Form (Kelly et al., 1989). We did this by either delivering the questionnaire in-person at the end of each participant's full consultation (i.e., after study completion, caregivers were subsequently trained to implement and evaluate the effect of trial-based intervention packages which included functional communication training and discrimination training; see Standish et al., in press). Questions on this assessment did not exclusively address the assessment process (i.e., they also assessed the

intervention process). Due to miscommunication, facilitators did not deliver questionnaires in-person to four caregivers. Rather, we mailed them to relevant caregivers after the consultation had already ended. No caregiver replied to surveys delivered in this way.

Results

Training outcomes are displayed graphically in Fig. 2 (Jay & Nora), Fig. 3 (Kristin & Tina), and Fig. 4 (Iman & Goldie). During Tier 1 pre-mastery probes, caregiver fidelity remained at low and stable levels. One participant (i.e., Goldie) displayed a small increasing trend prior to training. However, the overall level of her fidelity remained low. Likewise, all participants interpreted TBFA graphs and managed TBFA data (i.e., Tiers 2, 3) at moderate to low levels. Following training, all six caregivers demonstrated mastery of all targeted skills. Additionally, there were no increases in fidelity scores for lower-level tiers when upper-tiers were mastered. That is, when mastery of Tier 1 was demonstrated, there were no noticeable changes in the levels of fidelity for Tier 2 and Tier 3, and once Tier 2 was mastered, there was no noticeable changes in the levels of fidelity for Tier 3; thus confirming functional independence of skillsets targeted across tiers.

During Tier 1 self-evaluation and generalization probes, we did not analyse data from trials caregivers independently indicated were not conducted with fidelity, and

Fig. 2 Caregiver accuracy outcomes for Jay and Nora. Trials denoted by a “*” depict when participants correctly self-identified errors and re-implemented trials before reporting data. Trials denoted by semi-closed circle represent instances in which participants did not recognize an error they made and reported data from a flawed trial. *Pre* Pre-mastery probes, *Post* post-mastery probes, *Self-Eval* self-evaluation trials, *GN* generalization trials

Fig. 3 Caregiver accuracy outcomes for Kristin and Tina. Trials denoted by a “*” depict when participants correctly self-identified errors and re-implemented trials before reporting data. Trials denoted by semi-closed circle represent instances in which participants did not recognize an error they made and reported data from a flawed trial. *Pre* Pre-mastery probes, *Post* post-mastery probes, *Self-Eval* self-evaluation trials, *GN* generalization trials

Fig. 4 Caregiver accuracy outcomes for Iman and Goldie. Trials denoted by a “*” depict when participants correctly self-identified errors and re-implemented trials before reporting data. Trials denoted by semi-closed circle represent instances in which participants did not recognize an error they made and reported data from a flawed trial. *Pre* Pre-mastery probes, *Post* post-mastery probes, *Self-Eval* self-evaluation trials, *GN* generalization trials

only included data from the first trial caregivers reported conducting with 100% fidelity (even if we disagreed). We did this to reflect the fact that, in the absence of our direct support, caregivers would have reported these data to collaborating BCBAs (regardless of whether they had actually been conducted with fidelity). In these conditions, points denoted with an “*” indicate that a caregiver had identified an error during a previous attempt and had independently asked for a new opportunity to respond. Thus, values from these data points represent data from the caregiver’s final attempt, when they indicated they had accurately executed the targeted skill. Data points denoted by half circles represent trials in which the caregivers incorrectly identified a trial as being conducted to fidelity. Table 4 provides a brief summary of caregiver proficiency with identifying failed trials.

With the exception of *data management* for Kristin, Goldie, and Tina, mastery across tiers maintained during self-evaluation and generalization probes. For Tina, fidelity to data-management procedures fell during generalization (i.e., when she conducted a TBFA of her own child’s challenging behavior) and she required retraining. Retraining involved a series of structured practice opportunities with feedback until re-establishing 100% accuracy. Following retraining, this skill again improved to acceptable ranges.

Construct-validity data (i.e., child-outcome data, see Yoder et al., 2018) are briefly summarized in Table 5, and are fully displayed in this project’s companion piece, Standish et al. (in press). All participants who conducted a TBFA of actual challenging behavior during the generalization condition produced interpretable outcomes and either ruled out or confirmed functional relations for all conditions assessed. Further, across 100% of cases, a function-informed differential reinforcement procedure (e.g., trial-based functional communication training), which was designed based on the findings of each caregiver-conducted TBFA, successfully eliminated challenging behavior during four (or more) consecutive intervention trials, distributed to occur at the same times and in the same locations during which TBFA trials had previously established functional relations between relevant establishing operations and challenging behavior. Specifically, the mastery criteria for trial-based functional communication training entailed at least three consecutive trials, and at least 50% of total trials, with no challenging behavior and evidence of

Table 4 Independent and correct self-identification of trials conducted with errors

Participant	Total # incorrectly implemented trials	Total self-identified incorrect	Percentage self-identified incorrect (%)
Jay	4	3	75.0
Iman	25	19	76.0
Nora	9	9	100.0
Kristin	8	8	100.0
Goldie	32	29	90.6
Tina	18	17	94.4
Average	16.0	14.2	88.8

Table 5 Construct-validity outcomes

Participant	Function(s) identified	100% reduction during Tx?
Jay	N/A	N/A
Iman	Tan	Y
Nora	Esc	Y
Kristin	Attn, Tan	Y
Goldie	Attn, Tan	Y
Tina	Esc	Y

Tan Tangible, *Esc* escape, *Attn* attention, *Y* yes, *Tx* treatment, *N/A* not applicable (Jay dropped out of the study prior to completing all required TBFA trials)

independent manding. Social validity surveys were returned by 33% (two of six) of participants and suggest the training, assessment, and intervention procedures, and outcomes were acceptable (see Table 6).

Discussion

In this study, we sought to increase the accessibility of TBFA methodology by automating portions (i.e., instructions, modeling) of a historically effective approach for training naturalistic implementers, with limited behavior analytic training, to accurately conduct TBFA. Across 100% (18 of 18) of opportunities, our partially automated, BST-based, training package accomplished targeted objectives; suggesting that automation did not detract from the accomplishment of desired outcomes. This is meaningful because it presents a convenient (and potentially efficient) alternative to traditional BST approaches in which in-vivo trainers deliver all aspects of

Table 6 Social validity outcomes for TBFA and subsequent intervention procedures following training

Question 0 = strongly disagree; 4 = strongly agree	Mean (N = 2)
The TBFA is an acceptable way of assessing the child's problem behavior	4
I believe that it is acceptable to use this assessment without my child's consent	4
The TBFA provided valid results	4
The results of the TBFA led to an effective intervention for my child's problem behavior	4
I found the TBFA trainings helpful	4
I am glad I implemented the TBFA with my child	4
I would recommend other parents of children who engage in problem behavior complete these trainings	4
I found the intervention to be appropriate for treating my child's problem behavior	4
I found that my child can use the skills taught to him/her through the intervention	4
Overall, I have a positive reaction to participating in this research study	4

training content. It also expands the ways in which naturalistic implementers can be effectively incorporated into the assessment and treatment of challenging behavior (see also Benson et al., 2018; Suess et al., 2014; Vismara et al., 2018; Wacker et al., 2013). Because our materials are freely available (Bailey et al., 2020a, 2020b; Standish et al., 2020), it is our hope that this project increases the overall accessibility of TBFA methodology.

Limitations

A few of this study's limitations should be noted. First, due to the nature of our research design (i.e., multiple-probe across skills), assessment across tiers was cumbersome. Specifically, it took on average 31.2 (range, 16–44) days to progress from the onset of baseline to the final self-evaluation probe. This process considerably delayed practical action (i.e., TBFA of actual challenging behavior and the onset of function-based treatment) and may have affected caregiver retention. Related, it took an average of 11.8 days (range, 5–16) for caregivers to complete TBFA of challenging behavior during generalization. Although this is not a particularly prohibitive timeline for a TBFA, it is likely the assessment could have been completed much quicker if caregivers had been able to schedule trials without consideration for research-personnel availability (who needed to be present to collect fidelity data for three series of trials). Future researchers might explore more efficient and practitioner-friendly strategies for establishing experimental control.

A second limitation of this study is that fidelity to data-management procedures did not always maintain, once established (i.e., Tina). This suggests our data-management process is likely too complex for some caregivers. As a result, practitioners might consider drafting simpler procedures. Alternatively, practitioners might provide additional supports (e.g., periodic retraining), or complete all data-management tasks themselves.

Third, our generalization data should be interpreted with caution because we did not collect baseline data in generalization settings (i.e., during assessments of actual challenging behavior), and we only collected a single generalization data point for Tiers 2 and 3 (because only one data point per tier was available after caregivers conducted a full TBFA of their child's challenging behavior).

Fourth, caregivers were given the opportunity to pass on trials that they either did not know how to complete or did not feel comfortable attempting. Whereas this decision was intended to reduce the potentially degrading impact of repeatedly requiring someone to complete a task they do not know how to do, it limits the confidence with which we can say that baseline fidelity scores reflected true caregiver proficiency (e.g., it is possible that caregivers would have correctly executed the parts of trials they opted not to conduct, had they been required to). Although caregivers had the opportunity to pass on trials during all phases of the study, we did not collect data on the specific trials for which they opted to do so. Thus, we cannot assess whether this option was differentially employed during any given study phase.

A fifth limitation is that only two of six participants completed the social validity questionnaire. While those results were favorable, it is concerning that so few

participants returned the questionnaire. Related, although his reasons were personal, Jay's attrition also challenges the social validity of our project.

A sixth limitation is that there was some overlap in the skillsets required to master Tier 2 and Tier 3 probes (i.e., in both cases, participants needed to know rules for TBFA data-interpretation stipulated by Standish et al., *in press*). Notwithstanding, consistently poor performance during Tier 3 pre-mastery probes across participants demonstrates that the specific applications of these rules targeted during Tiers 2 and 3 remained functionally independent. Regardless, more compelling demonstrations across tiers would have included mastery of non-overlapping content. Related, because training content differed across tiers, our study only establishes experimental control over the impact of the framework (i.e., BST in which instruction and video models are delivered via partially automated modules) on skill acquisition. Content validation efforts require different methodology (cf. Yoder et al., 2018) and should serve as the focus of other research initiatives.

A final limitation may be associated with the population targeted for this study. Historically, naturalistic implementers have only needed to execute valid TBFA trials (e.g., Bloom et al., 2013; Lambert et al., 2012, 2017; LeJeune et al., 2019). In those studies, data interpretation and management were tasks relegated to experts in behavior analysis. Caregivers may neither have the time nor the inclination to engage in data interpretation and management tasks. However, our findings demonstrate that it is at least possible to train them to do so. In a climate in which it is difficult for some to find in-person behavior analytic services due to BCBA shortages (BACB, 2020), geographic barriers (e.g., Mello et al., 2016; Murphy & Ruble, 2012), and/or pandemic-related restrictions, the roles and responsibilities of caregivers and other naturalistic implementers (e.g., teachers, paraeducators) will likely continue to increase. Notwithstanding, future researchers might extend these findings to populations who interact with children with challenging behavior and who might more reasonably be expected to engage in tasks associated with data management and interpretation (e.g., Board Certified Behavior Analysts, Board Certified Assistant Behavior Analysts, Registered Behavior Technicians).

Future Directions

Importantly, in our study, we demonstrated that trainees who mastered our content could execute TBFA of actual challenging behavior and, combined, with the results of Standish et al. (*in press*) those outcomes could inform effective intervention. Although a number of previous researchers have demonstrated it is possible to train participants to execute operationalized TBFA protocols (cf. Alnarmay et al., 2017; Flynn & Lo, 2016; Kunnavatana et al., 2013a; Kunnavatana et al., 2013b; Lambert et al., 2013; Lambert et al., 2014; Rispoli et al., 2015; Rispoli et al., 2016; Vasquez et al., 2017), only a small subset validated their training content with interpretable outcomes from TBFA of actual challenging behavior (e.g., Flynn & Lo, 2016; Kunnavatana et al., 2013a, 2013b; Rispoli et al., 2015, 2016) or with effective interventions (e.g., Flynn & Lo). Thus, our findings replicate and extend this work by

making available to practitioners the materials they would need to replicate our findings (see Bailey et al., 2020a, 2020b; Standish et al., 2020).

To increase efficiency, future researchers might conduct component analyses of training content and evaluate which portions were critical to valued outcomes. They might also assess the degree to which outcomes can be replicated when trainings are incorporated into telehealth-based service-delivery models. As there is high a demand for behavior analytic services (cf. BACB, 2020; DiGennaro Reed & Henley, 2015), it could be prudent to explore strategies which empower naturalistic implementers to independently execute more of the assessment and treatment process without direct (i.e., in-person) oversight.

In that vein, it is important to note that previous research has demonstrated that automated TBFA trainings which do not include practice opportunities with corrective feedback are unlikely to consistently yield desired outcomes (e.g., Lambert et al., 2014). Thus, attempts to deliver all aspects of training from a distance should explore valid ways through which roleplaying can be accomplished virtually. Alternatively, researchers might leverage recent advances in technology to design versions of this training which fully automate roleplaying and corrective feedback opportunities (e.g., through virtual reality [VR]; cf. Clay et al., 2021; Vasquez et al., 2017). Such innovations, paired with efficient training frameworks (e.g., pyramidal training; Kunnavatana et al., 2013b; Lambert et al., 2013) could expand the reach of practitioners who serve clients in remote locations.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This research has been approved by the appropriate institutional research ethics committee and has been performed in accordance with the ethical standards as laid down in the 1964 Declaration of HELSINKI and its later amendments. This manuscript is not under review, nor has it been published, elsewhere. This submission has been approved by the responsible authorities where the work was carried out. The participant's guardians provided informed consent for participation before we initiated study-related activities.

References

- Alnernary, F., Wallace, M., Alnernary, F., Gharapetian, L., & Yassine, J. (2017). Application of a pyramidal training model on the implementation of trial-based functional analysis: A partial replication. *Behavior Analysis in Practice*, 10, 301–306. <https://doi.org/10.1007/s40617-016-0159-3>
- Behavior Analyst Certification Board. (BACB). (2020). *US employment demand for behavior analysts: 2010–2017*. Author. https://www.bacb.com/wp-content/uploads/2020/05/US-Employment-Demand-for-Behavior-Analysts_2020_.pdf
- Bailey, K. M., Lambert, J. M., Banerjee, I., Copeland, B. A., & Standish, C. M. (2020a). *Trial based functional communication training*. <https://lab.vanderbilt.edu/lambert-lab/trainings/trial-based-functional-communication-training/>
- Bailey, K. M., Lambert, J. M., Banerjee, I., & Standish, C. M. (2020b). *Trial based discrimination training*. <https://lab.vanderbilt.edu/lambert-lab/trainings/trial-based-discrimination-training/>

- Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis*, 46, 1–21. <https://doi.org/10.1002/jaba.30>
- Benson, S. S., Dimian, A. F., Elmquist, M., Simacek, J., McComas, J. J., & Symons, F. J. (2018). Coaching parents to assess and treat self-injurious behaviour via telehealth. *Journal of Intellectual Disability Research*, 62, 1114–1123.
- Bloom, S. E., Iwata, B. A., Fritz, J. N., Roscoe, E. M., & Carreau, A. B. (2011). Classroom application of a trial-based functional analysis. *Journal of Applied Behavior Analysis*, 44(1), 19–31. <https://doi.org/10.1901/jaba.2011.44-19>
- Bloom, S. E., Lambert, J. M., Dayton, E., & Samaha, A. L. (2013). Teacher-conducted trial-based functional analysis as the basis for intervention. *Journal of Applied Behavior Analysis*, 46, 208–218. <https://doi.org/10.1002/jaba.21>
- Clay, C. J., Schmitz, B. A., Balakrishnan, B., Hopfenblatt, J. P., Evans, A., & Kahng, S. (2021). Feasibility of virtual reality behavioral skills training for preservice clinicians. *Journal of Applied Behavior Analysis*, 54, 547–565. <https://doi.org/10.1002/jaba.806>
- Conroy, M., Fox, J., Crain, L., Jenkins, A., & Belcher, K. (1996). Evaluating the social and ecological validity of analog assessment procedures for challenging behaviors in young children. *Education and Treatment of Children*, 19, 233–256.
- DiGennaro Reed, F. D., & Henley, A. J. (2015). A survey of staff training and performance management practices: The good, the bad, and the ugly. *Behavior Analysis in Practice*, 8, 16–26. <https://doi.org/10.1007/s40617-015-0044-5>
- Drifke, M., Tiger, J. H., & Wierzbica, B. C. (2017). Using behavioral skills training to teach parents to implement three-step prompting: A component analysis and generalization assessment. *Learning and Motivation*, 57, 1–14. <https://doi.org/10.1016/j.lmot.2016.12.001>
- Fettig, A., & Barton, E. E. (2014). Parent implementation of function-based intervention to reduce children's challenging behavior: A literature review. *Topics in Early Childhood Special Education*, 34, 49–61. <https://doi.org/10.1177/0271121413513037>
- Flynn, S. D., & Lo, Y. Y. (2016). Teacher implementation of trial-based functional analysis and differential reinforcement of alternative behavior for students with challenging behavior. *Journal of Behavioral Education*, 25, 1–31. <https://doi.org/10.1007/s10864-015-9231-2>
- Fuhrman, A. M., Lambert, J. M., & Greer, B. D. (2021). A brief review of expanded-operant treatments for mitigating resurgence. *The Psychological Record*. <https://doi.org/10.1007/s40732-021-00456-z>
- Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2018). Multiple baseline and multiple probe designs. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology*. Routledge.
- Gerow, S., Hagan-Burke, S., Rispoli, M., Gregori, E., Mason, R., & Ninci, J. (2018). A systematic review of parent-implemented functional communication training for children with ASD. *Behavior Modification*, 42(3), 335–363. <https://doi.org/10.1177/001402918793399>
- Hahs, A. D., & Jarynowski, J. (2019). Targeting staff treatment integrity of the PEAK relational training system using behavioral skills training. *Behavior Analysis in Practice*, 12, 209–215. <https://doi.org/10.1007/s40617-018-00278-6>
- 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/10.1037/t24967-000>
- Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis*, 36, 147–185. <https://doi.org/10.1901/jaba.2003.36-147>
- Harvey, M. T., Lewis-Palmer, T., Horner, R. H., & Sugai, G. (2013). Trans-situational interventions: Generalization of behavior support across school and home environments. *Behavioral Disorders*, 28, 299–312.
- Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice*, 1(1), 3–9.
- Jeong, Y., & Copeland, S. R. (2020). Comparing functional behavior assessment-based interventions and non-functional behavior assessment-based interventions: A systematic review of outcomes and methodological quality of studies. *Journal of Behavioral Education*, 29, 1–41. <https://doi.org/10.1007/s10864-019-09355-4>
- Kelly, M., Heffer, R., Gresham, F., & Elliott, S. (1989). Development of a modified Treatment Evaluation Inventory. *Journal of Psychopathology and Behavioral Assessment*, 11, 235–247. <https://doi.org/10.1007/BF00960495>

- Kodak, T., Fisher, W. W., Paden, A., & Dickes, N. (2013). Evaluation of the utility of a discrete-trial functional analysis in early intervention classrooms. *Journal of Applied Behavior Analysis*, 24, 301–306. <https://doi.org/10.1002/jaba.2>
- Kunnavatana, S. S., Bloom, S. E., Samaha, A. L., & Dayton, E. (2013a). Training teachers to conduct trial-based functional analyses. *Behavior Modification*, 37, 707–722. <https://doi.org/10.1177/0145445513490950>
- Kunnavatana, S. S., Bloom, S. E., Samaha, A. L., Lingugariris-Kraft, B., Dayton, E., & Harris, S. K. (2013b). Using a modified pyramidal training model to teach special education teachers to conduct trial-based functional analyses. *Teacher Education and Special Education*, 36, 267–285. <https://doi.org/10.1177/0888406413500152>
- Lambert, J. M., Bloom, S. E., & Irvin, J. (2012). Trial-based functional analysis and functional communication training in an early childhood setting. *Journal of Applied Behavior Analysis*, 45, 579–584. <https://doi.org/10.1901/jaba.2012.45-579>
- Lambert, J. M., Bloom, S. E., Kunnavatan, S. S., Collins, S. D., & Clay, C. J. (2013). Training residential staff to conduct trial-based functional analyses. *Journal of Applied Behavior Analysis*, 46, 296–300. <https://doi.org/10.1002/jaba.17>
- Lambert, J. M., Lloyd, B. P., Staubitz, J. L., Weaver, E. S., & Jennings, C. M. (2014). Effect of an automated training presentation on pre-service behavior analysts' implementation of trial-based functional analyses. *Journal of Behavioral Education*, 23, 344–367. <https://doi.org/10.1007/s10864-014-9197-5>
- Lambert, J. M., Lopano, S. E., Noel, C. R., & Ritchie, M. N. (2017). Teacher-conducted, latency-based functional analysis as basis for individualized levels system in a classroom setting. *Behavior Analysis in Practice*, 10, 422–426. <https://doi.org/10.1007/s40617-017-0200-1>
- Lejeune, L. M., Lambert, J. M., Lemons, C. J., Mottern, R. E., & Wisniewski, B. T. (2019). Teacher-conducted trial-based functional analysis and treatment of multiply controlled challenging behavior. *Behavior Analysis: Research and Practice*, 19, 241–246. <https://doi.org/10.1037/bar0000128>
- Lloyd, B. P., Wehby, J. H., Weaver, E. S., Goldman, S. E., Harvey, M. N., & Sherlock, D. R. (2015). Implementation and validation of trial-based functional analyses in public elementary school settings. *Journal of Behavioral Education*, 24, 167–195. <https://doi.org/10.1007/s10864-014-9217-5>
- Lloyd, B. P., Torelli, J. N., & Pollack, M. S. (2021). Practitioner perspectives on hypothesis testing strategies in the context of functional behavior assessment. *Journal of Behavioral Education*, 30, 417–443. <https://doi.org/10.1007/s10864-020-09384-4>
- Lindberg, J. S., Iwata, B. A., Roscoe, E. M., Worsdell, A. S., & Hanley, G. P. (2003). Treatment efficacy of noncontingent reinforcement during brief and extended application. *Journal of Applied Behavior Analysis*, 36(1), 1–19. <https://doi.org/10.1901/jaba.2003.36-1>
- Mace, F. C. (1994). The significance and future of functional analysis methodologies. *Journal of Applied Behavior Analysis*, 27, 385–392. <https://doi.org/10.1901/jaba.1994.27-385>
- McClintock, K., Hall, S., & Oliver, C. (2003). Risk markers associated with challenging behaviours in people with intellectual disabilities: A meta-analytic study. *Journal of Intellectual Disability Research*, 47(6), 405–416. <https://doi.org/10.1046/j.1365-2788.2003.00517.x>
- Mello, M. P., Goldman, S. E., Urbano, R. C., & Hodapp, R. M. (2016). Services for children with autism spectrum disorder: Comparing rural and non-rural communities. *Education and Training in Autism and Developmental Disabilities*, 51, 355–365. <http://www.jstor.org/stable/26173863>
- Murphy, M. A., & Ruble, L. A. (2012). A comparative study of rurality and urbanicity on access to and satisfaction with services for children with autism spectrum disorders. *Rural Special Education Quarterly*, 31, 3–11. <https://doi.org/10.1177/875687051203100302>
- Miltenberger, R. G. (2012). *Behavior modification* (5th ed.). Cengage Learning.
- Oliver, A. C., Pratt, L. A., & Normand, M. P. (2015). A survey of functional behavior assessment methods used by behavior analysts in practice. *Journal of Applied Behavior Analysis*, 48, 817–829. <https://doi.org/10.1002/jaba.256>
- Rispoli, M., Ninci, J., Burke, M. D., Zaini, S., Hatton, H., & Sanchez, L. (2015). Evaluating the accuracy of results for teacher implemented trial-based functional analyses. *Behavior Modification*, 39, 627–653. <https://doi.org/10.1177/0145445515590456>
- Rispoli, M. J., Ninci, J., Neely, L., & Zaini, S. (2014). A systematic review of trial-based functional analysis of challenging behavior. *Journal of Development and Physical Disabilities*, 26, 271–283. <https://doi.org/10.1007/s10882-013-9363-z>

- Roane, H. S., Fisher, W. W., Kelley, M. E., Mevers, J. L., & Bouxsein, K. J. (2013). Using modified visual-inspection criteria to interpret functional analysis outcomes. *Journal of Applied Behavior Analysis*, 46, 130–146. <https://doi.org/10.1002/jaba.13>
- Roscoe, E. M., Phillips, K. M., Kelly, M. A., Farber, R., & Dube, W. V. (2015). A statewide survey assessing practitioners' use and perceived utility of functional assessment. *Journal of Applied Behavior Analysis*, 48, 830–844. <https://doi.org/10.1002/jaba.259>
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2012). Evidence-based staff training: A guide for practitioners. *Behavior Analysis in Practice*, 5, 2–11. <https://doi.org/10.1007/BF03391819>
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2013). Teaching practitioners to conduct behavioral skills training: A pyramidal approach for training multiple human service staff. *Behavior Analysis in Practice*, 6, 4–16. <https://doi.org/10.1007/BF03391798>
- Saini, V., Fisher, W. W., & Retzlaff, B. J. (2018). Predictive validity and efficiency of ongoing visual-inspection criteria for interpreting functional analyses. *Journal of Applied Behavior Analysis*, 51, 303–320. <https://doi.org/10.1002/jaba.450>
- Sigafoos, J., & Sagers, E. (1995). A discrete-trial approach to the functional analysis of aggressive behaviour in two boys with autism. *Australia & New Zealand Journal of Developmental Disabilities*, 20, 287–297.
- Standish, C. M., Lambert, J. M., Banerjee, I., Copeland, B. A., Bailey, K. M., Lloyd, B. P., Staubitz, J. L., & Weaver, E. S. (2020). *Trial based functional analysis*. <https://lab.vanderbilt.edu/lambert-lab/trainings/trial-based-functional-analysis/>
- Standish, C. M., Bailey, K. M., Lambert, J. M., Copeland, B. A., Banerjee, I., & Lamers, M. E. (2021). Formative applications of ongoing visual inspection for trial-based functional analysis: A proof of concept. *Journal of Applied Behavior Analysis*. <https://doi.org/10.1002/jaba.866>
- Suess, A. N., Romani, P. W., Wacker, D. P., Dyson, S. M., Kuhle, J. L., Lee, J. F., et al. (2014). Evaluating the treatment fidelity of parents who conduct in-home functional communication training with coaching via telehealth. *Journal of Behavioral Education*, 23, 34–59.
- Vasquez, E., III, Marino, M. T., Donehower, C., & Koch, A. (2017). Functional analysis in virtual environments. *Rural Special Education Quarterly*, 36, 17–24.
- Vismara, L. A., McCormick, C. E., Wagner, A. L., Monlux, K., Nadhan, A., & Young, G. S. (2018). Telehealth parent training in the early start Denver model: Results from a randomized controlled study. *Focus on Autism and Other Developmental Disabilities*, 33, 67–79.
- Wacker, D. P., Lee, J. F., Dalmau, Y. C. P., Kopelman, T. G., Lindgren, S. D., Kuhle, J., et al. (2013). Conducting functional communication training via telehealth to reduce the problem behavior of young children with autism. *Journal of Developmental and Physical Disabilities*, 25, 35–44.
- Yoder, P. J., Symons, F. J., & Lloyd, B. P. (2018). *Observational measurement of behavior* (2nd edn.). Brookes.

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