Piloting A Decision Tool to Guide Individualized Hypothesis Testing for Students with Severe and Complex Challenging Behavior

Blair P. Lloyd[1], Jessica N. Torelli[2], Marney S. Pollack[3], and Emily S. Weaver[4]

Peabody College, Vanderbilt University

[First published online May 2022 in the *Journal of Behavioral Education*]

Author Note

Support for this research was provided by the Institute of Education Sciences, U.S. Department of Education, through Grant R324B160010 to Vanderbilt University and Vanderbilt Institute for Clinical and Translational Research grant support (UL1 TR000445 from the National Center for Advancing Translational Sciences [NCATS]/National Institutes of Health [NIH]. We would like to thank Catherine Copeland, Sunya Fareed, and Hali Wilson for their support with data collection, as well as the school practitioners and specialists who offered valuable feedback as we developed and refined the decision tool. Jessica Torelli is now at Western Kentucky University. Emily Weaver is now at the Institute of Education Sciences.

Correspondence should be addressed to Blair Lloyd, Box 228 Peabody College, Nashville, Tennessee 37203. Telephone: (615) 343-9641; Email: blair.lloyd@vanderbilt.edu

[1] Blair P. Lloyd, Ph.D., BCBA-D; Department of Special Education, Peabody College, Vanderbilt University; Nashville, TN.
[2] Jessica N. Torelli, Ph.D., BCBA-D; Department of Special Education, Peabody College, Vanderbilt University; Nashville, TN.
[3] Marney S. Pollack, M.S., BCBA; Department of Special Education, Peabody College, Vanderbilt University; Nashville, TN.
[4] Emily S. Weaver, Ph.D., BCBA; Department of Special Education, Peabody College, Vanderbilt University; Nashville, TN.

Citation

Lloyd, B. P., Torelli, J. N., Pollack, M. S., & Weaver, E. S. (2024). Piloting a decision tool to guide individualized hypothesis testing for students with severe and complex challenging behavior. *Journal of Behavioral Education, 33*, 208–232. https://doi.org/10.1007/s10864-022-09478-1

The final publication is available at https://link.springer.com/article/10.1007/s10864-022-09478-1

Abstract

For students with severe or complex challenging behavior, incorporating hypothesis testing as a component of functional behavior assessment (FBA) is often warranted. Several hypothesis testing strategies (i.e., functional analysis, antecedent analysis, concurrent operant analysis) can confirm whether and how features of a student's environment impact their behavior to then inform effective intervention. Yet practitioners have limited guidance on how to select and individualize best-fit strategies for a given student and context. We developed a decision tool for behavior specialists and classroom teachers to collaboratively plan and implement individualized hypothesis testing strategies for students whose initial FBA was inconclusive. We piloted this tool with 12 practitioner teams and students, aiming to (a) identify which assessment strategies were indicated based on practitioner responses; (b) evaluate whether indicated assessments produced conclusive results; and (c) explore practitioner perceptions of the individualized assessment process. The most commonly indicated hypothesis testing strategy was functional analysis. Across teams, one or more hypothesis was successfully confirmed on the first or second assessment iteration. The assessment process was perceived positively by practitioners. Yet they reported feeling ill-equipped to complete the process independently, highlighting important next steps for training and technical assistance work.

Keywords: functional behavior assessment, decision tool, functional analysis, antecedent analysis, concurrent operant analysis

Piloting a Decision Tool to Guide Individualized Hypothesis Testing for Students with Severe and Complex Challenging Behavior

The rise of multi-tiered systems of behavior support has brought a shift away from traditional reactive and punitive approaches to address challenging behavior, and towards more proactive and positive ones (Horner et al., 2017; U.S. Department of Education, 2021). This shift has not only impacted the types and intensities of behavior supports provided in schools, but the process by which students are identified as needing them. Functional behavior assessments (FBAs) and individualized behavior supports, for example, are no longer reserved only for students with disabilities or cases in which they are legally mandated. Instead, FBA is incorporated in a continuum of behavioral intervention, with referrals based on a lack of response to universal or targeted supports (Collins & Zirkel, 2017). As a result, students referred for FBA represent an increasingly heterogeneous group with respect to disability status, behavior profiles, skill repertoires, and instructional settings (e.g., Anderson et al., 2015; Bruni et al., 2017; Oliver et al., 2015).

There is no single agreed-upon protocol for how to conduct an FBA. Rather, FBA encompasses a collection of assessment tools and strategies that can be used to generate or test hypotheses about when and why a student engages in challenging behavior (Collins & Zirkel, 2017). This information is then used to design an intervention matched to the behavior's function. In practice, behavior specialists and educators alike commonly rely on interviews and classroom observations as their primary assessment tools (Katsiyannis et al., 2008; Oliver et al., 2015; O'Neill et al., 2015). In cases of mild to moderate challenging behavior, these descriptive FBA methods are often sufficient to help teams generate hypotheses about behavior function and develop effective function-based interventions (e.g., Strickland-Cohen & Horner, 2015).

However, in cases of more severe, persistent, or complex challenging behavior, more rigorous assessment methods are often warranted (Anderson & St. Peter, 2013; Hanley, 2012).

Experimental analysis can be incorporated in FBAs to systematically test hypotheses related to challenging or replacement behaviors. While these strategies require planning and behavioral expertise, their increased rigor offers stronger conclusions concerning the environmental factors that impact student behavior (Hanley, 2012). The hypothesis testing strategy with the deepest evidence base is the functional analysis (FA; Iwata et al., 1994). Informed by a conceptual review of motivations for challenging behavior and an assumption that functions vary by individual (Carr, 1977), the FA was designed to systematically test hypotheses for when and why someone engages in challenging behavior. Levels of challenging behavior are compared between a series of test conditions and a control condition to determine which (if any) hypotheses are confirmed. The ability of this 'standard' FA to confirm functions of challenging behavior that lead to efficacious function-based interventions is supported by decades of research (Beavers et al., 2013). Yet several aspects of this approach make it difficult to incorporate in school-based FBAs. Notable challenges include limited resources (e.g., time, training, space, staff) and potential safety risks associated with repeatedly triggering challenging behavior and responding in ways that temporarily reinforce it (Deochand et al., 2020; Hanley, 2012; Lloyd et al., 2021; Oliver et al., 2015). Educational researchers have also expressed concerns regarding a lack of ecological validity of assessment conditions (e.g., Conroy et al., 1996).

The barriers to conducting FAs in schools have given rise to a related literature focusing on adaptations and alternatives to the standard FA (Lloyd et al., 2016). Several adaptations have been applied to maximize efficiency, safety, and ecological validity of FA procedures. For example, rather than conducting sessions lasting 10–15 min each, adaptations known as latency-based and trial-based FAs involve conducting brief trials that end on the first occurrence of challenging behavior (e.g., Lambert et al., 2017; Lloyd et al., 2015). These adaptations are designed to decrease total assessment time as well as the number of challenging behavior occurrences necessary to confirm hypotheses. Other adaptations include paring down the number of test conditions (Hanley, 2012; Iwata & Dozier, 2008); synthesizing multiple antecedents or consequences within a test condition to more closely approximate contingencies in the natural setting (e.g., Hanley et al., 2014; Santiago et al., 2016); and targeting lower-risk precursors to challenging behavior (e.g., whining, clenching fists) to avoid escalation to more dangerous topographies (Heath & Smith, 2019).

Other hypothesis testing strategies represent alternatives to the FA. The antecedent (or structural) analysis also involves test and control conditions, but only antecedent variables are systematically manipulated. Antecedent analyses have been used in classrooms to understand how various instructional and social variables can trigger challenging behavior (Stichter & Conroy, 2005). Identification of these triggering factors can then inform preventative strategies to incorporate in behavior support plans. Examples of preventative strategies informed by results of antecedent analysis include adjusting task difficulty (Umbreit et al., 2004), giving students opportunities to choose among tasks (Vaughn & Horner, 1997), and layering multiple instructional format variables (e.g., group work, task structure, teacher proximity) to increase the likelihood of appropriate behavior (Stichter et al., 2005).

Another alternative hypothesis testing strategy is a concurrent operant analysis (COA). This approach is distinct from functional and antecedent analysis in that it does not require evoking challenging behavior to interpret outcomes. In a COA, students are presented with a series of choices between two or more conditions that are simultaneously available (Harding et al., 1999). Rather than measuring challenging behavior, measures of choice allocation are used to interpret results. While this strategy does not directly test the function of challenging behavior, it can identify reinforcers for appropriate behavior (e.g., Lloyd et al., 2020), as well as inform sensitivity to certain parameters of reinforcement (i.e., quality, magnitude, immediacy, response effort), which can then be used to design interventions (e.g., Gardner et al., 2009; Romani et al., 2017). COAs also prove useful for students whose behavioral challenges are persistent yet passive in nature (e.g., not responding to instructional prompts; Quigley et al., 2013) and therefore more difficult to target via functional or antecedent analysis.

Despite there being several hypothesis testing strategies that show promise as a component of FBA, there is no existing framework to guide practitioner selection of a best-fit strategy for a given student and context. In fact, a recent survey of more than 600 practicing behavior analysts indicated a need for more decision tools to guide their practice, including those focused on risk-benefit analysis specific to FAs (Deochand et al., 2020). There is also limited research informing practitioner perspectives on these hypothesis testing strategies, including factors that might influence their choice of one strategy over another. In one recent survey study of school practitioners who supported students with challenging behavior, COAs were rated more favorably than functional and antecedent analysis with respect to acceptability of procedures, feasibility, and anticipated barriers (Lloyd et al., 2021). However, ratings were made based on short video overviews of each strategy—not based on direct experience and not with a specific student and context in mind.

The purpose of this study was to develop and pilot a decision tool to support behavior specialists and classroom teachers in collaboratively planning and implementing an individualized hypothesis testing strategy for students with severe or complex challenging behavior. We focused on partnerships between behavior specialists and teachers based on previous research suggesting collaboration between those with behavioral expertise and those who know the student and context best is critical for producing high quality function-based supports (Benazzi et al., 2006). The decision tool was intended for teams who had attempted a business-as-usual (i.e., descriptive) FBA, but whose results were inconclusive or insufficient in developing an effective behavior support plan. Based on a series of practitioner team responses to questions about their student, the decision tool indicated one of three hypothesis testing strategies (i.e., FA, antecedent analysis, COA) and provided a structure for teams to individualize assessment procedures. We addressed the following research questions: (1) Which hypothesis testing strategies were indicated by the response-guided decision tool? What practitioner responses drove these indications? (2) As guided by the decision tool, in what ways did practitioners individualize indicated hypothesis testing strategies? (3) Did results of individualized assessments successfully confirm one or more hypothesis? and (4) After playing active roles in planning and conducting hypothesis testing strategies, what were behavior specialist and classroom teacher perceptions of their acceptability, feasibility, and utility?

Method

Participants and Settings

After obtaining study approval from our university's Institutional Review Board and district-level approvals for research, we recruited behavior specialists by contacting behavior support team leads in two local school districts. We asked for nominations of students in Grades K–5 who engaged in severe, persistent, and/or complex challenging behavior and for whom an FBA had been attempted but had either (a) produced inconclusive results or (b) informed an intervention that had not been effective. We relied on reports from behavior specialists to make these determinations and collected completed FBAs and intervention data to confirm these reports when they were available. For each nominated student participant, we also required a behavior specialist and at least one classroom teacher who supported the student to participate. We obtained informed consent from students' caregivers, behavior specialists, and teachers, as well as verbal assent from students prior to collecting any study data.

Twelve students participated, as well as 21 school practitioners who supported them. Practitioners included seven behavior specialists (five of whom were Board Certified Behavior Analysts [BCBAs]), nine certified general education teachers, four certified special education teachers, and one paraeducator. Twenty practitioners identified as female (one male). Eighteen identified as White/Non-Hispanic, two as Black, and one as Hispanic.

Participating behavior specialists had between 1 and 13 years experience in their current role ($M = 5.0$). All behavior specialists reported having prior experience completing FBAs (i.e., participating on FBA teams, collecting observational data, completing interviews, completing record reviews, and reviewing and interpreting FBA data). Six of the seven behavior specialists reported previous experience training or supervising school staff on completing FBAs. All behavior specialists reported previous experience developing, implementing, and monitoring progress of individualized behavior support plans. Five specialists reported previous experience completing a functional analysis. Four of the seven behavior specialists participated on more than one student support team. Three specialists served on two teams (Teams 04 and 06; Teams 08 and 09; Teams 03 and 10). One specialist served on three teams (Teams 01, 05, and 07).

Participating classroom educators had between 1 and 20 years of experience in their current role ($M = 6.8$). Most of the 14 classroom educators reported experience participating on an FBA team ($n = 10$), reviewing FBA data ($n = 12$), collecting data on student behavior ($n =$

12), developing individualized behavior support plans (*n* = 10), implementing individualized behavior support plans (*n* = 12), and collecting progress monitoring data (*n* = 12). Three participating special educators reported previous experience completing a functional analysis.

Of the 12 participating students, 10 were male (two female). Seven students were Black/African American, four were White, and one was Hispanic. Seven received special education services at the time of participation; five did not, though two students were in the process of being evaluated. Special education eligibility categories included emotional disturbance (*n* = 3), autism spectrum disorder (*n* = 1), intellectual disability (*n* = 1), other health impairment (*n* = 1), and specific learning disability (*n* = 1). As rated by both behavior specialists and teachers, all participating students scored in the Abnormal range for Conduct Problems and/or Total Difficulties scores on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). Student challenging behaviors ranged across participants and included physical and verbal aggression, self-injury, major disruptive behavior, elopement, and persistent noncompliance/off-task behaviors that prevented them from accessing instruction. Challenging behavior topographies are listed by student team in Table 1.

Participating students' initial FBAs were completed in the same (*n* = 8) or previous (*n* = 4) academic year. For students whose initial FBA occurred in the previous academic year, participating behavior specialists were updating the FBA at the time of participation. With one exception (Team 11), all participating behavior specialists were part of the team that completed the initial FBA. All initial FBAs included at least one interview and two or more direct observations. Interviews were conducted with teachers, caregivers, and students (often some combination thereof) and observations were completed in students' general or special education classrooms (for a subset of students, additional observations took place in related arts or English

Language classes). During observations, behavior specialists reported collecting data on student challenging behavior, antecedent events, and consequences. The initial FBAs were either underway and inconclusive at the time of participation or had indicated multiple potential functions that had not yet informed an effective function-based intervention.

Participating students attended 10 schools in two public school districts. One district was large, urban, and served students in grades K-12. Sixty-nine percent of students in this district were Black, Hispanic, or Native American; 41% were economically disadvantaged; and 12% had disabilities (Tennessee Department of Education [TDOE], 2020). The second district was small, urban, and served students in grades PreK–6. Forty-eight percent of students in this district were Black, Hispanic, or Native American; 29% were economically disadvantaged; and 15% had disabilities (TDOE, 2020). Both districts had completed training and were receiving technical assistance on multi-tiered systems of behavior support; all schools reported having multi-tiered systems of support for instruction and behavior in place. With respect to FBAs, both district's behavior support teams followed a model of descriptive FBA (interviews, questionnaires, classroom observations) but expressed interest in incorporating experimental (i.e., hypothesis testing) strategies in complex cases.

Assessment planning meetings took place in conference rooms or empty classrooms at the school. Settings for completing assessments with students were determined through use of the decision tool and included general education classrooms (with and without peers present), special education classrooms (with and without peers present), and empty instructional spaces.

Measures and Materials

Hypothesis Testing Decision Tool

We iteratively developed the decision tool with the goal of guiding behavior specialists and classroom teachers through the process of selecting a hypothesis testing strategy that matched the team's priorities based on the student they supported, as well as individualizing critical features of the assessment. We developed the decision tool across a series of stages. First, we reviewed the empirical literature on FA approaches (e.g., Anderson et al., 2015; Beavers et al., 2013; Lloyd et al., 2016) as well as discussion articles by expert scientist-practitioners related to hypothesis testing (e.g., Hanley, 2012). Second, we sought and incorporated feedback from four nationally-recognized university researchers (each at different institutions) with distinct areas of expertise relating to the tool (e.g., severe and complex challenging behavior, assessment and intervention for students with emotional/behavioral disorders, assessment and intervention for students with intellectual/developmental disabilities, methods of training and coaching school personnel). Third, we sought and incorporated feedback from a team of state-funded technical assistance personnel who were training school practitioners to incorporate hypothesis testing strategies (i.e., FA, COA) in FBAs for students with the most intensive behavior support needs. Fourth, we conducted a state-wide survey evaluating practitioner perspectives on incorporating hypothesis testing in FBAs (Lloyd et al., 2021). Fifth, we conducted a series of nine focus groups with 30 behavior specialists and consultants working in public school settings. The final version of the decision tool was housed on a secure, online survey platform (REDCap; Harris et al., 2009) and is accessible via the following link: https://redcap.link/FBAHTdecisiontool. We used branching logic features to make all sections of the decision tool response guided, such that the total number and types of questions varied by user according to responses made throughout the survey. In addition, several of the questions included an optional response "I need more information." If users selected this response, additional descriptions or examples were provided to inform their selection. The hypothesis testing decision tool included three main sections:

hypothesis testing strategy selection, assessment context, and strategy-specific individualization.

Hypothesis Testing Strategy Selection. The initial set of questions was designed to indicate which of three hypothesis testing strategies (i.e., FA, antecedent analysis, COA) was most appropriate given the team's priorities and the student they supported. Users were asked to identify the primary behavior of concern, indicate how frequently the behavior occurred, and identify which of three sets of questions they wanted to prioritize addressing. Each set of questions corresponded to the hypothesis testing strategy that was best suited to answer them. These questions, as well as a summary of the branching logic, are depicted in Figure 1. After participants reached consensus on the question set they wanted to prioritize, information on advantages and disadvantages of the selected approach was presented, followed by a question of whether the team was comfortable proceeding with that approach. For example, if the selected questions corresponded to the FA, the next item emphasized that this method would require triggering and reinforcing challenging behavior, but that it would provide the most conclusive indication of behavior function. We included this step in the decision tool to allow teams the opportunity to select an alternative to the FA (or antecedent analysis) if they had concerns with procedures. At the end of this section, and based on user responses, the decision tool provided a recommendation for which of the three hypothesis testing strategies to use.

Assessment Context. The second set of questions was designed to inform contextual aspects of the assessment, including where to complete it and who should serve as assessment implementer. Users responded to questions about whether the student's challenging behavior was likely to vary by setting, adult implementer, or the presence of peers (if so, they were guided to conduct the assessment in the usual classroom with the teacher as implementer). Users also indicated whether the behaviors could be managed safely in the classroom (if not, they were guided to conduct the assessment in a separate location or when peers were not present).

Strategy-Specific Individualization. The third set of questions was specific to the hypothesis testing strategy selected. If users selected an FA or antecedent analysis, they were later prompted to respond to open-ended questions to identify specific antecedent or consequence events that might trigger or reinforce the student's challenging behavior. This information was used to design test and control conditions for each analysis type. We encouraged combining more than one antecedent and/or consequence variables in test conditions when doing so represented what was typical in the classroom setting (Hanley et al., 2014). Other questions were intended to inform procedural variations to maximize the safety, feasibility, and acceptability of assessment procedures. These procedural variations represented components of different empirically-supported models of FA, and included setting abbreviated session durations (i.e., 5 min; Wacker et al., 2004; Wallace & Iwata, 1999) and formats (i.e., latency-based or trial-based; Lambert et al., 2017; Lloyd et al., 2015); limiting the total number of test conditions according to the strongest hypotheses (Hanley, 2012; Iwata & Dozier, 2008); and identifying potential precursor behaviors to target during the analysis (Heath & Smith, 2019). If users selected a COA, they were prompted to answer questions informing which of two COA models to use (instruction-based vs. function-based) and how choice conditions should be introduced (i.e., verbal descriptions vs. pre-session exposure). An instruction-based COA was recommended when the primary concern was a behavioral deficit (i.e., persistent noncompliant or off-task behaviors) or when the team's priority was to identify conditions that would reinforce task engagement or work completion. A function-based COA (Casey, 2001) was recommended when the team's priority was to identify the function of challenging behavior, but other selections precluded conducting an FA (e.g., behaviors that were severe but occurred too infrequently to turn on and off within assessment sessions; discomfort with FA procedures).

Questionnaires

Participating teachers and behavior specialists completed an online demographic form that was housed on REDCap (Harris et al., 2009). The form included questions on demographic variables (e.g., gender, race/ethnicity, age, education), current professional roles and credentials, and previous experience with FBAs. To inform current levels of behavioral functioning, and ensure challenging behaviors were severe or complex enough to warrant hypothesis testing, practitioners were asked to complete the SDQ (Goodman, 2001) for the participating student.

Participating staff also completed a social validity questionnaire to inform their views on the assessment planning and implementation process. The social validity questionnaire was created by the research team for purposes of this study; several items were informed from a previous survey study evaluating potential barriers to hypothesis testing in the context of FBA (Lloyd et al., 2021). Participants rated the extent to which they agreed or disagreed with 16 statements relating to the acceptability, feasibility, and utility of the individualized assessment process (Likert-type scale ranging from 1 [*Strongly Disagree*] to 5 [*Strongly Agree*]; see Table 2). Items representing negative statements (e.g., *The assessment took too long to complete*) were reverse-scored, so that higher ratings reflected higher social validity across items. Questionnaires were completed anonymously to minimize the influence of social desirability on responses. Data from this social validity measure are missing from the last three participating teachers, as they completed an alternative social validity form focused on both assessment and intervention.

Procedures

Initial Interview

We conducted an initial 30- to 60-min phone interview with the participating behavior specialist to gather basic information about the student (e.g., grade level, special education status), the behavior(s) of concern, the initial FBA, and attempted behavior support strategies. We reviewed and confirmed conditions that warranted use of the decision tool (i.e., inconclusive FBA or ineffective support plan). We planned to exclude cases in which a behavior specialist attributed lack of intervention effectiveness to inconsistent implementation (though this did not occur). Depending on availability, the student's teacher also attended the initial interview. If the teacher was not available, we contacted them separately to address questions for which the behavior specialist suggested additional teacher input (e.g., student's preferred activities).

Decision Tool Overview

We met with behavior specialists in person to present an overview of the decision tool and orient them to using the tool. The goal of this 30- to 45-min meeting was to prepare behavior specialists to lead the assessment planning meeting. Behavior specialists accessed the decision tool on their own computer so that they could practice working through the tool. We guided them through each question set, pointed out the "need more information" options, and encouraged them to review the tool on their own prior to the assessment planning meeting. We solicited questions and discussed potential challenges they anticipated. We emphasized we would help facilitate the assessment planning meeting, but the priority was to solicit and consider all team members' input and reach consensus on each decision point.

Assessment Planning Meeting

During assessment planning meetings, the participating behavior specialist led the team to collaboratively work through the decision tool. Participating behavior specialists and classroom teachers were required to attend this meeting. We also encouraged participants to invite additional staff (e.g., administrators, school counselors, paraprofessionals) to attend these meetings if they wished. The first or second author facilitated all assessment planning meetings by stating the purpose and intended outcomes of the meeting, answering questions and providing clarification as needed, managing time to ensure all meeting goals were met, and summarizing planning outcomes and next steps. Behavior specialists led the team through the decision tool by presenting each question to the group, soliciting input and feedback from classroom teachers and other attending team members, contributing their own responses, and helping reach consensus on items under debate. Assessment planning meetings lasted 45–60 min. Following the meeting, the research team drafted an assessment protocol based on team selections and responses. Protocols included target behavior definitions, implementer roles, assessment setting and materials, and programmed procedures for each assessment condition. We solicited and incorporated feedback from both behavior specialists and classroom teachers on drafted assessment protocols before finalizing plans.

Hypothesis Testing Training and Implementation

Assessment training meetings were scheduled with assessment implementers (classroom teacher or behavior specialist) to review and practice assessment procedures. The primary purpose was to ensure the implementer understood the procedures, felt comfortable with them, and had opportunities to ask questions or express concerns prior to implementation. A research team member brought printed copies of the assessment protocol and a 1-page tip sheet of step-by-step instructions for each assessment condition. These meetings included brief reviews of assessment procedures, followed by opportunities for questions, models, and role-play with feedback. Though several implementers opted to role-play conditions, we did not require they do so, as in-the-moment coaching was available during actual implementation.

Each student's assessment was conducted as planned in the assessment protocol. At least two members of the research team were present during all assessments—one to serve a coaching role for the implementer, and another to collect data. Additional research team members were present for most assessments to collect reliability and/or fidelity data. When behavior specialists were not primary implementers, they also collected data on student behavior and helped coach the implementer. The research team member serving as coach provided reminders to the implementer prior to each session and delivered within-session prompts on an as-needed basis. Each assessment was completed during one 30- to 75-min assessment visit; for four students, we returned on a different day to complete a modified assessment plan. The research team created modified assessment plans following collaborative debriefs with participating behavior specialists and teachers, and with reference to the decision tool.

Reliability data on student behavior were collected for 68.4% of assessment sessions (participant range, 31.0%–100%). When latency to target behavior was the dependent variable (functional and antecedent analyses), sessions in which each observer's latency was within 5 s of the other were scored as 100% agreement (latencies outside the 5-s window of agreement were scored as 0%). When total duration of time spent in each condition was the dependent variable (COA), we calculated a total percentage agreement by dividing the smaller duration by the larger duration and multiplying by 100 (per choice condition). We also calculated total percentage agreement (smaller/larger*100) on frequency of target behavior during COA sessions.

Fidelity data on implementer behavior were collected for 85.1% of assessment sessions (participant range, 33.0%–100%). For functional and antecedent analyses, our fidelity data collection form included a combination of checklist items (scored as *yes* or *no*; e.g., reinforcer delivered within 3 s of target behavior) and event tallies for 'repeated use' behaviors (e.g., delivers attention at least once per 30 s) or other implementer behaviors that varied by opportunity (e.g., ignore target behavior). Tallies were entered in either a *correct* or *incorrect* column. Percentages of fidelity were calculated as the sum of checklist items marked *yes* and tallies in the *correct* column divided by the sum of all checklist items and all tallies (multiplied by 100). For COAs, we indicated *yes* or *no* for initial set-up behaviors (e.g., explained rules for each choice condition) and scored student access to consequences programmed for each choice condition (e.g., preferred items, adult attention), as well as implementer interactions with the student (e.g., prompts, rule reminders) on an interval-by-interval basis. Fidelity percentages were calculated as the number of intervals with correct implementation divided by total intervals and multiplied by 100. Means and ranges of inter-observer agreement and procedural fidelity for FAs and the antecedent analysis are reported by student in Figure 2; similar summaries of agreement and fidelity for COAs are included as supplemental material.

Following the assessment, the research team summarized and graphed assessment data and scheduled a follow-up meeting to review results, discuss corresponding behavior support strategies, and create a plan for intervention support. Intervention support from the research team included (a) preparing written summaries of intervention strategies that incorporated feedback and input from participating practitioners; (b) sharing brief tip-sheets to support implementation; and (c) completing follow-up visits where we modeled and/or observed strategy implementation and provided performance feedback.

Results

Which Hypothesis Testing Strategies Were Indicated by the Response-Guided Decision Tool?

The response-guided decision tool indicated FA as the best-fit hypothesis testing strategy for eight of the 12 participating teams; COA for three teams; and antecedent analysis for one team. Figure 1 depicts four distinct paths illustrating team responses that informed the indicated hypothesis testing strategy (Paths A, B, C, and D). Path A depicts the most common series of responses that indicated FA. Across teams who decided to conduct an FA, the primary concern was the presence or excess of one or more challenging behavior (rather than topographies representing behavioral deficits; i.e., persistent noncompliant/off-task behaviors). These challenging behaviors also were reported to occur frequently. When presented with three sets of questions to prioritize answering for their student, these teams selected questions that the FA was uniquely designed to answer (i.e., why the challenging behavior was occurring, how to respond to it, what replacement behaviors to teach). Further, when provided information about what procedures would be required to answer these questions (i.e., triggering challenging behavior and responding in ways that would temporarily reinforce it), teams reached consensus on their willingness to proceed.

Path B depicts the series of responses that indicated an antecedent analysis. This team also described the primary concern as the presence or excess of one or more challenging behavior. These behaviors were reported to occur less frequently overall but reliably in certain instructional contexts. This team indicated a priority to address a set of questions for which the antecedent analysis was uniquely aligned. That is, they wanted to identify specific instructional prompts or tasks that triggered the student’s challenging behavior. When provided information on the procedures that would be required to answer this question (i.e., triggering challenging behavior), they also reached consensus to proceed.

Indications of COA are depicted by Paths C and D. For one team, the COA (function-based) was indicated because the challenging behavior they wanted to target was severe but did not occur frequently enough to reliably evoke via functional or antecedent analysis (Path C). For the other two teams, the COA (instruction-based) was indicated because the primary concern was persistent noncompliant/off-task behaviors (Path D). In these cases, the goal was to identify instructional conditions that would most likely reinforce task engagement and completion.

In What Ways Did Practitioners Individualize the Selected Assessment Strategy?

Assessment characteristics are summarized by student team in Table 1. Among the nine teams for whom functional or antecedent analysis was indicated, seven decided to have the classroom teacher serve as implementer, and six decided to implement the assessment in the usual classroom setting. These decisions were based on expectations that challenging behavior was more likely to be evoked under these conditions, and that it could be managed safely in the usual classroom if certain precautionary measures were in place (e.g., having an extra adult in the room). All nine teams identified at least one precursor to the primary behavior of concern (e.g., clenching fists, stomping feet, leaving work area) that could be included as target behavior to decrease the likelihood of behavior escalation. Finally, with only one exception, teams completing an FA opted to conduct a latency-based variation, in which each test session ended upon the first instance of precursor or challenging behavior, with a maximum session duration of 5 min. This selection was based on an expectation that challenging behavior could be triggered relatively quickly (i.e., within a few minutes) in the evocative context and a preference for completing the assessment in a single visit (as opposed to distributing brief trials throughout and/or across school days, as is recommended for the trial-based variation; Sigafoos & Saggers, 1995). One team selected the trial-based FA.

All three teams for whom COA was indicated decided to have the teacher serve as implementer based on expectations that choices among conditions that included an adult (e.g., attention or assistance with instruction) would vary if unfamiliar adults were involved. All three teams decided to implement the assessment in an empty classroom for reasons of convenience (e.g., to avoid interrupting ongoing instruction) and because choice-making behavior was not expected to vary by setting. Finally, all three teams selected verbal descriptions of choice presentations over experiential presentations based on student communication skills.

A mean of 2.0 test conditions were included across functional and antecedent analyses (range, 1–3). Test conditions are labeled in the FA and antecedent analysis graphs depicted in Figure 2. Most test conditions focused on variables that commonly reinforce challenging behavior in school (e.g., access to adult attention, escape from task demands; Anderson et al., 2015). However, test conditions were individualized with respect to the types of attention (e.g., redirections, statements of concern) and task demands (e.g., independent math work, 1:1 reading instruction) programmed. A few FAs tested idiosyncratic functions (e.g., escape peer noise, Team 01) and several teams synthesized multiple antecedents or consequences in single test conditions (e.g., escape to attention, escape to attention and preferred activities) to align with typical antecedent and consequence sequences encountered in the classroom.

Did Results of Individualized Assessments Successfully Confirm One or More Hypotheses?

Graphed results of functional and antecedent analyses are depicted in Figure 2; total sessions, assessment durations, and target (i.e., challenging or precursor) behaviors evoked during assessment sessions are reported in the last two columns of Table 1. Student-level COA data summaries are included as supplemental material. For eight of 12 teams, we confirmed one or more hypothesis in the first attempt of the selected assessment, which included five FAs (Teams 02, 05, 10, 11, and 12) and three COAs (Teams 03, 04, and 09). In Figure 2, test condition labels that are underlined indicate confirmed hypotheses. Among these assessments, the mean assessment duration was 27.5 (range, 13–41) min; the mean number of targeted precursor and/or challenging behavior occurrences was 2.6 (range, 0–7). As expected, no challenging behavior occurred during COAs.

For the other four teams (Teams 01, 06, 07, and 08), confirming one or more hypothesis required a second assessment attempt. For these teams, we held a collaborative debrief after the first assessment attempt, and the research team drafted a modified assessment plan informed by the discussion and with reference to the decision tool. For two of these teams (Teams 01 and 07), we completed a second attempt of the same selected hypothesis testing strategy (FA). For Team 01, we completed a second FA where we changed the variation from trial-based to latency-based (due to challenging behavior carrying over from test to control segments) and further specified the hypothesis based on practitioner input. For Team 07, we completed a second FA in a different context. The first FA was completed in the student's classroom with peers present. The student did not engage in challenging behavior during test conditions, which the team attributed to his access to attention from preferred peers (a variable that was not captured by our fidelity measure, which focused on teacher implementation). We completed a second FA with peers absent, and confirmed an escape-to-attention hypothesis.

For two other teams (Teams 06 and 08), the second attempt required a change in assessment strategy. For Team 06, we first attempted a latency-based FA but observed no target behaviors. In the post-assessment debrief, there was not an obvious procedural adjustment that would make target behaviors more likely to occur. Thus, consistent with the decision tool, we then conducted a function-based COA (indicated when a team prioritizes identifying function but challenging behaviors do not occur frequently enough to evoke via FA). The function-based COA confirmed access to preferred items as the most likely reinforcer for appropriate behavior. For Team 08, we initially attempted an antecedent analysis but results were undifferentiated between test and control conditions. Because we were unable to identify reliable antecedent triggers (and team members did not identify alternative antecedent hypotheses to test during the post-assessment debrief), the team decided to complete a latency-based FA that incorporated the hypothesized antecedent trigger (i.e., tasks requiring open-ended written responses) in the escape condition. The FA confirmed that the student engaged in target behaviors to escape these tasks. Among the four cases requiring a second assessment attempt, the mean total assessment duration was 47.5 (range, 34–66) min; the mean total occurrence of targeted precursor and/or challenging behavior was 4.8 (range, 0–9).

What Were Practitioner Perceptions of Acceptability, Feasibility, and Utility?

The mean agreement score across social validity items was 4.02 (item range, 2.61–4.81) on a 5-point scale, representing broad agreement that the individualized assessment process was acceptable, feasible, and useful. Means and ranges of ratings are listed by item in Table 2. Items with the highest mean ratings reflected practitioner views on the feasibility of the process (given the supports in place), the acceptability of the process for the participating student, the extent to which each team member played a meaningful role, and their perceived safety of those involved in the assessment. Items with the lowest mean ratings reflected concerns about not having sufficient training and expertise to use the assessment process without additional supports from a research team, and a perceived lack of support from other school team members (not involved in the study) to use the assessment process. In addition to variation by item, ratings also varied by practitioner role. Relative to teachers, mean ratings by behavior specialists were more favorable toward the assessment process for 13 of the 16 items. For example, relative to teachers, behavior specialists agreed more strongly that the individualized assessment was a good fit for the student. Additionally, relative to teachers, behavior specialists disagreed more strongly that the assessment took too long to complete and that they lacked sufficient supports from team members to conduct similar assessments without the research team.

Discussion

While many adaptations and alternatives to standard FA have shown promise for school application, practitioners who support students with severe and complex challenging behavior lack guidance on how to select and individualize hypothesis testing strategies for each unique student they serve. Additionally, because hypothesis testing is rarely incorporated in FBAs in practice, we know little about practitioner perspectives and priorities when it comes to implementing these assessments in schools. We designed and piloted a response-guided decision tool to promote collaborative planning of individualized hypothesis testing strategies for students whose initial FBAs were inconclusive or did not lead to effective behavior support plans. Through piloting this decision tool with a series of practitioner teams, we found that FA was still the most commonly indicated strategy based on practitioner responses. Importantly, however, the FAs that were collaboratively designed and implemented incorporated several procedural variations that made them more ‘school friendly’ relative to standard FA models. Across practitioner teams, we also found that the decision tool supported selection and individualization of assessments that (a) prioritized safety and efficiency; (b) successfully confirmed one or more hypothesis on the first or second attempt; and (c) were perceived to be socially valid by those directly involved in assessment planning and implementation.

With respect to the common indication for FA, we were not surprised that most practitioner teams chose to prioritize addressing questions aligned with confirming the function of their student’s challenging behavior. Because we recruited students with severe and complex challenging behavior, most practitioner teams had already attempted strategies aligned with the antecedent analysis (i.e., how to prevent or minimize classroom triggers). We were surprised, however, that teams were not deterred by the procedures required of an FA. Results of a previous practitioner survey suggested that relative to antecedent analysis and COA, FA procedures were perceived as less acceptable and/or feasible (Lloyd et al., 2021). Based on these data and other research highlighting barriers around practitioner acceptability of FA procedures (e.g., Oliver et al., 2015), we designed the decision tool to include opportunities for teams to select alternative assessment strategies if they expressed discomfort with FA procedures. Yet teams in the current study never chose alternatives to FA based on concerns with procedures. Practitioners' willingness to complete FAs in classrooms may have been influenced by adaptations and safeguards incorporated in the decision tool to make the procedures more 'school friendly' (e.g., targeting precursors, brief trials, few test conditions). Another likely influential factor was that practitioners made these selections knowing they would have support from the research team when conducting the assessment. Even so, our results suggest school practitioners—including classroom teachers with varied training or experience in behavioral assessment—were open to completing FAs for students with severe and complex challenging behavior.

It is also worth noting that while the decision tool framed the antecedent analysis as addressing *when* challenging behavior occurred (and how to prevent it), and framed the FA as addressing *why* challenging behavior occurred (and how to respond to and replace it), the FAs we conducted addressed both when *and* why challenging behavior occurred. With this in mind, our results suggest a simplified assessment selection process may be warranted. That is, FA may be the recommended strategy for hypothesis testing (unless counter-indicated by the student's behavioral profile) with planning time allocated to designing FA procedures that maximize safety, efficiency, and ecological validity.

On the whole, the decision tool supported design and implementation of hypothesis testing strategies that were safe, efficient, and produced conclusive results. For 10 of 12 cases, the selected assessment strategy confirmed one or more hypothesis on the first ($n = 8$) or second ($n = 2$) attempt. For comparison, research on standard and latency-based FAs designed and conducted by trained research teams suggest that interpretable results are obtained on the first attempt in fewer than half of cases (Hagopian et al., 2013; Lambert et al., 2017). Though more time was devoted to the assessment selection and planning process in this study, the hypothesis testing strategies themselves were highly efficient ($M = 34$ min) and evoked relatively few occurrences of precursor and challenging behavior ($M = 3.3$; range, 0–9). For context, in a systematic review of 39 studies involving FAs conducted in public school settings, Lloyd et al. (2016) estimated a mean assessment duration of 4 hr (representing session time alone).

Additionally, based on mean social validity ratings exceeding 4.0 on a 5-point scale, our results suggested participating practitioners thought this assessment process (a) was socially and ethically appropriate and “a good fit” for the student; (b) maintained safety for the student and others involved; (c) resulted in each team member playing a meaningful role in assessment; and (d) produced interpretable results. Because our only point of comparison was the initial descriptive FBAs completed by these teams, we are unable to isolate the impact of the decision tool on assessment outcomes. We speculate, however, that the decision tool offered a structure for facilitating collaboration across team members, individualization of assessment procedures, and, particularly in the case of FAs, offering procedural adaptations to maximize safety and efficiency. Our social validity data also extend research from previous survey studies evaluating practitioner perspectives on FA and other hypothesis testing strategies based only on written descriptions or brief video overviews of what each strategy entails (e.g., Lloyd et al., 2021;

O'Neill et al., 2015). Results from this study suggest participating practitioners considered their indicated hypothesis testing strategy to be acceptable, feasible (with available supports), and useful after experiencing them directly and playing active roles in their design and implementation.

While most of the social validity data were encouraging, our results also revealed participating practitioners did not feel equipped to complete a similar assessment process without the supports provided by the research team. Indeed, we supported teams throughout this process to ensure the decision tool was applied as intended and to evaluate its promise. As described above, we (a) answered questions and offered input (when solicited) during initial planning meetings; (b) drafted assessment protocols; (c) led training meetings on selected assessment strategies prior to implementation; and (d) helped coach, collect data, and trouble-shoot on assessment visits. This raises the important question of whether and how similar supports can be transferred to school or district personnel. Consistent with developers of other related decision frameworks (e.g., Deochand et al., 2020), our intention for this decision tool was not to replace clinical expertise or judgement related to hypothesis testing. Rather, we developed the decision tool to offer a structure for student support teams to collaboratively select and plan a rigorous and individualized assessment when such efforts were warranted. Given FAs should be reserved for students with the most severe and complex challenging behavior, training one or more specialist per district on FA technology—including 'school friendly' adaptations—may be an appropriate first step. Train-the-trainer models focusing on district behavior support teams may have the most potential for building district capacity in this area, and will be important to evaluate in future research. In fact, similar models have been applied in the context of state-funded technical assistance projects aimed at building district capacity to serve students with the most intensive behavioral needs (e.g., Bassingthwaite et al., 2018).

Our results should be considered with the following limitations in mind. First, we relied on participating behavior specialist reports with respect to whether initial FBAs were inconclusive or did not lead to effective interventions. The data we collected on initial FBAs suggested they were of adequate quality (e.g., incorporated both indirect and direct assessment methods). However, the availability and quality of initial intervention data varied among teams. Thus, it is possible some students met our inclusion criteria due to initial interventions that were of low quality or not implemented with sufficient consistency. Second, we did not systematically evaluate effects of interventions matched to hypothesis testing outcomes across participants. The goals of this stage of the project were to evaluate (a) usability of the decision tool across multiple and varied support teams and student profiles and (b) promise of the decision tool by determining whether individualized hypothesis testing strategies produced interpretable results (i.e., confirmed one or more hypothesis). While controlled intervention evaluations would have informed treatment utility of this assessment process, a substantial evidence base suggests hypotheses confirmed via experimental analysis lead to efficacious function-based interventions (Gage et al., 2012; Goh & Bambara, 2012; Walker et al., 2018). Third, the social validity data we presented do not include views of three participating teachers, due to their completion of an alternative social validity form. Their ratings of the assessment process, however, mirrored patterns identified from the other participating teachers. Finally, due to the relatively small number of behavior analysts and specialists in each participating district, a subset of behavior specialists participated across more than one student team. Five of seven participating behavior specialists also reported they had completed an FA before, though it was unclear whether this experience was from a graduate training program or their professional practice. Thus, the individualized assessment process may have been influenced by these specialists' familiarity and/or experience with hypothesis testing.

Conclusion

For students with severe and complex challenging behavior, business-as-usual FBAs consisting of interviews and classroom observations are often insufficient to identify behavior function(s) and design effective interventions. In these cases, practitioners need supports to select and individualize more rigorous assessment strategies that will guide individualized interventions. The decision tool in the current study offered a supportive framework for behavior specialists and teachers to collaboratively plan and implement individualized hypothesis testing strategies. Patterns of practitioner selections, assessment characteristics, hypothesis testing outcomes, and social validity ratings suggest a focus on transfer of FA technology to district personnel may be warranted. Additionally, results suggest that when teams prioritize collaboration, individualization, and safety of assessment procedures, traditional barriers to conducting FAs in schools can be overcome.

References

Anderson, C. M., Rodriguez, B. J., & Campbell, A. (2015). Functional behavior assessment in schools: Current status and future directions. *Journal of Behavioral Education, 24*(3), 338–371. https://doi.org/10.1007/s10864-015-9226-z

Anderson, C. M., & St. Peter, C. C. (2013). Functional analysis with typically developing children: Best practice or too early to tell?: In response to Hanley (2012). *Behavior Analysis in Practice, 6*(2), 62–76. https://doi.org/10.1007/BF03391806

Bassingthwaite, B. J., Graber, J. E., Weaver, A. D., Wacker, D. P., White-Staecker, D., Bergthold, S., & Judkins, P. (2018). Using teleconsultation to develop independent skills of school-based behavior teams in functional behavior assessment. *Journal of Educational and Psychological Consultation, 28*(3), 297–318. https://doi.org/gzds

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), 1–21. https://doi.org/f46jxq

Benazzi, L., Horner, R. H., & Good, R. H. (2006). Effects of behavior support team composition on the technical adequacy and contextual fit of behavior support plans. *The Journal of Special Education*, *40*(3), 160–170. https://doi.org/10.1177/00224669060400030401

Bruni, T. P., Drevon, D., Hixson, M., Wyse, R., Corcoran, S., & Fursa, S. (2017). The effect of functional behavior assessment on school-based interventions: A meta-analysis of single-case research. *Psychology in the Schools, 54*(4), 351–369. https://doi.org/f94xnm

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

Casey, S. D. (2001). *Comparing functional outcomes between brief functional analyses and* *concurrent operants assessments* (Publication No. 3050783) [Doctoral dissertation, University of Iowa]. ProQuest Dissertations and Theses Global.

Collins, L. W., & Zirkel, P. A. (2017). Functional behavior assessments and behavior intervention plans: Legal requirements and professional recommendations. *Journal of Positive Behavior Interventions, 19*(3), 180–190. https://doi.org/gbmkpq

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.

Deochand, N., Eldridge, R. R., & Peterson, S. M. (2020). Toward the development of a functional analysis risk assessment decision tool. *Behavior Analysis in Practice, 13*(4), 978–990. https://doi.org/10.1007/s40617-020-00433-y

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. *Behavioral Disorders, 37*, 55–77. https://doi.org/10.1177/019874291203700201

Gardner, A. W., Wacker, D. P., & Boelter, E. W. (2009). An evaluation of the interaction between quality of attention and negative reinforcement with children who display escape-maintained problem behavior. *Journal of Applied Behavior Analysis, 42*, 343–348. https://doi.org/10.1901/jaba.2009.42-343

Goh, A. E., & Bambara, L. M. (2012). Individualized positive behavior support in school settings: A meta-analysis. *Remedial and Special Education*, *33*(5), 271–286. https://doi.org/10.1177/0741932510383990

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire.

Journal of the American Academy of Child & Adolescent Psychiatry, 40(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

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

Hanley, G. P. (2012). Functional assessment of problem behavior: Dispelling myths, overcoming implementation obstacles, and developing new lore. *Behavior Analysis in Practice, 5*(1), 54–72. https://doi.org/10.1007/bf03391818

Hanley, G. P., Jin, C. S., Vanselow, N. R., & Hanratty, L. A. (2014). Producing meaningful improvements in problem behavior of children with autism via synthesized analyses and treatments. *Journal of Applied Behavior Analysis, 47*, 16–36. https://doi.org/f5tt9t

Harding, J. W., Wacker, D. P., Berg, W. K., Cooper, L. J., Asmus, J., Mlela, K., & Muller, J. (1999). An analysis of choice making in the assessment of young children with severe behavior problems. *Journal of Applied Behavior Analysis, 32*(1), 63–82. https://doi.org/d5n4ms

Harris, P., Taylor, R., Thielke, R., Payne, J., Gonzalez, J., & Conde, G. (2009). Research Electronic Data Capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Heath, H., & Smith, R. G. (2019). Precursor behavior and functional analysis: A brief review. *Journal of Applied Behavior Analysis, 52*(3), 804–810. https://doi.org/10.1002/jaba.571

Horner, R. H., Sugai, G., & Fixsen, D. L. (2017). Implementing effective educational practices at scales of social importance. *Clinical Child and Family Psychology Review, 20*, 25–35.

https://doi.org/10.1007/s10567-017-0224-7

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., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice*, *1*(1), 3–9. https://doi.org/f8b6

Katsiyannis, A., Conroy, M., & Zhang, D. (2008). District-level administrators' perspectives on the implementation of functional behavior assessment in schools. *Behavioral Disorders, 34*(1), 14–26. https://doi.org/10.1177/019874290803400102

Lambert, J. M., Staubitz, J. E., Torelli, J., Houchins-Juarez, N. J., Juarez, A. P., Sanders, K. B., & Warren, Z. E. (2017). Outcome summaries of latency-based functional analyses conducted in hospital inpatient units. *Journal of Applied Behavior Analysis, 50*(3), 487–494. https://doi.org/10.1002/jaba.399

Lloyd, B. P., Randall, K. R., Weaver, E. S., Staubitz, J. L., & Parikh, N. (2020). An initial evaluation of a concurrent operant analysis framework to identify reinforcers for work completion. *Behavioral Disorders, 45*(2), 85–102. https://doi.org/gh2xfx

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

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*(2), 167–195. https://doi.org/f8vj

Lloyd, B. P., Weaver, E. S., & Staubitz, J. L. (2016). A review of functional analysis methods conducted in public school classroom settings. *Journal of Behavioral Education*, *25*(3), 324–356. https://doi.org/gh868v

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*(4), 817–829. https://doi.org/10.1002/jaba.256

O'Neill, R. E., Bundock, K., Kladis, K., & Hawken, L. S. (2015). Acceptability of functional behavior assessment procedures to special educators and school psychologists. *Behavioral Disorders, 41*(1), 51–66. https://doi.org/10.17988/0198-7429-41.1.51

Quigley, S. P., Peterson, S. M., Frieder, J. E., Brower-Breitwieser, C., Mullins, S., Ivers, H., & Acker, N. (2013). Example of a function-based choice assessment for a student with "passive" problem behavior. *Journal of Evidence-Based Practices for Schools, 14*(1), 81–102.

Romani, P. W., Alcorn, A. S., Miller, J. R., & Clark, G. (2017). Preference assessment for dimensions of reinforcement to inform token economies targeting problem behavior. *Journal of Behavioral Education, 26*, 221–237. https://doi.org/gch3rg

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/gdtkzn

Sigafoos, J., & Saggers, E. (1995). A discrete-trial approach to the functional analysis of aggressive behaviour in two boys with autism. *Australia and New Zealand Journal of Developmental Disabilities*, *20*(4), 287–297. https://doi.org/fvnqrb

Stichter, J. P., & Conroy, M. A. (2005). Using structural analysis in natural settings: A responsive functional assessment strategy. *Journal of Behavioral Education, 14*, 19–34.

https://doi.org/10.1007/s10864-005-0959-y

Stichter, J. P., Hudson, S., & Sasso, G. M. (2005). The use of structural analysis to identify setting events in applied settings for students with emotional/behavioral disorders. *Behavioral Disorders, 30*, 403–420. https://doi.org/10.1177/019874290503000410

Strickland-Cohen, M. K., & Horner, R. H. (2015). Typical school personnel developing and implementing basic behavior support pans. *Journal of Positive Behavior Interventions, 17*(2), 83–94. https://doi.org/10.1177/1098300714554714

Tennessee Department of Education (2020). *Tennessee Department of Education Report Card (2020-2021)*. https://reportcard.tnedu.gov

Umbreit, J., Lane, K. L., & Dejud, C. (2004). Improving classroom behavior by modifying task difficulty. *Journal of Positive Behavior Interventions, 6*, 13–20. https://doi.org/dfnpdc

Vaughn, B. J., & Horner, R. H. (1997). Identifying instructional tasks that occasion problem behaviors and assessing the effects of student versus teacher choice among these tasks. *Journal of Applied Behavior Analysis, 30*, 299–312. https://doi.org/czmcbt

Wacker, D., Berg, W., Harding, J., & Cooper-Brown, L. (2004). Use of brief experimental analyses in outpatient clinic and home settings. *Journal of Behavioral Education, 13*, 213–226. https://doi.org/10.1023/B:JOBE.0000044732.42711.f5

Wallace, M. D., & Iwata, B. A. (1999). Effects of session duration on functional analysis outcomes. *Journal of Applied Behavior Analysis, 32*, 175–183. https://doi.org/bhcc7m

Walker, V. L., Chung, Y.-C., & Bonnet, L. K. (2018). Function-based intervention in inclusive school settings: A meta-analysis. *Journal of Positive Behavior Interventions, 20*(4), 203–216. https://doi.org/10.1177/1098300717718350

U.S. Department of Education, Office of Special Education and Rehabilitation Services. (2021).

Supporting child and student social, emotional, behavioral, and mental health needs. Washington, DC. https://www.pbis.org/resource/supporting-child-and-student-social-emotional-behavioral-and-mental-health-needs

Table 1

Summary of Target Behaviors and Hypothesis Testing Characteristics by Student Team

Student Team	Primary Behavior(s) of Concern	Precursors (FA/AA only)	HT Strategy	Variation	Setting	Implementer(s)	Total Sessions (Duration)	Total TBs
01	Physical aggression (peer)	Clenching fists; sudden movement toward specific peer	FA	Trial	SE Classroom (peers present)	SE teacher Behavior specialist	6 (18 min)	4
			FA	Latency	SE Classroom (peers present)	SE teacher Behavior specialist	6 (16 min)	3
02	Verbal and physical aggression	Head down on desk; verbal refusals; leaving assigned area; forcefully pushing away from desk	FA	Latency	SE Classroom (peers absent)	SE teacher Paraeducator	9 (21 min)	3
03	Physical aggression (peers + adults)	n/a	COA	Function	GE Classroom (peers absent)	GE teacher Research staff	5 (13 min)	0
04	Persistent noncompliance/ off-task	n/a	COA	Instruction	GE Classroom (peers absent)	GE teacher Research staff	9 (25 min)	0
05	Verbal and physical aggression	Forceful nose blowing	FA	Latency	SE Classroom (peers present)	SE teacher Paraeducator	9 (27 min)	6
06	Self-injury	Grunting, screaming, verbal threats of self-harm, forceful contact between clenched fists and hard surfaces	FA	Latency	GE Classroom (peers present)	GE teacher Behavior specialist Research staff	6 (30 min)	0
			COA	Function	GE Classroom (peers absent)	GE teacher Behavior specialist	6 (18 min)	0
07	Physical aggression (adults)	Leaving assigned area; stomping feet; pounding fists on hard surface	FA	Latency	SE Classroom (peers present)	GE teacher	9 (44 min)	1
			FA	Latency	SE Classroom (peers absent)	Paraeducator	6 (22 min)	2
08	Major disruptive behavior	Whining; verbal refusals; putting head inside shirt	AA	Latency	GE Classroom (peers absent)	GE teacher	8 (16 min)	6
			FA	Latency	Multipurpose room (peers absent)	Research staff Behavior specialist	7 (26 min)	3
09	Persistent noncompliance/ off-task	n/a	COA	Instruction	Multipurpose room (peers absent)	Reading teacher Research staff	13 (38 min)	0

10	Elopement from classroom	Leaving assigned area	FA	Latency	GE classroom (peers present)	Behavior specialist GE teacher	9 (41 min)	2
11	Major disruptive behavior	Verbal refusals; hiding under table; rolling on floor	FA	Latency	GE/SE classrooms (peers present; absent)	SE teacher	11 (32 min)	7
12	Physical aggression (adults)	Verbal refusals; swiping materials off table; forcefully banging objects	FA	Latency	GE classroom (peers present)	GE teacher	6 (23 min)	3

Note. HT = hypothesis testing; FA = functional analysis; AA = antecedent analysis; COA = concurrent operant analysis; SE = special education; GE = general education; TBs = Target (i.e., primary behaviors of concern or precursor) behaviors. Precursor behaviors are labeled n/a for teams that only completed a concurrent operant analysis, as precursors were only identified and targeted for functional and antecedent analysis.

Table 2

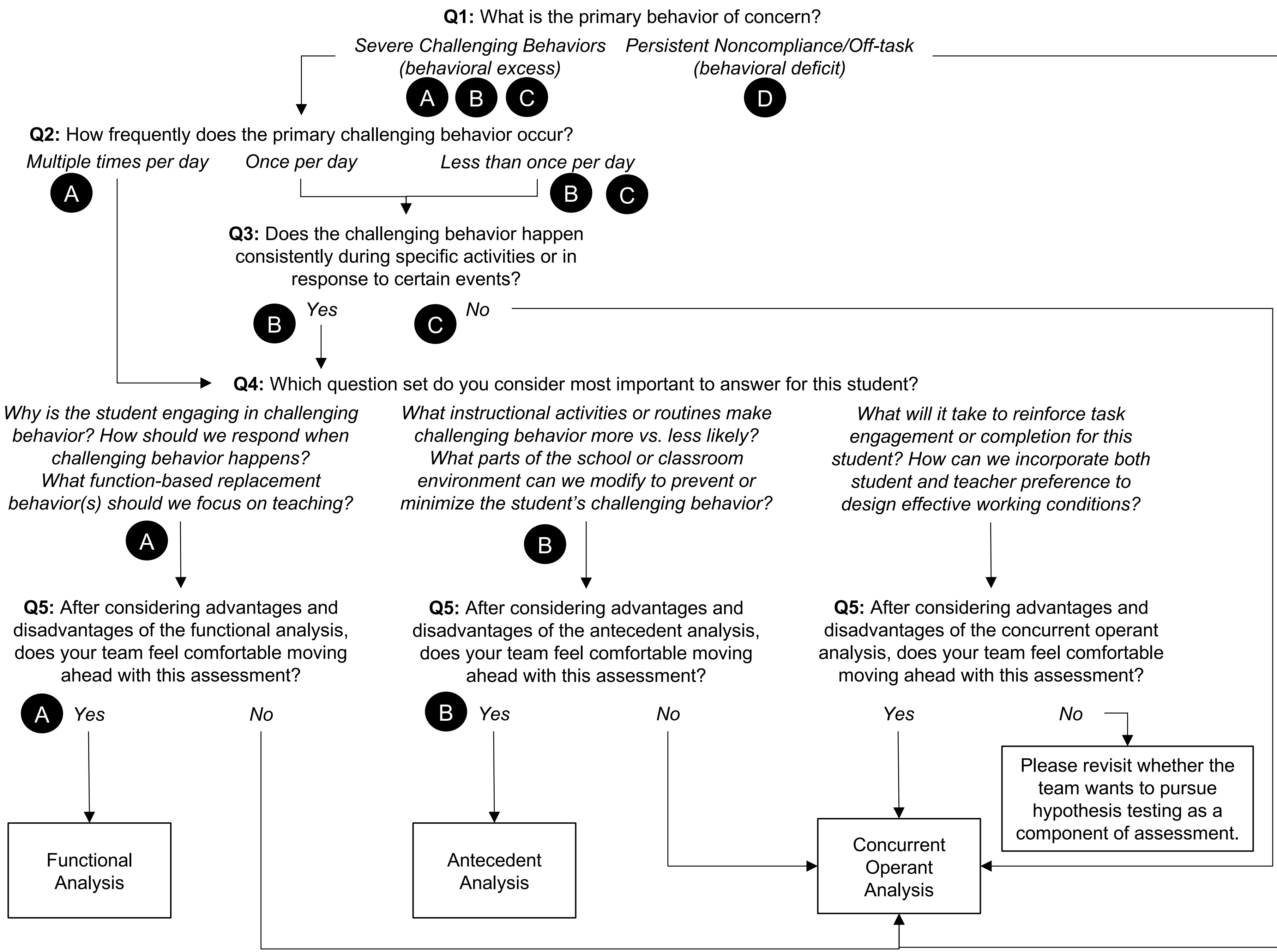
Social Validity Rating Means and Ranges by Item and Professional Role

	Mean (Range)		
Social Validity Item *Likert-type Scores from 1 (Strongly Disagree) to 5 (Strongly Agree)*	All ratings (n = 21)	Behavior Specialist (n = 12)	Classroom Teacher (n = 9)
Overall, the individualized assessment process was feasible to complete given the supports that were available (research team, collaborating school staff).	4.81 (3–5)	5.00	4.56 (3–5)
Overall, the individualized assessment process was acceptable (i.e., socially and ethically appropriate) for the participating student.	4.67 (3–5)	4.75 (4–5)	4.56 (3–5)
During the assessment, I felt that safety was maintained for the student(s) in the room.	4.57 (2–5)	4.75 (4–5)	4.33 (2–5)
Everyone on the team (teacher, behavior support specialist, research staff) played a meaningful role in the assessment process.	4.57 (2–5)	4.50 (2–5)	4.67 (4–5)
During the assessment, I felt that safety was maintained for the adults in the room.	4.50 (1–5)	4.75 (4–5)	4.22 (1–5)
We didn't have access to an appropriate space to complete the assessment.*	4.48 (3–5)	4.42 (3–5)	4.56 (4–5)
The assessment we completed was a good fit for the participating student.	4.38 (2–5)	4.67 (4–5)	4.00 (2–5)
The assessment took too long to complete.*	4.33 (3–5)	4.58 (4–5)	4.00 (3–5)
Results of the assessment were difficult to interpret.	4.29 (3–5)	4.42 (3–5)	4.11 (3–5)
The assessment procedures were too difficult to implement correctly.*	4.14 (4–5)	4.17 (4–5)	4.11 (4–5)
I feel confident assessment results will lead to an effective behavior support plan for this student.	4.05 (3–5)	4.08 (3–5)	4.00 (3–5)
Overall, results of the individualized assessment were useful in informing a behavior support plan for the participating student.	3.95 (2–5)	4.08 (2–5)	3.78 (2–5)
The student's problem behavior worsened in the days or weeks following the assessment.*	3.19 (1–5)	3.42 (2–5)	2.89 (1–5)
It would be possible to complete a similar assessment without the support of a research team.	2.95 (1–4)	2.83 (1–4)	3.11 (2–4)
I don't have enough support from other school team members to use an assessment like this without the additional support of a research team.*	2.81 (1–5)	3.25 (2–5)	2.22 (1–4)
I don't have the necessary training/expertise to use an assessment without support of a research team.*	2.61 (1–4)	2.78 (1–4)	2.44 (1–4)

Note. Asterisks indicate reverse-scored items. Items are listed in order of highest to lowest mean scores for total practitioner ratings. n reflects number of ratings across individualized assessment cases (the number of behavior specialist ratings exceeds the total number of participating behavior specialists, as a subset of behavior specialists participated on more than one student team).

Figure 1

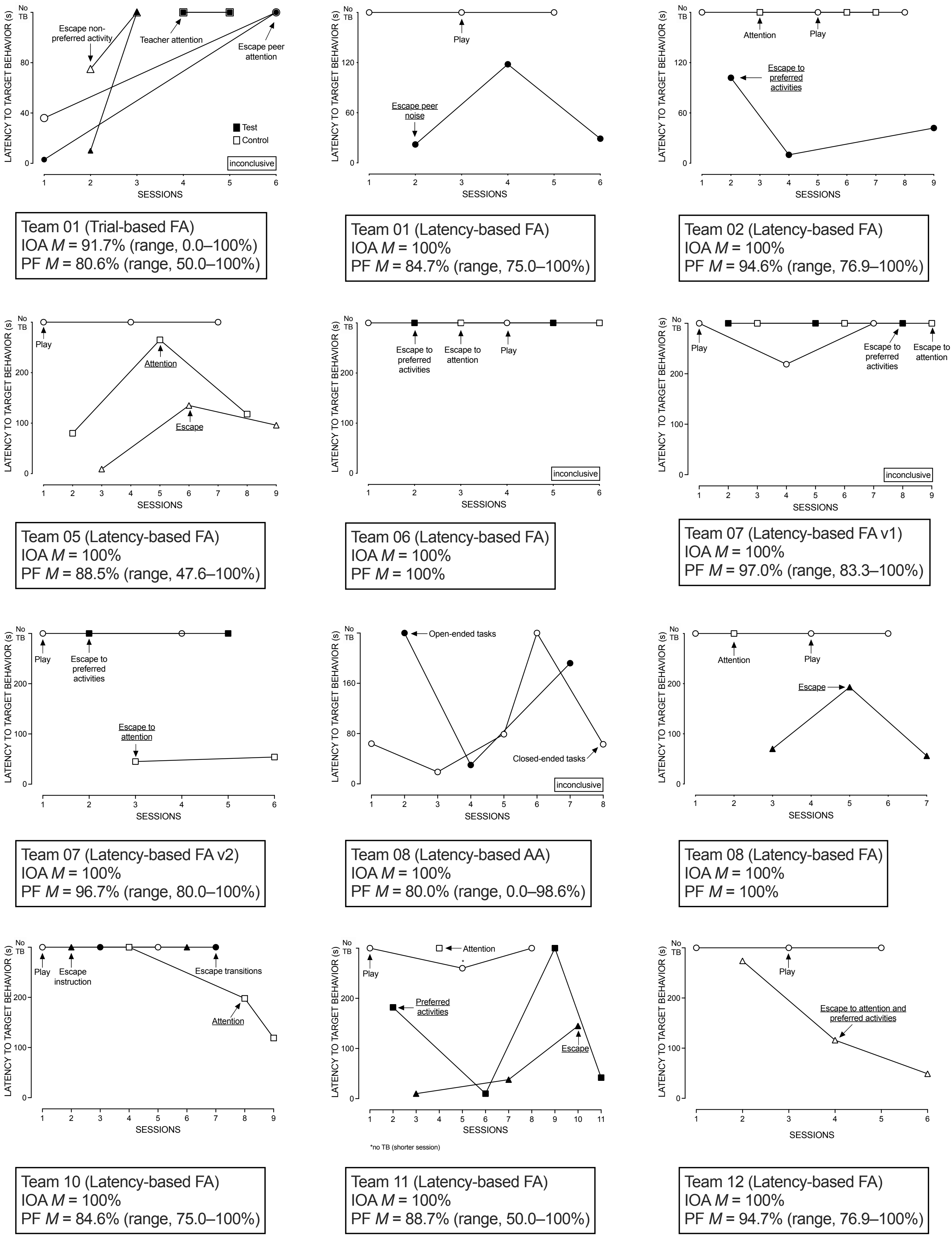
Summary of Branching Logic for Hypothesis Testing Decision Tool



Note. Path A represents the decision factors for 8 teams (Teams 01, 02, 05, 06, 07, 10, 11 12); path B for 1 team (Team 08); path C for 1 team (Team 03); path D for 2 teams (Teams 04, 09).

Figure 2

Graphed Results of Functional and Antecedent Analyses



Note. TB = target behavior; FA = functional analysis; AA = antecedent analysis; IOA = interobserver agreement; PF = procedural fidelity. Underlined test condition(s) were confirmed during the assessment.

Supplementary Material: Concurrent Operant Analysis (COA) Outcome Summaries

Function-based COAs

Team 03

Session	Choice A (% time selected)	Choice B (% time selected)
1	Demand without attention (100%)	Free play with attention and preferred items (0%)
2	Demand without attention (27%)	Free play with attention and preferred items (73%)
3	Directed play with preferred items (100%)	Free play with preferred items and without attention (0%)
4	Demand with attention (100%)	Free play with preferred items and without attention (0%)
5	Free play with attention and low-preferred items (100%)	Free play with preferred items and without attention (0%)

Note: Gray shading indicates the preferred condition, defined as the condition in which the student spent the majority of session time. Percentages of session time allocated to each condition are indicated in parentheses. The condition types and sequences followed a model outlined by Casey (2001) and Lloyd et al. (2020).

Inter-observer agreement: *M* = 99.5% (range, 99.2–100%)
Procedural fidelity: *M* = 95.0% (range, 80.0–100%)

<u>Interpretation</u>: With the exception of the first session, this student consistently allocated her time to conditions involving adult attention, even when that attention came in the form of directions on how to play (Session 3) or instructional demands (Session 4), and even without preferred items (Session 5). These results suggest *adult attention* (relative to access to preferred items and escape from instructions) as the most likely reinforcer for appropriate behavior.

Team 06

Session	Choice A (% time selected)	Choice B (% time selected)
1	Demand without attention (0%)	Free play with attention and preferred items (100%)
2	Free play with attention and preferred items (100%)	Demand without attention (0%)
3	Directed play with preferred items (100%)	Free play with preferred items and without attention (0%)
4	Demand with attention (0%)	Free play with preferred items and without attention (100%)
5	Directed play with preferred items (100%)	Alone (0%)
6	Free play with attention and low-preferred items (0%)	Free play with preferred items and without attention (100%)

Note: Gray shading indicates the preferred condition, defined as the condition in which the student spent the majority of session time. Percentages of session time allocated to each condition are indicated in parentheses. The condition types and sequences followed a model outlined by Casey (2001) and Lloyd et al. (2020).

Inter-observer agreement: *M* = 97.2% (range, 92.7–99.4%)
Procedural fidelity: *M* = 97.3% (range, 92.0–100%)

Interpretation: The student consistently allocated his time to conditions involving access to preferred items, even when the adult directed the student on how to play with these items (Sessions 3 and 5) and even in the absence of adult attention (Sessions 4 and 6). These results suggest *access to preferred items* (relative to access to adult attention or escape from instructions) as the most likely reinforcer for appropriate behavior.

Instruction-based COAs

Team 04

Session	Choice A (% time selected)	Choice B (% time selected)	Choice C (%)
1	Adult attention (no preferred items; 100%)	Free play with preferred items (no attention; 0%)	N/A
2	Break alone (no attention or materials; 0%)	MP task to earn preferred activity (100%)	N/A
3	LP task with adult attention (100%)	MP task to earn preferred activity (0%)	Break alone (0%)
4	LP task with adult attention (100%)	MP task to earn preferred activity (0%)	Break alone (0%)
5	1 min LP task to earn 1 min attention (0%)	3 MP tasks to earn 1 min preferred activity (100%)	Break alone (0%)
6	1 min LP task to earn 1 min attention (0%)	8 MP tasks to earn 2 min preferred activity (100%)	Break alone (0%)
7	1 min LP task to earn 1 min attention (56.8%)	4 hard MP tasks to earn 2 min preferred activity (43.2%)	Break alone (0%)
8	1 min LP task to earn 1 min attention (0%)	1 hard MP task to earn 2 min preferred activity (100%)	Break alone (0%)
9	1 min LP task to earn 1 min attention (0%)	2 hard MP tasks to earn 2 min preferred activity (100%)	Break alone (0%)

Note: Gray shading indicates the preferred condition, defined as the condition in which the student spent the majority of session time. Percentages of session time allocated to each condition are indicated in parentheses. LP = less preferred; MP = more preferred. The condition types and sequences were determined individually for the student, and with input from participating teachers.

Inter-observer agreement: *M* = 97.4% (range, 94.4–99.6%)
Procedural fidelity: *M* = 100%

Interpretation: Across Sessions 1–4, the student consistently allocated his time to conditions involving adult attention. He avoided choosing breaks from work when no attention was available; and he chose to work on a less-preferred task (over a more-preferred task) when attention was also available. These initial results suggested *adult attention* as the most likely reinforcer for appropriate behavior. Based on input from the teacher, the next series of sessions was designed to determine whether and how to shift the student's preference from working to earn attention to working to earn preferred activities (a consequence that was more feasible for the teacher to deliver on a rich schedule in the general education classroom). Results of Sessions

5–9 suggested the magnitude (i.e., duration) of earned reinforcement, as well as response effort (i.e., task type and difficulty) influenced his preference. Additionally, results suggested the student consistently chose to complete work (over taking a break alone) when individualized reinforcers were programmed for work completion. Conditions in which the student chose to work to earn preferred activities were shared with the teacher as potential starting points for expectations and schedules of reinforcement that would motivate engagement and work completion in the general education setting.

Team 09

Session	Choice A (% time selected)	Choice B (% time selected)	Choice C (%)
1	LP instructional task (0%)	MP instructional task (100%	Break alone (0%)
2	LP tasks to earn break alone (0%)	LP tasks to earn adult attention (100%)	Break alone (0%)
3	LP tasks to earn preferred activity (100%)	LP tasks to earn adult attention (0%)	Break alone (0%)
4	LP tasks (fixed number) to earn preferred activity (100%)	MP tasks (fixed time) to earn preferred activity (0%)	Break alone (0%)
5	LP tasks (fixed number) to earn preferred activity (0%)	MP tasks (fixed time) to earn preferred activity (100%)	Break alone (0%)
6	LP tasks (fixed number) with adult help to earn preferred activity (0%)	MP tasks (fixed time) to earn preferred activity (100%)	Break alone (0%)
7	LP task with lower work requirement (100%)	MP task with higher work requirement (0%)	Break alone (0%)
8	LP task with increased difficulty, lower work requirement (100%)	MP task with higher work requirement (0%)	Break alone (0%)
9	LP task for fixed amount of time (100%)	LP task for fixed amount of work (0%)	Break alone (0%)
10	LP task for fixed amount of work (100%)	LP task for fixed amount of (increased) time (0%)	Break alone (0%)
11	LP task (higher work requirement) to earn preferred activity (0%)	LP task (lower work requirement) to earn adult attention (0%)	Break alone (100%)

Note: Gray shading indicates the preferred condition, defined as the condition in which the student spent the majority of session time. Percentages of session time allocated to each condition are indicated in parentheses. LP = less preferred; MP = more preferred. The condition types and sequences were determined individually for the student, and with input from participating teachers.

Inter-observer agreement: *M* = 96.0% (range, 90.7–99.4%)
Procedural fidelity: *M* = 99.0% (range, 91.0–100%)

<u>Interpretation</u>: Session 1 served as a confirmation of task preference (less vs. more preferred). Response patterns in Sessions 2 and 3 indicated the student was willing to complete a small amount of a low-preferred task for adult attention, but preferred to work to earn preferred activities. Response patterns in Sessions 4–6 suggested the student preferred to work on the more-preferred task for a fixed amount of time, relative to the less-preferred task for a fixed number of tasks, to earn preferred activities. Response patterns in Sessions 7–8 suggested the student's preference shifted to completing the less-preferred task when the work requirement was

lower than for the more-preferred task, even when the difficulty of the less-preferred task increased. Response patterns in Sessions 9–10 indicated the student preferred working on the low-preferred task for a fixed amount of time (rather than a fixed number of tasks), though when the time was increased, his preference shifted to the fixed number of tasks. Session 11 was designed to assess whether the student would choose a higher work requirement to earn the most preferred consequence (over a lower work requirement to earn a less-preferred consequence); during this session, the student chose to escape work to take a break on his own—indicating neither instructional condition was preferred. Overall, results suggested access to preferred activities as the most likely reinforcer for appropriate behavior; and that student preference and work completion was impacted by task type, response effort, and work requirement (time-based vs. response-based). Additionally, results suggested that with only one exception, the student consistently chose to complete work (over taking a break alone) when individualized reinforcers were programmed for work completion.

References

Casey, S. D. (2001). *Comparing functional outcomes between brief functional analyses and concurrent operants assessments* (Publication No. 3050783) [Doctoral dissertation, University of Iowa]. ProQuest Dissertations and Theses Global.

Lloyd, B. P., Randall, K. R., Weaver, E. S., Staubitz, J. L., & Parikh, N. (2020). An initial evaluation of a concurrent operant analysis framework to identify reinforcers for work completion. *Behavioral Disorders, 45*(2), 85–102. https://doi.org/gh2xfx