INCORPORATING DATA TRIANGULATION TO PROMOTE GENERALIZABLE OUTCOMES FOR FUNCTION-BASED TREATMENT OF SEVERE CHALLENGING BEHAVIOR

By

Bailey Andressen Copeland

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

May 10, 2024

Nashville, Tennessee

Approved:

Joseph M. Lambert, Ph.D., BCBA-D

Elizabeth E. Biggs, Ph.D.

Tara A. Fahmie, PD.D., BCBA-D

Jennifer R. Ledford, Ph.D., BCBA-D

Johanna L. Staubitz, Ph.D., BCBA-D

To Dad for teaching me how to be brave.
To Mom for loving me even when I'm not.

ACKNOWLEDGEMENTS

I want to thank several people for supporting me throughout this process. I first and foremost need to thank my dedicated research team- Taylor Anderson, Dakota Kanae-Harris, Alysa Longo, and Caitlin Price. I would have never survived the last year without your tireless effort and unwavering support. I could not have dreamed up a better group of teammates. I need to thank Janey Macdonald and Dr. Jessica Paranczak; I have learned more from the two of you than I could ever express in this brief paragraph. I want to thank the members of my committee including Drs. Elizabeth Biggs, Tara Fahmie, Jennifer Ledford, and Johanna Staubitz for their mentorship and guidance. I want to thank my family and friends for providing unwavering support and sanity outside of this project. Mimi, Poppa, Jess, Lucy, Will, Charlie, and Sasa- I never would have made it without you. Dr. Lambert, thank you for being my advisor, mentor, and friend. What will I do without you?

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Children with intellectual and developmental disabilities (IDD) are at increased risk for developing patterns of challenging behavior (Kurtz et al., 2020; McClintock et al., 2003). Estimates vary widely, but approximately 50% of children with IDD develop problem behaviors above and beyond that of their typically developing peers (Farmer et al., 2011). Left untreated, these patterns of behavior may persist or worsen (Dunlap & Fox, 2011).

A smaller group of children (i.e., 5 - 10%) develop severe behaviors with extreme consequences (Dekker et al., 2002; Emerson et al., 2001; Sturmey et al., 2008). When this happens, the challenging behaviors pose a threat not only to the child but to those around them (e.g., risk of injury, disfigurement, or death; Kurtz et al., 2020). Perhaps unsurprisingly, these behaviors have been linked with impaired social relationships (Fitzpatrick et al., 2016), increased risk of abuse (Stith, et al., 2009), and restricted community access (Kanne & Mazurek, 2011).

Function-Based Assessment

Fortunately, individualized, function-based treatments can decrease dangerous behavior and simultaneously increase socially appropriate behavior (Ghaemmaghami et al., 2021; Hurl et al., 2016; Lambert, Copeland et al., 2022). This model of service delivery relies heavily on functional behavior assessment to identify the variables that occasion and maintain challenging behavior. By conducting a functional analysis (FA; Iwata et al., 1994), behavior analysts can directly examine how problem behavior changes as environmental antecedents and consequences are systematically altered.

For example, if a child's problem behavior is hypothesized to be maintained by access to preferred items, the therapist may set up a scenario (i.e., a control condition) in which the child has unrestricted access to preferred items. No problem behavior would be expected to occur in this condition. In the test condition, the therapist would remove the preferred items, returning them contingent on challenging behavior. If challenging behavior is repeatedly observed in the test condition (i.e., above and beyond levels documented during control sessions), the team can confirm their hypothesis. Importantly, designing effective assessment and treatment programs requires fluent application of basic behavioral principles paired with real-time reactions to data (e.g., Hagopian et al., 2013; Lambert, Copeland et al., 2022).

Function-informed Mechanism-based (FIMB) Treatment

As highlighted in the APA Task Force on Evidence-Based Practice (2006), achieving desirable outcomes requires researchers to engage in value-informed, data-based reactions to all accumulated sources of evidence. In response, the field of behavior analysis has begun to embrace a new view of evidence-based practice that prioritizes judgment as a central component

of every decision-making process (Contreras et al., 2021; Slocum et al., 2014). Lambert, Copeland, et al. (2022) described one such approach embracing this flexible view of evidence-based practices.

Specifically, Lambert, Copeland, et al. (2022) evaluated the effectiveness of the FIMB framework in guiding the design of all assessment and treatment procedures used in a university-based outpatient clinic across a 6-year time span. All assessment and treatment decisions were made by taking stock of the natural processes operating on the client's behavior and reorganizing the circumstances under which they occurred. As such, decisions were informed by the anticipated impact of known operant-learning mechanisms (e.g., reinforcement, punishment, extinction). Standard goals of the FIMB framework included (a) confirming or ruling out functions for challenging behavior, (b) reducing challenging behavior in historically evocative contexts, (c) establishing a tolerance for these evocative contexts when functional reinforcers were unavailable, and (d) generalizing effects across implementers and settings.

To standardize assessments of treatment efficacy across cases, Lambert, Copeland, et al. (2022) compared each client's performance during treatment against their performance during baseline and calculated a mean baseline reduction (cf. Greer et al., 2016; Rooker et al., 2013; Saini et al., 2019). Any reduction at or above 90% corresponded to a highly effective treatment. Reductions between 70-89% corresponded to effective treatments, and reductions from 50-69% corresponded to marginally effective treatments. Any reduction that fell below 50% corresponded to a treatment with equivocal results.

Overall, outcomes were favorable, suggesting that this formative process translated into meaningful reductions in challenging behavior. Specifically, 79% of cases analyzed (i.e., 42 of 53) had effective or highly effective reductions in behavior from baseline to the first phase of

treatment (i.e., response elimination). Further, 72% of cases (i.e., 38 of 53) maintained or increased their effectiveness score when moving to the next phase of treatment, which involved tolerating periods of delay to functional reinforcers.

Contextualizing Treatment Gains

Despite these promising clinical outcomes, some researchers (e.g., Lambert, Sandstrom, et al., 2022; Schwartz & Kelly, 2021) have begun to question whether gains documented through treatment plans will translate to meaningful improvements in life skills (e.g., choice-making or friendship-building). In their 2022 study, Lambert, Sandstrom, et al. contacted 53 stakeholders (e.g., parents, guardians, teachers) who originally referred the participants served through the FIMB framework as detailed in Lambert, Copeland et al. (2022). A total of 29 stakeholders responded and completed the survey.

Lambert, Sandstrom, et al. (2022) analyzed survey results to explore the extent to which treatment effects (described in Lambert, Copeland et al., 2022) promoted desirable, distal changes in behavior. As part of the survey, respondents completed the Problem Behavior subscale of the *Scales of Independent Behavior – Revised* (SIB-R; Bruininks et al., 1996). Researchers conducted a paired-sample *t*-test and found a statistically significant decrease in the reported prevalence (t [28] = 5.44, $p < .001$) and intensity (t [28] = 5.03, $p < .001$) of challenging behavior from pre- to post-service delivery.

However, approximately half of respondents (51.7%) reported a return of challenging behavior after discharge, and there was no statistically significant difference identified in participant access to inclusive environments before and after treatment (t [28] = -0.60, ns). As Lambert, Sandstrom, et al. discussed their own conflicting results, they emphasized the critical

need to integrate comprehensive assessment of social validity into behavior analytic treatment models. Interestingly, this issue was highlighted in the literature over 30 years ago when Kennedy conducted a systematic review in 1992 and found that only 20% of articles published in *Journal of Applied Behavior Analysis (JABA)* and *Behavior Modification* reported findings related to social validity. When a similar review was conducted over two decades later, Ferguson et al. (2018) found that only 15% of articles published in JABA reported on social validity. Taken together, these reviews suggest a persistent underrepresentation of social validity reports in the behavior analytic research base.

This gap in our literature base becomes even more significant when considering the foundational characteristics of the field. In the seminal article outlining the dimensions of applied behavior analysis, Baer et al. (1968) defined the applied dimension of ABA as "the interest society shows in the problem behavior being studied" (p. 92). Then, in 1978 Wolf formally introduced the issue of social validity to the field, defining it as the use of evaluative feedback from recipients or consumers of ABA services to guide program planning and evaluation. Schwartz and Baer (1991) expanded on this definition by suggesting that social validity assessment should (a) be a standard part of behavior analytic practice and (b) be administered throughout the treatment evaluation (i.e., rather than once at the conclusion of services).

Quality of Life

As we work to create socially valid treatment models, it may be helpful to consider the existing research on quality of life (QoL) as a framework for evaluating whether we are in fact achieving socially significant outcomes following behavior-analytic treatment. QoL is not widely measured in the behavior analytic research base; however, Schwartz and Kelly (2021) argue that

an enhanced QoL is the defining, critical outcome to consider for individuals receiving behavior analytic services. Looking outside of our own field, existing research supports use of the QoL measure to evaluate program-level outcomes and satisfaction with services.

Discussion of QoL first emerged in the IDD literature in the 1980s with hundreds of proposed definitions and models of QoL (Lyons, 2011). Across all proposed models and definitions, researchers largely agreed that the conceptualization should not change based on disability status (Turnbull & Brunk, 1990). Thus, it is perhaps fitting to consider the definition proposed by the World Health Organization (2012), in which QoL refers to an individual's perception of their position within the cultural context of their community. This perception must be considered in relation to the individual's goals, expectations, values, and standards.

Family Quality of Life

Knowing that disability impacts the whole family (Turnbull et al., 2007), researchers have started considering the experiences of the larger family unit when assessing quality of life (Lyons, 2011). The family unit can be viewed as a dynamic, interconnected, and self-regulating system (Turnbull et al., 1988) within which everyone is influenced by both shared and individual contextual factors (Gardiner & Iarocci, 2012). This creates a complicated construct for measurement that is perhaps best explained by the theory of change posed by Zuna et al. (2011). Family quality of life (FQoL) is positioned as a dynamic measure of the family's well-being (i.e., defined collectively by its members) as a reflection of the interaction of family- and individual-level needs (Zuna et al., 2011).

Leaders in the field of disability have called for assessments of FQoL to evaluate disability-related policies and services (Dunst & Brudger, 2002). One such measure, the Beach Center Family Quality of Life Scale (Hoffman et al., 2006) was developed following extensive

qualitative investigation and three national field tests of family satisfaction levels with various aspects of FQoL (Park et al., 2003; Summers et al., 2005; Hoffman et al., 2006). Additionally, this scale was designed as an outcome measure that would be useful to policymakers, service providers, and families looking to evaluate program quality (Summers et al., 2005).

Data Triangulation

Despite their utility, formal measures of QoL are not often included in behavior analytic research. This is perhaps because of the field's reliance on rigorous, highly controlled single-case logic. Because we are not able to repeatedly administer the FQoL it cannot be used as a primary outcome measure for a single case design. However, the variables that control responding in the context of a highly structured research or treatment context may not be present in less controlled spaces (Ledford et al., 2016). This creates a problem for evaluating the extent to which our outcomes generalize outside the context of the original treatment setting. Specifically, there is a need to contextualize the more molar outcomes of improvement that cannot be assessed in an ongoing manner (e.g., FQoL scores).

Formative triangulation may offer a solution. Formative data triangulation refers to a research method used in research and treatment evaluation designed to enhance the validity and reliability of findings by integrating and comparing data from multiple and varied sources of data at the formative stage of treatment (Bekhet et al., 2012). By engaging in formative triangulation, researchers can identify points of divergence and convergence across data sources as they gain a more dynamic understanding of the dependent variable. Lambert et al. (2024) used formative triangulation by combining quantitative and qualitative data sources to adjust within-session treatment procedures based on parent reports of out-of-session behavior. In this case study,

researchers documented generalized behavior change for a child with severe challenging behavior. However, the researchers did not control these effects.

Rationale

The current project was designed to extend the work of Lambert et al. (2024) across two studies. In the first study, researchers used a changing criterion design to evaluate the effectiveness of the FIMB framework to increase levels of prosocial replacement behaviors and decrease levels of challenging behavior for three children (i.e., a single sibling set) with IDD and a history of severe challenging behavior. During this study, researchers sought to answer the following research question:

1. Does the FIMB intervention (a) increase latencies to challenging behavior and (b) decrease latencies in prosocial behaviors observed in formal appointments for three participants with IDD who engage in severe challenging behavior?

During the second study, researchers analyzed caregiver reports of the children's behavior outside the highly controlled setting in which the behavior analytic treatment was occurring. This dynamic view of participant performance informed treatment decisions made across both phases of the study. By expanding the number and type of data sources considered, researchers sought to answer the following questions:

2. Does the FIMB+ intervention decrease levels of caregiver-reported challenging behavior for three participants with IDD outside of formal sessions?
3. Does the family's quality of life improve as the children's challenging behavior improves in the unsupervised context?

CHAPTER II

GENERAL METHOD

Recruitment

With approval from Vanderbilt University's Institutional Review Board, researchers recruited one caregiver and three school-aged children. Participants were recruited from a bank of referrals submitted to a university-based clinic as described in Lambert, Copeland, et al. (2022). To recruit participants, the first author shared information about the study to all community providers (i.e., 3 behavior analysts and 5 principals) who had submitted pending referrals pending assignment. Two providers (i.e., both behavior analysts) replied, indicating a likely fit between the study aims, inclusion criteria, and a family referred to the clinic.

The first author then contacted the parents or legal guardians listed on the referral (either via phone or email) to describe the study and answer questions about participation. The research team obtained informed consent from the caregiver of the first family to respond and proceeded to the intake process. Child assent was obtained during the first intake meeting and evaluated continuously. The second family was thanked for their time and returned to the wait list for the university-based clinic.

Caregiver Inclusion Criteria

To be eligible to participate in the study, designated caregivers were required to (a) consent to participate in the study and (b) spend at least 1 hr per day supervising participating children. Caregivers were excluded if they did not directly supervise the children or were not willing to modify their schedule to do so for both baseline and treatment conditions. Caregiver interest and availability for participation were obtained via self-report.

Child Inclusion Criteria

To be eligible to participate in the study, children must (a) have a documented intellectual or developmental disability (IDD), (b) be school-aged at the time they are enrolled in the study, (c) engage in severe and persistent challenging behavior as reported by their parent or guardian, (d) have parental consent to participate, and (e) assent to participate in the study. Notably, any nominated child who did not engage in challenging behavior during intake assessments would have been excluded. Assent was evaluated on an ongoing basis and is described below.

Information on each child's age, diagnoses, and patterns of challenging behavior were obtained through caregiver report. To confirm a documented IDD, researchers conducted a record review of each child's previous psychological evaluations and most recent individualized educational plans. Finally, to confirm patterns of challenging behaviors, researchers completed a series of intake assessments along with a functional analysis (i.e., FA; Iwata et al; 1994).

Participating Family

Kim was a 66-year-old White woman with a graduate degree who was the sole caregiver for six children, all of whom had identified disabilities. She was fluent in both spoken English and American sign language (ASL). Kim had three teenage daughters (i.e., Kyra, Jasmine, and Harmony) who were adopted internationally. Kyra (i.e., age 17) and Harmony (i.e., age 14) were biological siblings and were adopted at ages 6 and 4 respectively. Two years later, Jasmine (i.e., age 15) was adopted at age 7. All three girls were Black.

Both Kyra and Jasmine were diagnosed with profound Deafness and communicated in American sign language (ASL). They also attended a residential school for the deaf, where they boarded for the week, spending weekends and school breaks at home. Harmony was hearing and communicated in both spoken English and ASL. Harmony was diagnosed with autism, post-traumatic stress disorder (PTSD), and a mild intellectual disability. Harmony lived at home and attended her zoned public school, where she received special education services. None of the three older girls engaged in challenging behavior and thus were excluded from participation in the study. However, Kim also had three younger children (i.e., Jerome, Serenity, and Zeke) who were identified as eligible participants.

Jerome

Jerome was a 9-year-old Black boy enrolled in third grade who Kim described as "contagiously enthusiastic" and "willing to try anything twice." Jerome was adopted internationally from Africa when he was 3 years old. At the start of the study, Jerome was enrolled in a special education school for students exhibiting severe and persistent challenging behavior. Jerome had a medical diagnosis of profound Deafness and autism. Kim reported that Jerome frequently (i.e., at least once per hour) engaged in episodes of physical aggression, property destruction, and elopement.

Kim also reported that Jerome had recently experienced a two-year regression in his skills. During the regression, Jerome stopped communicating in ASL, lost interest in all preferred toys and activities, and required support with previously mastered daily skills such as dressing and bathing. Jerome experienced three hospitalizations for his aggression throughout the regression period. He was taking daily medications that were managed by a psychiatrist.

Serenity

Serenity was a 10-year-old child enrolled in fourth grade. Kim described her as a "very bright and compassionate" child who is "always a step ahead of everyone around her." Serenity was initially placed into Kim's care as an infant in the foster care system. She was ultimately adopted into the family at age 4. Serenity is biracial (i.e., White and Indigenous) but often self-identified as Indigenous. When researchers asked Serenity to self-report her gender, she indicated that she is nonbinary and prefers to use she/her pronouns. Serenity attended a private school for children diagnosed with autism. Kim reported that Serenity frequently (i.e., at least daily) engaged in episodes of physical aggression, property destruction, and verbal aggression.

Kim also reported concerns related to frequent (i.e., hourly) statements in which Serenity threatened to take her own life or the lives of her family members. She had been discharged from the care of several psychologists and therapists due to concerns related to her behavior. She was taking daily medications that were managed by a psychiatrist who was in frequent communication with Kim but was not able to meet with Serenity due to safety concerns. Serenity spoke English only and communicated vocally using complete sentences.

Zeke

Zeke was an 8-year-old Black boy enrolled in third grade. Kim described Zeke as an "earnest" and "silly" child with "an immense willingness to please." Zeke was placed into Kim's care at 18 months of age when he entered foster care. He has remained with the family and is currently undergoing the process of adoption. Zeke had a medical diagnosis of attention deficit disorder (ADHD), a visual impairment (i.e., Cortical Visual Impairment and Nystagmus), a traumatic brain injury (TBI), and autism. Zeke attended his zoned public elementary school, where he received special education services under the eligibility category of "other health impairments." Kim reported that Zeke frequently (i.e., daily) engaged in episodes of physical aggression, property destruction, and verbal threats.

She also reported concerns related to toileting, academic progress, and his independence with their daily routines at home (e.g., dressing, bathing, teeth brushing). Zeke spoke English only and communicated vocally using complete sentences. Zeke took daily medication that was managed by a psychiatrist and weekly support from a mental health counselor over telehealth.

STUDY I METHOD

Setting

With the exception of one phase of treatment (i.e., generalization), all appointments occurred in the children's bedrooms. Each child worked separately with a team of therapists (i.e., 2-3), including a supervising BCBA. For safety reasons, therapists initially removed loose items and furniture. During maintenance and generalization phases, all furniture remained in the room.

Jerome's bedroom (10 ft x 6 ft) contained a lofted twin-sized bed, window, swing, and small closet filled with his toys and clothes. Jerome typically chose to work and play on the floor beside his bed. Serenity's bedroom (10 ft x 10 ft) contained a lofted twin sized bed, window, two bookshelves, a Lego® corner, two buckets of toys, and two small closets. Serenity chose to work and play in front of the window. Her room was located directly beside Jerome's on the main floor of the house. Zeke's bedroom (10 ft x 6 ft) was located on the bottom floor of the home. It also contained a lofted twin sized bed, a window, a small table, several drawers of clothes, and two buckets of toys. He chose to work and play on the floor beside his bed.

Jerome's bedroom connected to a time out space (10 ft x 6 ft), which the family used for crisis management for two children (i.e., Serenity and Jerome). The space was kept clear of all objects and furniture. A video camera was mounted to the ceiling to allow for uninterrupted visual monitoring when the room was in use.

The door was controlled by a safety mechanism that could only be activated (i.e., locked) if it was actively held down by a human operator. If the person operating the switch moved their hand or walked away from monitoring the space, the door was automatically unlocked. The switch controlling the safety mechanism was mounted on the wall at a height of approximately 5 ft to prevent the three younger children from accessing the switch. Figure 1 includes a layout of these bedrooms and the time out room.

Figure 1

Diagram of Home Layout

Materials

Observers used handheld computers (e.g., tablets or smartphones) equipped with Countee® to collect data on all dependent variables. Procedural fidelity data sheets were set up on a clipboard, with a writing utensil, before each session began.

Across all appointments, therapists brought a bin for each child containing preferred toys. A full list of included toys and activities presented to each child is included in Appendix A. The therapists also had access to a separate bin containing a set of school supplies for each child (i.e., dry erase markers, dry erase board, glue sticks, crayons, pencils, pens, safety scissors) and academic tasks matched to their instructional level. Additional materials were developed as part of each child's behavior plan. Specifically, therapists used a visual timer during delay tolerance training for Jerome and a brightly colored star (3" x 3") for Serenity and Zeke.

Dependent Variables and Metrics

During each appointment, trained observers collected timed-event data in-vivo, using continuous count coding procedures. All assessment sessions were a fixed duration of 5 min across participants. During the treatment evaluation phase, researchers moved to a trial-based format, in which session duration varied. Each trial ended contingent on the participant (a) meeting pre-specified schedule requirements to access reinforcement and (b) experiencing the programmed dosage of reinforcement. Then, to evaluate a more global impact of the FIMB framework across clients, we evaluated the extent to which case conceptualization facilitated the FIMB framework's four primary objectives (i.e., conceptual clarify of functional relations, response elimination, EO tolerance, and generalization).

Reinforcement

Observers measured time in reinforcement (SR) by activating the SR key when the therapist implemented procedures designed to abolish challenging behavior. When the SR key was activated, coders were presented with three qualifier keys (i.e., attention, tangible, escape). This required the observer to indicate whether the therapist had (a) provided access to preferred items or activities, (b) delivered attention, or (c) provided a break free of task demands. Multiple SR keys could be selected (i.e., if both attention and tangibles were provided, the coder would activate both SR Tan and SR Attn). At the end of each trial, researchers calculated the total number of seconds spent in SR (i.e., both within and across SR keys). The total duration of time in SR was then divided by the total duration of the trial and multiplied by 100 (i.e., resulting in a percentage of time spent in SR for each trial).

Establishing Operations

Observers measured time spent in programmed establishing operations (EO) by activating the EO key when the therapist implemented any procedure designed to evoke challenging behavior. When the EO key was activated, coders were presented with three qualifier keys (i.e., attention, tangible escape). This required the observer to indicate whether the therapist had (a) restricted access to attention, (b) restricted access to preferred items or activities, or (c) began issuing demands.

Multiple EO keys could be selected (e.g., if both attention and tangibles were restricted). EO and SR codes were mutually exclusive. In other words, activating EO automatically deactivated the corresponding SR key and vice versa. Notably, functional reinforcers remained available when the EO key was activated. At the end of each trial, researchers calculated the total number of seconds spent in EO (i.e., both within and across EO keys). The total duration of time in EO was then divided by the total trial duration and multiplied by 100, resulting in a percentage of time spent in EO.

Stimulus-Delta

To capture the time in EO during which functional reinforcers were not available, observers activated a stimulus-delta (S^{Δ}) key. When the S^{Δ} key was activated, coders were presented with three qualifier keys that were identical to the ones described above in the EO section. The app was programmed to prevent the coder from selecting the EO code while the S^{Δ} code was activated. When the therapist removed the S^{Δ} the observer deactivated the S^{Δ} key and immediately activated the EO key. After activating the EO key the observer was free to toggle between SR and EO in response to therapist behavior. Researchers reported time in S^{Δ} per trial to facilitate comparisons in changes in level across trials.

Challenging Behavior

Operational definitions of targeted challenging behavior for each child were developed based on the results of intake assessments. All three children were reported to engage in *tantrum* behaviors that included physical aggression and property destruction. However, Jerome's tantrums were reported to escalate to include disrobing or urinating on the floor. Both Serenity

and Zeke typically began their tantrums by exhibiting verbal aggression (e.g., growling, hissing, yelling, or cursing) before escalating to property destruction or physical aggression. Serenity's tantrums often terminated with an occurrence of elopement. Jerome also engaged in elopement, but this behavior was purported to serve a different function and was treated as a separate operant outside the context of this study. Each participant's operational definitions are described in detail in Appendix C.

During assessment, observers coded each occurrence of tantrum behavior observed within a 5-min session. This was then converted to a rate, which was used as the primary dependent variable to guide decision making during FA. During treatment, observers coded the first instance of challenging behavior following trial onset. To analyze treatment effects, researchers calculated the latency to challenging behavior observed during each trial.

Mands

Observers tracked multiple types of manding (i.e., requesting) for each child. An *independent mand* (I-Mand) was coded when a child independently requested a functional reinforcer in the presence of an EO for a targeted challenging behavior (e.g., asking for a break when a demand was presented). Notably, this request had to occur during an EO period (i.e., not in the presence of a S^{Δ}). A *prompted mand* (P-Mand) was coded when a child requested a functional reinforcer following a therapist prompt. A *delta mand* (S^{Δ} Mand) was coded when an independent mand occurred when the S^{Δ} code was activated (i.e., when the EO was implemented by the therapist, but functional reinforcers were not available). When any of these keys were activated, the observer was required to specify the associated reinforcer by selecting if the child requested attention, tangibles, or a break. Prompted mands and independent mands were

exclusive codes (e.g., if a trial included a P-Mand it could not include an I-Mand). This is because either form resulted in immediate delivery of reinforcement and termination of the trial. However, each trial may include a S^{Δ} Mand in addition to either a P-Mand or I-Mand (i.e., following a S^{Δ} Mand there was a subsequent opportunity for either a P-Mand or I-Mand during time in EO).

During assessment, observers coded each I-Mand that occurred within a 5-min session, allowing researchers to calculate a rate of I-Mands per session. Neither P-Mands nor S^{Δ} Mands were applicable during assessment. During treatment, observers coded (a) the first mand (i.e., either I-Mand or P-Mand) to occur outside of the S^{Δ} period and (b) the first mand to occur within the S^{Δ} period (i.e., S^{Δ} Mand). To analyze treatment effects, researchers calculated the latency to manding (i.e., I-Mands, P-Mands, and S^{Δ} Mands) coded during each trial. Latency to independent manding was used as the primary dependent variable to guide decision-making during attention and tangible treatment evaluations.

Compliance

Each time the therapist issued a demand, the observer selected the *demand* key on Countee. Two qualifier keys were programmed to appear (i.e., comply and noncomply). *Comply* was coded when a child responded in a contextually appropriate manner within 5 s of the initial demand or within 5 s of the model prompt. *Noncomply* was coded if more than 10 s passed from the initial demand (i.e., or more than 5 s from the model prompt) without a contextually appropriate response from the child. *Noncomply* was also coded if the child was manually guided to complete the response before 10 s had elapsed (i.e., which would have been simultaneously noted as a procedural fidelity error).

By coding each time a demand was presented, observers were simultaneously required to code participant response to the demands. This allowed researchers to analyze compliance at the level of both the participant and trial. Compliance was reported as a count per trial and a percentage of opportunities. To calculate the percentage of demands that resulted in compliance, researchers divided all instances of compliance by the number of demands presented. The resulting quotient was multiplied by 100 and converted to a percentage. However, a count of instances of compliance was used as the primary dependent variable to guide decision making during escape treatment evaluations.

Model Evaluation

After analyzing trial-level outcomes for each treatment evaluation, we evaluated the extent to which FIMB case conceptualization facilitated four main objectives. We accomplished this by calculating categorical impact scores into a single FIMB score for each participant. FIMB scores of 10 to 12 signified strong evidence of the model's efficacy. Scores of 7 to 9 produced moderate evidence, and scores of 4 to 6 produced limited evidence. If a participant's treatment evaluation was scored in the range of 0 to 3, we suggested the approach was contraindicated (e.g., Lambert, Copeland, et al., 2022).

Conceptual clarity. Across every FA conducted, each test condition was compared to its corresponding control and coded for demonstrations of effect. A demonstration of effect was defined as an observed difference in level, trend, or variability following a change from test to control. Detailed rules for coding individual tests as confirmed, suggested, or ruled out are outlined in Lambert, Copeland, et al. (2022). We scored the FA as providing full conceptual clarity if all tests were confirmed or ruled out. It was coded as providing partial clarity if at least

one functional relation was confirmed, but other tests produced outcomes that were not resolved. The assessment was scored as actionable if challenging behavior was observed, but no functional relations were confirmed. The assessment was coded as providing no clarity if challenging behavior was never observed. A single categorical impact score was assigned to each participant based on the overall clarity provided by the conglomerate of assessments conducted.

Response elimination. To standardize evaluations of efficacy across participants, we compared each child's performance during *each* treatment against their performance during that treatment's baseline condition (e.g., Hagopian et al., 2013). Specifically, researchers mean occurrences of challenging behavior during the final three trials from baseline (MBL-3) and response elimination (MRE-3). We then subtracted the MRE-3 from the MBL-3, divided the difference by the MBL-3, and multiplied the divided by 100 to produce a mean baseline reduction (MBLR-3; e.g., Greer et al, 2016). Treatments with scores of at least 90% reductions in challenging behavior were determined to be highly effective. Treatments with scores between 70-89% were determined to be effective. Treatments with scores between 50-69% were scored as marginally effective. Any treatment with a score below 50% was coded as equivocal. After calculating the impact score for each individual treatment evaluation, researchers averaged all relevant MBLR-3 scores to determine the global categorical impact score of the FIMB framework for each participant.

EO tolerance. For all treatments that progressed to EO tolerance, researchers calculated an MBLR-3 and evaluated it using the same criteria described above. To evaluate the maintenance of treatment effects, researchers compared the categorical impact scores calculated for response elimination and EO tolerance. If the categorical score remained the same or improved, the treatment was coded the treatment as having strong maintenance effects. If the

score deteriorated by one level, maintenance effects were scored as moderate. If the categorical score dropped by two levels, we coded the treatment as having limited maintenance effects. Any treatment that dropped by three levels was scored as having poor maintenance effects.

Generalization. Using the criteria described in Lambert, Copeland, et al. (2022), treatment evaluations were scored as having comprehensive generalization programming if the team incorporated (a) endogenous implementers, (b) endogenous settings, and (c) a distributed-trial format (i.e., extending beyond formal therapy sessions). Researchers documented partial programming if only two techniques were attempted. If researchers attempted one technique, it was coded as having limited generalization programming. If none of these techniques were documented, the coders indicated generalization programming was not attempted.

Observers and Implementers

The research team included a male Latino BCBA-D, two doctoral candidates in special education programs, and several graduate-level students seeking certification in behavior analysis. Prior to working with the family, all members of the research team participated in 18 hr of behavioral skills training (Finn, 2020; Sun, 2020). During this training, the team received direct instruction on data collection, data management, functional analysis, function-based intervention, and crisis management.

Training Procedures

Master's level research assistants served as the primary coders for all sessions in the supervised context. Before collecting data in the field, all observers were trained according to the video-based protocol outlined by Dempsey et al. (2012). Each team member was required to achieve agreement of at least 90% with the master code for each video. The first author then trained the coders during a 1 hr in-person training session on participant-specific definitions. After the training session, all coders took a paper-pencil quiz on the codebook definitions. They were required to score 100% on the quiz before moving to in-vivo coding.

Interobserver Agreement (IOA)

A second trained observer (often the first author) collected reliability data on 43% of trials across all participants and phases. Using a 3-s window of agreement, the researcher calculated point-by-point agreement by (a) counting the number of agreements coded for each variable, (b) counting the number of disagreements coded for each variable, (c) dividing the number of agreements by the total number of agreements and disagreements, and (d) multiplying by 100 (Ledford & Lane, 2024).

Agreements included all instances in which observers coded the same variable within the 3-s time window. Disagreements were coded in two circumstances. If only one observer coded a variable, it was marked as a disagreement. If both observers coded the same variable outside the 3-s window of agreement, this was also coded as a disagreement. A single score was first calculated for each variable and then aggregated across variables and reported as a single agreement score for each session.

Experimental Design and Data Analysis

Several different single-case designs were used over the course of the study, each paired with distinct data analysis techniques and considerations (Ledford & Gast, 2024).

Function Identification

Researchers used multielement-logic to evaluate FA outcomes (Barlow & Hayes, 1979). By rapidly alternating across four conditions (e.g. attention, tangible, escape, play) during a single phase, researchers analyzed the extent to which the data from the test conditions differed in level from that of the data in the control condition (Ledford et al., 2024). As researchers analyzed these differentiated levels of responding, we consulted the data interpretation rules formalized by Roane et al. (2013) to decide when to discontinue each test condition (i.e., when to identify or rule out possible functional relations).

In the final iteration of Jerome's FA the team transitioned to a sequential introduction and withdrawal across adjacent phases to address concerns about (a) multitreatment interference and (b) a potential cyclical pattern of problem behavior (e.g., Kim hypothesized that tantrums happened every 30 min regardless of circumstance). In this iteration, we analyzed response patterns for changes in level associated with the identified phase changes.

Treatment Evaluation

During treatment evaluation, researchers used a changing criterion design (Hartman & Hall, 1976). The team made two decisions intended to reduce exposure to counter-therapeutic contingencies. First, researchers used data from relevant test conditions of the FA (i.e., those with confirmed functions) as baseline data. To do this, researchers re-analyzed all FA data and

calculated the latency to the first occurrence of target behavior. Second, researchers embedded a countertherapeutic criterion change (i.e., we went back to a previous criterion level) to strengthen the validity of the design and avoid the need to revert back to reinforcing challenging behavior (Ledford & Tuck, 2024).

Decisions about phase changes and criteria changes were both response-guided (i.e., made in response to graphed data) and grounded in a few pre-existing rules. A minimum of three data points needed to be collected in each condition before considering any changes. Each phase of treatment had a corresponding mastery criterion (described below). As the number of data points required to reach mastery increased, we required additional data points to be collected in each phase. Specifically, half of the data points collected in any treatment condition needed to meet mastery requirements. Additionally, 50% of the data points contributing to mastery criteria must be in the second half (i.e., temporally) of the data set.

General Procedures

To structure appointments, the researchers assembled two teams (i.e., three therapists and one supervising BCBA per team). Each team attended two appointments per week for approximately 2 hr per appointment. During this time, teams conducted multiple sessions per appointment to evaluate responsiveness to programmed contingencies in the supervised context. All children were invited to work with the researchers at the start of each appointment. If they declined, their dissent was documented. The child then rejoined the typically scheduled household activities. Although we did not use data for research purposes when dissent was recorded, the team implemented the child's behavior plan whenever tantrums occurred.

All appointments were divided into three, 30-min segments to allow all therapists to work with all three children. For example, if a therapist started the appointment with Jerome, they would then transition to Serenity, and finish the appointment with Zeke. The remaining 30-min of the appointment was allocated to transitions between rotations (i.e., 5- min between each or 15 min of transition) or appointment-specific jobs (i.e., setting up materials, post-session clean up, caregiver debrief).

Intake Assessments

Researchers completed a series of intake assessments to learn more about each child. Two team members completed open-ended interviews with Kim (e.g., FAI; O'Neill et al., 1997) to operationalize each child's challenging behavior and identify an array of consequences reported to follow these behaviors. Two separate team members simultaneously conducted a series of preference and demand assessments. This information was used to help the team design an individualized functional analysis for each child.

Function Identification

Functional analyses were completed just before treatment began for each child. Across all test conditions, programmed consequences were delivered contingent on challenging behavior according to a fixed-ratio 1 schedule of reinforcement (i.e., therapists reinforced every instance of target behavior). Researchers included attention, tangible, and escape test sessions for all three children. Additionally, FA conditions were presented in the order suggested by Hammond et al. (2013) to control for carry-over effects and capitalize on relevant motivating operations.

Attention Condition

Prior to beginning an attention session, therapists provided high-quality attention for a minimum of 30 s. At session onset, the therapist removed attention, ignoring all nontarget behavior. Contingent on tantrum behaviors, the therapist provided 30 s of attention in the form of reprimands (i.e., Jerome and Zeke) or statements of concern (i.e., Serenity). Access to moderately preferred items (i.e., determined during preference assessments) remained available throughout the session. No demands were presented.

Tangible Condition

Therapists provided brief access (30 s) to highly preferred items before beginning tangible sessions. At session onset, the therapist removed the items, returning them contingent on tantrum behavior. Access to attention was available throughout the session, and no demands were presented.

Escape Condition

After ensuring the child had access to a break from all demands (including questions) for 30 s, the therapist began an escape session by presenting a task demand (i.e., gross-motor tasks for Jerome; household chores for Zeke and Serenity). All demands were presented vocally for Serenity and Zeke and in ASL for Jerome. Compliance with demands produced a subsequent demand. If the child did not complete the demand within 5 s, the therapist provided a model prompt, which was followed by manual guidance to complete the task after an additional 5 s. Contingent on tantrum behavior, the therapist provided a 30-s break from demands. No preferred tangible items were present during this condition.

Play Condition (Control)

During play sessions, therapists avoided delivering demands, ensured continuous access to high-preferred tangible items, and provided attention at least once every 30 s. Tantrum behavior produced no programmed consequences.

Response Elimination

During this phase, all three participants were repeatedly exposed to therapeutic contingencies in the presence of challenging behavior's establishing operations (i.e., confirmed functions). To select techniques, researchers consulted the hierarchically established considerations of precision outlined in Lambert, Copeland, et al. (2022). As stipulated by the framework, therapists began therapy with functional communication training for all identified functions of challenging behavior (FCT; Carr & Durand, 1985)

Functional Communication Training (FCT)

Therapists structured motivating operations for attention (i.e., all children), escape (i.e., all children), and tangible (i.e., Serenity and Zeke) trials as described in FA procedures. However, tantrums were placed on extinction (i.e., did not produce any programmed consequences). Each trial began with therapists implementing the programmed EO. Therapists taught a functional communication response using a 5-s progressive time delay (Touchette & Howard, 1984).

During the first trial, a separate therapist immediately prompted the child to request a break (i.e., at a 0-s delay) using a 3-step least-to-most prompting procedure. For Serenity and Zeke, the 3-step procedure included (a) providing an expectant look, (b) providing a verbal cue

(e.g., “you can ask”), and (c) vocally modeling the expected mand. Jerome had a similar 3-step prompting procedure that included (a) providing an expectant look paired with the “ask me” in ASL, (b) modeling the mand in ASL, and (c) manually guiding the mand in ASL. If any child responded incorrectly or did not respond to a prompt within 5 s, the therapist increased the prompt level. The delay to the initial prompt was gradually increased by 5 s after every trial until the team reached the terminal prompt delay of 30 s.

Prompted and independent mands were immediately followed by access to reinforcement. The session then ended with the conclusion of the reinforcement period. Any session with an independent mand and no tantrums was considered a demonstration of mastery (i.e., contributed to satisfaction of the mastery criteria).

Synthesized Functional Communication Training (FCT)

Only Jerome participated in this phase of treatment. Therapists synthesized EOs and required Jerome to request both a break and attention within the same reinforcement period (i.e., instead of practicing each response in isolation). Therapists now provided high quality attention in addition to a break from all demands prior to starting each session. When the session began, the therapist shifted to a more neutral “business-like” tone and delivered a demand.

During the first trial, a separate therapist immediately prompted Jerome to request a break (i.e., at a 0-s delay) using the same progressive time delay, 3-step least-to-most prompting procedure described above. Following either a prompted or independent escape mand, the therapist removed all demands and began the SR period. Using a constant time delay prompting procedure, the second therapist immediately prompted Jerome to request attention (i.e., 0-s delay) after the escape SR period began. For all subsequent sessions, the prompt for the attention

mand occurred 5-s after the SR period for escape began. All prompted and independent mands for attention (i.e., that followed a mand for escape) resulted in access to high-quality attention for the remainder of the 30-s of SR. Notably, the second request did not reset the clock or extend time in SR. Additionally, therapists required the mands to occur in this order (i.e., request a break first and then request attention). Any trial both independent mands (i.e., I-Mand Esc and I-Mand Attn) and no tantrum behavior was considered a demonstration of mastery.

Building Tolerance for Establishing Operations

Through two-component chained schedules arrangements (Ferster & Skinner 1957), therapists varied the availability of functional reinforcers by schedule components based on each child's performance. Each component was signaled by a unique stimulus (i.e., S^{Δ} for extinction, S^{D} for reinforcement). Trials always in S^{Δ} with both mands and tantrums placed on extinction in the presence of programmed EOs. Therapists held up a brightly colored paper star while exposing participants to their identified EOs. Any mands that occurred during this time resulted in the therapist gesturing to the timer to remind the child that reinforcement was unavailable.

For both Serenity and Zeke, this cue was gradually faded out and replaced by any item the child selected as "the star" paired with a verbal reminder of the rules previously paired with the star (e.g., "Okay this Lego is our star. Remember, when the Lego man is on the shelf, I can't talk to you and you need to wait patiently."). Jerome required a more dynamic visual support to indicate the passage of time, so therapists provided him with a token board and visual timer. The number of tokens he was required to earn corresponded to the programmed response requirement of the S^{Δ} schedule component. Any mands for a break that occurred during this time resulted in the therapist gesturing to the token board to remind him that reinforcement was unavailable.

In the second component (i.e., S^D), therapists maintained the relevant EO but removed the signal (i.e., "star" for Serenity and Zeke; delivered the final token for Jerome). Transitions from S^Δ to. S^D were contingent on the satisfaction of a conjunctive contingency which required (a) the absence of tantrum behavior for a pre-established duration and (b) satisfaction of a pre-established performance criterion for a prosocial, alterative behavior. The requirements for transitions during treatment evaluations involving escape functions for each participant are outlined in Table 1. Notably, Kim selected the terminal schedule value for each child. Researchers then calculated the number of demands presented between trial onset and the first instance of tantrum behavior during baseline to select the first schedule requirement. Response requirements were doubled between criterion phases for Jerome. Requirements increased by two for Serenity and Zeke, with two exceptions (i.e., both described in the results section).

Table 1

EO Tolerance Requirements for S^Δ Transitions During Escape

	Jerome	Serenity	Zeke
Phase 1	FR-1	FR 2	FR-1
Phase 2	FR 2	FR 4	FR 2
Phase 3	FR 4	FR 6	FR 4
Phase 4	FR 8	FR 8	FR 3
Phase 5	FR 16	FR 10	FR 5
Phase 6	FR 1	FR 12	FR 7
Phase 7	FR-16	FR 22	FR 9
Phase 8	-	FR 2	FR 11
Phase 9	-	FR 22	FR 1
Phase 10		-	FR11

The requirements for transitions during treatment evaluations involving attention nad tangible functions are outlined in Table 2. Because Jerome's treatment was synthesized with escape at this stage, he is not included in this table. Again, Kim selected the terminal schedule value for each child. Researchers used the latency to the first occurrence of tantrum behavior during baseline to set the initial schedule requirement. Response requirements were increased either by 1.5 times the requirement for the previous criterion or doubled (i.e., per a randomized number generator) number generator. Requirements decreased on at least one occasion per evaluation to enhance the internal validity of the design (i.e., additional detail seen in results).

Table 2

EO Tolerance Requirements for S^{Δ} Transitions for Attention and Tangible

	Attention		Tangible	
	Serenity	Zeke	Serenity	Zeke
Phase 1	30 s	10 s	20 s	10 s
Phase 2	45 s	20 s	40 s	20 s
Phase 3	30 s	40 s	80 s	40 s
Phase 4	45 s	80 s	160 s	80 s
Phase 5	90 s	160 s	320 s	160 s
Phase 6	180 s	320 s	180 s	240 s
Phase 7	360 s	640 s	320 s	360 s
Phase 8	720 s	10 s	540 s	540 s
Phase 9	30 s	640 s	20 s	810 s
Phase 10	720 s-	-	540 s-	10 s-
Phase 11	-	-	-	810 s

When transitioning to the second schedule component after meeting the requirements outlined in the tables above, tantrums remained on extinction, but mands were now reinforced on a fixed-ratio 1 schedule. If no manding occurred within 10 s of the transition between schedule components, the therapist delivered a controlling prompt.

Generalization

Researchers trained Kim on the final iterations of all treatment procedures using a behavioral skills training model. After demonstrating mastery of intervention procedures in a role play scenario with a member of the research team, Kim implemented the treatment with her children. One therapist provided feedback and coaching throughout while the remaining team members collected data.

Procedural Fidelity

A research team member who was not implementing treatment collected procedural fidelity data across 37% of sessions to measure adherence to study procedures. Some therapist behaviors were binary (i.e., coded as correct or incorrect). Other behaviors were dependent on child behaviors and required only under certain circumstances. For these opportunity-bound behaviors, observers coded each opportunity as correct or incorrect. Researchers then calculated a percentage adherence score. To do this, we calculated the proportion of correct responses for any opportunity-bound behaviors that were coded more than once (e.g., compliance). Then, researchers added those proportions to the number of correct binary codes to produce the total number of correct behaviors. We then divided the total number of correct behaviors by the total number of possible behaviors and multiplied the quotient by 100. Mean percentage adherence exceeded 90% across all contexts and conditions (see Table 3).

Table 3

Mean (Range) Percentage Adherence to Procedures by Condition and Context

	FA	RE	EO	Gen
Jerome	98.7 (97-100)	100	100	100
Serenity	100	94.3 (80.0–100)	100	100
Zeke	100	100	100	100

STUDY I RESULTS

Results for all phases of treatment are first discussed at a summative level using the FIMB framework categorical evaluations (as proposed in Lambert et al., 2022). Participant-level results follow.

Functional Analyses

Both Serenity and Zeke's FAs were coded as providing full conceptual clarity (i.e., all included test conditions were either confirmed through functional relations or ruled out by the end of the assessment). Jerome's FA was scored as having partial clarity, meaning at least one functional relation was confirmed, but other tests produced suggested outcomes.

Jerome's Functional Analysis

Jerome's FA was completed in five appointments across four iterations (Figure 2). The first iteration was completed across 14 sessions (11 test and 3 control), lasting a total of 70 min. Approximately 27.1% of this time was spent in SR, with the remaining time spent in purported EO. Therapists presented 91 demands across three escape sessions, and Jerome complied with 26 demands presented (i.e., 28.6%).

The comparison included an adequate number of data points for comparing tantrum rates across conditions. Because data patterns were not consistent (i.e., differentiated responding was initially observed during tangible and escape sessions before responding dropped to zero for all data paths), additional data collection would be needed to draw conclusions about functional relations. However, the team shifted to a pairwise assessment to address concerns related to multitreatment interference (i.e., alternating between a single test and control).

The second iteration was completed across 15 sessions (11 escape and 4 control), lasting a total of 75 min. Approximately 30.1% of this time was spent in SR, with remaining time spent in purported EO. Therapists presented a total of 270 demands across all escape sessions in this iteration, and Jerome complied with 116 (73%) of the demands presented.

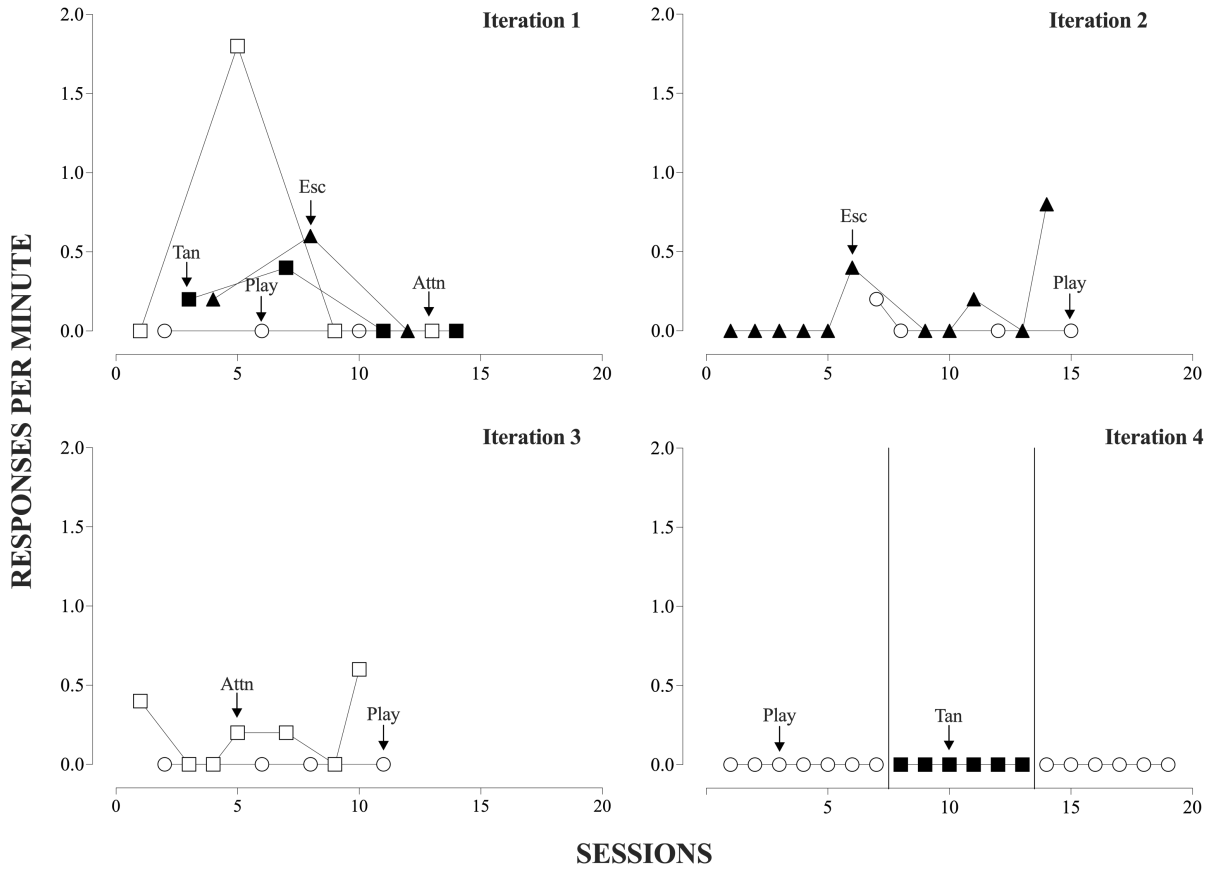
The comparison included an adequate number of data points from both conditions. However, data patterns were inconsistent, with differentiated responding observed in two of the three final condition pairings (i.e., escape session 11 vs. play session 12; escape session 14 vs. play session 15). Low levels of tantrum behavior were recorded during across control sessions (i.e., except for the single occurrence documented in the first play session). Tantrum behavior was initially at zero levels during escape sessions, before increasing in level and variability. A functional relation determination could not be made without additional data. However, due to time constraints, the team coded escape as a suggested function of Jerome's tantrums and proceeded to the third iteration.

The third iteration was completed across 11 sessions (7 attention and 4 control) lasting a total of 55 min. Approximately 43.6% of this iteration was spent in SR, with the remaining time spent in purported EO. This comparison included an adequate number of data points, with four instances of differentiated responding. Two comparisons resulted in identical levels of responding (i.e., attention session 5 vs. play session 6; attention session 7 vs. play session 8). No tantrums were observed across play sessions. Tantrums occurred at low, variable rates during attention sessions, with four sessions containing low (i.e., but comparatively elevated) levels of tantrum behavior. A functional relation was confirmed for attention.

The fourth and final iteration was completed across 19 sessions (6 tangible and 13 control). The total duration of this iteration was 95 min with 72% of the time spent in SR. By setting up a withdrawal design, researchers alternated between a series of play and tangible sessions. This A-B-A comparison allows for only two potential demonstrations. However, no tantrums were observed, and the team discontinued the assessment due to time constraints. The team did rule out tangibles as a possible functional reinforcer for Jerome's tantrums after completing six consecutive sessions in which Jerome was exposed to the establishing operation without engaging in challenging behavior.

Figure 2

Functional Analysis of Jerome's Tantrums



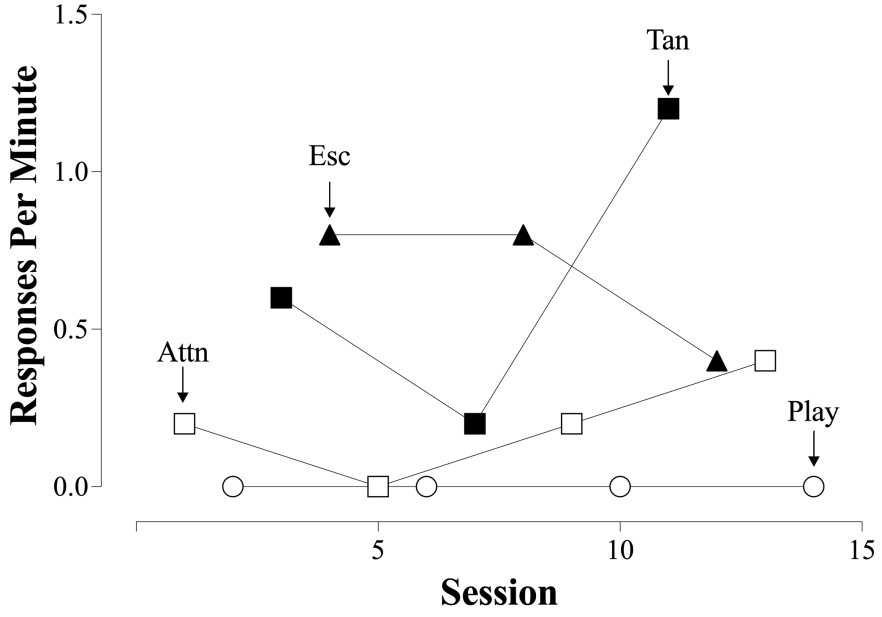
Serenity's Functional Analysis

Serenity's FA was completed across three appointments in a single iteration (Figure 3). It included 14 sessions (10 test and 4 control) that were completed in 70 min (i.e., with 42.6% of the time spent in reinforcement). Therapists presented 27 demands across all escape sessions, and Serenity complied with 15 of the presented demands (i.e., 55%). The comparison included an adequate number of data points across four conditions. Observers recorded elevated levels of tantrum behavior during all tangible and escape sessions. All paired sessions for these conditions were documented to have consistently differentiated levels of responding during control. Of the

four possible comparisons for the attention condition, three resulted in differentiated responding with elevated levels of tantrums observed during attention sessions as compared to play sessions. Researchers identified a functional relation between attention, tangible, and escape conditions and Serenity's challenging behavior.

Figure 3

Functional Analysis of Serenity's Tantrums



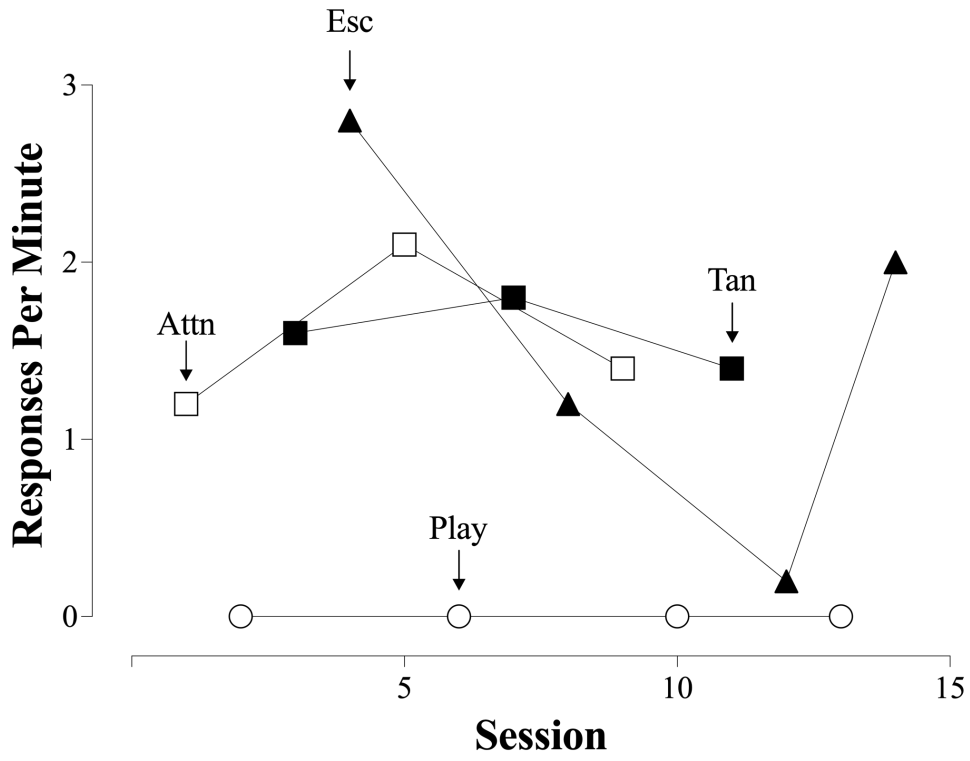
Zeke's Functional Analysis

Zeke's functional analysis was completed across four appointments in a single iteration (Figure 4). The FA was completed over 14 sessions (10 test and 4 control) in 70 min (i.e., with 46.5% of the time spent in reinforcement). Therapists presented a total of 58 demands across escape sessions, and Zeke complied with 54 of the demands (i.e., 93.1%). Observers recorded elevated levels of tantrum behavior during all attention, tangible, and escape sessions. No

tantrums were observed during play sessions. Due to the decreasing trend documented across the first three escape sessions (i.e., with similar levels of responding in the third comparison), we conducted an additional comparison for this condition before confirming attention, tangible and escape functions for Zeke's tantrum behavior.

Figure 4

Functional Analysis of Zeke's Tantrums



Treatment Outcomes

Treatment was highly effective (i.e., resulted in at least a 90% decrease in challenging behavior) for all nine treatment evaluations. Eight of the treatments progressed to EO tolerance (i.e., Jerome's attention and escape treatments were synthesized at this stage). All eight EO tolerance treatments were scored as highly effective and displayed strong maintenance (i.e., there was no degradation in categorical impact scores). Generalization outcomes are discussed in the results section of the second study.

Jerome's Treatment Evaluation

Jerome's treatment evaluation was completed across two phases. The phase (as seen in Figure 5) included 17 trials (i.e., 5 attention and 12 escape) of treatment completed across ten appointments. Notably, for six of these appointments, Jerome dissented from participating in study procedures after a single trial. In total, Jerome spent 30 min in this phase of treatment (i.e., with 87.5% of the time spent in reinforcement). This treatment plan was discontinued after documenting (a) little change in independent manding and (b) the frequency with which Jerome would dissent from study procedures following treatment onset. These data are included in Appendix D

During the second phase of his treatment evaluation, therapists addressed both attention and escape functions across 32 treatment trials and five appointments. Jerome did not dissent from study procedures. The results can be seen in Figure 5 across four panels. Of note, the changing criterion design shown in the top panel (i.e., count of instances of demand presentation and compliance) included six opportunities for demonstrations of effect, including one instance of reverting to a previous criterion (i.e., moving from FR-16 to FR-1 and back to FR-16 in the

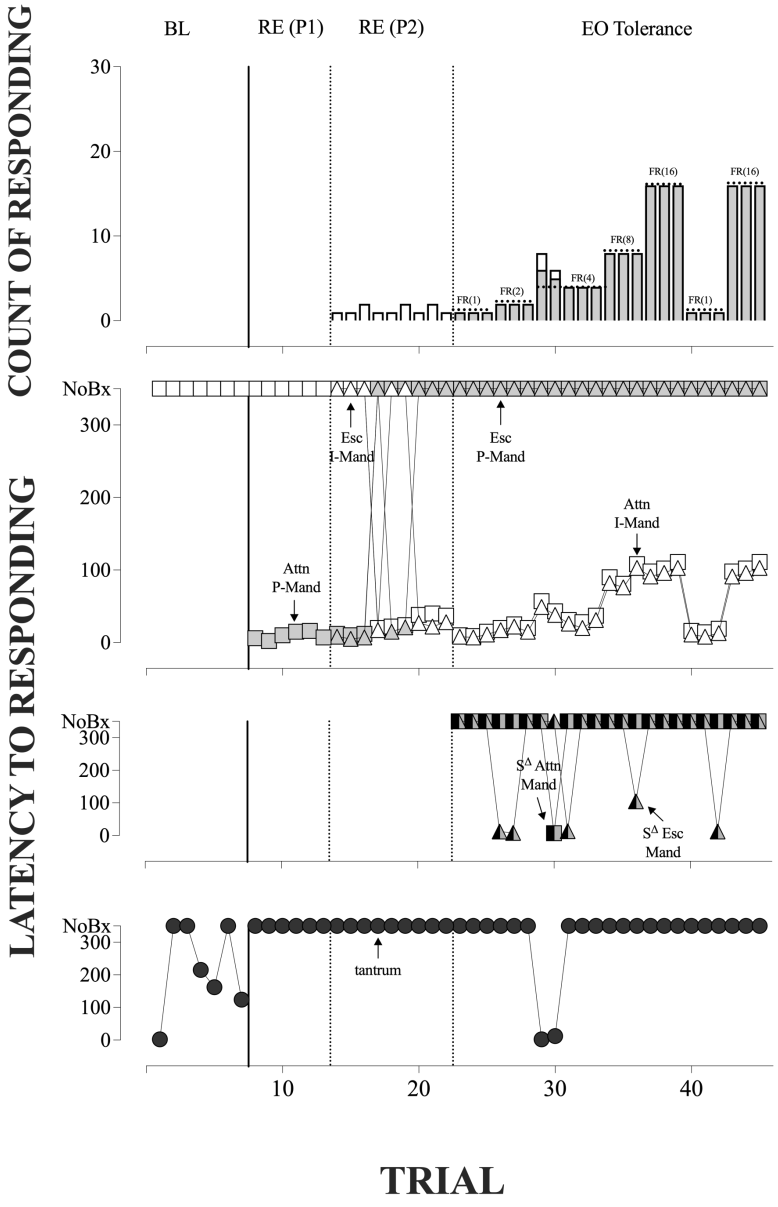
final three phases). White columns represent trials with noncompliance. Gray columns indicate instances of compliance. Columns with gray and white signify that both compliance and noncompliance behaviors were observed. The height of each column corresponds to the total number of demands presented during that trial. Stable data patterns were observed within each criterion of this panel before researchers discontinued data collection, ensuring the level of compliance closely aligned with the criterion for most data points. Changes between criteria led to immediate differences in the level of responding. Taken together, researchers identified a functional relation between the FIMB framework and increased levels of compliance.

Researchers simultaneously monitored Jerome's latency to responding in other panels to ensure he satisfied the mastery criteria for each phase and to calculate an efficacy score for the treatment evaluation. Specifically, in panel two (i.e., independent and prompted mands), we observed a decrease in the latency to independent requests (i.e., though slightly delayed, around trial 16) and increase in latency to prompted mands following the introduction of synthesized FCT (i.e., RE[P2]). These results maintained during the EO tolerance phase of treatment.

In the third panel of Figure 5, researchers documented variable levels of S^{Δ} mands through the EO tolerance phase of treatment. However, prompted mands did not resurgence during this stage. In the bottom panel of Figure 5, researchers documented decreased levels of tantrum behavior throughout all phases of treatment (i.e., with the exception of trials 29-30).

Figure 5

Jerome's Treatment Evaluation



Note. BL = baseline; RE = response elimination; P1 = Phase 1; P2 = Phase 2; EO = establishing operations; criterion requirements are indicated in the top panel above each marked goal line; the white columns in the top panel indicate the total number of demands presented and gray columns indicate demands that resulted in compliance

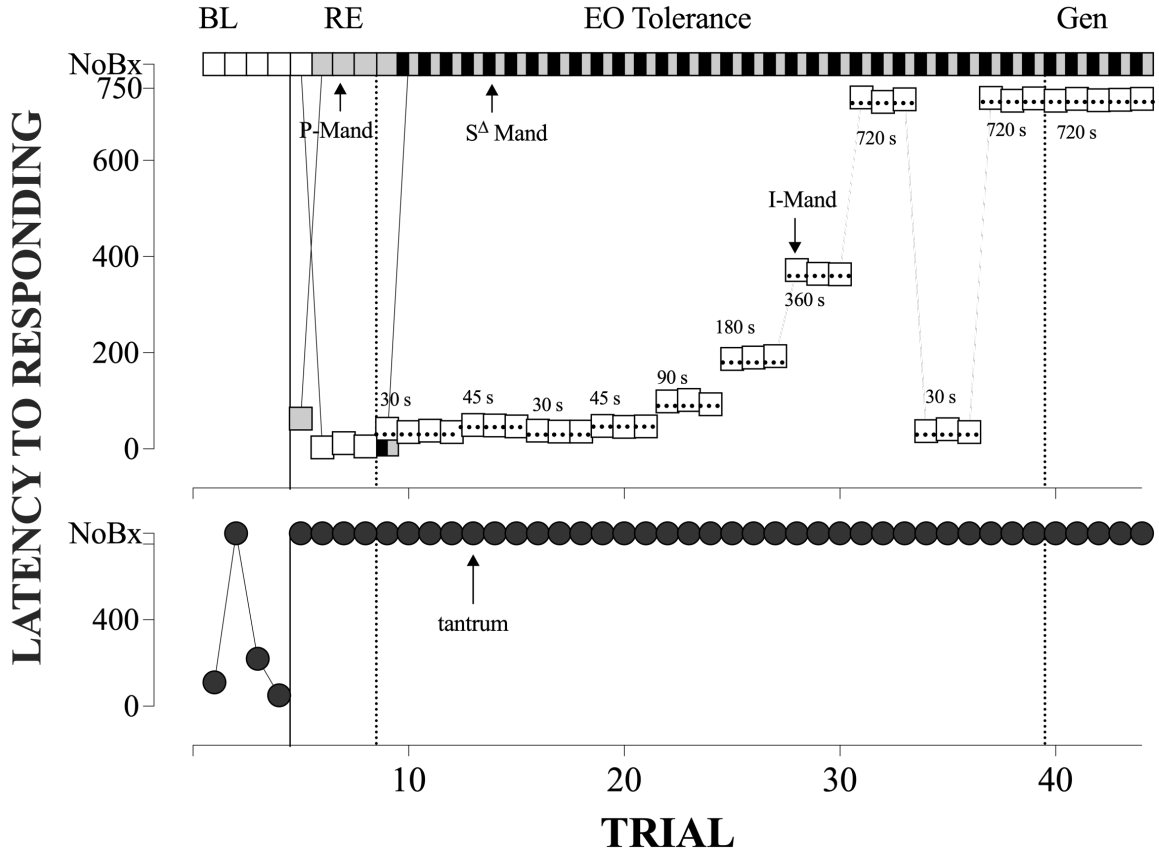
Serenity's Treatment Evaluation

Researchers completed three distinct treatment evaluations with Serenity. Her attention treatment evaluation (Figure 6) was completed across 44 trials and nine appointments. Serenity dissented from one appointment. In all, the evaluation took approximately 171 min with 21.6% of session time spent in SR. The changing criterion design depicted in the top panel of Figure 6 included nine opportunities for demonstrations of effect, including two instances of reverting to a previous criterion (i.e., the first and last changes in criteria requirements. Specifically, we first reverted to a 30-s S^{Δ} schedule component after Serenity mastered the 45-s criterion before reverting to the 30-s criterion after she mastered the 720-s criterion. Stable data patterns were observed within each criterion before researchers discontinued data collection, ensuring latencies to attention I-Mands closely aligned with the criterion. Further, changes between criteria led to immediate differences in level. Taken together, researchers identified a functional relation between the FIMB framework and increases in appropriate requests for attention.

Data in the bottom panel of Figure 6 were analyzed concurrently as researchers monitored Serenity's tantrum behaviors to ensure she satisfied the mastery criteria for each phase and to calculate an efficacy score for the treatment evaluation. No instances of tantrum behavior were documented after baseline concluded.

Figure 6

Serenity's Attention Treatment Evaluation



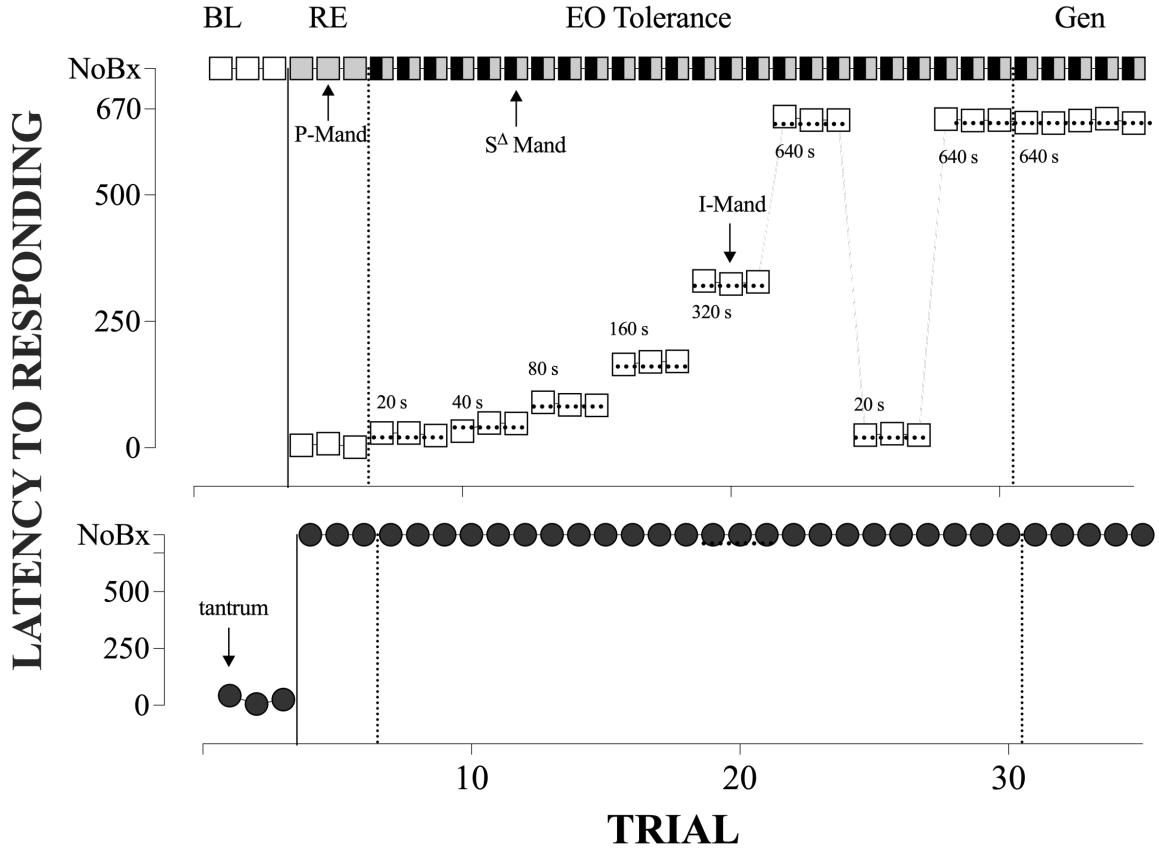
Note. BL = baseline; RE = response elimination; EO = establishing operations; GEN = generalization; criterion requirements are indicated in the top panel near the marked goal line for each criterion

Serenity's tangible treatment evaluation (Figure 7) was completed across 32 trials and 14 appointments. She did not dissent from participating in any appointments linked to the tangible treatment evaluation. In all, the evaluation took approximately 203 min with 24.8% of session time spent in SR. The changing criterion design depicted in the top panel of Figure 7 included seven opportunities for demonstrations of effect, including one instance of reverting to a previous criterion. Specifically, we reverted to a 20-s S^{Δ} schedule component after Serenity

mastered the 640-s criterion. Stable data patterns were observed within each criterion before researchers discontinued data collection, ensuring latencies to tangible I-Mands closely aligned with the criterion. Changes between criteria led to immediate differences in level. Researchers identified a functional relation between the FIMB framework and increases in appropriate requests for tangible items. Data in the bottom panel of Figure 7 were analyzed as described above. No instances of tantrum behavior were documented after baseline concluded.

Figure 7

Serenity's Tangible Treatment Evaluation



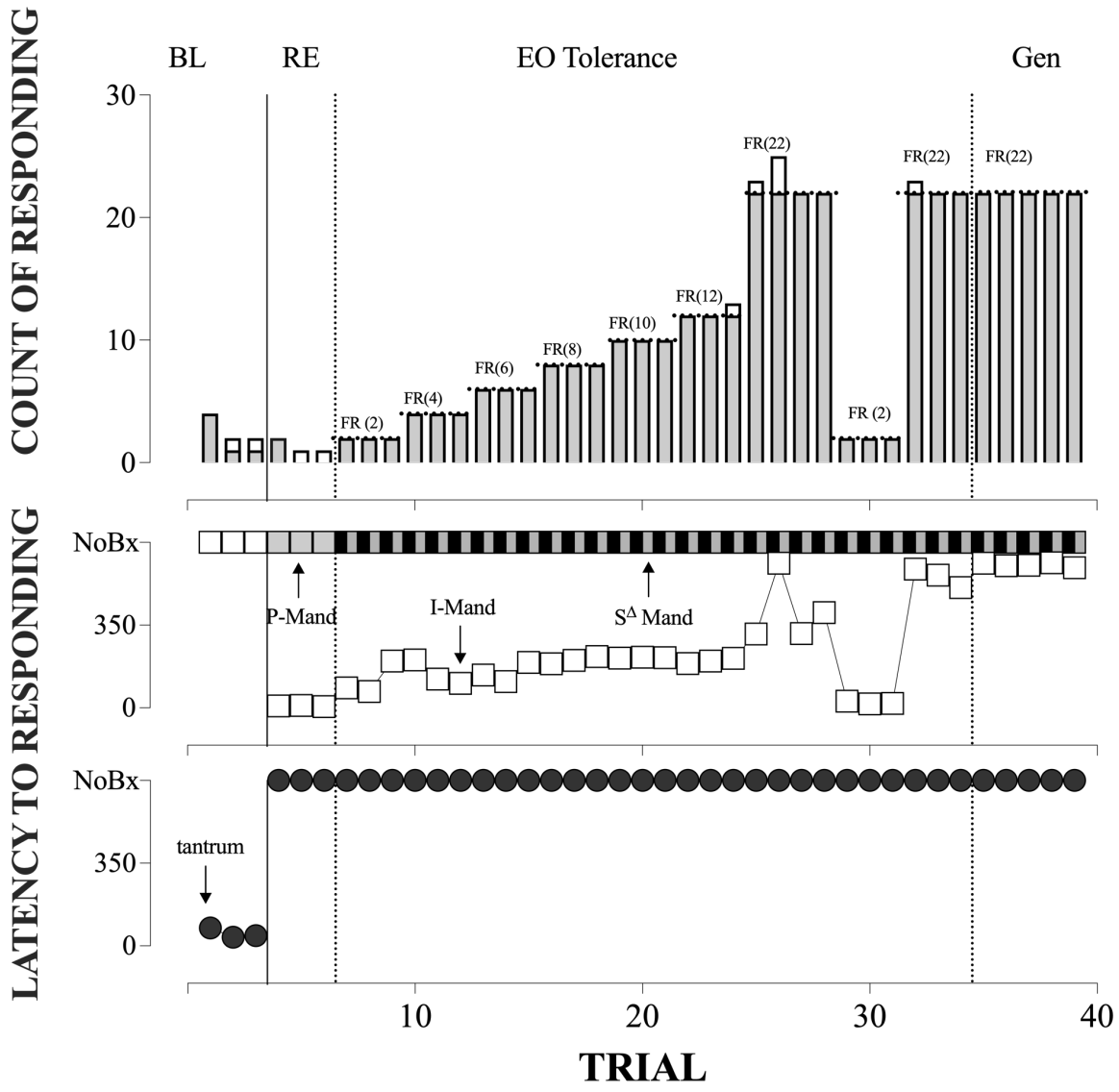
Note. BL = baseline; RE = response elimination; EO = establishing operations; GEN = generalization; criterion requirements are indicated in the top panel near the marked goal line for each criterion

The escape treatment evaluation (Figure 8) was completed across 39 trials and 18 appointments. Serenity dissented from three appointments after therapists initiated an escape trial. In all, the evaluation took approximately 753 min with 18.1% of time spent in SR. This changing criterion design included nine opportunities for demonstrations of effect, including one instance of reverting to a previous criterion. Specifically, we reverted to a FR(2) S^{Δ} schedule component after Serenity mastered the FR(22) criterion. Stable data patterns were observed within each criterion before researchers discontinued data collection. Changes between criteria led to immediate differences in level. Researchers identified a functional relation between the FIMB framework and increases in compliance for Serenity.

Data in the bottom panel of Figure 8 were analyzed concurrently. As seen in the middle panel, researchers observed levels of independent mands that corresponded with increasing schedule requirements across the EO tolerance phase. No prompted mands nor S^{Δ} mands were observed. Additionally, no tantrum behavior was documented after baseline concluded.

Figure 8

Serenity's Escape Treatment Evaluation



Note. BL = baseline; RE = response elimination; EO = establishing operations; GEN = generalization; criterion requirements are indicated in the top panel above each marked goal line; white columns in the top panel indicate the total number of demands presented and gray columns indicate demands that resulted in compliance

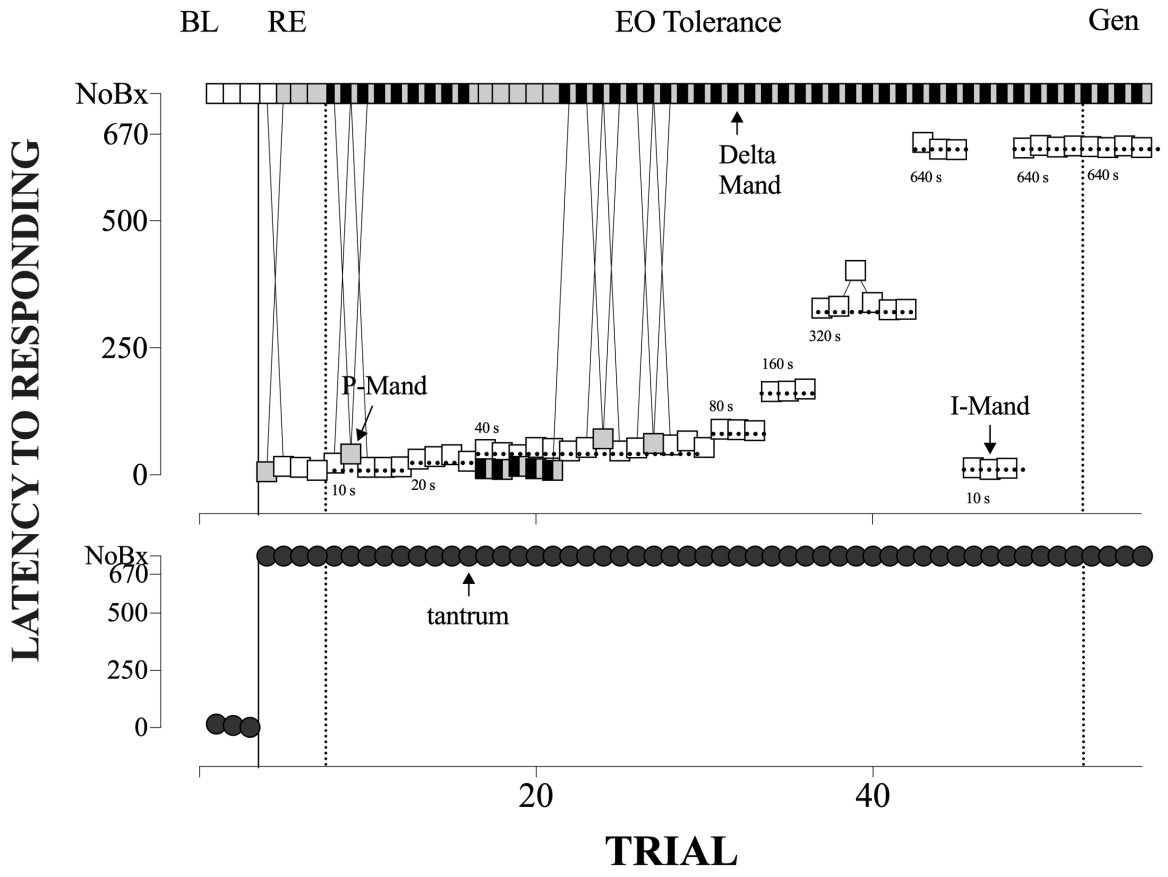
Zeke's Treatment Evaluation

Three treatment evaluations were completed to assess Zeke's progress. The attention treatment evaluation (Figure 9) was completed across 36 trials and 17 appointments. Zeke did not dissent during this treatment evaluation. In all, the evaluation took approximately 820 min with 26.2% of trial time spent in SR. The changing criterion design depicted in the top panel of Figure 9 included nine opportunities for demonstrations of effect, including one instance of reverting to a previous criterion. Specifically, we reverted to a 10-s S^{Δ} schedule component after Zeke mastered the 640-s criterion.

Stable data patterns were observed within each criterion before researchers discontinued data collection. During the second criterion requirement (i.e., when Zeke was required to wait 20 s without tantrum behavior before manding for attention), researchers observed low latencies to S^{Δ} mands (i.e., Zeke was asking before the therapist transitioned to the schedule component in which attention was available upon request). This prompted the team to extend data collection for this phase. As latencies to S^{Δ} mands increased, the latencies to prompted mands decreased for two trials before stable, low levels of I-Mands were observed again. All other phase changes between criteria led to immediate differences in the level of independent mands. Thus, researchers identified a functional relation between the FIMB framework and increases in independent mands for attention.

Figure 9

Zeke's Attention Treatment Evaluation



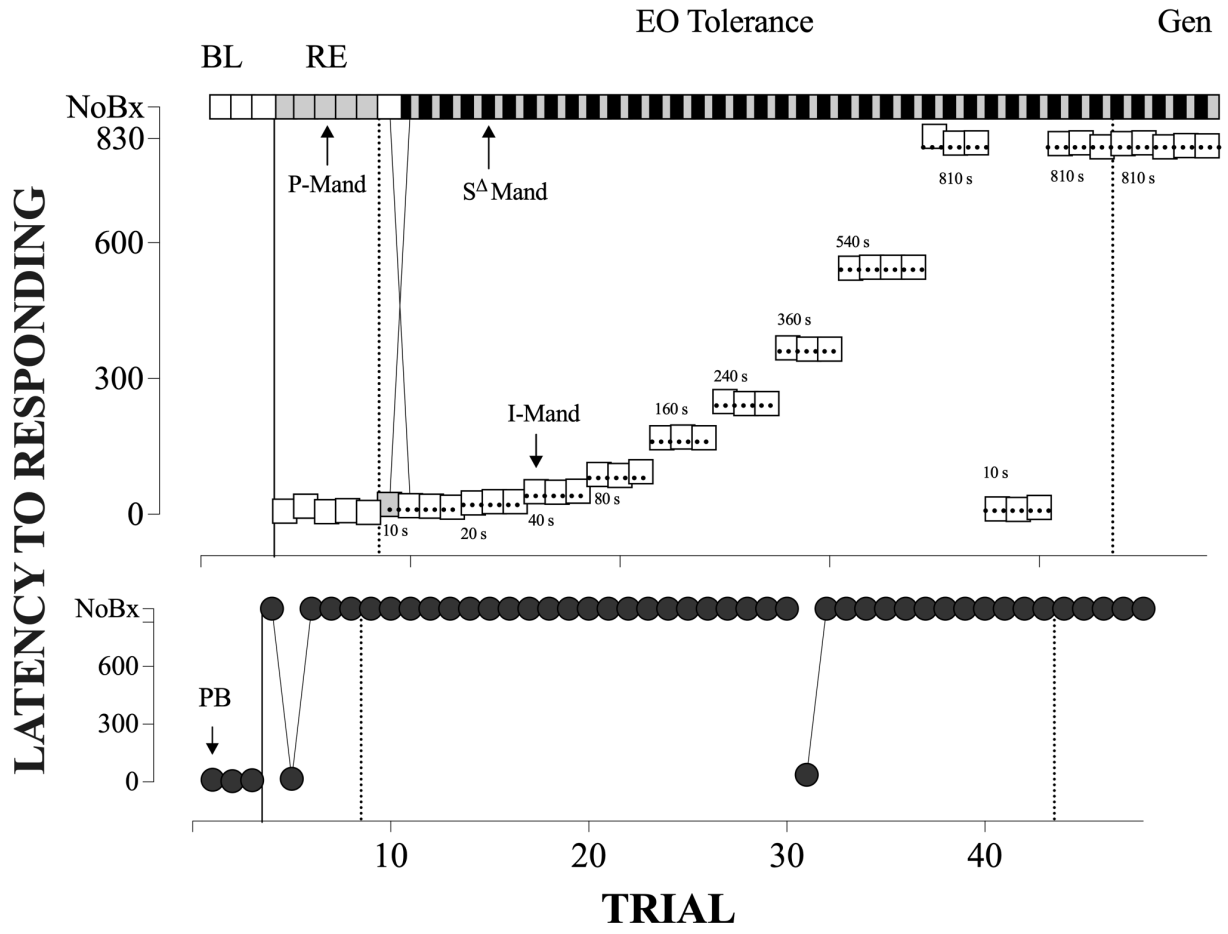
Note. BL = baseline; RE = response elimination; EO = establishing operations; GEN = generalization; criterion requirements are indicated in the top panel near the marked goal line for each criterion

The tangible treatment evaluation (Figure 10) was completed across 48 trials and 12 appointments. In all, the evaluation took approximately 713 min with 20% of trial time spent in SR. This changing criterion design included ten opportunities for demonstrations of effect, including one instance of reverting to a previous criterion. Specifically, we reverted to a 10-s S^{Δ} schedule component after Zeke mastered the 810-s criterion. Stable data patterns were observed within each criterion before researchers discontinued data collection. All changes between

criteria led to immediate differences in level for independent requests for tangibles. Thus, researchers identified a functional relation between the FIMB+ treatment model and increases in I-Mands for tangible for Zeke.

Figure 10

Zeke's Tangible Treatment Evaluation



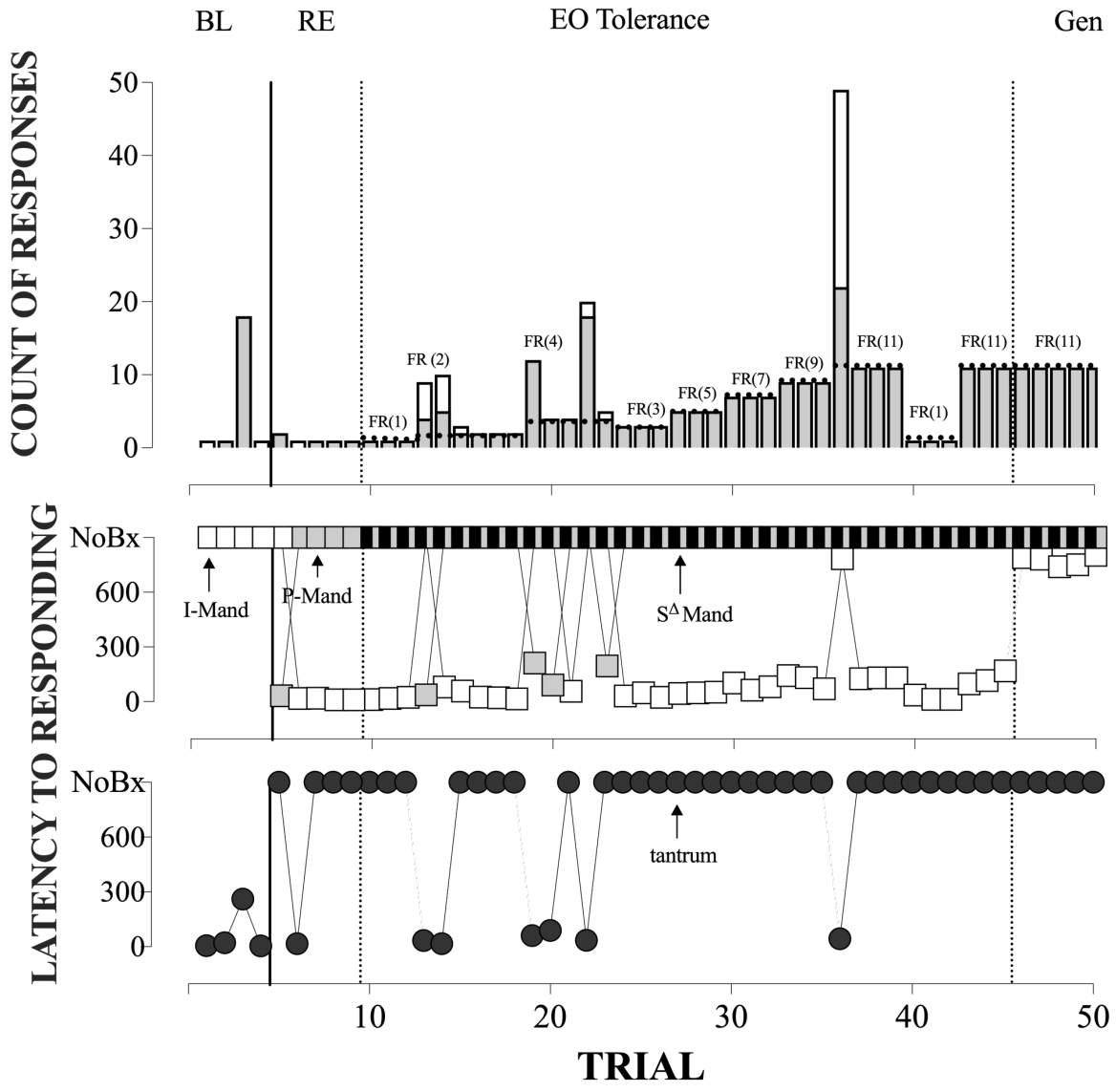
Note. BL = baseline; RE = response elimination; EO = establishing operations; GEN = generalization; criterion requirements are indicated in the top panel near the marked goal line for each criterion

The escape treatment evaluation (Figure 11) was completed across 50 trials and 21 appointments. Zeke dissented from three appointments during this phase of treatment. In all, the evaluation took approximately 912 min with 17.3% of session time spent in SR. The changing criterion design depicted in the top panel of Figure 11 included nine opportunities for demonstrations of effect, including one instance of reverting to a previous criterion. Specifically, we reverted to a FR(1) S^{Δ} schedule component after Zeke mastered the FR(11) criterion. Stable data patterns were observed within each criterion before researchers discontinued data collection, with one exception. After observing decreased latencies to tantrum behaviors and variable levels of compliance during the third criterion (i.e., FR 4) the team lowered the response requirement. All other changes between criteria led to immediate differences in levels of compliance that remained stable across the phase. Thus, researchers identified a functional relation between the FIMB framework and increased levels of compliance for Zeke.

Data in the bottom panel of Figure 11 were analyzed concurrently. As seen in the middle panel, researchers observed levels of independent mands that corresponded with increasing schedule requirements across the EO tolerance phase. Researchers never observed S^{Δ} mands but did observe variable levels of prompted mands in the first several trials of both phases of treatment. Additionally, observed a resurgence of tantrum behavior during the early stages of EO tolerance. By the later trials of this phase, tantrum behavior had returned to zero levels.

Figure 11

Zeke's Escape Treatment Evaluation



Note. BL = baseline; RE = response elimination; EO = establishing operations; GEN = generalization; criterion requirements are indicated in the top panel above each marked goal line; white columns in the top panel indicate the total number of demands presented and gray columns indicate demands that resulted in compliance

Interobserver Agreement

Mean IOA across all baseline and treatment trials is displayed in Table 1. Mean IOA for Jerome was 96% (range, 95.8-96.2%) across 43.2% of baseline (i.e., FA sessions), 97.1% (range, 95-100%) across 36% of response elimination trials, and 99% (range, 98.-100%) across 33% of EO tolerance and generalization trials. Mean IOA for Serenity was 96.3% across 50% of baseline, 98.2% across response elimination trials, and 97.8% across EO tolerance and generalization trials. Mean IOA for Zeke was 94.7% across 40% of baseline, 94.3% (range, 90.7-98.14%) across 35.3% of response elimination trials, and 90.3% (range, 88.1-99.3%) across 35.3% of EO tolerance and generalization trials.

CHAPTER III

STUDY 2 METHOD

The research team designed Study 2 to contextualize the data collected during the structured appointments described in Study 1. Despite the overwhelmingly positive outcomes achieved in Study 1, researchers were initially spending a significant amount of time at the end of every appointment responding to crisis situations and managing severe behavior. Further, these treatment gains were not generalizing to the children's behavior outside of the structured appointments (i.e., in the researchers' absence). The purpose of Study 2 was to evaluate the effect of embedding formative data triangulation into the FIMB framework on caregiver reports of tantrum behaviors. We also sought to evaluate the effects of this treatment model on improvements in the family's QoL.

Setting

Because Study 2 took outside of structured appointments, the primary setting was the family's home (i.e., not just the children's bedroom). The home itself encompassed four common living areas (i.e., the kitchen, a sitting area attached to the kitchen, a family exercise room, and a den with a television in the basement) in addition to the children's bedrooms (described in Study 1). However, some sessions occurred across community locations such as the local dance studio, a neighborhood park, a grocery store, or a children's museum.

All mealtime routines and food preparation occurred in the kitchen (i.e., food was not allowed in other areas of the home). The den area included a couch, rug, and large television equipped with several streaming services. The sitting area was attached to the kitchen and included two couches along with all "family" level books, board games, and toys (i.e., anything that was supposed to be shared across siblings). The family exercise room was situated in front

of the sitting area and kitchen. It encompassed a stationary bike, indoor trampoline, weightlifting equipment, several jump ropes, and a hula hoop. Kim did not report on the children's behavior if she was not directly supervising them during that time (i.e., across the school day).

Materials

Kim completed all data collection and video coding using her smartphone. No paper and pencil data collection were used. During this phase of the project, supplemental materials were developed as part of each child's behavior plan and in response to Kim's feedback. The team developed visual supports depicting intervention contingencies for Serenity (see Appendix E) in addition to an interactive visual schedule for each child to use across the day. As the children learned to use the visual schedule, the team created a task analysis with visual supports for the family's morning and evening routines (Appendix F). Additionally, the research team provided Kim with individualized parent training that required both a computer and a training binder.

Dependent Variables and Metrics

As a pre-post assessment, researchers distributed a survey to assess changes in the family's quality of life. As a more formative assessment, Kim reported data daily on each child's tantrum behavior (i.e., outside the context of appointments) across all phases of the study.

Family Quality of Life

Kim completed the 25-item Beach Center Family Quality of Life Scale (Hoffman et al., 2006). This self-report questionnaire contained five subscales (i.e., family interaction, parenting, emotional well-being, physical and material well-being, and disability-related support). In response to each item, Kim rated her satisfaction level using a 5-point Likert scale (e.g., 1 = very dissatisfied, 3 = neither satisfied nor dissatisfied, 5 = very satisfied). The FQoL scale was administered before and after the intervention evaluation. Pre-test results were used to create the template used to structure all formative triangulation meetings (described below). Changes from pre- to post-test were evaluated descriptively in conjunction with the daily parent reports.

Daily Reports of Tantrum Behavior

Researchers set up a daily data collection system through the HIPAA-compliant MyCap mobile application (Harris et al., 2022), which was integrated with a secure web platform (i.e., REDCap; Harris et al., 2009). At the end of each day, Kim received a push notification on her phone, indicating it was time to report information about the children. A separate survey was completed for each child.

Each survey began by asking Kim to report a count of tantrum behavior that had occurred throughout the day. If Kim reported zero occurrences, the survey was programmed to congratulate the family on a great day. If she reported any number above zero, a series of follow-up questions was enabled through branching logic. Kim was asked to recall the longest episode of challenging behavior from the day and rank the duration in comparison to the child's average tantrum length (i.e., score it as below average, average, or above average). She was then asked to recall the child's most intense episode of challenging behavior from the day and rank it in

comparison to the child's average tantrum intensity. Two open-ended questions were included. The first question asked if there was anything Kim wanted to share about the child's day. This question appeared on every survey, regardless of whether tantrums were reported. The second open-ended question asked if there was anything Kim wanted to share specifically related to the child's tantrum. The second question only appeared if Kim reported that tantrums had occurred. To promote reliable reporting, Kim was provided with a summary sheet of each child's operational definitions and customized intensity and duration scales. An example summary sheet is provided in Appendix G.

Researchers analyzed these daily reports and extracted information on (a) the count of tantrums reported for each child, (b) the reported intensity, and (c) the reported duration. The count of tantrums was used as the primary dependent variable for the treatment evaluation conducted in the unsupervised context. Kim's responses to the open-ended questions were considered during data triangulation meetings (described below).

Reporter

Kim served as the primary reporter for all of Study 2. Kim received approximately 5 hr of training on reporting procedures across 5 appointments (i.e., 1hr per appointment). During the first training session, researchers met with Kim to create operational definitions for each child's tantrum behavior. During this meeting, Kim was asked to list examples of behaviors she considered to be indicative of an average intensity tantrum for each child. Using this list as an anchor, she was then asked to create a list of below and above average examples of tantrum behavior. This process was repeated for setting the benchmarks for tantrum duration.

Between the first and second sessions, Kim was asked to save several video samples of the children engaging in tantrum behaviors through the family's Ring™ camera. Researchers directed Kim to record a minimum of three separate samples of tantrum behavior for each child (i.e., an average example, an above average example, and a below average example). Using the video samples that Kim had collected, a researcher modeled the process of reporting each child's behavior for all collected samples.

Kim collected a new set of videos between the next two appointments. During the third training session, Kim and the researcher completed the data collection processes separately before comparing results. This process was repeated until they reached full agreement on two separate video sets for each child. This occurred after the fifth training session.

Checks for Observer Drift

Checks for observer drift were repeated biweekly (i.e., every two weeks) using a novel set of video samples. During these biweekly check-ins, Kim and the researcher independently coded video samples as described above during training. If a disagreement occurred, the first author retrained the caregiver on the relevant definitions.

Experimental Design and Data Analysis

To evaluate the results of Study 2, researchers used a concurrent, multiple baseline across participants design (Baer et al., 1968) to assess the effects of embedding formative data triangulation into the FIMB framework on daily reports of tantrums for three children. To move between tiers, we used a combination of response-guided decisions and pre-established rules about treatment dosage.

Researchers required each participant to complete all phases of FIMB before beginning treatment with the next participant in the supervised context. Because of the potential for one child's behavior change to impact that of their siblings, we required a minimum lag of five data points (i.e., days) between tiers to conduct a vertical analysis for unplanned changes across tiers.

Formative Triangulation Meetings

Twice per week, all team members met for an hour to review all data (i.e., quantitative and qualitative data) collected across both studies. The team identified points of convergence (or divergence) across data sources and evaluated whether the documented trends aligned with anticipated patterns of behavior. During each meeting, the team completed a review of (a) graphed data from Study 1, (b) relevant FIMB categorical impact scores, and (c) Kim's daily reports of the children's tantrum behaviors.

When the team identified unanticipated data patterns, a problem-solving discussion began. Team members proposed adaptations to treatment based on family values, available evidence, and professional judgment. These proposed adaptations were discussed in relation to the identified critical domains of family quality of life measured with the Beach FQoL scale. Specifically, the five subscales (i.e., family interaction, parenting, emotional well-being, physical well-being, and disability-related support) were discussed in relation to identified focus items (i.e., items that received a low pre-test score). The team brainstormed action items to improve the FQoL that corresponded to the active phase of treatment. All procedural adjustments were documented and discussed with Kim prior to implementation.

Behavioral Skills Training

After all children had completed their treatment evaluations with researchers as therapists, the team shifted their focus to family-level support by providing behavioral skills training based on the guidelines of Latham (1994). Researchers (a) provided instruction on a targeted set of parenting skills, (b) modeled the skills for Kim, (c) provided Kim structured opportunities to practice the skills, and (d) provided immediate feedback for her use of the skills. As Kim mastered the skill in the role-play context, she was then provided with an opportunity to practice the skill with her children (i.e., with researcher support). Over the course of this training, we covered (a) foundations of behavior, (b) setting expectations, (c) creating structure, (d) building appropriate behaviors, (e) responding to challenging behavior, and (f) creating connections.

STUDY 2 RESULTS

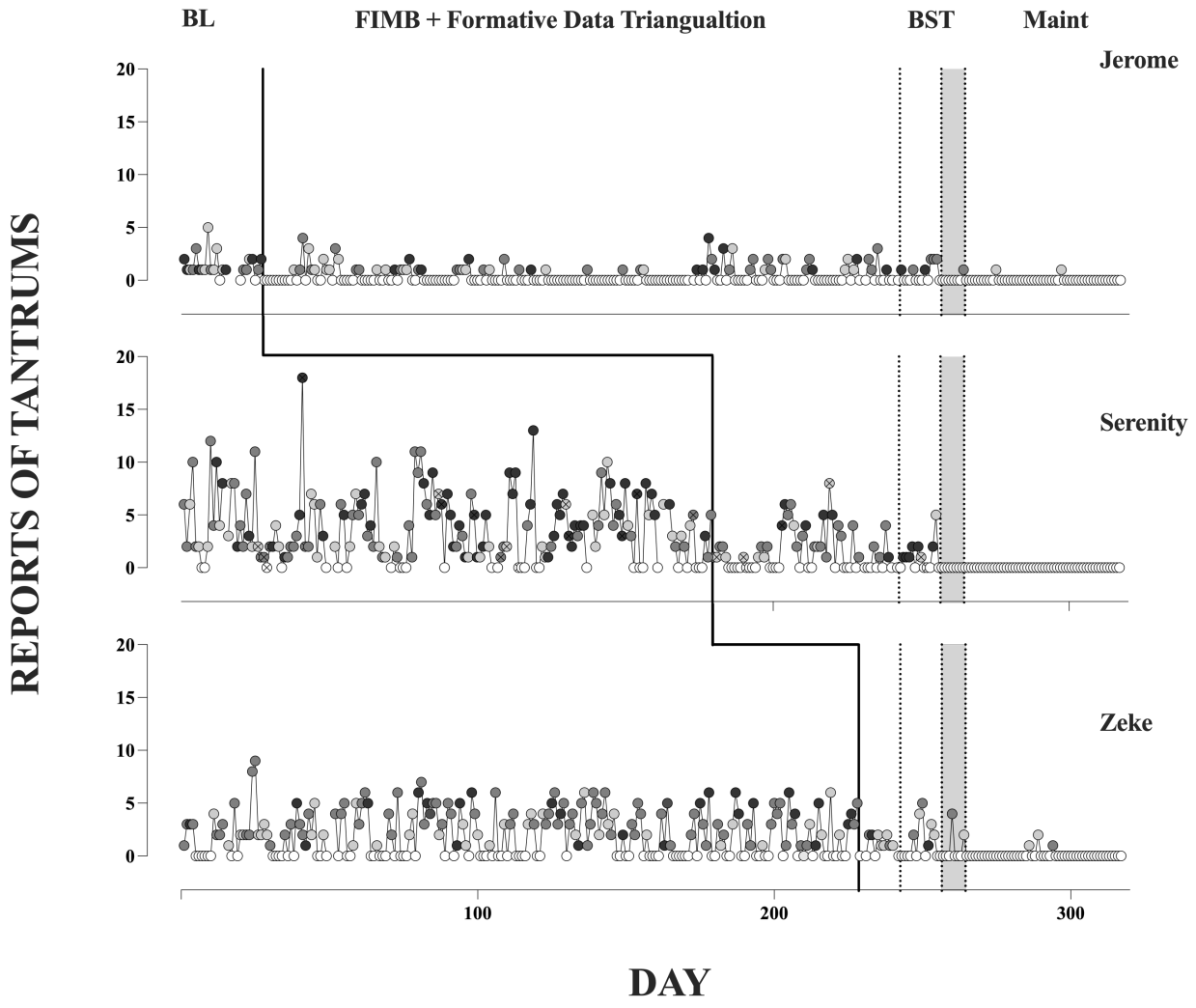
Reports of Generalized Behavioral Outcomes

Kim's daily reports of the children's tantrum behaviors are depicted in Figure 12. Variable and elevated levels of reported tantrums were observed across baseline conditions. Shifting from baseline to treatment led to immediate decreases in tantrum behaviors for both Jerome and Zeke. However, many data points overlapped with baseline for all participants in the initial phase of treatment. Changes were somewhat consistent across participants, with decreases in reports of tantrums observed for all children. No covariation was identified between tiers contingent with the introduction of treatment.

As the team shifted from focusing on individual child outcomes to parent and family training (i.e., starting with the shift to the BST phase on Day 242), reports of tantrum behavior dropped to zero levels. Starting on Day 257 (the shaded region), Kim was implementing the full treatment plan without support across the day for all children. Researchers continued to monitor Kim's fidelity to implementation for approximately ten days before shifting to maintenance for the final phase depicted on this graph. Levels of reported tantrums remained low for all children across the maintenance phase.

Figure 12

Family Level Treatment Evaluation



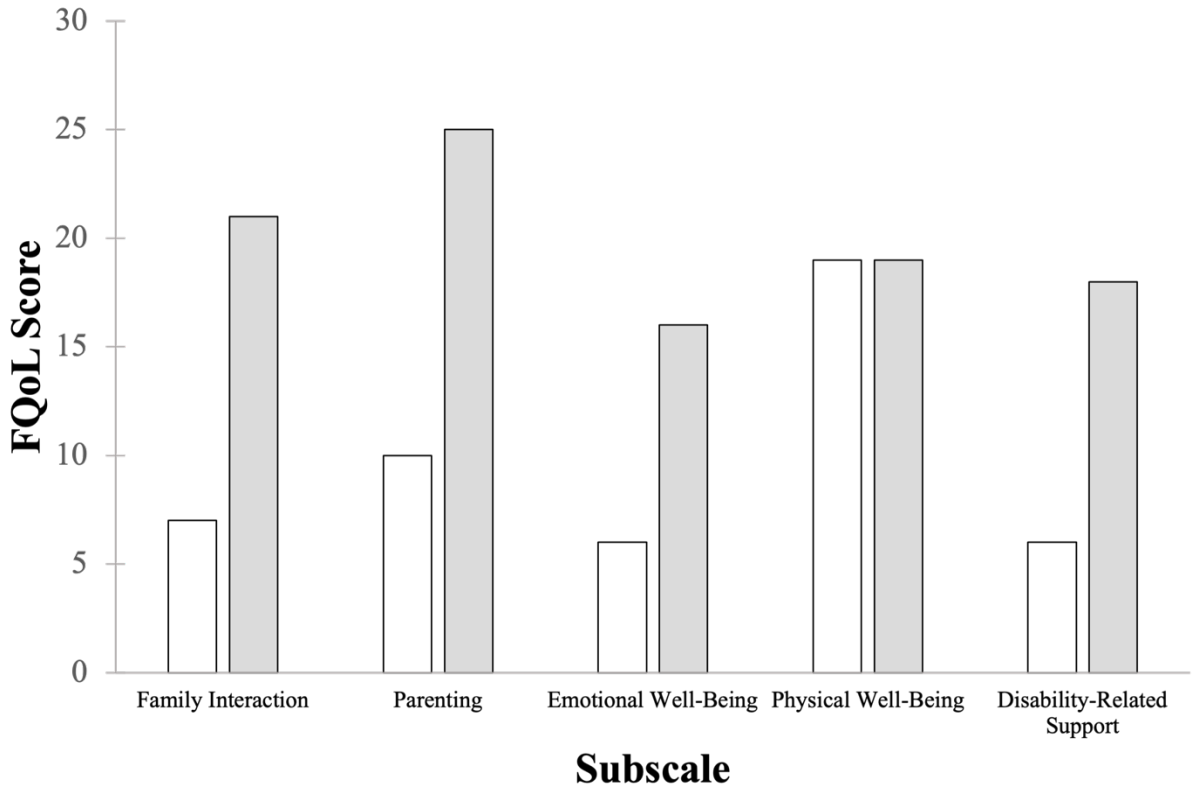
Notes. The intensity of each episode is reflected by the color of the data point (i.e., black = high intensity tantrum; dark gray = average intensity, light gray = below average intensity). BL = baseline; BST = behavioral skills training; Maint. = maintenance

Family Quality of Life

Results from the Beach Center Family Quality of Life Scale are summarized across five subscales (i.e., family interaction, parenting, emotional well-being, physical well-being, and disability-related support) in Figure 13. Kim reported increases in FQoL across all but one domain (i.e., physical well-being) from pre-test to post-test. Kim's responses to specific items included in the assessment are included in Appendix E.

Figure 13

Family Quality of Life



Note. White columns represent pre-test scores and gray columns represent post-test scores.

CHAPTER IV

DISCUSSION

The goal of this study was to evaluate the effectiveness of an iterative treatment model that prioritized family quality of life for three children diagnosed with IDD and a history of severe challenging behavior. In our first study, we used a changing criterion design to evaluate the effectiveness of using the Function-Informed Mechanisms-Based (FIMB) framework to increase levels of replacement behaviors and decrease tantrum behavior. Eight of nine treatment evaluations included in Study 1 progressed to EO tolerance (i.e., Jerome's attention and escape treatments were synthesized at this stage). All eight EO tolerance treatments were scored as highly effective and displayed strong maintenance (i.e., there was no degradation in categorical impact scores. Functional relations were identified between increased prosocial behaviors and implementation of the FIMB framework for 8 of the 9 treatment evaluations.

In a second study, we interpreted these results based on caregiver reports of the children's behavior outside of appointments. Through a process of formative data triangulation that culminated in comprehensive parent training, researchers expanded the existing treatment plan to promote generalizable decreases in challenging behavior and socially significant improvements in the family's quality of life. These results provide a demonstration of socially valid outcomes following intensive, individualized support provided to three children with IDD. While it is not surprising that we achieved clear demonstration of effect in Study 1, outcomes of Study 2 represent a novel contribution to the behavior analytic literature base.

Even though disability impacts the whole family (Turnbull et al., 2006), behavior analytic studies often target the needs of the individual engaging in challenging behavior independent of the needs of the family. In fact, intensive interventions most likely to address severe challenging behavior are typically guided by single case logic, which often involves context-bound demonstrations of behavior change. To address this gap in the literature, researchers completed an analysis of both context-dependent behavior change (i.e., performance within sessions) and what was conceptualized as more generalized changes in challenging behavior (i.e., tantrum behaviors reported across the day, outside of sessions). The subsequent sections include a summative evaluation of results across measures, a description of explanations for the findings, an overview of limitations, and finally a discussion implications for future research and practice.

Summary of Results Across Cases

A functional relation was identified between implementation of FIMB and increased replacement behaviors for all participants. Within sessions, we observed increased levels compliance and mands in the presence of variables that historically evoked tantrum behavior. These results maintained when Kim was taught to implement the treatment during the generalization phase of the supervised context.

When comparing these results with Kim's daily reports of tantrum behavior outside of sessions, data suggest a possible demonstration of generalized change in behavior that maintained and improved following treatment in the supervised plane. Researchers continued to collect daily reports for approximately two months following Zeke's final session in the supervised context. Effects maintained for all participants. Despite the limitations associated with the data collected in the supervised context (summarized below), the improvements in the FQoL scores (i.e., in four out of five domains) bolster our confidence about drawing conclusions from the parent reports of daily tantrums.

Limitations

The results of this study must be interpreted with a few limitations in mind. Despite frequent assessment of interobserver agreement in the supervised context, it was not possible to replicate this process in the unsupervised plane. Kim was sole caregiver in the household; however, even with an additional caregiver the requirements and logistics would create an unnecessary burden for the family. To mitigate the limitations the data collected in this context, we structured the MyCap app to disallow retroactive reporting and provided ongoing checks of Kim's knowledge of the definitions (i.e., when Kim and the researcher met and coded video sample as described in the training sections). Two instances of re-training were conducted.

Despite including three children and numerous treatment evaluations, this study included only one family. It is impossible to know if the effects demonstrated in Study 2 would be replicated by another caregiver. Additionally, the unique dynamics of the included family (i.e., six adopted children with disabilities and a single elderly caregiver) raises questions about the generalizability of the findings.

Because we were only able to administer the FQoL assessment at two time points for a single family, it is not possible to identify when the scores changed in the treatment process. To mitigate this limitation, the research team collected and consumed daily qualitative reports from Kim about the children's days. While these reports were instrumental in making clinical decisions, it should be noted that the research team consists of novice qualitative analysts. Thus, a more complex evaluation was not conducted at this time.

Implications for Research and Practice

Not long ago, ABA was considered a mechanistic intervention for ensuring that people with ASD left treatment indistinguishable from their neurotypical peers (Lovaas, 1987). This idea aligns with an outdated, medical model of disability that aims to cure individual pathology (Baker, 2011). The medical model of disability does not acknowledge how culture may define "normal" ability, nor does it distinguish between conditions that result from neurological pathology and those that reflect a misalignment between an individual and their environment (Kapp et al., 2013; Wolfensberger, 1970). As the field transitions, it is fitting to focus on evaluating dependent variables that prioritize and contextualize socially valid outcomes.

Despite several proposed models to improve collaborative processes designed to expand the social validity and generality of treatment effects (e.g., Moore & Amado, 2021), few are accompanied with empirical demonstrations of impact (see Lambert et al., 2024 for a notable exception). Embedding opportunities for formative data triangulation, centered around the tenets of FQoL, created an opportunity to expand the problem-solving framework proposed in Lambert, Copeland, et al. (2022). In light of our promising results, we are reminded of the need for additional measures of treatment effectiveness in our daily practice (i.e., not just in the context of

research). As was highlighted by Study 2, we are otherwise at risk of declaring success before our treatment plan results in generalized outcomes of behavior change.

The time and effort required for the family to contact this level of success was substantial and should be noted. For example, consider the cost of providing direct services to three children (i.e., each requiring at least two trained therapists) across four appointments per week, with each appointment lasting a minimum of 2 hr. However, appointments often lasted longer if the team was responding to dangerous behaviors. In total, we spent XYZ hours at the family's house for just under a year (i.e., 317 days). Further, the caregiver had continuous access to a behavior analyst via phone or text for questions and concerns about each child, their goals, and the treatment plans. When considering viable options for treatment of severe challenging behavior, these costs (e.g., both financial and to the FQoL), along with the documented benefits, should be weighed against those of business-as-usual procedures (e.g., residential treatment). Future research should explore all of these variables in an effort to impact families' access to services through common funding agencies (e.g., health insurance).

In future research, each decision point made in the problem-solving process should be coded and analyzed with a qualitative lens by a team of experts. This may promote ongoing efforts to improve nuanced frameworks designed for practitioner use. The framework described in the current paper requires advanced training to navigate decision points and may not be replicable for practitioners without additional guidance.

Conclusion

Researchers estimate that anywhere from 17% to 69% of individuals with IDD engage in challenging behaviors including aggression (e.g., Fitzpatrick et al. 2016; Kanne & Mazurek, 2011), self-injury (Richards et al., 2012; Soke et al., 2016), and elopement (Anderson et al., 2012). These dangerous behaviors have been linked with negative outcomes including impaired social relationships (Fitzpatrick et al., 2016), increased risk of abuse (Stith, et al., 2009), and restricted access to community settings (Guercio, 2022). Collectively, these risk factors can have a significant impact on the quality of life of individuals with IDD (e.g., Biggs & Carter, 2016). Researchers have found that children and adolescents with IDD who engage in persistent challenging behavior often report a lower QoL as compared to typically developing peers, or even peers with IDD who do not engage in challenging behavior (Clark et al., 2015; de Vries & Geurts 2015; Tavernor et al., 2013).

One child's severe challenging behavior creates an immense risk factor for the quality of life for an entire family unit. Thus, it is essential to continue researching highly effective, function-based treatments that promote generalized changes in patterns of behavior. As Scott et al. (2023) recently emphasized, the most crucial aspect of a behavioral intervention may be the ability to sustain its results over time. As the field moves towards more global outcomes of behavioral treatments, we must consider data sources in addition to our typical, context-bound demonstrations of behavior change. It may be that incorporating formative data triangulation into an existing, iterative decision making framework is the key to promoting generality.

References

Allen, K. D., & Warzak, W. J. (2000). The problem of parental nonadherence in clinical behavior analysis: effective treatment is not enough. Journal of Applied Behavior Analysis, 33(3), 373–391. https://doi.org/10.1901/jaba.2000.33-373

APA Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. The American Psychologist, 61(4), 271-285.

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. (2002, amended effective January 1, 2017). https://www.apa.org/ethics/code/

Anderson, C., Law, J. K., Daniels, A., Rice, C., Mandell, D. S., Hagopian, L., & Law, P. A. (2012). Occurrence and Family Impact of Elopement in Children with Autism Spectrum Disorders. Pediatrics, 130(5), 870–877. https://doi.org/10.1542/peds.2012-0762

Ayres, M., Parr, J. R., Rodgers, J., Mason, D., Avery, L., & Flynn, D. (2018). A systematic review of quality of life of adults on the autism spectrum. Autism, 22(7), 774–783. https://doi.org/10.1177/1362361317714988

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1(1), 91–97. https://doi.org/10.1901/jaba.1968.1-91

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behavior Analysis, 12(2), 199-210. https://doi.org/10.1901/jaba.1979.12-199

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. http://dx.doi.org/10.1002/jaba.30.

Beaulieu, L., Addington, J., & Almeida, D. (2019). Behavior analysts' training and practices regarding cultural diversity: The case for culturally competent care. Behavior Analysis in Practice, 12(3), 557–575. https://doi.org/10.1007/s40617-018-00313-6

Behavior Analyst Certification Board. (2020). Ethics code for behavior analysts. https://bacb.com/wp-content/ethics-code-for-behavior-analysts

Berg, W. K., Wacker, D. P., Ringdahl, J. E., Stricker, J., Vinquist, K., Salil Kumar Dutt, A., Dolezal, D., Luke, J., Kemmerer, L., & Mews, J. (2016). An integrated model for guiding the selection of treatment components for problem behavior maintained by automatic reinforcement. Journal of Applied Behavior Analysis, 49(3), 617-638. https://doi.org/10.1002/jaba.303

Biggs, E. E., & Carter, E. W. (2016). Quality of life for transition-age youth with autism or intellectual disability. Journal of Autism and Developmental Disorders, 46, 190–204. https://doi.org/10.1007/s10803-015-2563-x.

Brannan, A.M., Heflinger, C.A., Bickman, L. (1997). The caregiver strain questionnaire: Measuring the impact on the family of living with a child with serious emotional disturbance. Journal of Emotional and Behavioral Disorders, 5(4) 212-222. https://doi.org/10.1177/106342669700500404

Bruininks, R., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1996). Scales of Independent Behavior–Revised (SIB-R). Riverside.

Byles, J., Byrne, C., Boyle, M. H., & Offord, D. R. (1988). Ontario Child Health Study: reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. Family Process, 27(1), 97-104. https://doi.org/10.1111/j.1545-5300.1988.00097.x

Carr, J. E., Austin, J. L., Britton, L. N., Kellum, K. K., & Bailey, J. S. (1999). An assessment of social validity trends in applied behavior analysis. Behavioral Interventions, 14, 223–231. https://doi.org/10.1002/(SICI)1099-078X(199910/12)14:4<223::AID-BIN37>3.0.CO;2-Y

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. Journal of Applied Behavior Analysis, 18(2), 111-126 https://doi.org/10.1901/jaba.1985.18-111

Clark, B. G., Magill-Evans, J. E., & Koning, C. J. (2015). Youth with autism spectrum disorders: Self- and proxy-reported quality of life and adaptive functioning. Focus on Autism and Other Developmental Disabilities, 30, 57–64. https://doi.org/10.1177/1088357614522289

Contreras, B. P., Hoffmann, A. N., & Slocum, T. A. (2022). Ethical Behavior Analysis: Evidence-Based Practice as a Framework for Ethical Decision Making. Behavior Analysis in Practice, 15(2), 619–634. https://doi.org/10.1007/s40617-021-00658-5

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). Applied behavior analysis. Pearson UK.

D'Agostino, S. R., Douglas, S. N., & Duenas, A. D. (2019). Practitioner-implemented naturalistic developmental behavioral interventions: Systematic review of social validity practices. Topics in Early Childhood Special Education, 39(3), 170-182. https://doi.org/10.1177/0271121419854803

Dempsey, C. M., Iwata, B. A., Fritz, J. N., & Rolider, N. U. (2012). Observer training revisited: A comparison of in vivo and video instruction. Journal of Applied Behavior Analysis, 45(4), 827-832. https://doi.org/10.1901/jaba.2012.45-827

Dekker, M. C., Koot, H. M., Ende, J. V. D., & Verhulst, F. C. (2002). Emotional and behavioral problems in children and adolescents with and without intellectual disability. Journal of

Child Psychology and Psychiatry, 43(8), 1087-1098. https://doi.org/10.1111/1469-7610.00235

De Vries, M., & Geurts, H. M. (2015). Influence of autism traits and executive functioning on quality of life in children with an autism spectrum disorder. Journal of Autism and Developmental Disorders, 45(9), 2734–2743. https://doi.org/10.1007/s10803-015-2438-1

Dunlap, G., & Fox, L. (2011). Function-based interventions for children with challenging behavior. Journal of Early Intervention, 33(4), 333-343. https://doi.org/10.1177/105381511142997

Emerson, E., Kiernan, C., Alborz, A., Reeves, D., Mason, H., Swarbrick, R., Mason, L., & Hatton, C. (2001). The prevalence of challenging behaviors: a total population study. Research in Developmental Disabilities, 22(1), 77–93. https://doi.org/10.1016/S0891-4222(00)00061-5

Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster family assessment device. Journal of Marital and Family Therapy, 9(2), 171–180. https://doi.org/10.1111/j.1752-0606.1983.tb01497.x

Farmer, C. A., & Aman, M. G. (2011). Aggressive behavior in a sample of children with autism spectrum disorders. Research in Autism Spectrum Disorders, 5(1), 317-323. https://doi.org/10.1016/j.rasd.2010.04.014

Ferguson, J. L., Cihon, J. H., Leaf, J. B., Van Meter, S. M., McEachin, J., & Leaf, R. (2018). Assessment of social validity trends in the Journal of Applied Behavior Analysis. European Journal of Behavior Analysis, 20(1), 146–157. https://doi.org/10.1080/15021149.2018.1534771

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement. Appleton-Century-Crofts.

Finn, L. L. (2020). Improving interactions between parents, teachers, and staff and individuals with developmental disabilities: a review of caregiver training methods and results. International Journal of Developmental Disabilities, 66(5), 390–402. https://doi.org/10.1080/20473869.2020.1830460

Fitzpatrick, S. E., Srivorakiat, L., Wink, L. K., Pedapati, E. V., & Erickson, C. A. (2016). Aggression in autism spectrum disorder: presentation and treatment options. Neuropsychiatric Disease and Treatment, 12(1), 1525-1538. https://doi.org/10.2147/NDT.S84585

Geiger, K. B., Carr, J. E., & LeBlanc, L. A. (2010). Function-based treatments for escape-maintained problem behavior: A treatment-selection model for practicing behavior analysts. Behavior Analysis in Practice, 3(1), 22-32. https://doi.org/10.1007/BF03391755

Ghaemmaghami, M., Hanley, G. P., & Jessel, J. (2021). Functional communication training: From efficacy to effectiveness. Journal of Applied Behavior Analysis, 54(1), 122–143. https://doi.org/10.1002/jaba.762

Greer, B. D., Fisher, W. W., Saini, V., Owen, T. M., & Jones, J. K. (2016). Functional communication training during reinforcement schedule thinning: An analysis of 25 applications. Journal of Applied Behavior Analysis, 49(1), 105-121. https://doi.org/10.1002/jaba.265

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

Hammond, J. L., Iwata, B. A., Rooker, G.W., Fritz, J. N. & Bloom, S. E. (2013). Effects of fixed versus random condition sequencing during multielement functional analyses. Journal of Applied Behavior Analysis, 46(1), 22-30. https://doi.org/10.1002/jaba.7

Hanley, G. P., Jin, C. S., Vanselow, N. R., & Hanratty, L. A. (2014). Producing meaningful improvements in problem behavior of children with autism via synthesized analyses and treatments. Journal of Applied Behavior Analysis, 47(1), 16-36. https://doi.org/10.1002/jaba.106

Hartman, D. P., & Hall, R. V. (1976). The changing criterion design. Journal of Applied Behavior Analysis, 9(1), 527–532. https://doi.org/10.1901/jaba.1976.9-527

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. 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

Harris, P. A., Swafford, J., Serdoz, E. S., Eidenmuller, J., Delacqua, G., Jagtap, V., Taylor, R. J., Gelbard, A., Cheng, A. C., & Duda, S. N. (2022). MyCap: a flexible and configurable platform for mobilizing the participant voice. JAMIA Open, 5(2), 1-6. https://doi.org/10.1093/jamiaopen/ooac047

Hoffman, L., Marquis, J., Poston, D., Summers, J. A., & Turnbull, A. (2006). Assessing family outcomes: Psychometric evaluation of the beach center family quality of life scale. Journal of Marriage and Family, 68(4), 1069-1083. https://doi.org/10.1111/j.1741-3737.2006.00314.x

Hurl, K., Wightman, J., Haynes, S. N., & Virues-Ortega, J. (2016). Does a pre-intervention functional assessment increase intervention effectiveness? A meta-analysis of within-

subject interrupted time-series studies. Clinical Psychology Review, 47, 71–84. https://doi.org/10.1016/j.cpr.2016.05.003

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

Kanne, S. M., & Mazurek, M. O. (2011). Aggression in children and adolescents with ASD: Prevalence and risk factors. Journal of Autism and Developmental Disorders, 41(7), 926–937. https://doi.org/10.1007/s10803-010-1118-4.

Kennedy, C. H. (1992). Trends in the measurement of social validity. The Behavior Analyst, 15, 147–156. https://doi.org/10.1007/BF03392597

Kurtz, P. F., Leoni, M., & Hagopian, L. P. (2020). Behavioral approaches to assessment and early intervention for severe problem behavior in intellectual and developmental disabilities. Pediatric Clinics, 67(3), 499-511. https://doi.org/10.1016/j.pcl.2020.02.005

Lambert, J. M., Copeland, B. A., Paranczak, J. L., Macdonald, M. J., Torelli, J. N., & Houchins-Juarez, N. J. (2022). Description and evaluation of a function-informed and mechanisms-based framework for treating challenging behavior. Journal of Applied Behavior Analysis, 55(4), 1193–1219. https://doi.org/10.1002/jaba.940

Lambert, J. M., Sandstrom, A. L., Hodapp, R. M., Copeland, B. A., Paranczak, J. L., Macdonald, M. J., & Houchins-Juarez, N. J. (2022). Revisiting the social validity of services rendered through a university-based practicum addressing challenging behavior. Journal of Applied Behavior Analysis, 55(4), 1220-1238. https://doi.org/10.1002/jaba.939

LeBlanc, L. A., Raetz, P. B., Sellers, T. P., & Carr, J. E. (2016). A proposed model for selecting measurement procedures for the assessment and treatment of problem behavior. Behavior Analysis in Practice, 9(1), 77-83. https://doi.org/10.1007/s40617-015-0063-2

Ledford, J. R., Bailey, K. M., Pak, N. S. (2024). Analyzing data from studies using rapid iterative alternation. In J. R. Ledford & D. L. Gast (Eds.), Single case research methodology (4th ed., pp. 247-261). Routledge. http://10.4324/9781003294726-16

Ledford, J. R., & Gast, D. L. (2024). Single case research methodology (4th ed). Routledge.

Ledford, J. R., & Lane, J. D. (2024). Reliability and validity of dependent variables. In J. R. Ledford & D. L. Gast (Eds.), Single case research methodology (4th ed., pp. 69-82). Routledge. http://doi.org/10.4324/9781003294726-6

Ledford, J. R., Hall, E., Conder, E., & Lane, J. D. (2016). Research for young children with autism spectrum disorders: Evidence of social and ecological validity. Topics in Early Childhood Special Education, 35(4), 223–233. https://doi.org/10.1177/0271121415585956

Ledford, J. R., Tuck K. N. (2024). Conducting studies using sequential introduction and withdrawal of conditions. In J. R. Ledford & D. L. Gast (Eds.), Single case research methodology (4th ed., pp. 137-160). Routledge. http://10.4324/9781003294726-13

Ledford, J. R., Tuck K. N. (2024). Conducting studies using time-lagged condition ordering. In J. R. Ledford & D. L. Gast (Eds.), Single case research methodology (4th ed., pp. 179-203). Routledge. https://doi.org/10.4324/9781003294726-11

Lyons, G. (2011). Quality of life for persons with intellectual disabilities: A review of the literature. In R. Kober (Ed.), Enhancing the quality of life of people with intellectual

disabilities: From theory to practice (73-126). Springer. https://doi.org/10.1007/978-90-481-9650-0_15

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

O'Neill, R. E., Horner, R. H., Albin, R. W., Sprague, J. R., Storey, K., & Newton, J. S. (1997). Functional assessment and program development for problem behavior: A practical handbook (2nd ed.). Belmont, CA: Wadsworth

Richards, C., Oliver, C., Nelson, L., & Moss, J. (2012). Self-injurious behaviour in individuals with autism spectrum disorder and intellectual disability: Self-injury in autism spectrum disorder. Journal of Intellectual Disability Research, 56(5), 476–489. https://doi.org/10.1111/j.1365-2788.2012.01537.x

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(1), 130-146. https://doi.org/10.1002/jaba.13

Rooker, G. W., Jessel, J., Kurtz, P. F., & Hagopian, L. P. (2013). Functional communication training with and without alternative reinforcement and punishment: An analysis of 58 applications. Journal of Applied Behavior Analysis, 46(4), 708-722. https://doi.org/10.1002/jaba.76

Saini, V., Jessel, J., Iannaccone, J. A., & Agnew, C. (2019). Efficacy of functional analysis for informing behavioral treatment of inappropriate mealtime behavior: A systematic review

and meta-analysis. Behavioral Interventions, 34(2), 231-247. https://doi.org/10.1002/bin.1664

Schaefer, J. M., & Andzik, N. R. (2021). Evaluating behavioral skills training as an evidence-based practice when training parents to intervene with their children. Behavior Modification, 45(6), 887-910. https://doi.org/10.1177/0145445520923996

Schalock, R., Baker, A., Claes, C., Gonzalez, J., Malatest, R., Van Loon, J., Verdugo, M. A., & Wesley, G. (2018). The use of quality-of-life scores for monitoring and reporting, quality improvement, and research. Journal of Policy and Practice in Intellectual Disabilities, 15(3), 176–182. https://doi.org/10.1111/jppi.12250

Schalock, R. L., & Verdugo, M. A. (2002). Handbook on quality of life for human service practitioners. Washington, DC: American Association on Mental Retardation.

Schwartz, I. S., & Kelly, E. M. (2021). Quality of life for people with disabilities: Why applied behavior analysts should consider this a primary dependent variable. Research and Practice for Persons with Severe Disabilities, 46(3), 159-172. https://doi.org/10.1177/15407969211033629

Scott, V., Saini, V., Busch, L., & Solomon, N. (2023). On the longevity of behavioral interventions for challenging behavior. Behavioral Interventions. https://doi.org/10.1002/bin.1929

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. Annual Review of Clinical Psychology, 4(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Slocum, T. A., Detrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The evidence-based practice of applied behavior analysis. The Behavior Analyst, 37(1), 41–56. https://doi.org/10.1007/s40614-014-0005-2

Snodgrass, M. R., Chung, M. Y., Meadan, H., & Halle, J. W. (2018). Social validity in single-case research: A systematic literature review of prevalence and application. Research in Developmental Disabilities, 74, 160–173. https://doi.org/10.1016/j.ridd.2018.01.007

Soke, G. N., Rosenberg, S. A., Hamman, R. F., Fingerlin, T., Robinson, C., Carpenter, L., Giarelli, E., Lee, L.-C., Wiggins, L. D., Durkin, M. S., & DiGuiseppi, C. (2016). Brief Report: Prevalence of Self-injurious Behaviors among Children with Autism Spectrum Disorder—A Population-Based Study. Journal of Autism and Developmental Disorders, 46(11), 3607–3614. https://doi.org/10.1007/s10803-016-2879-1

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., Som, A., McPherson, M., & Dees, J. E. M. E. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. Aggression and Violent Behavior, 14(1), 13–29. https://doi.org/10.1016/j.avb.2006.03.006

Sturmey, P., Seiverling, L., & Ward-Horner, J. (2008). Assessment of Challenging Behaviors in People with Autism Spectrum Disorders. In Clinical Assessment and Intervention for Autism Spectrum Disorders (pp. 131–163). Elsevier Inc. https://doi.org/10.1016/B978012373606-2.50007-3

Sun, X. (2022). Behavior skills training for family caregivers of people with intellectual or developmental disabilities: a systematic review of literature. International Journal of Developmental Disabilities, 68(3), 247–273. https://doi.org/10.1080/20473869.2020.1793650

Tavernor, L., Barron, E., Rodgers, J., & McConachie, H. (2013). Finding out what matters: Validity of quality of life measurement in young people with ASD. Child, 39(4), 592–601. https://doi.org/10.1111/j.1365-2214.2012.01377.x.

The WHOQOL Group: WHOQOL User Manual, and annexes (WHO/MSA/MHP/98.3). World Health Organization; 2012.

Touchette, P. E., & Howard, J. S. (1984). Errorless learning: Reinforcement contingencies and stimulus control transfer in delayed prompting. Journal of Applied Behavior Analysis, 17(2), 175-188. https://doi.org/10.1901/jaba.1984.17-175

Turnbull, H. R., & Brunk, G. L. (1990). Quality of life and public philosophy. In R. Schalock (Ed.), Quality of life: Perspectives and issues (193-209). American Association on Mental Retardation.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis, 11(2), 203–214. https://doi.org/101901/jaba.1978.11-203

Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., & Berkoff, K. A. (1990). Psychometric Characteristics of the Multidimensional Scale of Perceived Social Support. Journal of Personality Assessment, 55(3-4), 610–617. https://doi.org/10.1080/00223891.1990.9674095

Zuna, N., Summers, J. A., Turnbull, A. P., Hu, X., & Xu, S. (2010). Theorizing About Family Quality of Life. In Enhancing the Quality of Life of People with Intellectual Disabilities (pp. 241–278). Springer Netherlands. https://doi.org/10.1007/978-90-481-9650-0_15

Appendix A

Session Materials

Bin	Jerome	Serenity	Zeke
Yellow Bin	• Play-Doh™ • parachute • basketball • ponies • Uno™ • Batman™ cave • nail polish • comic book • mini golf • magic kit	• 100-piece puzzle • Pokémon cards • tattoos • art supplies • word search • LEGO™ kits • toy dragons • comic book • science kit • PollyPocket™	• Ninjago™ • slime kit • LiteBrite™ • SpotIt! ™ • MadLibs ™ • action figures • Magnatiles™ • Hot Wheels™ • chalk • marble run
Blue Bin	• slime kit • sticker book • PollyPocket™ • chalk • Marble Run • crayons • hair supplies • action figures • Magnatiles™ • Hot Wheels™	• bubbles • digital camera • painting kit • nature kit • toy cats • walkie talkies • DroneHome • farting pen • Kinetic sand • LiteBrite™	• airplane kit • HungryHippos • push pop game • LEGO™ kits • art supplies • knight storybook • pirate ship • parachute • comic book • Batman™ cave
Green Bin	• dress up kit • doll house • pirate ship • nature kit • HungryHippos • Kinetic Sand • digital camera • mini drone • push pop game • painting kit	• Bluey figures • Parachute • slime skit • mini golf • magic kit • Ninjago™ • Magnatiles™ • slime skit • marble run • mini drone	• bubbles • digital camera • painting kit • nature kit • space ship • walkie talkies • DroneHome • farting pen • Kinetic sand • basketball

Appendix B

Example of Procedural Fidelity Data Sheet

Session: **Therapist:** **Prim/Reli:** **Phase:**		
Materials: "work" items for therapist to engage with (e.g., phone, papers), signal for delta (e.g., star)		
Pre-session tasks	**Yes**	**No**
1. Remove highly preferred items.		
2. Provide 30 s of high-quality attention.		
Steps to complete sessions	**Yes**	**No**
1. Remove attn within 5 s of starting session and engage in an alternate activity. (e.g., pretend to send emails, organize papers)		
2. Present visual signal for delta (e.g., red star) for predetermined duration. _____ *s delay*		
3. If needed, prompt mand following delta at: *0-s delay 10-s delay 20-s delay 30-s delay*		
4. Deliver 30 s of SR following mand. (i.e., prompted or independent)		
5. End session after delivering SR.		
Total session requirements	**Yes**	**No**
1. Delta reset contingent on target PB.		
2. Provide access moderately preferred items.		
3. No demands presented.		
Total		
Yes / [Yes + No] * 100 =		

Appendix C

Visual for Serenity's Behavior Plan

Magic Token

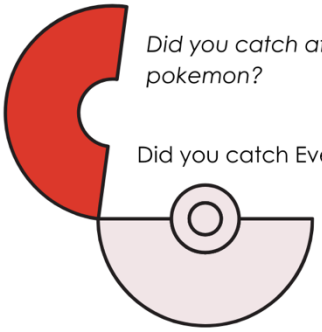
Morning Routine
Did I complete all steps?

Chores
Did I help my family?

Sibling TV Time
Did I respect Zeke and Jerome's time?

Self Control
No threats or aggression?

Bedtime Routine
Did I complete all steps?



Did you catch at least 4 pokemon?

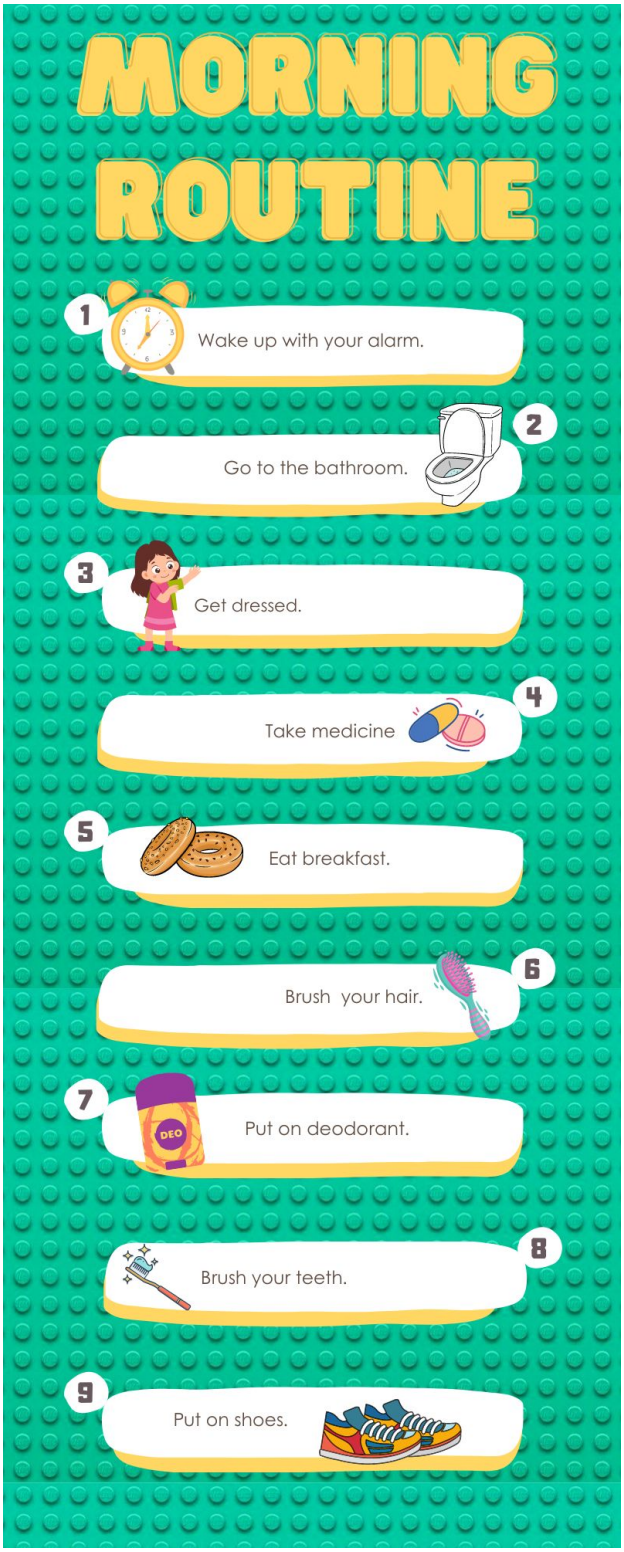
Did you catch Evee?

If yes, you have earned a magic token for tomorrow! Enjoy your bonus tech time!

If no, you let's come up with a plan of what we will do differently tomorrow.

Appendix D

Visual of Morning Routine



Appendix E

Summary Sheet for Parent Reported Data

Serenity Definitions

Tantrum: One episode begins any time Serenity starts to engage in physical aggression, property destruction, or verbal aggression. It ends when Serenity has been calm for 5 consecutive minutes, meaning she has not engaged in these targeted problem behaviors.

Risky Behavior: Code any instance of suicidal behavior, homicidal behavior, or other behaviors that could pose a threat to Serenity or another person or animal

Please indicate whether the behavior was **Below Average, Average,** or **Above Average** with respect to intensity and duration. The scale for both **intensity** and **duration** measures are described below.

Intensity Scale (Examples)

Below Average	Average	Above Average
• 1-2 occurrences of hitting/kicking property without causing damage • Isolated incident of throwing a small item at someone • Isolated incident of cursing at someone • Isolated incident of negative self-talk	• Dropping to the floor • Repeatedly hitting/kicking property without causing major damage • Repeatedly throwing small items at someone • Repeated incidents of cursing • Screaming or repeating self-deprecating statements • Threatening to harm herself of harm others	• Hitting, kicking, punching, scratching another person • Pulling hair • Throwing an item at someone with the potential to cause harm • Spitting at someone • Throwing or banging items and damaging property • Engaging in acts of risky behaviors

Duration Scale

Below Average	Average	Above Average
Less than 30 minutes	30 minutes to 1 hour	More than 1 hour

Appendix F

Family Quality of Life Scores

Item	Pre	Post
1. My family enjoys spending time together	1	3
2. My family members help the children learn to be independent	1	4
3. My family has the support we need to relieve stress	1	4
4. My family members have friends or others who provide support.	2	4
5. My family members help the children with schoolwork and activities.	1	3
6. My family has transportation to get to the places they need to be.	4	4
7. My family members talk openly with each other.	1	3
8. My family members teach the children how to get along with others.	1	4
9. My family members have some time to pursue our own interests.	1	4
10. Our family solves problems together.	1	4
11. My family members support each other to accomplish goals	1	4
12. My family members show that they love and care for each other.	2	4
13. My family has outside help available to us to take care of special needs of all family members.	2	4
14. Adults in our family teach the children to make good decisions.	3	5
15. My family gets medical care when needed.	5	5
16. My family has a way to take care of our expenses.	5	5
17. Adults in my family know other people in the children's lives (friends, teachers, etc.).	3	5
18. My family is able to handle life's ups and downs.	1	3
19. Adults in my family have time to take care of the individual needs of every child.	1	4
20. My family gets dental care when needed.	5	5
21. My family feels safe at home, work, school, and in our neighborhood.	2	4
22. My family member with a disability has support to accomplish goals at school or at workplace.	1	4
23. My family member with a disability has support to accomplish goals at home.	1	4
24. My family member with a disability has support to make friends.	1	5
25. My family has good relationships with the service providers who provide services and support to our family member with a disability.	3	5

Appendix G

Codebook of Jerome's Operational Definitions

Topography	Definition	Measurement	Example	Nonexample
Aggression				
Kicking	Contextually inappropriate, forceful physical strike of Jerome's foot or leg that contacts another person	code each strike	striking therapist in leg with foot	tripping over someone
Hitting	Contextually inappropriate, forceful physical strike of Jerome's hand, or an object he is controlling, that contacts another person	code each strike	punching, striking with an object	belly bump, hugs, high fives
Throwing	Contextually inappropriate instance of propelling an object through the air the direction of another person	code each item thrown	hurling/flinging a toy at someone	throwing a ball and catching it
Pushing	Any instance of applying contextually inappropriate, forceful physical contact against another person (e.g., using the hands or upper body) that displaces the person or creates resistance	code each time the person is displaced, or resistance is created	shoving a sibling off a chair,	bumping into someone in the hallway
Grabbing	Any contextually inappropriate application of a forceful grip onto another person's limb or article of clothing	code each time Jerome renews his grip	seizing a person by the arm/collar	grabbing a hand to guide another
Pinching	Any contextually inappropriate act of gripping/compressing a small area of another person's skin, typically using a thumb and fingers, to tightly squeeze a localized area of the skin	code each time Jerome applies pressure to the skin	squeezing another's skin	lightly touching another's body
Scratching	Using nails or another sharp object to scrape the surface of another person's skin typically by applying pressure or friction through repetitive back-and forth or circular motions	code each swipe (i.e., each new contact with the person's skin)	swiping another person's arm with his nails	impaling another with a pencil
Hairpulling	Any instance of grasping, tugging, or pulling another person's hair away from their scalp or body	code each time Jerome renews his grasp as a new instance	grabbing ponytail and yanking	brushing someone's hair
Biting	Any instance using his teeth to contact another person's body part (or clothing item) and applying pressure (i.e., closing the jaws/teeth together)	code each time Jerome closes his mouth around another person	biting someone's arm or shirt	biting a toy or his own arm
Choking	Any instance of Jerome applying pressure to another person's neck or throat with his hands/arms (or with an object) that impedes another person's airflow/breathing	code each time Jerome closes his hands/arms around another person's neck	squeezing hand(s) around throat, headlock	tightly squeezing around therapist waist/chest

Topography	Definition	Measurement	Example	Nonexample
Property Destruction				
Kicking	Contextually inappropriate, forceful physical strike of Jerome's foot or leg that contacts an object/surface	code each strike	striking wall with leg or foot	tripping over object
Hitting	Contextually inappropriate, forceful physical strike of Jerome's hand, or an object he is controlling, that contacts an object/surface	code each strike	punching, striking with object with an object	belly bump, hugs, high fives
Throwing	Any contextually inappropriate instance of propelling an object through the air the direction of another object/surface	code each item thrown	hurling/flinging a toy at wall	throwing a ball and catching it
Swiping	Any contextually inappropriate sweeping or arching movement of Jerome's hands/arms that displaces items from their original location(s)	code each new swipe and each time he changes directions of his arm movements	knocking items off the table with forearm	knocking over a glass by mistake when reaching for his plate
Self-Injury				
Biting	Any instance using his teeth to contact his own body part and applying pressure (i.e., closing the jaws/teeth together)	code each time Jerome closes his mouth around his arm	biting his own arm	chewing on a toy or his shirt
ISB				
Disrobing	Pulling pants/shorts/underwear more than 6" below his hips or unzipping/buttoning his pants and revealing his bottom or genitals	code each article of clothing that is removed or each time he reveals a specified body part	removing shorts in the middle of a work task	using the restroom or changing for swimming in a private area
Voiding	voiding urine outside of an appropriate receptacle or designated area	code each occurrence of releasing urine	urinating on the floor	urinating in the toilet

Appendix H

Codebook of Serenity's Operational Definitions

Topography	Definition	Measurement	Example	Nonexample
Aggression				
Kicking	Contextually inappropriate, forceful physical strike of Serenity's foot or leg that contacts another person/animal	code each strike	striking therapist in leg with foot	tripping over someone
Hitting	Contextually inappropriate, forceful physical strike of Serenity's hand, or an object she is controlling, that contacts another person/animal	code each strike	punching, striking with an object	belly bump, hugs, high fives
Throwing	Any contextually inappropriate instance of propelling an object through the air the direction of another person/animal	code each item thrown	hurling/flinging a toy at someone	throwing a ball and catching it
Pushing	Any instance of applying contextually inappropriate, forceful physical contact against another person (e.g., using the hands or upper body) that displaces the person or creates resistance	code each time the person is displaced, or resistance is created	shoving a sibling off a chair,	bumping into someone in the hallway
Grabbing	Any contextually inappropriate application of a forceful grip onto another person's limb or article of clothing	code each time Serenity renews her grip	seizing a person by the arm/collar	grabbing a hand to guide another
Pinching	Any contextually inappropriate act of gripping/compressing a small area of another person's skin, typically using a thumb and fingers, to tightly squeeze a localized area of the skin	code each time Serenity applies pressure to the skin	squeezing another's skin	lightly touching another's body
Scratching	Using nails or another sharp object to scrape the surface of another person's skin typically by applying pressure or friction through repetitive back-and forth or circular motions	code each swipe (i.e., each new contact with the person's skin)	swiping another person's arm with his nails	impaling another with a pencil
Hairpulling	Any instance of grasping, tugging, or pulling another person's hair away from their scalp or body	code each time she renews her grasp as a new instance	grabbing ponytail and yanking	brushing someone's hair
Biting	Any instance using his teeth to contact another person's body part (or clothing item) and applying pressure (i.e., closing the jaws/teeth together)	code each time Serenity closes her mouth as a separate instance	biting someone's arm or shirt	biting a toy or her own arm

Topography	Definition	Measurement	Example	Nonexample
Property Destruction				
Kicking	Contextually inappropriate, forceful physical strike of Serenity's foot or leg that contacts an object/surface	code each strike	striking wall with leg or foot	tripping over object
Hitting	Contextually inappropriate, forceful physical strike of Serenity's hand, or an object he is controlling, that contacts an object/surface	code each strike	punching, striking with object with an object	belly bump, hugs, high fives
Throwing	Any contextually inappropriate instance of propelling an object through the air the direction of another object/surface	code each item thrown	hurling/flinging a toy at wall	throwing a ball and catching it
Swiping	Any contextually inappropriate sweeping or arching movement of Serenity's hands/arms that displaces items from their original location(s)	code each new swipe and each time she changes directions of her arm movements	knocking items off the table with forearm	knocking over a glass by mistake when reaching for her plate
Verbal Aggression				
	Statements or gestures that are delivered in a hostile, forceful, or confrontational manner using a raised voice, shouting, or rapid speech	code each occurrence separately (e.g., "you fucking bitch" should include two instances PB)	"fucking damn ass idiot" (e.g., 3 instances)	dropping something and saying "shit"

Topography	Definition	Measurement	Example	Nonexample
Threats to Others	Any instance of making an oral statement or physical gesture that suggests an intent or desire to cause physical harm, pain, injury, or fear toward others	code each occurrence separately	grabbing knife and saying I'm going to kill him! (e.g., two instances)	"I wish I didn't have any brothers."
Threats to Self	Any instance of making an oral statement or physical gesture that suggests an intent or desire to cause physical harm, pain, or injury toward herslf	code each occurrence separately	threatening to jump off of a balcony, laying in front of a car	"I am so stupid."

Appendix I

Codebook of Zeke's Operational Definitions

Topography	Definition	Measurement	Example	Nonexample
Aggression				
Kicking	Contextually inappropriate, forceful physical strike of Xavier's foot or leg that contacts another person/animal	code each strike	striking therapist in leg with foot	tripping over someone
Hitting	Contextually inappropriate, forceful physical strike of Xavier's hand, or an object he is controlling, that contacts another person/animal	code each strike	swatting with an open hand	belly bump, hugs, high fives
Throwing	Any contextually inappropriate instance of propelling an object through the air the direction of another person/animal	code each item thrown	hurling/flinging a toy at someone	throwing a ball and catching it
Pushing	Any instance of applying contextually inappropriate, forceful physical contact against another person (e.g., using the hands or upper body) that displaces the person or creates resistance	code each time the person is displaced, or resistance is created	shoving someone out of the way	bumping into someone in the hallway
Property Destruction				
Kicking	Contextually inappropriate, forceful physical strike of Xavier's foot or leg that contacts an object/surface	code each strike	striking wall with leg or foot	tripping over object
Hitting	Contextually inappropriate, forceful physical strike of Serenity's hand, or an object he is controlling, that contacts an object/surface	code each strike	knocking over chairs	belly bump, hugs, high fives
Throwing	Any contextually inappropriate instance of propelling an object through the air the direction of another object/surface	code each item thrown	tossing a toy at wall	throwing a ball and catching it
Swiping	Any contextually inappropriate sweeping or arching movement of Xavier's hands/arms that displaces items from their original location(s)	code each new swipe and each time he changes directions of his arm movements	knocking items off the table with forearm	knocking over a glass by mistake when reaching for his plate
Verbal Aggression	Statements or gestures that are delivered in a hostile manner using a tone of high-pitched voice or rapid speech	code each occurrence separately (e.g., "I hate you. I want a new family." - 2 instances)	"Serenity is the devil"	"Math is so dumb"