Received: 26 March 2024 | Revised: 8 September 2024 | Accepted: 17 September 2024

DOI: 10.1002/bin.2059

RESEARCH ARTICLE

WILEY

Program evaluation of a specialized treatment home for adults with severe challenging behavior

Thurka Thillainathan[1] | Bruce Linder[2] | Alison D. Cox[1]

[1]Department of Applied Disability Studies, Brock University, St. Catharines, Ontario, Canada

[2]Safe Management Group, Pryor, Linder and Associates, Ontario, Canada

Correspondence

Thurka Thillainathan, Department of Applied Disability Studies, Brock University, St. Catharines, ON L2S 3A1, Canada.
Email: mj20lr@brocku.ca

Abstract

Individuals with intellectual and developmental disabilities who engage in severe challenging behavior may comprise 5%–10% of the clinical population. Unfortunately, challenging behavior literature tends to underrepresent adult participants and emphasize efficacy (Does the intervention work?) more often than effectiveness (Does the intervention work in *real world* settings?). We conducted a systematic program evaluation to examine the *effectiveness* of a comprehensive behavioral treatment package using a hybrid quasi-experimental consecutive case series design featuring eight adults who experienced the treatment package. The results depicted a substantial decrease in challenging behavior and increase in adaptive behavior across sessions for most participants. Impressive treatment integrity outcomes ($M = 84\%$, range, 82%–90%) showcased that the intervention was implemented as intended. Social validity surveys administered to participants, caregivers, and case managers provide support for the acceptability of the treatment package. Project limitations, clinical considerations, and future directions are discussed.

KEYWORDS

consecutive case series, effectiveness, intellectual and

developmental disabilities, program evaluation, severe challenging behavior, treatment recipient social validity

1 | INTRODUCTION

Approximately 50% of individuals with intellectual and developmental disabilities engage in challenging behavior and 5%–10% of these individuals' challenging behavior may be described as *severe* (Bowring et al., 2017; Lowe et al., 2007). Challenging behavior can take several forms, including aggression, property destruction, and self-injury, although the term *severe* has been applied inconsistently in the literature. That is, the decision to apply it as a descriptor has been based on referencing topography, injury, frequency, and/or risk, to name a few (Fahmie & Iwata, 2011; Lowe et al., 2007; Oropeza et al., 2018; Poppes et al., 2010). As a result, it has been used as a descriptor for vastly different participant profiles. Support for individuals engaging in severe challenging behaviors can be costly, while the dangers associated with ineffective treatment are intense. As such, research on intervention *effectiveness* is vital.

1.1 | Applied behavior analytic interventions for severe challenging behavior

Applied behavior analysis [ABA] may be considered an evidence-based treatment approach for challenging behavior in persons with intellectual and developmental disabilities among non-pharmacological interventions (Kurtz et al., 2021). At the same time, current research evaluating interventions informed by this discipline tend to place greater emphasis on efficacy. That is, a plethora of research examining treatment effects under highly controlled conditions to demonstrate a causal relation between the treatment and behavior change exists (Ghaemmaghami et al., 2021; Singal et al., 2014). By comparison, few studies examine treatment effects under real-world circumstances to demonstrate a treatment's generality, acceptability, and clinical utility, also known as *effectiveness* (Singal et al., 2014). As it stands, several behavior analytic intervention options have been employed to reduce challenging behavior, though few have an established effectiveness evidence-base (Chambless & Hollon, 1998; Ghaemmaghami et al., 2021).

1.1.1 | Function-based interventions

Function-based interventions, defined as interventions that directly address consequences maintaining challenging behavior (e.g., access to preferred items/activities, escape from instructions), are predominantly implemented within the field of ABA to decrease challenging behavior (Geiger et al., 2010). The following "overarching" categories of function-based interventions may have established efficacy, but not effectiveness, including: (1) modification of motivating operations, such as noncontingent reinforcement (NCR; e.g., Saini et al., 2017); (2) extinction procedures, such as escape-extinction, as one component of a treatment package in accordance with behavior-analytic best practice (e.g., Newcomb & Hagopian, 2018); and (3) differential reinforcement (DR) procedures, such as functional communication training (FCT; e.g., Chowdury & Benson, 2011). Those that endeavored to examine function-based intervention effectiveness have described a myriad of limitations, including restricted participant samples (Phillips et al., 2017) or barriers to collecting response measures (e.g., interobserver agreement [IOA] on staff recorded behavior observations; Chartier et al., 2020).

1.1.2 | Punishment-based interventions

When function-based interventions in isolation fail to produce meaningful challenging behavior change (e.g., Foxx & Garito, 2007), *additional* consequence-based procedures, such as punishment, may need to be implemented to produce clinically meaningful outcomes and achieve socially significant goals (Newcomb & Hagopian, 2018). Such cases typically include circumstances in which: (1) the challenging behavior presents a serious risk of harm (e.g., physical injury; Busch et al., 2018); (2) extinction- and reinforcement-based interventions have not resulted in challenging behavior reduction to clinically meaningful levels (Behavior Analyst Certification Board, 2020); and (3) the reinforcer maintaining challenging behavior cannot be identified with certainty (i.e., through functional analysis) or withheld due to behavioral and/or environmental conditions (Cooper et al., 2020; Lerman & Vorndran, 2002).

Prior literature has demonstrated the efficacy of punishment-based interventions in decreasing challenging behavior, such as: (1) response blocking (e.g., Ahearn et al., 2007; DeRosa et al., 2019); (2) overcorrection in isolation (e.g., Anderson & Le, 2011; Peters & Thompson, 2013) as well as in conjunction with reinforcement-based procedures (e.g., Steinhauser et al., 2021); and (3) response cost in isolation (Bartlett et al., 2011; DeJager et al., 2020) as well as in conjunction with reinforcement-based procedures (e.g., DeJager et al., 2020).

As with function-based research, there appears to be a relative lack of effectiveness research. For instance, research evaluating response blocking have primarily focused on children (defined as less than 18 years of age) with autism spectrum disorder (ASD) who engage in either motor or vocal stereotypy (Ayvaci et al., 2024). Ayvaci et al. (2024) also describe several other important limitations that interfere with drawing definitive conclusions around combined interventions (i.e., punishment plus reinforcement) efficacy, including treatment effect maintenance. Finally, little is known regarding treatment social validity (Kliebert et al., 2011). The relative dearth examining the effectiveness of punishment-based interventions *in conjunction with reinforcement-based procedures* (e.g., overcorrection as one component in a multi-component package; Williams et al., 2009) suggests it may be prudent to continue evaluating these interventions for a contemporary and comprehensive understanding of effectiveness (i.e., feasibility, generality, acceptability; DeJager et al., 2020).

1.2 | Efficacy versus effectiveness of interventions for severe challenging behavior

Despite the established prevalence and resultant need for *effective* interventions, adults with intellectual and developmental disabilities who engage in severe challenging behavior are largely underrepresented in the behavior analytic literature (Cox et al., 2021). Research patterns also suggest an emphasis on antecedent and reinforcement-based strategies which, for individuals displaying truly severe challenging behavior (see for example, Foxx, 2003), may prolong treatment. That is, it may take longer to achieve treatment gains with extremely harmful behaviors continuing to occur at the peril of the individual with intellectual and developmental disabilities and their caregivers. Relying solely on antecedent and reinforcement-based strategies may also produce muted treatment effects in comparison to what other approaches could achieve (i.e., less overall reduction in challenging behavior; Gover et al., 2019; Hanley et al., 2005).

In general, research examining treatment outcomes in the context of large, aggressive clients implemented by natural change agents in treatment contexts (i.e., effectiveness; Geiger et al., 2010) is relatively limited; whether the interventions are function-based, punishment-based, or multi-faceted programming that includes an escape-extinction element.

1.3 | Program evaluations

Enacting program evaluations can support an independent review initiative aiming to evaluate treatment package effectiveness. In other words, given program evaluations consider ongoing interventions that continue to be in

place in real-life applied settings, it can contribute meaningful information pertaining to program inputs, outcomes, and feasibility (Miller, 2017). Program evaluations can thereby improve program accountability, inform practitioners and other stakeholders of program effectiveness, and promote adherence to best-practice treatments (Shepley et al., 2021).

1.4 | Research purpose, rationale and hypothesis

We conducted a program evaluation to assess the overall effectiveness of a comprehensive behavioral treatment package comprised of DR, escape-extinction, and punishment. The partnering agency's clinical implementers of this unique program aimed to promote marked behavioral improvements in adult clients with intellectual and developmental disabilities who engaged in severe challenging behavior. It follows that our purpose was three-fold. First, to add to a relatively understudied area (i.e., severe challenging behavior in adults with intellectual and developmental disabilities). Second, to speak to program effectiveness (e.g., under which parameters and/or for whom a comprehensive behavioral intervention package featuring core tenants of ABA may work for). Finally, to inform other works interested in program evaluations.

2 | MATERIALS AND METHODS

This program evaluation project received ethics clearance from Brock University's Research Ethics Board (file number 22–333) to: (1) invite all past and current clients who experienced the featured treatment package to participate in the program evaluation; (2) compile existing clinical data for consenting participants for secondary use; (3) conduct treatment integrity; (4) collect IOA data for dependent variables and treatment integrity; and (5) administer social validity surveys to participants, caregivers, and case managers.

2.1 | Participants and recruitment

Participants included eight adults diagnosed with intellectual and/or developmental disability who provided consent (or had substitute decisions makers provide consent) and were enrolled in the treatment program at the specialized treatment home between October 2017 and July 2023, regardless of their success. See Table 1 for a detailed description of participant demographics. Psychotropic medications were stable during proxy baseline and medication changes did not coincide with this study's featured phase change from baseline to intervention (see Supporting Information S1: Appendix A). Overall, participants experienced primarily psychotropic medication reductions and/or withdrawals (see Table 2). This meant that the psychotropic medication regime that coincided with the program evaluation's end featured either lower psychotropic medication dosages, fewer concurrently prescribed psychotropic medications, or both. At admission, all participants were considered in-crisis, as evidenced by the specific severe challenging behavior exemplars described below for each participant.

2.1.1 | Lily

Lily was a 28-year-old female diagnosed with ASD, mild intellectual disability, generalized anxiety disorder, oppositional defiant disorder, disruptive behavior disorder—not otherwise specified (NOS), and factitious disorder. She had a longstanding history of severe challenging behavior, such that prior to admission to the specialized treatment home, she required emergency medical intervention at least twice per week. That is, she would routinely

TABLE 1 Participant demographics.

Application	Name	Challenging behaviors	Hypothesized behavior function
1	Lily	ED, elopement, flopping, ingesting foreign objects, NC, PA, SIB, VA, VD	Attention, escape, access to tangibles
2	Kevin	Agitation, ED, elopement, NC, PA, RAF, SIB, social fantasizing/play, suicidal ideologies	Escape, attention, access to preferred items and activities
3	Ginny	ED, elopement, fixation on items, NC, PA, SIB, VA	Automatic
4	Riley	ED, elopement, NC, PA, perseverating and whining, SIB, VA	Escape, access to negative social attention and tangibles
5	Micaela	ED, elopement, NC, PA, requesting emergency services, SIB, VA, VD	Access to tangibles, attention, escape
6	Anthony	ED, elopement, inappropriate Internet/electronic usage, ISB, MB, NC, PA, SIB VA, VC	Escape, access to preferred social interactions and tangibles
7	Taylor	AB, ED, elopement, perseveration, NC, SB, SIB, VA	Access to tangibles, attention, escape
8	Oliver	Compulsive spending, deceptive statements, ED, elopement, food stealing, inappropriate social behavior, intrusive thoughts and fixations, NC, PA, VA, VC	Escape, attention, access to tangibles

Abbreviations: AB, antagonizing behavior; ED, environmental destruction; ISB, inappropriate and sexualized behaviors; MB, mischief behavior; NC, non-cooperation; PA, physical aggression; RAF, requesting, arguing, and fabricating; SB, sexualized behavior; SIB, self-injurious behavior; VA, verbal aggression; VC, verbal complaining; VD, verbal disruption.

ingest razor blades, poisonous berries, and other harmful substances that required group home staff to call an ambulance or rush her to emergency where she would need gastric suction.

2.1.2 | Kevin

Kevin was a 29-year-old male diagnosed with mild developmental disability, reactive mood disorder, paraphilia–NOS, and attention deficit hyperactivity disorder (ADHD). Due to severe challenging behavior over several years, he had been extensively involved with the judicial system. At one point, he had experienced 10 formal charges in a single year. Prior to admission to the specialized treatment home, Kevin experienced police intervention at least three times per week for several consecutive months due to the severity of his aggression with staff. As an example, he would throw his feces at staff, ejaculate and throw semen at them, and would also find items in his home to break and attempt to use as weapons. He was also so self-injurious that he had caused irreparable damage to his body.

2.1.3 | Ginny

Ginny was a 23-year-old female with a global developmental delay (GDD). Due to severe property destruction and aggressive behavior toward family members, including several younger siblings, Ginny had to move into a group home as a young teen. She stayed at the youth group home until its closure after which she transitioned to adult services. As a result of ongoing excessive aggressive behavior, she was repeatedly evicted from previous adult placements prior to her admission to the featured specialized treatment home.

TABLE 2 Medication conditions summary.

Participant	Number of medication changes	Baseline		Final	
		Medication name	Total dosage of medication (mg)	Medication name	Total dosage of medication (mg)
Lily	7	Carbamazepine	700	Taro-carbamazepine	300
		Duloxetine	60		
Kevin	3	Apo-divalproex	1500	Olanzapine	10
		Clonidine	0.1	Paroxetine	60
		Olanzapine	10		
		Paroxetine	60	Mylan-divalproex	1500
		PMS-clonazepam	1		
Ginny	5	Apo-aripiprazole	22.5	Trazodone	100
		Clonidine	2.5	Mint-clonidine	0.1
Riley	6	Clonidine	0.4	Clonidine	0.3
		Aripiprazole	15	Aripiprazole	15
		Chlorpromazine	200	Trazodone	25
		Lithium carbonate	450		
Micaela	–	–	–	–	–
Anthony	2	Aripiprazole	12	Aripiprazole	12
				Clonidine	0.075
		Clonidine	0.075	Venlafaxine	150
		Venlafaxine	187.5	Gabapentin	600
		Gabapentin	600	Lorazepam	1
Taylor	1	Carbamazepine	400	Carbamazepine	400
		Clonidine	0.3	Clonidine	0.3
		Olanzapine	7.5	Olanzapine	7.5
		Quetiapine	400	Quetiapine	400
		Sertraline	125	Sertraline	125
		Lorazepam	3	Lorazepam	3
		Haloperidol	15		
Oliver	0	Risperidone	6	Risperidone	6
		Venlafaxine	375	Venlafaxine	375

Note: Dashed cells indicate that data were not obtained. Baseline refers to medication regime at admission and final refers to medication regime at the end of data collection.

2.1.4 | Riley

Riley was a 20-year-old male diagnosed with ASD and moderate intellectual disability with an extensive history of severe challenging behavior. For instance, Riley consistently engaged in severe property destruction (e.g., smashing glass, punching through drywall, breaking laptops), physical aggression (e.g., biting family members such that they had to wear long sleeves to protect themselves), disruptive behaviors (e.g., urinating or defecating outside the toilet such that school staff wore goggles and lab coats to protect themselves from his bodily fluids), and elopement that frequently required police and other emergency services. Prior to admission to the specialized treatment home, he resided alone in a ply-wall reinforced locked unit that was video monitored by staff.

2.1.5 | Micaela

Micaela was a 26-year-old female diagnosed with borderline personality disorder (BPD) and developmental delay. She had a history of severe aggressive and destructive behaviors resulting in repeated placement breakdowns, hospitalizations, police involvement, and incarceration. To exemplify, prior to admission to the specialized treatment home, she was charged with assault and released on probation under the condition that she abide by treatment programming wherever she resides.

2.1.6 | Anthony

Anthony was a 21-year-old male diagnosed with alcohol related neurodevelopmental disorder, ADHD, sensory processing disorder, anxiety disorder—NOS, depressive disorder—NOS, developmental coordination disorder, and a learning disability in reading, writing, and math. Prior to admission, he had a history of severe aggressive and destructive behaviors that resulted in repeated placement breakdowns, hospitalizations, and police involvement. One challenging behavior episode required police intervention and resulted in Anthony being charged with two counts of assault and assault of a peace officer. During another episode that took place at school, he was arrested and charged with mischief. He was ultimately suspended from school, and never able to return to school, despite programming aimed at helping him meet the school's prerequisites for him to return.

2.1.7 | Taylor

Taylor was a 29-year-old female diagnosed with ASD, BPD, schizophrenia, and developmental delay. She had a history of severe aggressive and destructive behaviors that resulted in placement breakdowns, hospitalizations, and police involvement. To exemplify, one episode of severe physical aggression resulted in staff injury requiring medical attention, and included her attempting to injure herself by running onto the streets. Following this episode, she was admitted to a hospital, charged with three counts of assault, and placed on 12 months of probation. This subsequently led to a short stay at a correction center as well as at brief shelters and respite placements, which she was ultimately evicted from.

2.1.8 | Oliver

Oliver was a 24-year-old male with a diagnosis of unspecified impulse control disorder (sexual urges), major depressive disorder, anxiety disorder–NOS, conduct disorder, ADHD, fetal alcohol spectrum disorder, borderline

intellectual functioning as well as global and severe adaptive limitations. He had a history of engaging in severe challenging behavior in the form of property destruction, elopement, and inappropriate sexual behaviors. To exemplify, at his previous placement, police were contacted when he threatened staff with a hammer and then proceeded to destroy the environment. Following this incident, he was admitted to a hospital where he was assessed and found to be a risk to himself and others by the physician, which ultimately resulted in his eviction from his previous placement.

2.2 | Setting

The featured treatment package was implemented in three specialized treatment homes. The homes were designed to treat challenging behavior, prevent hospital admission, and serve as a transition placement where participants could develop appropriate replacement behaviors and adaptive skills to facilitate successful relocation to community placements. The treatment team at each treatment home was comprised of direct-care staff, clinical staff (i.e., Board Certified Behavior Analysts [BCBAs], behavior therapists, and behavior therapist assistants), and a clinical house manager. All three homes were supervised by a Certified Psychologist who was also a BCBA–Doctoral (BCBA-D). Each treatment home housed two to three participants who received care 24 h a day for 7 days a week. All three treatment homes consisted of an open concept layout with individualized programming areas, a kitchen, and a bathroom comprising the shared spaces as well as separate bedroom units for each participant.

2.3 | Program evaluation timeline

The research team conducted the program evaluation from May 2022 to July 2023. The duration of the program evaluation (i.e., 14 months) involved the execution of program evaluation measures, namely IOA and treatment integrity data collection as well as social validity survey administration to clients, caregivers, and case managers (described below).

2.4 | Measurement

2.4.1 | Response measures

Participants' challenging behavior topographies (i.e., negative target behaviors) were operationally defined as per their existing behavior support plan (BSP; see Table 3). Frontline staff collected data on daily episode frequency (primary dependent variable), which was reviewed weekly and used to inform clinical decisions. The secondary dependent variable, adaptive behavior, was also individualized across participants (see Table 4). For each participant, a skill acquisition program (SAP) outlining target adaptive behaviors was developed. The measurement system and data collection schedule was individualized (e.g., a cold probe once a week, a cold probe twice a week, discrete trial training [DTT] four times a week). Frontline and clinical staff collected data. For all participants, adaptive behavior skills progress was measured by percentage program mastery across months in a skills supervision report, which was used to inform clinical decisions.

2.4.2 | Interobserver agreement

As part of the program evaluation, IOA data was collected live for both dependent variables. For challenging behavior, trained, independent observers: (1) recorded the start and end time of each observation period to

TABLE 3 Sample negative target behaviors and operational definitions.

Target behavior	Operational definition
Verbal aggression	Any instance of yelling, swearing, insulting or criticizing staff, threatening, accusing others, and loud complaining (i.e., speaking above conversational level). This does not include crying or complaining at a conversational or quiet level unless swearing or threatening/attacking staff is observed.
Physical aggression	Any attempted or accomplished instance of biting, pinching, scratching, flailing at others, punching, hitting, kicking, hair pulling, spitting directly at/onto another person, and throwing items at others.
Ingesting foreign objects	Any witnessed or disclosed instance of intentionally seeking out and swallowing non-food items for self-harm, including unsuccessful attempts. Items may include, but are not limited to bobby pins, nails, screws, hooks, light bulb pieces, pop cans, license plates, mailboxes, eaves troughs, batteries, coil springs, razor blades, toilet bowl cleaning, and other hazardous liquids.
Self-injurious behavior	Any attempted or accomplished instance of causing harm to self. This may include, but is not limited to biting, head-banging on hard surfaces, body-slamming, wrist-cutting with sharp and broken items, and self-asphyxiation.
Environmental destruction	Any attempted or accomplished instance of destroying items in the environment. This includes punching/kicking holes in walls, destroying furniture, and picking at baseboards, electric sockets, vents, etc.
Elopement	Any instance of leaving designated areas without permission, accompaniment, or staff knowledge.
Requesting, arguing, and fabricating (RAF)	Any instance in which: (a) a request is made over and above what the current behavior support plan/daily activity schedule permits (i.e., wanting to cook during skills interval as opposed to prescribed cooking interval, wanting an extra cigarette, asking staff to bring coffee from Tim Horton's, etc.); (b) arguing with staff when told "no" or when advised to discuss it during "daily concerns" time; and (c) saying that a BT or other person approved the request earlier (at a different time) when they did not.
Non-cooperation	Any instance of not initiating a task within 30 s of instruction, which may be accompanied by vocalizations of refusal.

facilitate accurate IOA calculations; and (2) stood off to the side of the room and recorded challenging behavior frequency each time it occurred during IOA observation sessions. We used total count IOA, wherein reliability between staff recorded data and independent observer data was calculated by dividing the smaller count by the larger count and multiplying that by 100 (Cooper et al., 2020). For adaptive behavior, trained, independent observers: (1) were onsite when cold probes or DTT were conducted by frontline or clinical staff; and (2) stood off to the side of the room and recorded data as per the measurement system outlined for the skill being observed. For DTT, trial-by-trial IOA was used (Cooper et al., 2020). Reliability between staff recorded data and the researcher was calculated by dividing the number of trials that were in agreement by the total number of trials and multiplying that by 100 (Ledford & Gast, 2018). Agreement was described as both staff and the researcher recording the occurrence of a correct (or incorrect) response for a given trial as defined by the program (e.g., tacts correct picture independently). For cold probes, researchers used occurrence IOA. Specifically, each observation opportunity constituted an occurrence trial. Reliability between staff recorded data and researcher data was calculated by dividing the number of occurrence trials (i.e., observation opportunities) that agreed by the total number of occurrence trials and multiplying by 100 (Ledford & Gast, 2018). Agreement was described as both staff and researcher recording the occurrence of a correct (or incorrect) response for the probe as defined by the program (e.g., states correct answer to question about collaborative problem solving).

TABLE 4 Sample adaptive target behaviors and response requirements.

Target behavior	Correct response	Incorrect response
Completes single chores independently	A correct response is defined as completing a step as outlined (e.g., open lid of the garbage, remove full garbage bag, hold both sides of the garbage bag, cross both sides of the garbage bag, loop the ends of the bag threading one side of the garbage bag through the hole, etc.) independently and without any prompts.	An incorrect response is defined as requiring more than one gestural prompt or a higher level of prompting (i.e., verbal, partial, or full physical) for a step.
Reading	A correct response is defined as reading a target word (e.g., chamber, nobleman, utterly, etc.) accurately, independently, and without any prompts.	An incorrect response is defined as reading a target word inaccurately or not knowing how to say a target word.
Spelling	A correct response is defined as spelling a target word (e.g., search, check, quick, etc.) accurately and within 5 s of the instruction being delivered (i.e., "spell ___).	An incorrect response is defined as spelling a target word inaccurately or after 5 s of the instruction being delivered.
Maintains a conversation	A correct response is defined as responding within 10 s and independently meeting each of the following criteria: (1) comments on the staff's questions; (2) uses a filler word; and (3) asks staff a WH-question specific to the topic being discussed.	An incorrect response is defined as not responding within 10 s or erroring on one or more of the following criteria: (1) comments on the staff's questions; (2) uses a filler word; and (3) asks staff a WH-question specific to the topic being discussed.
Performs basic microsoft word functions	A correct response is defined as accurately and independently performing the action (e.g., creating a bulleted, numbered, or lettered list).	An incorrect response is defined as requiring a reminder on how to perform the action.
Yoga	A correct response is defined as watching the YouTube video and trying to complete the moves for the entire duration of the interval.	An incorrect response is defined as not trying to watch the YouTube video and not trying to complete the moves for the entire duration of the interval.

IOA sampling probes

Participants resided in the home 24 h a day for 365 days a year. Programming was applied from when the participants woke up to when they went to sleep across the homes. Given our small research team, we could not be onsite for the entire active programming duration which constituted up to 16 h each day. Therefore, the team endeavored to collect IOA sampling probes for both dependent variables one to two times per week for each participant. To accomplish this, each day of the week was considered one data collection opportunity. This meant that there were seven data collection opportunities per week per participant. As such, conducting IOA data collection for both dependent variables once per week per participant was equivalent to sampling 14% of sessions (i.e., days), while collecting IOA data for both dependent variables twice per week per participant was equivalent to sampling 29% of sessions (i.e., days). Through this approach, we sampled IOA for up to 15% of sessions.

IOA data for adaptive behavior was not collected for Oliver as he did not have a formal SAP at the time of data collection while, for Micaela, it was halted at approximately 1 month after initiation of data collection due to extenuating circumstances. For these reasons, outcomes for Micaela and Oliver are not included. See Table 5 for IOA outcomes for both dependent variables across participants.

TABLE 5 Interobserver agreement data for dependent variables across participants.

	Challenging behavior			Adaptive behavior		
Participant	% Session	% Agreement	% Range	% Session	% Agreement	% Range
Lily	8	95	17–100	7	94	63–100
Kevin	9	100	100	8	98	86–100
Ginny	11	86	33–100	6	100	100
Riley	9	95	75–100	9	92	0–100
Micaela	15	100	100			
Anthony	8	100	100	5	97	75–100
Taylor	5	100	100	6	95	50–100
Oliver	7	100	100			

Note: Blank cells indicate that data were not collected.

2.5 | Comprehensive treatment package procedures

Below is a general summary of the featured comprehensive treatment package that was implemented for each participant by the staff at the treatment homes.

2.5.1 | Pre-admission

The intake process was conducted by agency behavior therapists prior to implementing the program evaluation. Thus, it could not be a part of the current project. In the interest of transparency (i.e., technological; Baer et al., 1968), we elected to briefly describe this process alongside corresponding Intervention phases. The intake process included interviewing previous support staff and caregivers (Daily Activity Schedule-Questionnaire [DAS-Q]; Linder, 2014), completing a comprehensive file review to obtain relevant past assessment and supports provided, obtaining any existing target behavior data, and conducting functional behavior assessment on negative target behavior to hypothesize behavior function (e.g., Questions About Behavioral Function [QABF]; Matson & Vollmer, 1995; Functional Assessment Screening Tool [FAST]; Iwata & DeLeon, 2005). Information garnered from this process resulted in the development of a behavior assessment report and informed each individualized initial BSP.

2.5.2 | Intervention

The intervention consisted of three overarching phases: (1) Behavioral Stabilization; (2) Skill Acquisition; and (3) and Generalization and Maintenance. Importantly, the treatment package components (e.g., DR, response cost, etc.) implemented within each phase were individualized to match each participant's unique profile (e.g., challenging behavior, adaptive functioning, preferred items/activities). In other words, although the intervention goals and underlying behavioral approaches informing the comprehensive behavioral treatment package (i.e., phases 1–3) were consistent across participants, individualized programming and data-based clinical decisions guided its application for each participant.

Phase 1

Behavioral Stabilization was the focus in this phase given skills are difficult to teach when frequent, severe challenging behaviors occur. Stabilization was achieved via a combination treatment package that included DR, escape-extinction (see below for fulsome description), token economy, response cost, overcorrection, and response blocking (where applicable). As part of programming, each participant was expected to complete their daily activity schedule (DAS). The DAS is an individualized, highly structured routine consisting of 22–28 activity intervals per day with embedded skill-building opportunities. These activities included task sequences, activities of daily living (ADLs), household chores, community outings (where appropriate), physical activity, leisure time, and staff or peer interaction times. Token economy and response cost systems augmented by social positive reinforcement (e.g., verbal praise, physical praise) were incorporated to encourage participation and adaptive behavior. Points, awarded contingent on the absence of severe challenging behavior and ongoing active participation in their DAS, were redeemed for predetermined daily and/or weekly awards.

If participants did not cooperate with their DAS or reasonable staff instructions (e.g., "It's not time for leisure right now, let's get back to finishing your laundry because your housemate needs to use the machine soon"), a graduated least-to-most prompt hierarchy was implemented. For instance, for Oliver, upon task refusal, staff offered three verbal prompts spaced 2 min apart followed by a total duration of 5 min to allow him to consider his actions. Following this, collaborative problem solving was enacted—if Oliver was calm enough to engage. If these strategies did not promote activity completion, then staff were expected to provide a gestural prompt followed by gentle physical prompting. Escape-extinction was enacted in response to extreme physical aggression or other severe, dangerous behavior (e.g., SIB causing lacerations) permitting implementers to effectively withhold reinforcement for challenging behavior. Specifically, escape-extinction was delivered in a DTT format. Brief contingent physical holds were used, when necessary, to facilitate implementers in expeditiously redirecting participants back to task (activities that can be reinforced) while simultaneously minimizing risk of injury to implementers and participants. At times, mechanical restraints were used for prevention of non-task related escape behavior (e.g., dangerous elopement). Individualized restraint release criteria were developed to minimize application frequency and/or duration. For example, for Taylor, one strap was faded for every 10 correct and independent responses of the task at hand in a prespecified order as outlined in her BSP (i.e., second hand, shoulders, leg, waist). Importantly, one hand was always unrestricted so that participants could complete tasks. Restraint application permitted the DR programming element by supporting return to task and/or appropriate, alternative replacement behavior, which were then reinforced.

Participants' physicians provided medical clearance for restraint application processes. The following measures were meticulously observed by the agency to ensure ongoing compliance with legal and ethical standards: (1) all BSPs were written in accordance with province mandated Quality Assurance Measures (2011); (2) all BSPs were evaluated on a yearly basis by qualified independent reviewers; and (3) all staff received competency-based physical and mechanical intervention training with formal ongoing implementation monitoring by supervisors, complimented by an annual physical and mechanical intervention renewal training session delivered by certified instructors. Finally, clinical oversight and best-practices on the application of restraint as described by Vollmer et al. (2011) were followed precisely.

Phase 2

The second phase, Skill Acquisition, became the focus once sufficient learning had occurred in that severe challenging behaviors do not result in access to reinforcement. Individualized skill building programs for target adaptive skills were developed, augmented by least-to-most prompting procedures according to the participants' strengths. For example, for Ginny, toileting was identified as a target adaptive skill and an individualized program outlining training steps, as well as a reinforcement hierarchy (i.e., behavior specific verbal praise, photos on iPad for 2 min, photos on iPad for 4 min, and iPad access for 5 min), was established. Participants also began to learn greater independence in working through their day with support from the team. For example, at the time of the program

evaluation, Kevin was being taught to move through his day on time with few verbal prompts from staff. Contingent physical and mechanical intervention supports continued to be in place across all treatment phases, if required.

Phase 3

In the third phase, Generalization and Maintenance, the skills that were taught in phase 2 continued to be generalized across a wide variety of individuals, times, and settings. Generalization and Maintenance were the focus, while the phase also involved preparation for discharge from the treatment setting. For example, during this phase, participants may decide to take part in volunteer or paid work activities outside of the treatment home thereby broadening their opportunities for generalized responding.

2.5.3 | Treatment integrity

A behavior therapist or BCBA created treatment integrity checklists (TICs) in accordance with each participant's clinical programming (e.g., BSP, Task Guidelines, Reinforcer Guidelines). Relevant modifications to TICs were made on an ongoing basis in accordance with programming changes (e.g., BSP changes). Treatment integrity data were typically collected by clinical staff who observed staff onsite and scored each item on the TIC as a correct response or an incorrect response. Of note, for the project's purpose, the frequency of treatment integrity collected by the clinical team did not align with how often research standards dictate these should be completed (see Cooper et al., 2020). Therefore, the research team collected additional TICs to better align with these standards. So, the clinical staff enacted treatment integrity as normal, and the research team collected treatment integrity data between one to two times per week for each participant for up to 20% of sessions. Treatment integrity was calculated by dividing the total number of correct responses by the total number of items on the checklist and multiplying that by 100 (Cooper et al., 2020).

To lend further confidence to the reliability of the data collected, treatment integrity IOA was conducted once a month for randomly selected participants. Reliability between both researchers (either two research assistants, or the research lead [BCBA–D] and a research assistant) was calculated by dividing the number of items on the TIC that agreed by the total number of items and multiplying that by 100. Agreement was defined as: (1) both researchers scoring an item on the TIC as a correct (or incorrect) response. The standards for staff set out by the agency state that treatment integrity scores should yield a minimum score of 80% before directed trainings are required by the clinical team and management. See Table 6 for treatment integrity outcomes across participants, which reflect that these standards were being met.

Error analysis

The global treatment integrity analysis and outcomes (see Table 6) provide an overall quantification of integrity. However, as with any approach that includes aggregating data, it is possible that these overall scores might have masked poorer performance on the execution of individual treatment package elements (Cook et al., 2015). Poor performance on individual components may differentially affect treatment outcomes (e.g., Carroll et al., 2013). Therefore, we conducted an item-by-item error analysis for each participant.

2.6 | Social validity

Social validity questionnaires were administered to participants, caregivers, and case managers to assess treatment goals, procedures, and effects by consumers (Ferguson et al., 2019). More specifically, two social validity assessment processes were enacted: one for participants and the other for caregivers and case managers. The survey was administered electronically via Qualtrics and was formatted to reduce response effort for participants. This was

TABLE 6 Treatment integrity data across participants.

	Treatment integrity			Treatment integrity IOA		
Participant	% Session	% Correct	% Range	% Session	% Agreement	% Range
Lily	14	93	71–100	7	99	97–100
Kevin	22	86	52–100	5	100	100
Ginny	23	90	60–100	8	100	100
Riley	10	82	22–100	14	96	92–100
Micaela	17	82	50–100	20	95	90–100
Anthony	7	82	37–99	11	93	85–100
Taylor	10	88	56–100	8	96	88–100
Oliver	9	80	63–90	16	92	83–100

Abbreviation: IOA, interobserver agreement.

accomplished in part by: (1) ensuring the survey font was large; (2) bolding important keywords to draw attention to them; (3) including a progress bar at the top of the survey page; and (4) not requiring the completion of all questions in one sitting. Participants could also request support from a trusted staff member to help complete the measure. Given participants were diagnosed with some level of intellectual and developmental disability, if they agreed to participate, a pre-test consisting of five orienting questions was provided at the onset of the survey to ensure the validity of their answers (e.g., What month is it?; Wells & Ruesch, 1972). A percentage correct for the pre-test was generated following completion and a score of 80% (i.e., 4/5) or greater on the pre-test suggested that participant responses may be considered valid (Wells & Ruesch, 1972). All participants, regardless of their score on the pre-test proceeded to the social validity questions. The social validity survey was separated into three domains: (1) Before Services ($n = 6$ questions); (2) During Services ($n = 5$ questions); and (3) Final Thoughts ($n = 6$ questions; Lambert et al., 2022; Luiselli et al., 2015; Luiselli et al., 2017). Questions were either open-ended, yes/no, or on a five-point Likert scale. The five-point Likert scale constituted the following: Strongly Agree ("I really agree with this"), Agree ("I sort of agree with this"), Disagree ("I sort of disagree with this"), Strongly Disagree ("I really disagree with this"), and Prefer Not to Answer (Sturgis et al., 2012). See Table 7 for a fulsome description of social validity questionnaire statements, domains, and formats for participants.

2.7 | Research design

We applied a quasi-experimental consecutive case series (CCS) design, which is described as a research design that involves "the application of a defined treatment to a series of cases presenting with similar problems" (Hagopian, 2020, p. 614). In this context, a single-case experimental design with each case is not *required* and, as such, A-B designs are often used (e.g., Lomas Mevers et al., 2018). Given the program evaluation did not coincide with participants' admission to the treatment homes, participant data had been collected before the study began and continued to be collected upon study completion. Thus, the current project may be better described as a *hybrid* quasi-experimental CCS.

2.8 | Data evaluation and analysis

Due to the severity of the challenging behavior engaged in by the participants, phase 1 of the intervention (i.e., Behavioral Stabilization) was implemented immediately upon admission. This approach is not uncommon in the

TABLE 7 Social validity statements, domains, and formats for participants.

Statements	Domain	Format
S1: The services at this agency are the first behavioral services I have ever received.	Before services	Y/N
S2: Other services (like behavioral interventions) before coming to this agency have not been helpful.	Before services	Y/N
S3: Other services (like behavioral interventions) before this agency were helpful, but did not reduce my challenging behavior. For example, breaking things, physically hurting other people, swearing at people, getting arrested.	Before services	Likert scale
S4: Other services (like behavioral interventions) before this agency were helpful in reducing my challenging behavior, but it came back.	Before services	Likert scale
S5: Before this agency, it was important to me to stop engaging in challenging behavior.	Before services	Likert scale
S6: Before this agency, it was important to me to learn new things and learn new skills.	Before services	Likert scale
S7: The services at this agency are necessary to make sure I do not get hurt and do not hurt others.	During services	Likert scale
S8: All of the items in my behavior support plan are helpful.	During services	Likert scale
S9: My behavior support plan is helping me decrease how often I try to hurt myself or others. That is, engage in challenging behavior.	During services	Likert scale
S10: My skill building program is helping me learn new skills.	During services	Likert scale
S11: This agency has helped me progress and work toward a better life. I feel better now.	During services	Likert scale
S12: I feel like I am getting the best service here.	Final thoughts	Y/N
S13: I feel supported here.	Final thoughts	Y/N
S14: I feel safe where I am currently living.	Final thoughts	Y/N
S15: Did the program help you?	Final thoughts	Y/N
S16: Can you list one (or more) things that you feel have helped you the most?	Final thoughts	Open-ended
S17: If you could change one thing about the services you are getting at this agency, what would it be?	Final thoughts	Open-ended

Abbreviations: Y/N, yes or no.

treatment of severe challenging behavior (e.g., Pritchard et al., 2018). However, it can complicate the prediction and verification of resultant intervention outcomes due to a lack of baseline. To address this, baseline and intervention were delineated as early treatment and late treatment, respectively, as is often observed in cardiology research (e.g., Homorodean et al., 2019; O'Hara et al., 1995).

To generate individualized early treatment phases for each participant, quantitative criteria described by Schoenfeld et al. (1956) utilized in a program titled *Stability Check* were employed (Costa & Cançado, 2012). The Schoenfeld et al. (1956) criterion is represented by the following formula.

$$\left| \frac{\left[\left(\frac{D_1+D_2+D_3}{3}\right)\right] - \left[\left(\frac{D_4+D_5+D_6}{3}\right)\right]}{\left(\frac{D_1+D_2+D_3+D_4+D_5+D_6}{6}\right)} \right| \times 100\%$$

According to this criterion, the first 7 days on any schedule are not considered in calculating stability. For the next 6 days, the mean of the first 3 days (i.e., D_1, D_2, and D_3, respectively) of the six is compared with that of the last 3 days (i.e., D_4, D_5, and D_6, respectively). If the difference between these means is less than 5% of the 6 days' mean, the individual is considered to have stabilized and would have been shifted to the next condition. If the difference between sub means is greater than 5% of the grand mean, another experimental day is added (i.e., D_7) and similar calculations are made for that day as well as the 5 days immediately preceding it (D_2, D_3, D_4, D_5, and D_6). The following formula is to be used:

$$\left|\frac{\left[\left(\frac{D_2+D_3+D_4}{3}\right)\right]-\left[\left(\frac{D_5+D_6+D_7}{3}\right)\right]}{\left(\frac{D_2+D_3+D_4+D_5+D_6+D_7}{6}\right)}\right| \times 100\%$$

Such extensions of the experiment and calculations of stability are continued daily until the prespecified 5% criterion is reached (Schoenfeld et al., 1956).

The Schoenfeld et al. (1956) criterion was employed with the data compiled for each participant to delineate early treatment. Stability was established at 88 days for Lily, 14 days for Kevin, 26 days for Ginny, 34 days for Riley, 15 days for Micaela, 17 days for Anthony, and 24 days for Taylor. However, for Oliver, the Schoenfeld et al. (1956) criterion was unsuccessful in determining stability. Existing research exemplars suggest a viable solution may be averaging across participants (Anderson et al., 2023). Thus, rather than exclude Oliver's data, early treatment was delineated at 27 days by averaging the days at which stability was established for the other seven participants. The term "baseline" hereafter will be used to refer to early treatment intervals (i.e., *proxy* baseline) delineated for each participant using the Schoenfeld et al. (1956) criterion. Data that fell outside of the baseline period was considered *late treatment* data; hereafter referred to as *intervention*. Intervention intervals (i.e., late treatment) across participants were comprised of data collected across phases 1 and 2, as well as phase 3, where applicable—as a whole. This was done so that we could examine *overall* intervention impact.

2.9 | Modified Brinley plot

Given featured participants experienced the comprehensive treatment package every day for up to 6 years, the resultant datasets consisted of an extensive number of sessions (range, 591 sessions–2085 sessions). This rendered interpreting the ensuing time-series graphs difficult (see Supporting Information S1: Appendices B and C). As such, we elected to use a modified Brinley plot for several reasons. First, employing this approach affords greater efficiency. That is, this plot allows users to clearly depict results across several comparisons within and across participants. Moreover, its visual aids (e.g., identity line, gray dashed diagonal line) allow for a quick evaluation of multiple aspects (Manolov et al., 2022). Second, past research has found insufficient agreement between visual analysts inspecting time-series plots (Ninci et al., 2015). In fact, when using time-series graphs, variations in the ratio between the *x*-axis and *y*-axis have been found to lead to potential distortion during visual analysis (Kubina et al., 2017). Given a modified Brinley plot is square, it is unaffected by graphical features like the ratio between the *x*-axis and *y*-axis. Thus, it can mitigate visual inspector bias during visual analysis (Manolov et al., 2022). A comprehensive list of the advantages and disadvantages of this approach goes beyond the scope of the current paper. Interested readers may refer to Blampied (2017), Manolov and Tanious (2022), and Manolov et al. (2022) for further information.

2.9.1 | Interpreting a modified Brinley plot

Challenging behavior data (i.e., daily episode frequency across negative target behaviors listed in the BSP) for all participants ($N = 8$) is depicted using a modified Brinley plot in Figure 1. A modified Brinley plot provides a graphical

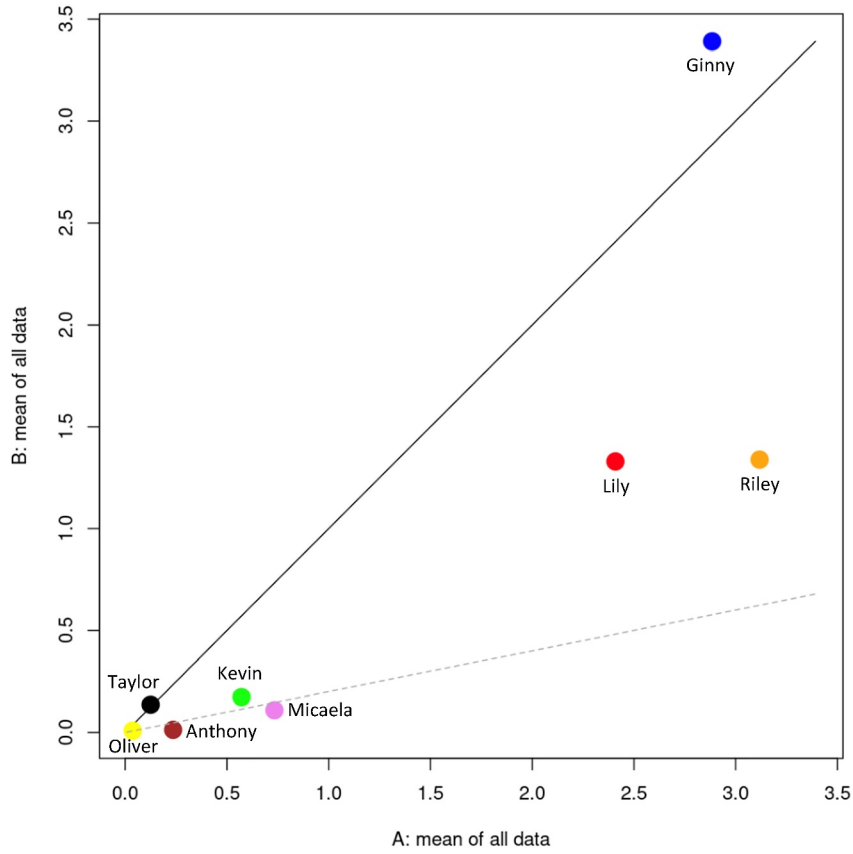

FIGURE 1 Modified Brinley Plot for Challenging Behavior across Baseline and Intervention. This figure depicts challenging behavior data across baseline and intervention for all participants on a modified Brinley plot. The *x*-axis represents the daily baseline mean and the *y*-axis represents the daily intervention mean. The black solid diagonal line signifies the equivalence between baseline and intervention and the gray dashed diagonal line indicates the desired magnitude of change from the baseline level. The data points are designated as follows: red = Lily, green = Kevin, blue = Ginny, orange = Riley, pink = Micaela, brown = Anthony, black = Taylor, and yellow = Oliver.

display of phase means so that any treatment effect may be clearly observed. That is, a comparison between *baseline* (phase A) and *intervention* (phase B) is defined by the phase A mean and the phase B mean (Blampied, 2017; Manolov et al., 2022). A diagonal line, known as an identity line, in which the intercept is zero and the slope is one, is included to represent the equivalence between baseline and intervention (i.e., no difference; Manolov et al., 2022). If the improvement is an increase, then all data points should be above the diagonal line and if the improvement is a reduction, then all data points should be below the diagonal line (Manolov et al., 2022). For the current project, given this is challenging behavior data, improved outcomes would be depicted by the data points falling below the identity line (Blampied, 2017; Manolov et al., 2022).

An additional visual aid in the form of a gray dashed diagonal line can be included on modified Brinley plots. This line represents the desired *magnitude* of change from the baseline level (Manolov et al., 2022). Often, this line is defined using expert judgment as it pertains to applied significance. That is, domain-specific knowledge can be drawn upon to determine when a sufficient departure from challenging behavior has occurred (Manolov et al., 2022). Existing literature suggests that an 80% reduction in challenging behavior may be indicative of an effective intervention (Rooker et al., 2013). For the current project, an 80% reduction in challenging behavior would be depicted by the data points falling on or below the gray dashed diagonal line.

3 | RESULTS

3.1 | Challenging behavior outcomes

Challenging behavior outcomes are depicted for all consecutive applications ($N = 8$) of the comprehensive treatment package in Figure 1. Data points for Lily, Kevin, Riley, Micaela, Anthony, and Oliver appeared below the identity line while the data point for Ginny fell above the identity line and the data point for Taylor is on the identity line. Thus, therapeutic outcomes (i.e., a decrease in challenging behavior) may be observed for 75% of participants ($n = 6$), neutral outcomes (i.e., no change) may be observed for 13% of participants ($n = 1$), and contra-therapeutic outcomes (i.e., an increase in challenging behavior) may be observed for 13% of participants ($n = 1$).

Data points for Micaela and Anthony are found below the gray-dashed line while the data point for Oliver is depicted on the line and the data points for Lily, Kevin, Ginny, Riley, and Taylor are above the line. Thus, the desired magnitude of change from the baseline level (i.e., 80% reduction) may have been achieved by 38% of participants ($n = 3$) and may not *yet* have been achieved by 63% of participants ($n = 5$).

3.2 | Adaptive behavior outcomes

Baseline data for adaptive behavior are not reported in part because skill acquisition data was not being collected upon intake across participants; this is partly because the focus early on was Behavioral Stabilization (i.e., phase 1) due to challenging behavior severity (a process that was independent of the current program evaluation; see Materials and Methods section).

Intervention data for adaptive behavior (i.e., cumulative number of mastered skills targets per month) is depicted for all consecutive applications ($n = 7$) of the comprehensive treatment package in Figures 2–8. We observed an accelerating, therapeutic trend across most of the intervention phase for all participants. That is, a steep slope indicating an increase in the number of cumulative mastered skills targets was observed for all participants; albeit learning (as evidenced by skills mastered) began at different times. A detailed visual inspection description for each participant may be obtained by emailing the first author.

3.3 | Treatment integrity error analysis outcomes

The results of the treatment integrity item-by-item error analysis demonstrated two general patterns across participants. For Lily and Ginny, percentage error was generally lower and, for these two participants as well as Riley, greater error rates tended to coincide with items outside of the management of behavior domain (range, 30%–81%; e.g., overnight protocol, urine accident after mechanical chair release, preventative strategies, daily activity schedule, bathroom, evening/bedtime). For Kevin, Micaela, Anthony, Taylor, and Oliver, percentage error was generally higher for items within the preventative strategies (range, 68%–88%) and management of behavior domains (range, 42%–88%). Additionally, across all eight participants, minimal error percentages were observed for all items in the data collection domain (0%–37%). Specific individualized error patterns are available for review by emailing the first author.

3.4 | Social validity outcomes

Six of eight participants elected to complete the social validity survey, while seven caregivers or case managers elected to complete the survey.

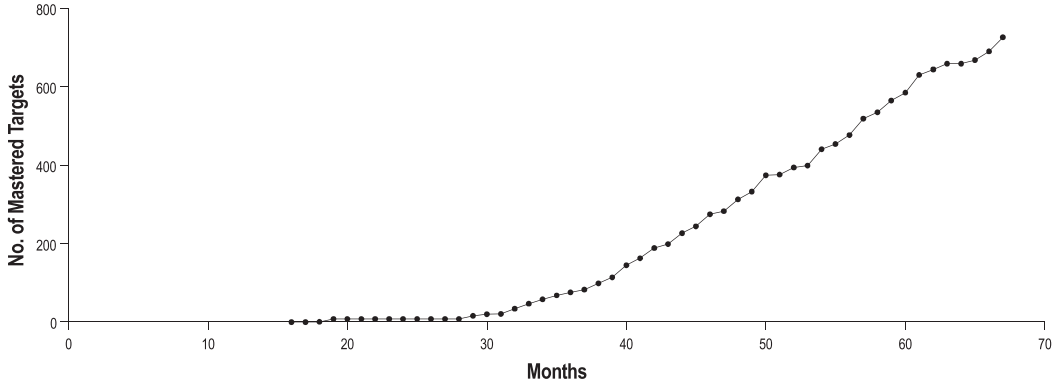

FIGURE 2 Intervention Data for Lily's Adaptive Behavior across Months. This graph depicts intervention data for Lily's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 16 months after admission.

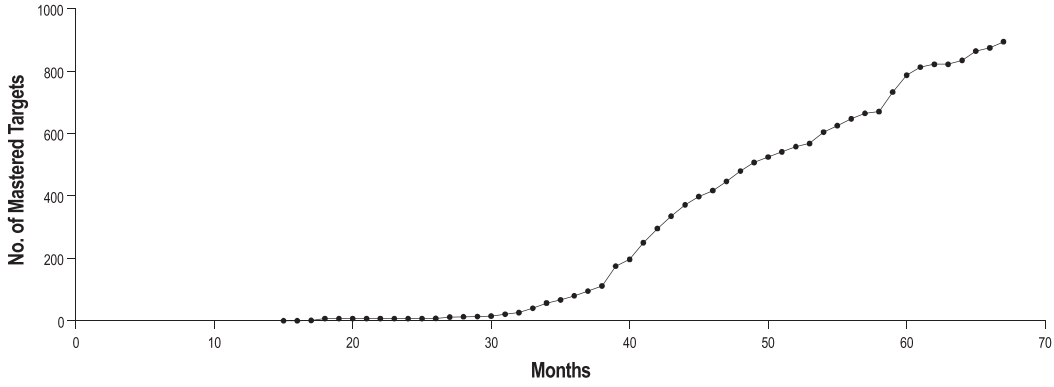

FIGURE 3 Intervention Data for Kevin's Adaptive Behavior across Months. This graph depicts intervention data for Kevin's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 15 months after admission.

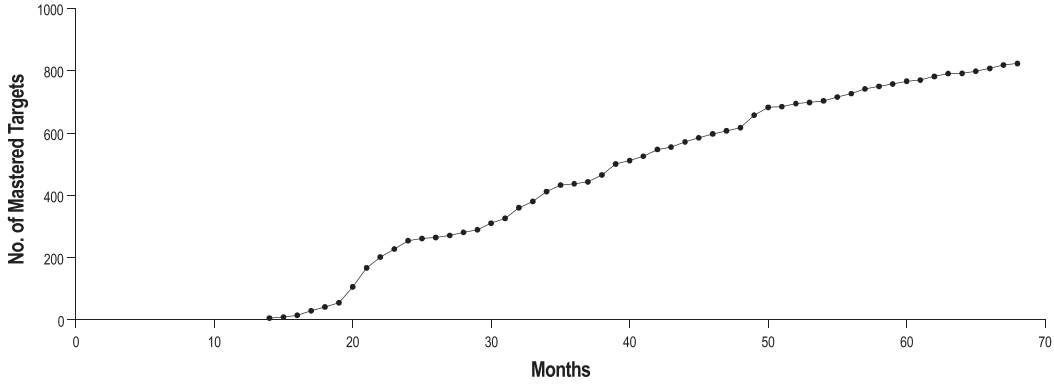

FIGURE 4 Intervention Data for Ginny's Adaptive Behavior across Months. This graph depicts intervention data for Ginny's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 14 months after admission.

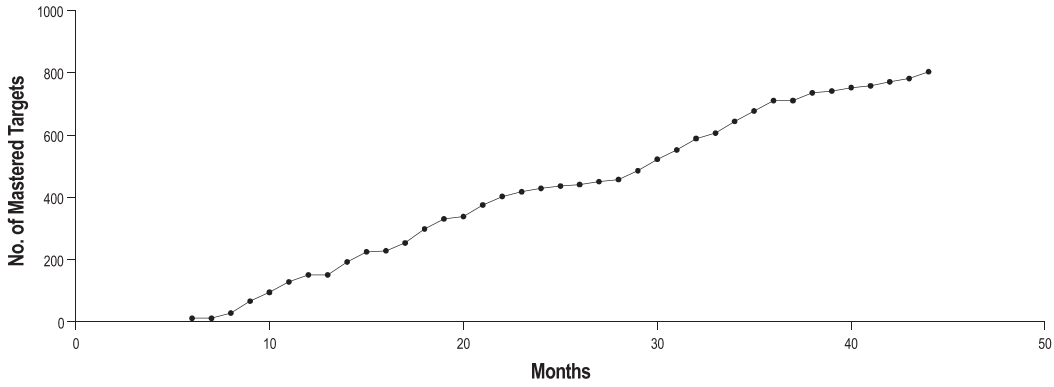

FIGURE 5 Intervention Data for Riley's Adaptive Behavior across Months. This graph depicts intervention data for Riley's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 6 months after admission.

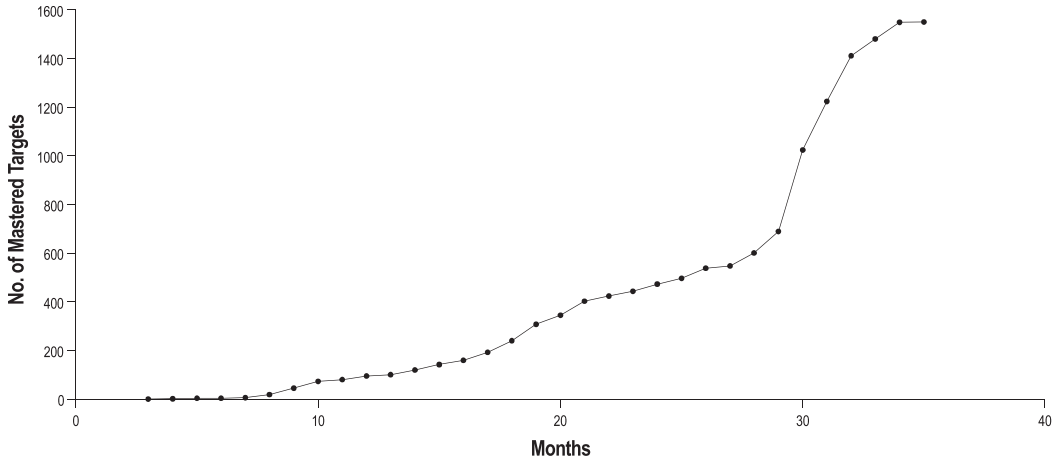

FIGURE 6 Intervention Data for Micaela's Adaptive Behavior across Months. This graph depicts intervention data for Micaela's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 3 months after admission.

3.4.1 | Participants

The pre-test score of 80% (i.e., correct responding on 4/5 questions) or greater outlined a priori was achieved by five out of the six participants who completed the survey. Their answers were thus considered valid for inclusion (Wells & Ruesch, 1972). Among this sample, 100% (5/5) felt that the treatment package was instrumental in helping them achieve a better quality of life while 80% (4/5) reported that the comprehensive treatment package was helpful. Four of five participants answered open-ended questions regarding aspects of the treatment package that have helped them the most (see Discussion).

3.4.2 | Caregivers and case managers

Among the caregivers and case managers who completed the survey, 86% (6/7) reported that they felt their client (loved one) was receiving the best treatment at the specialized treatment home and 100% (7/7) reported that they felt

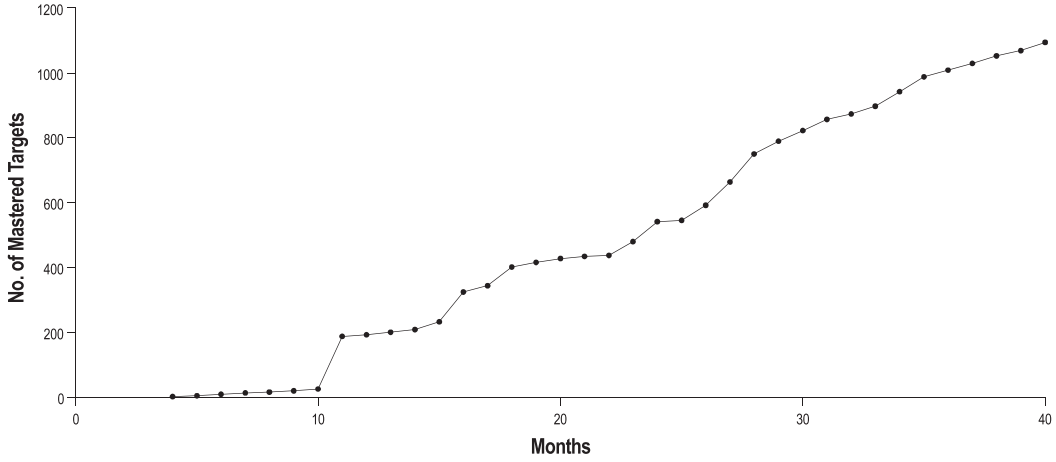

FIGURE 7 Intervention Data for Anthony's Adaptive Behavior across Months. This graph depicts intervention data for Anthony's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 4 months after admission.

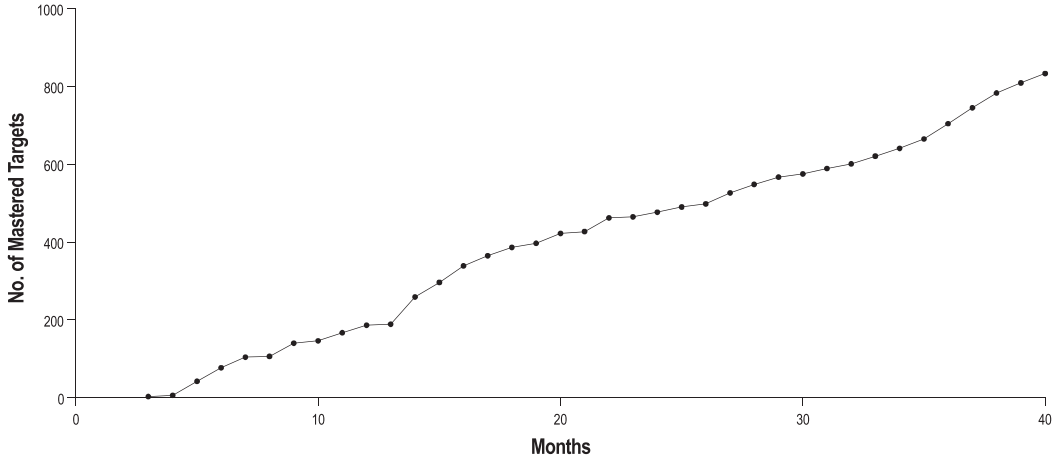

FIGURE 8 Intervention Data for Taylor's Adaptive Behavior across Months. This graph depicts intervention data for Taylor's adaptive behavior. The *y*-axis represents the number of mastered skills targets and the *x*-axis represents months in treatment. Data collection for adaptive behavior commenced at 4 months after admission.

confident in the supports, safety, and overall helpfulness of the current placement. Five of seven respondents answered open-ended questions regarding aspects of the treatment package that have helped their client (loved one) the most while four of seven respondents commented on items they would like to see changed (see Discussion).

4 | DISCUSSION

The purpose of the current project was to conduct a program evaluation on the use of a comprehensive behavioral treatment package encompassing DR, escape-extinction, and strategies featuring decelerative properties (e.g., overcorrection, response cost) to produce marked improvements in adults with intellectual and developmental

disabilities who engage in severe challenging behavior. Project outcomes indicated treatment integrity was greater than 80% across all participants, which met the agency's standards. Importantly, the treatment integrity outcomes also met the threshold for acceptable integrity performance outlined in the research literature (80%; e.g., Cook et al., 2015; Hagermoser Sanetti & Kratochwill, 2008). Project outcomes also suggest the application of the comprehensive intervention resulted in a substantial decrease in challenging behavior from baseline to intervention for most participants. Finally, all participants appeared to be steadily learning adaptive skills, as evidenced by an increasing trend of cumulative mastered target skills coinciding with the end of the program evaluation.

4.1 | A demonstration of effectiveness

Specific to the current project, a demonstration of effectiveness may have been accomplished in two ways. First, we evaluated the application of the comprehensive behavioral treatment package with "relatively *heterogeneous* populations, in typical settings where these treatments occur" (Ghaemmaghami et al., 2021, p. 124; Singal et al., 2014). So, although all participants were adults with intellectual and/or developmental disabilities, other characteristics suggested heterogeneity in sex (i.e., four males and four females), diagnoses (e.g., ASD, ADHD, BPD, generalized anxiety disorder, schizophrenia), and function of behavior (e.g., automatic, multiply controlled). Second, the quasi-experimental CCS afforded determining the proportion of cases in which the comprehensive treatment package produced a positive outcome (Hagopian, 2020). Specifically, the results indicated that therapeutic outcomes (i.e., challenging behavior reduction) were observed for 75% of participants ($n = 6$). These six participants (i.e., Lily, Kevin, Riley, Micaela, Anthony, Oliver) had comorbid diagnoses as well as multiply controlled function of behavior. Neutral outcomes (i.e., no change) were observed for 13% of participants ($n = 1$; Taylor), although the data point for Oliver bordered the no change line. Notably, Taylor and Oliver are the latter two cases in the series of consecutively encountered cases and therefore have experienced the comprehensive treatment package for only 2–3 years compared to 3–6 years by all other participants.

Contra-therapeutic outcomes (i.e., an increase in challenging behavior) may have been observed for 13% of participants ($n = 1$; Ginny). Interestingly, Ginny was unique in that she was only diagnosed with GDD (as opposed to comorbid diagnoses as with all other participants) and her support plan documents indicated challenging behavior may have been maintained solely by automatic function (as opposed to being multiply controlled as with all other participants). Collectively, these findings could suggest that the featured comprehensive treatment package may be effective for adults with intellectual and developmental disabilities who engage in truly severe challenging behavior and have: (1) comorbid diagnoses; (2) multiply controlled function of behavior; and (3) experienced the treatment for at least 3 years.

4.2 | A demonstration of program evaluation processes

Another project objective aimed to showcase the barriers and corresponding processes to overcome them so that others interested in executing a program evaluation may be better able to do this. First, setting access and subsequent project coordination between the researchers, participants, and agency staff was enabled and supported by the program's Clinical Director. This support ultimately ensured the project could be executed. Transparency, respect, prioritizing shared goals, and bi-weekly communications (i.e., virtual meetings) were vital contributors to seeing the project to completion (Rycroft-Malone et al., 2015). A second barrier to evaluating clinical research outcomes is not being able to collect IOA with staff data (Chartier et al., 2020). Although participants featured in the current program evaluation were in treatment 24 h a day for 365 days a year, IOA with staff data were collected by adjusting accordingly for the applied context (e.g., use of total count IOA).

A third barrier commonly associated with treating severe challenging behavior is that intervention is implemented immediately upon admission (Pritchard et al., 2018). To address the absence of a *true* baseline, the current project demonstrated the viability of using innovative tools such as *Stability Check* based on the Schoenfeld et al. (1956) criterion to delineate individualized *proxy* baselines (Costa & Cançado, 2012).

4.3 | Project strengths

In addition to demonstrating effectiveness and program evaluation processes, the inclusion of the treatment integrity error analysis and social validity measure involving *participants* and stakeholders are project strengths. The former facilitated generating an empirical answer as to how well the program is doing so that the agency may improve service delivery, which is a primary purpose of program evaluations (Miller, 2017). The latter served to address an under-incorporated area in behavior-analytic literature. That is, social validity systematic reviews have found that social validity measures are infrequently reported (i.e., 12%; Ferguson et al., 2019), goals selected for treatment were the least likely to be assessed (i.e., 12%; Ferguson et al., 2019), and direct recipients of behavior-change intervention are underrepresented (i.e., less than 3% of applications; Hanley, 2010). Regarding social validity outcomes, caregivers and case managers provided some recommendations, such as their desire for the agency to reduce staff turnover rates. Standard staff turnover rates in the disabilities sector (not to mention clients who engage in severe challenging behavior) is approximately 43% (National Core Indicators, 2022). Thus, although suggestions around reducing staff turnover may be tricky to address, it may be important for the agency to communicate their staff retention efforts to stakeholders. Outlining these efforts clearly to stakeholders may be a way to acknowledge this recommendation, maintain transparency, and ultimately foster rapport between parties.

Respondents, both participants and caregivers, also made note of several program strengths. Participants stated that staff and the clinical team were instrumental in supporting their success and that programming was directly useful to skills they believe is important for an effective lifestyle (e.g., time management, productivity). Caregivers and case managers reported that the structure and consistency offered by the program (e.g., DAS, staff expectations, staff training, communication between all teams) played a substantial role in the success of their client (loved one).

4.3.1 | Treatment integrity error analysis

The treatment integrity error analysis permitted the team to highlight some areas for improvement and potential next steps. Notably, for participants who engaged in challenging behavior less often (i.e., Kevin, Micaela, Anthony, Taylor, Oliver), percentage error was generally higher in the management of behavior domain. It is possible that because these participants infrequently engaged in challenging behavior, staff were unaccustomed to executing (or describing) certain behavior management items and were therefore not able to recall the protocol readily. Moving forward, it may be important for the clinical team to automate error analyses (i.e., create a computer program that automatically graphs where errors are most commonly being made) so that they can review the errors informed by all the TICs collected in a month. Collectively, these data could inform targeted trainings that could be conducted once per month on the items associated with high percentage errors.

Importantly, treatment integrity error analysis outcomes also showcased program areas of strength. For instance, consistent low percentages of error were observed for participants who engaged in higher frequencies of challenging behavior (i.e., Lily, Ginny). This indicates that staff who regularly implemented management of behavior domain components did so with high levels of integrity, lending support for programming effectiveness. Further, the low error rates observed within the data collection domain for all eight participants specifically pertained to whether data were recorded: (1) as required; (2) to its entirety with no information left blank or missing; and (3)

accurately. Thus, the high integrity levels for these items are particularly impressive as it lends confidence to the accurateness and completeness of the data reported.

4.4 | Project limitations

There are several noteworthy project limitations. First, although we were able to conduct IOA sampling probes, we were unable to procure the recommended amount due to the uncontrolled applied setting (i.e., treatment home). Further, between May 2022 to July 2023 several factors restricted onsite observation for up to weeks at a time including: (1) participants testing positive for COVID-19 resulting in treatment home lockdowns; and (2) participants becoming sick in general. Ultimately, we prioritized minimizing staff and participant disruptions that may have resulted from procuring a larger research team, so that we could spend more time onsite collecting data. Importantly, in quantitative statistics, researchers are rarely able to collect data from an entire population (Emmert-Streib & Dehmer, 2019). Instead, they rely on sampling as a mechanism for understanding phenomena (Emmert-Streib & Dehmer, 2019). So, although we were only able to sample a portion of time, it may be reasonable to apply this sampling logic to the current project. It follows that the IOA, and treatment fidelity sample obtained may be representative of what is happening on an ongoing basis.

A second limitation was that the design was not a *controlled* CCS because only A-B designs were used, and with these designs full experimental control cannot be demonstrated. However, the current project does showcase replication of the A-B design across six of the eight participants, which may lend credibility to the outcomes observed (i.e., clinically important challenging behavior reduction; Lomas Mevers et al., 2018). Further, quasi-experimental designs have been purported to be appropriate for examining questions of clinical utility at the effectiveness stage of research (Chambless & Hollon, 1998). Finally, it is possible that ongoing psychotropic medications may have impacted intervention outcomes. However, psychotropic medications were stable during proxy baseline and medication changes did not coincide with this study's featured phase change from baseline to intervention. Further, medication changes were primarily reductions. This is noteworthy because research on this topic suggests prescribing practitioners tend to refrain from decreasing medications for long-term users for fear of destabilization (i.e., increase in challenging behavior following a reduction; de Kuijper & Hoekstra, 2017). Since this was largely not the case in the current sample; that is, most participants experienced medication decreases, it could suggest consulting prescribers may have felt confident in the comprehensive behavioral treatment package. As such, participant outcomes may be considered even more impressive.

4.5 | Future research

Future research could explore addressing some of this project's limitations, as well as build on the featured processes. For instance, while a comprehensive examination of psychotropic medication impact on challenging behavior in the current context went beyond the scope of the current program evaluation, we recommend future research examine this phenomenon (e.g., Cox et al., 2022) in the context of severe challenging behavior. Additionally, given proxy baselines were generated retrospectively for all participants, future research may look to conducting a Monte Carlo analysis for each participant to compare clinical data for challenging behavior to simulated samples of behavioral data in order to discern the likelihood of the clinically obtained challenging behavior data occurring due to chance (Friedel et al., 2021). This analysis could be conducted using readily available Applications (see Friedel et al., 2021; https://shiny.georgiasouthern.edu/BA_Monte_Carlo/). Finally, future research examining comprehensive behavioral treatment packages in the context of severe, dangerous challenging behavior in adult participants may consider examining the added value of escape-extinction, perhaps through a dropout or add in component analysis. A better understanding of which treatment package element (or combination of

elements) are vital to promoting success (and ensuring staff and client safety) may be helpful because it could mean reduced resource requirements for program implementation. That is, escape-extinction required several staff members to execute. It follows that if this programming aspect coincided with little added value, removing it may reduce staff to client ratios, which could improve feasibility and increase the number of potential recipients (i.e., funding could be spread amongst more service users).

5 | CONCLUSION

Successfully conducting and disseminating effectiveness research may be partially barred by the continued perpetuation of the "conjoint-set perspective" (Critchfield & Reed, 2017, p. 123). That is, the seven dimensions described by Baer et al. (1968) are held as necessary components for *good* research. However, there needs to be a balance between "striving for as much analytic rigor as possible while acknowledging the constraints of the applied setting" (Bruni & Lancaster, p. 8). The results of this project will hopefully inform treatment implementation for individuals with similar profiles (i.e., adults with intellectual and/or developmental disabilities who engage in severe challenging behavior) and encourage continued collaboration with community partners to evaluate and disseminate programming outcomes in applied settings.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to the partnering agency, clinical team, frontline staff, research assistants, participants, and stakeholders who made this project possible. This project was not funded.

CONFLICT OF INTEREST STATEMENT

The first (TT) and third (AC) authors have no conflicts of interest to declare. The second author (BL) is the cofounder and one of three co-owners of the organization supporting these individuals. He is the Clinical Director of the program that was featured in this program evaluation and receives direct compensation for his role. All participants receiving services at the agency received services free-of-charge, in that government funding permitted their admission and ongoing participation in the program.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Thurka Thillainathan https://orcid.org/0009-0003-8621-1625

REFERENCES

Ahearn, W. H., Clark, K. M., MacDonald, R. P. F., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with Autism. *Journal of Applied Behavior Analysis*, *40*(2), 263–275. https://doi.org/10.1901/jaba.2007.30-06

Anderson, B. M., Kozluk, A., Morgan, M. C. M., MacDonald, M. A., Friedel, J., & Cox, A. D. (2023). Exploring characteristics of interventionist training associated with improved learner outcomes: A meta-analysis. *Journal of Behavioral Education*. https://doi.org/10.1007/s10864-022-09504-2

Anderson, J., & Le, D. D. (2011). Abatement of intractable vocal stereotypy using an overcorrection procedure. *Behavioral Interventions*, *26*(2), 134–146. https://doi.org/10.1002/bin.326

Ayvaci, A. S., Cox, A. D., & Dimopoulos, A. (2024). A quantitative systematic literature review of combination punishment literature: Progress over the last decade. *Behavior Modification*, *0*(0). https://doi.org/10.1177/01454455241262414

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

Bartlett, S. M., Rapp, J. T., Krueger, T. K., & Henrickson, M. L. (2011). The use of response cost to treat spitting by a child with autism. *Behavioral Interventions*, *26*(1), 76–83. https://doi.org/10.1002/bin.322

Behavior Analyst Certification Board. (2020). Ethics code for behavior analysts. Retrieved from https://www.bacb.com/wp-content/uploads/2022/01/Ethics-Code-for-Behavior-Analysts-220316-2.pdf

Blampied, N. M. (2017). Analyzing therapeutic change using modified Brinley plots: History, construction, and interpretation. *Behavior Therapy*, *48*(1), 115–127. https://doi.org/10.1016/j.beth.2016.09.002

Bowring, D. L., Totsika, V., Hastings, R. P., Toogood, S., & Griffith, G. M. (2017). Challenging behaviours in adults with an intellectual disability: A total population study and exploration of risk indices. *British Journal of Clinical Psychology*, *56*(1), 16–32. https://doi.org/10.1111/bjc.12118

Bruni, T. P., & Lancaster, B. M. (2019). Applied behavior analysis in pediatric primary care: Bringing ABA to scale. *Behavior Analysis: Research and Practice*, *19*(1), 5–13. https://doi.org/10.1037/bar0000152

Busch, L. P. A., Saini, V., Zorzos, C., & Duyile, L. (2018). Treatment of life-threathening pica with 5-year follow-up. *Advances in Neurodevelopmental Disorders*, *2*(4), 335–343. https://doi.org/10.1007/s41252-018-0053-9

Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis*, *46*(2), 379–394. https://doi.org/10.1002/jaba.49

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7–18. https://doi.org/10.1037/0022-006X.66.1.7

Chartier, K., McGowan, N., Ng, O., Makela, T., Legree, M., & Feldman, M. (2020). Impact of legislated quality assurance measures on interventions and challenging behaviour in adults with intellectual disabilities. *Journal on Developmental Disabilities*, *25*(2), 1–17.

Chowdury, M., & Benson, B. A. (2011). Use of differential reinforcement to reduce behavior problems in adults with intellectual disabilities: A methodological review. *Research in Developmental Disabilities*, *32*(2), 383–394. https://doi.org/10.1016/j.ridd.2010.11.015

Cook, J. E., Subramaniam, S., Brunson, L. Y., Larson, N. A., Poe, S. G., & St. Peter, C. C. (2015). Global measures of treatment integrity may mask important errors in discrete-trial training. *Behavior Analysis in Practice*, *8*(1), 37–47. https://doi.org/10.1007/s40617-014-0039-7

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson Education.

Costa, C. E., & Cançado, C. R. X. (2012). Stability Check: A program for calculating the stability of behavior. *Mexican Journal of Behavior Analysis*, *38*(1), 61–71.

Cox, A., Leung, J., Anderson, B. M., & Morgan, M.-C. (2021). Examining research trends in the treatment of adults with problem behavior and intellectual and developmental disabilities: A brief report. *Behavioral Development*, *26*(1), 14–28. https://doi.org/10.1037/bdb0000100

Cox, A. D., Pritchard, D., Penney, H., Eiri, L., & Dyer, T. J. (2022). Demonstrating an analyses of clinical data evaluating psychotropic medication reductions and the ACHIEVE! Program in adolescents with severe problem behavior. *Perspectives on Behavior Science*, *45*(1), 125–151. https://doi.org/10.1007/s40614-020-00279-3

Critchfield, T. S., & Reed, D. D. (2017). The fuzzy concept of applied behavior analysis research. *The Behavior Analyst*, *40*(1), 123–159. https://doi.org/10.1007/s40614-017-0093-x

DeJager, B., Houlihan, D., Filter, K. J., Mackie, P. F. E., & Klein, L. (2020). Comparing the effectiveness and ease of implementation of token economy, response cost, and a combination condition in rural elementary school classrooms. *Journal of Rural Mental Health*, *44*(1), 39–50. https://doi.org/10.1037/rmh0000123

de Kuijper, G. M., & Hoekstra, P. J. (2017). Physicians' reasons not to discontinue long-term used off-label antipsychotic drugs in people with intellectual disability. *Journal of Intellectual Disability Research*, *61*(10), 899–908. https://doi.org/10.1111/jir.12385

DeRosa, N. M., Novak, M. D., Morley, A. J., & Roane, H. S. (2019). Comparing response blocking and response interruption/redirection on levels of motor stereotypy: Effects of data analysis procedures. *Journal of Applied Behavior Analysis*, *52*(4), 1021–1033. https://doi.org/10.1002/jaba.644

Emmert-Streib, F., & Dehmer, M. (2019). Understanding statistical hypothesis testing: The logic of statistical inference. *Machine learning and Knowledge Extraction*, *1*(3), 945–961. https://doi.org/10.3390/make1030054

Fahmie, T. A., & Iwata, B. A. (2011). Topographical and functional properties of precursors to severe problem behavior. *Journal of Applied Behavior Analysis*, *44*(4), 993–997. https://doi.org/10.1901/jaba.2011.44-993

Ferguson, J. L., Cihon, J. H., Leaf, J. B., Van Meter, S. M., McEachin, J., & Leaf, R. (2019). 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

Foxx, R. M. (2003). The treatment of dangerous behavior. *Behavioral Interventions*, *18*(1), 1–21. https://doi.org/10.1002/bin.127

Foxx, R. M., & Garito, J. (2007). The long term successful treatment of the very severe behaviors of a preadolescent with autism. *Behavioral Interventions*, *22*(1), 69–82. https://doi.org/10.1002/bin.232

Friedel, J. E., Cox, A., Galizio, A., Swisher, M., Small, M. L., & Perez, S. (2021). Monte Carlo analyses for single-case experimental designs: An untapped resource for applied behavioral researchers and practitioners. *Perspectives on Behavior Science*, *45*(1), 209–237. https://doi.org/10.1007/s40614-021-00318-7

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

Gover, H. C., Fahmie, T. A., & McKeown, C. A. (2019). A review of environmental enrichment as treatment for problem behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, *52*(1), 299–314. https://doi.org/10.1002/jaba.508

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy*, *4*(1), 95–114. https://doi.org/10.1037/h0100835

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

Hanley, G. P. (2010). Towards effective and preferred programming: A case for the objective measurement of social validity with recipients of behavior-change programs. *Behavior Analysis in Practice*, *3*(1), 13–21. https://doi.org/10.1007/BF03391754

Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis*, *38*(1), 51–65. https://doi.org/10.1901/jaba.2005.6-04

Homorodean, C., Iancu, A. C., Lecuta, D., Balanescu, S., Dregoesc, I. M., Spinu, M., Ober, M., Tataru, D., Olinic, M., Bindea, D., & Olinic, D. (2019). New predictors of early and late outcomes after primary percutaneous coronary intervention in patients with ST-segment elevation myocardial infarction and unprotected left main coronary artery culprit lesion. *Journal of Interventional Cardiology*, *2019*(2), 1–9. https://doi.org/10.1155/2019/8238972

Iwata, B., & DeLeon, I. (2005). The functional analysis screening tool. In *The Florida center on self-injury*. University of Florida.

Kliebert, M. L., Tiger, J. H., & Toussaint, K. A. (2011). An approach to identifying the conditions under which response interruption will reduce automatically reinforced problem behavior. *Behavior Analysis in Practice*, *4*(1), 17–26. https://doi.org/10.1007/BF03391771

Kubina, R. M., Kostewicz, D. E., Brennan, K. M., & King, S. A. (2017). A critical review of line graphs in behavior analytic journals. *Educational Psychology Review*, *29*(3), 583–598. https://doi.org/10.1007/s10648-015-9339-x

Kurtz, P. F., Leoni, M., & Hagopian, L. P. (2021). Behavioral approaches to assessment and early intervention for severe problem behavior in intellectual and developmental disabilities. *Pediatric Clinics of North America*, *67*(3), 499–511. https://doi.org/10.1016/j.pcl.2020.02.005

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

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology* (3rd ed.). Routledge.

Lerman, D. C., & Vorndran, C. M. (2002). On the status of knowledge for using punishment implications for treating behavior disorders. *Journal of Applied Behavior Analysis*, *35*(4), 431–464. https://doi.org/10.1901/jaba.2002.35-431

Linder, B. (2014). Daily activity schedule-quality rating scale.

Lomas Mevers, J., Muething, C., Call, N. A., Scheithauer, M., & Hewett, S. (2018). A consecutive case series analysis of a behavioral intervention for enuresis in children with developmental disabilities. *Developmental Neurorehabilitation*, *21*(5), 336–344. https://doi.org/10.1080/17518423.2018.1462269

Lowe, K., Allen, D., Jones, E., Brophy, S., Moore, K., & James, W. (2007). Challenging behaviours: Prevalence and topographies. *Journal of Intellectual Disability Research*, *51*(8), 625–636. https://doi.org/10.1111/j.1365-2788.2006.00948.x

Luiselli, J. K., Sperry, J. M., & Draper, C. (2015). Social validity assessment of physical restraint intervention by care providers of adults with intellectual and developmental disabilities. *Behavior Analysis in Practice*, *8*(2), 170–175. https://doi.org/10.1007/s40617-015-00820z

Luiselli, J. K., Sperry, J. M., Draper, C., & Richards, C. (2017). Parent-guardian evaluation of physical restraint among adults with intellectual disability: A social validity assessment. *Advances in Neurodevelopmental Disorders*, *1*(2), 73–78. https://doi.org/10.1007/s41252-017-0013-9

Manolov, R., & Tanious, R. (2022). Assessing consistency in single-case data features using modified Brinley plots. *Behavior Modification*, *46*(3), 581–627. https://doi.org/10.1177/0145445520982969

Manolov, R., Tanious, R., & Fernandez-Castilla, B. (2022). A proposal for the assessment of replication of effects in single-case experimental designs. *Journal of Applied Behavior Analysis*, *9999*(3), 1–28. https://doi.org/10.1002/jaba.923

Matson, J. L., & Vollmer, T. R. (1995). *The questions about behavioral function (QABF) user's guide*. Scientific Publishers.

Miller, K. L. (2017). The use of evaluation in treatment programs for children with autism. *Behavior Analysis in Practice*, *10*(1), 35–44. https://doi.org/10.1007/s40617-016-0130-3

National Core Indicators. (2022). National core indicators intellectual and developmental disabilities 2020 staff stability survey report. Retrieved from https://legacy.nationalcoreindicators.org/upload/core-indicators/2020StaffStabilitySurveyReport_FINAL.pdf

Newcomb, E. T., & Hagopian, L. P. (2018). Treatment of severe problem behaviour in children with autism spectrum disorder and intellectual disabilities. *International Review of Psychiatry*, *30*(1), 96–109. https://doi.org/10.1080/09540261.2018.1435513

Ninci, J., Vannet, K. J., Willson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification*, *39*(4), 510–541. https://doi.org/10.1177/0145445515581327

O'Hara, P. J., Hertzer, N. R., Krajewski, L. P., Tan, M., Xiong, X., & Beven, E. G. (1995). Ten-year experience with abdominal aortic aneurysm repair in octogenarians: Early results and late outcomes. *Journal of Vascular Surgery*, *21*(5), 830–837. https://doi.org/10.1016/s0741-5214

O. Reg. 299/10: Quality Assurance Measures. under Services and supports to promote the social inclusion of persons with developmental disabilities Act, 2008, S. O. 2008, c. 14(2011). Retrieved from https://ontario.ca/laws/regulation/100299

Oropeza, M. E., Fritz, J. N., Nissen, M. A., Terrel, A. S., & Phillips, L. A. (2018). Effects of therapist-worn protective equipment during functional analysis of aggression. *Journal of Applied Behavior Analysis*, *51*(3), 681–686. https://doi.org/10.1002/jaba.457

Peters, L. C., & Thompson, R. H. (2013). Some indirect effects of positive practice overcorrection. *Journal of Applied Behavior Analysis*, *46*(3), 613–625. https://doi.org/10.1002/jaba.63

Phillips, C. L., Iannaccone, J. A., Rooker, G. W., & Hagopian, L. P. (2017). Noncontingent reinforcement for the treatment of severe problem behavior: An analysis of 27 consecutive applications. *Journal of Applied Behavior Analysis*, *50*(2), 357–376. https://doi.org/10.1002/jaba.376

Poppes, P., van der Putten, A., & Vlaskamp, C. (2010). Frequency and severity of challenging behaviour in people with profound intellectual and multiple disabilities. *Research in Developmental Disabilities*, *31*(6), 1269–1275. https://doi.org/10.1016/j.ridd.2010.07.017

Pritchard, D., Penney, H., & Mace, F. C. (2018). The achieve! Program: A point and level system for reducing severe problem behavior. *Behavioral Interventions*, *33*(1), 41–55. https://doi.org/10.1002/bin.1506

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 application. *Journal of Applied Behavior Analysis*, *46*(4), 708–722. https://doi.org/10.1002/jaba.76

Rycroft-Malone, J., Burton, C., Wilkinson, J. E., Harvey, G., McCormack, B., Baker, R., Dopson, S., Graham, I., Staniszewska, S., Thompson, C., Ariss, S., Melville-Richards, L., & Williams, L. (2015). Collective action for knowledge mobilisation: A realist evaluation of the collaborations for leadership in applied health research and care. *Health Services and Delivery Research*, *3*(44), 1–166. https://doi.org/10.3310/hsdr03440

Saini, V., Fisher, W. W., & Pisman, M. D. (2017). Persistence during and resurgence following noncontingent reinforcement implemented with and without extinction. *Journal of Applied Behavior Analysis*, *50*(2), 377–392. https://doi.org/10.1002/jaba.380

Schoenfeld, W. N., Cumming, W. W., & Hearst, E. (1956). On the classification of reinforcement schedules. *Proceedings of the National Academy of Sciences of the United States of America*, *42*(8), 5630570. https://doi.org/10.1073/pnas.42.8.563

Shepley, C., Shepley, S. B., Allday, R. A., Tyner-Wilson, M., & Larrow, D. (2021). Evaluation of a brief family-centred service provision model for treating children's severe behavior: A retrospective consecutive case series analysis. *Behavior Analysis in Practice*, *14*(1), 86–96. https://doi.org/10.1007/s40617-020-00487-y

Singal, A. G., Higgins, P. D. R., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, *5*(1), e45. https://doi.org/10.1038/ctg.2013.13

Steinhauser, H. M. K., Ahearn, W. H., Foster, R. A., Jacobs, M., Doggett, C. G., & Goad, M. S. (2021). Examining stereotypy in naturalistic contexts: Differential reinforcement and context-specific redirection. *Journal of Applied Behavior Analysis*, *54*(4), 1420–1436. https://doi.org/10.1002/jaba.847

Sturgis, P., Roberts, C., & Smith, P. (2012). Middle alternatives revisited: How the neither/nor response acts as a way of saying "I don't know"? *Sociological Methods & Research*, *43*(1), 15–38. https://doi.org/10.1177/0049124112452527

Vollmer, T. R., Hagopian, L. P., Bailey, J. S., Dorsey, M. F., Hanley, G. P., Lennox, D., Riordan, M. M., & Spreat, S. (2011). The Association for Behavior Analysis International position statement on restraint and inclusion. *The Behavior Analyst*, *34*(1), 103–110. https://doi.org/10.1007/bf03392238

Wells, F. L., & Ruesch, J. (1972). In *The mental examiners' handbook comp.* Psychological Corporation.

Williams, D. E., Kirkpatrick-Sanchez, S., Enzinna, C., Dunn, J., & Borden-Karasack, D. (2009). The clinical management and prevention of pica: A retrospective follow-up of 41 individuals with intellectual disabilities and pica. *Journal of Applied Research in Intellectual Disabilities*, *22*(2), 210–215. https://doi.org/10.1111/j.1468-3148.2008.00490.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Thillainathan, T., Linder, B., & Cox, A. D. (2024). Program evaluation of a specialized treatment home for adults with severe challenging behavior. *Behavioral Interventions*, e2059. https://doi.org/10.1002/bin.2059