WHITEPAPER

Measuring What Matters: Preliminary Evidence for Caregiver Competence as a Clinical Outcome in Applied Behavior Analysis

AUTHORS

Cate Davis PhD, BCBA-D, LBA

Sean Yocum MA, BCBA, LBA

Michael Carrero MS, BCBA, LBA

Hi Rasmus®

HICKORY LEARNING GROUP

Abstract

The field of Applied Behavior Analysis (ABA) is undergoing increasing scrutiny regarding the sustainability, ethical alignment, and long-term effectiveness of its service delivery models. Historically, quality in ABA has been operationalized as the number of direct intervention hours delivered to the learner, often reinforcing a clinician-dependent model that prioritizes intensity over durability (Dixon et al., 2016; Linstead et al., 2017). This paradigm conflicts with foundational principles of behavior analysis, which define meaningful behavior change as that which generalizes across contexts, persists over time, and is maintained by natural contingencies (Baer, Wolf, & Risley, 1968; Stokes & Baer, 1977). Emerging outcome literature further suggests that early gains achieved during intensive intervention do not reliably predict long-term adaptive functioning, raising concerns regarding the efficacy of hour-based treatment frameworks (Jónsdóttir et al., 2018).

An important gap within current practice is the insufficient measurement and development of caregiver competence as a primary clinical outcome (Bearss et al., 2015; Brookman-Frazee et al., 2006). Without deliberate programming for caregiver implementation fidelity, families frequently experience a "graduation cliff," wherein progress diminishes or regresses following the reduction or termination of services. This phenomenon aligns with research demonstrating that treatment effectiveness is significantly mediated by caregiver adherence, confidence, and skill acquisition (Allen & Warzak, 2000; Strauss et al., 2012; Martínez-Rico et al., 2025). Although parent-mediated intervention (PMI) literature has demonstrated that caregivers can implement interventions with high fidelity when properly trained (Kasari et al., 2015; Conrad et al., 2021), the absence of standardized tools to measure caregiver competence limits the integration of these approaches into routine clinical practice.

This paper presents a retrospective review of archival clinical data collected across a sample of 20 families, with treatment durations ranging from 3 to 24 months, to evaluate the impact of the Generalization Assessment for Independence & Nurturing (GAIN) Protocol within a hybrid ABA service model.

Findings from the archival review indicate statistically significant increases in caregiver competence over time, with the large majority of families moving into a higher caregiver-independence tier by the current assessment. Given the retrospective design and absence of a comparison group, these patterns should be interpreted as preliminary rather than confirmatory; in particular, because service-hours data were not available for quantitative analysis in this sample, the relationship between rising GAIN scores and reduced service intensity is described qualitatively here and warrants direct measurement in future work. Nonetheless, systematic caregiver training paired with

ongoing performance measurement via the GAIN Protocol appears associated with improved generalization and movement toward discharge readiness, supporting continued investigation of caregiver-mediated outcomes as a measurable indicator of treatment effectiveness alongside traditional service-intensity models.

Introduction

Applied Behavior Analysis (ABA) is a scientific discipline focused on understanding and improving socially significant behavior through the systematic application of behavioral principles. Rooted in the philosophy of behaviorism, ABA emphasizes observable and measurable behavior, environmental determinants, and data-driven decision-making. Core components of ABA include functional assessment, reinforcement-based intervention, skill acquisition programming, behavior reduction strategies, and continuous measurement of outcomes (Baer et al., 1968). Much of the field's evidence base for intervention intensity traces to early outcome studies demonstrating substantial gains from high-hour, clinician-delivered intervention (Lovaas, 1987; Sallows & Graupner, 2005), findings that helped establish direct service hours as a proxy for treatment quality.

A defining characteristic of effective ABA intervention is not simply behavior change in controlled environments, but the generalization and maintenance of those behaviors across time, people, and settings. Generalization refers to the transfer of learned skills beyond the training context, while maintenance reflects the persistence of behavior change after intervention is reduced or removed (Stokes & Baer, 1977; Stokes & Osnes, 1989). Despite these foundational principles, contemporary service delivery models often emphasize high-intensity, clinician-led intervention, which can inadvertently foster dependency on professional support rather than independence within the natural environment. This concern is not new: clinic-based behavior change has long been recognized as insufficient on its own, since gains secured in treatment settings frequently fail to persist when caregivers are not equipped to sustain them at home (Allen & Warzak, 2000).

Hybrid ABA models have emerged as an alternative framework, integrating direct intervention with structured caregiver training and environmental supports. Within this approach, Parent-Mediated Interventions (PMIs) play a central role, though the field has used inconsistent terminology—spanning parent training, parent coaching, and parent-mediated intervention—to describe closely related approaches (Bearss et al., 2015). PMIs involve training caregivers to implement behavioral strategies within daily routines, thereby increasing opportunities for learning and promoting ecological validity. Meta-analytic research demonstrates that PMIs can produce significant improvements in communication, social behavior, and adaptive functioning in children

with autism spectrum disorder (Conrad et al., 2021; Tabatabaei et al., 2022), a pattern substantiated in randomized comparative trials of parent-mediated approaches for young children (Kasari et al., 2015). Caregiver-implemented interventions have also been associated with increased generalization and reduced reliance on clinical services (Strauss et al., 2012).

Parent-Implemented Intervention has itself been formally recognized as an evidence-based practice by the National Clearinghouse on Autism Evidence and Practice (NCAEP), the field's most comprehensive ongoing review of the autism intervention literature. NCAEP defines Parent-Implemented Intervention as parent delivery of an intervention to a child that promotes social communication or other skills, or decreases challenging behavior, and the practice met NCAEP's evidence-based criteria on the strength of 55 qualifying studies—28 group-design and 27 single-case-design—across both review periods (Hume et al., 2021; Steinbrenner et al., 2020). The evidence base grew substantially between reviews, from 13 qualifying studies in the 1990–2011 period to 42 in the 2012–2017 period, reflecting a rapidly maturing area of the field.

Several manualized parent-implemented models have themselves accumulated sufficient evidence to be separately classified as evidence-based within this category, including Project ImPACT (Ingersoll & Dvortcsak, 2019) and Stepping Stones Triple P (Turner et al., 2010), underscoring that structured, manualized approaches to parent-implemented intervention are empirically supported rather than merely feasible (Hume et al., 2021). At the same time, NCAEP's companion review of null findings identified four qualifying studies in which parent-implemented intervention did not produce statistically or experimentally significant effects on outcomes including joint attention, social behavior, and challenging behavior (Nowell et al., 2021), a reminder that, like any evidence-based practice, PII does not produce uniform effects across all children, caregivers, or outcome domains.

What existing reviews of Parent-Implemented Intervention do not evaluate, however, is caregiver outcome itself—that is, whether caregivers' own skill, confidence, and independence in implementing strategies change over the course of treatment. NCAEP's evidence base, like most of the literature it synthesizes, is organized around child outcomes; caregiver competence is treated as a delivery mechanism rather than a measured result. This gap is precisely where the present study is positioned: rather than asking only whether parent-implemented intervention improves child outcomes, the GAIN Protocol asks whether caregivers themselves are becoming more capable and independent as a result of treatment.

However, a significant limitation within the PMI literature is the lack of standardized measurement systems to evaluate caregiver competence as a clinical outcome. While fidelity checklists and coaching models exist, few frameworks integrate caregiver performance into treatment decision-making processes such as titration or discharge planning. Evidence-based staff training models offer a template for what such measurement could look like, emphasizing performance feedback and direct observation over passive instruction (Parsons et al., 2012), but comparable frameworks for caregivers remain underdeveloped. This gap contributes to inconsistencies in implementation and limits the scalability of caregiver-centered models. More broadly, emerging evidence suggests that treatment hours alone are a poor predictor of functional outcomes once care is individualized, raising questions about service models that rely primarily on service volume rather than individualized, capacity-building approaches (Ostrovsky et al., 2022). This pattern is echoed in findings that supervision quality and caseload, rather than treatment hours alone, predict outcomes (Dixon et al., 2016), and that the relationship between intervention intensity and duration and treatment outcomes varies substantially across domains (Linstead et al., 2017).

Relevance to the Field of Applied Behavior Analysis

The integration of caregiver competence as a primary outcome variable represents a meaningful evolution in ABA practice. Comprehensive reviews of evidence-based practice in autism intervention have consistently emphasized generalization and family involvement as markers of intervention quality (Wong et al., 2015; Smith & Iadarola, 2015), and accreditation standards within the field increasingly call for outcome measurement beyond service volume alone (Autism Commission on Quality, 2022). As the field faces increasing demands for cost-effectiveness, ethical accountability, and long-term impact, traditional hour-based models face growing scrutiny. A caregiver-centered, hybrid approach may align more closely with the defining dimensions of applied behavior analysis by emphasizing generality, effectiveness, and technological application.

Additionally, this model has the potential to address systemic challenges within the field, including workforce shortages, clinician burnout, and disparities in service access. By shifting the role of the clinician from direct implementer to trainer and coach, hybrid models expand service capacity while maintaining treatment integrity. This approach also aligns with value-based care initiatives, which prioritize outcomes over service volume.

Methodology

Design

This paper presents a retrospective review of archival clinical data collected during the routine provision of ABA services; no prospective experimental manipulation was conducted. This retrospective review was employed to evaluate the effectiveness of the GAIN Protocol within a hybrid Applied Behavior Analysis (ABA) service delivery model. This should be considered a preliminary, exploratory analysis rather than a confirmatory evaluation of the GAIN Protocol.

Population

Records were reviewed for 38 client entries in the clinical database. Clients were receiving ABA services through a hybrid, caregiver-centered service model at the time of this study. Of these 38 available records, 20 had a complete GAIN-P score at both baseline and a current/reassessment timepoint and comprise the final analytic sample reported below. Of the remaining 18, thirteen had a baseline score only, with no recorded current score at the time of this review, and were excluded on that basis; the other five had no GAIN-P data recorded at all. Clients missing either timepoint were excluded listwise.

The sample consisted of 20 individuals diagnosed with Autism Spectrum Disorder (ASD) who received services through Hickory Learning Group between 2024 and 2026. The cohort was predominantly male (n = 18, 90.0%) and included two females (n = 2, 10.0%), with participant ages ranging from 2 years, 10 months to 20 years (M = 8.53 years, SD = 4.86). Documented DSM-5 ASD severity levels included Level 1 (n = 3, 15.0%), Level 2 (n = 6, 30.0%), and Level 3 (n = 7, 35.0%), while the remaining participants presented with unstratified or unspecified support levels in clinical records (n = 4, 20.0%).

Table 1

Participant Demographic Characteristics (N = 20)

Characteristic	n	%
Sex		
Male	18	90.0
Female	2	10.0
ASD Severity Level (DSM-5)		
Level 1	3	15.0
Level 2	6	30.0
Level 3	7	35.0
Unspecified/unstratified	4	20.0

Note. Age ranged from 2 years, 10 months to 20 years (M = 8.53 years, SD = 4.86).

In addition to ASD, 75.0% of the sample (*n* = 15) presented with at least one co-occurring medical, behavioral health, neurodevelopmental, or genetic condition, whereas 25.0% (*n* = 5) had no secondary diagnoses documented. Because many clients presented with more than one co-occurring diagnosis, these categories are not mutually exclusive, and the reported percentages therefore sum to more than the overall 75.0% comorbidity rate. Speech, language, and motor impairments were the most prevalent comorbidities (*n* = 8, 40.0%), including mixed receptive-expressive language disorder (F80.2, *n* = 4, 20.0%), unspecified speech/language delay (F80.89, F80.9, *n* = 4, 20.0%), articulation disorder (F80.0, *n* = 1, 5.0%), specific developmental disorder of motor function (F82, *n* = 1, 5.0%), and other communication-related presentations including nonverbal status and an unspecified childhood social-communication disorder (R47.01, F94.9, *n* = 2, 10.0%). Intellectual and developmental delays were identified in 50.0% of participants (*n* = 10), encompassing Global Developmental Delay (F88, *n* = 4, 20.0%), unspecified developmental delay (R62.0, R62.50, *n* = 2, 10.0%), and Intellectual Developmental Disorder ranging from mild to severe (F70, F71, F72; *n* = 5, 25.0%).

Table 2

Co-occurring Speech/Language/Motor and Developmental Diagnoses (n = 20)

Diagnostic Category	n	%
Any co-occurring condition	15	75.0
No secondary diagnosis documented	5	25.0
Speech, language, and motor impairments	8	40.0
Mixed receptive-expressive language disorder (F80.2)	4	20.0
Unspecified speech/language delay (F80.89, F80.9)	4	20.0
Articulation disorder (F80.0)	1	5.0
Specific developmental disorder of motor function (F82)	1	5.0
Other communication-related presentations (nonverbal status, unspecified social-communication disorder; R47.01, F94.9)	2	10.0
Intellectual and developmental delays	10	50.0
Global Developmental Delay (F88)	4	20.0
Unspecified developmental delay (R62.0, R62.50)	2	10.0
Intellectual Developmental Disorder, mild–severe (F70–F72)	5	25.0

Note. Categories and subtypes are not mutually exclusive; clients may present with more than one co-occurring diagnosis, so subtype counts within a category can sum to more than the category total.

Co-occurring Attention-Deficit/Hyperactivity Disorder (ADHD) was documented in 25.0% of the cohort (*n* = 5), presenting as combined type (F90.2, *n* = 3, 15.0%) and predominantly inattentive type (F90.0, *n* = 2, 10.0%). Genetic, congenital, and neurological conditions (*n* = 5, 25.0%) included Down syndrome/Trisomy 21 (Q90.9, *n* = 2, 10.0%), DiGeorge syndrome (D82.1, *n* = 1, 5.0%), Prune Belly Syndrome and MMIHS (Q87.89, Q43.1, *n* = 1, 5.0%), idiopathic/generalized epilepsy (G40.909, G40.309, *n* = 3, 15.0%), congenital cerebral anomalies including cerebral cysts and ventriculomegaly (G93.0, Q04.4, *n* = 1, 5.0%), and a history of craniosynostosis repair (*n* = 1, 5.0%). Additional psychiatric, behavioral, and somatic presentations (*n* = 7, 35.0%) included separation anxiety disorder (F93.0, *n* = 1, 5.0%), unspecified depression (F32.A, *n* = 1, 5.0%), Tourette's syndrome (F95.2, *n* = 1, 5.0%), conduct/disruptive behavior disorders (F91.8, F91.9, *n* = 2, 10.0%), sleep disturbance (G47.9, *n* = 1, 5.0%), sensory food aversion/feeding difficulties (R63.30, R63.39, *n* = 2, 10.0%), aggressive or oppositional behavior (R46.89, *n* = 1, 5.0%), sensory processing difficulties and bilateral Eustachian

tube dysfunction (R20.8, R20.9, H69.93, *n* = 1, 5.0%), hypoglycemia (E16.2, *n* = 1, 5.0%), and other somatic or medical conditions including chronic kidney disease, thyroid dysfunction, gastroesophageal reflux, allergic rhinitis, and intestinal malrotation (*n* = 1, 5.0%).

Table 3

Co-occurring ADHD, Genetic/Congenital/Neurological, and Psychiatric/Behavioral Diagnoses (n = 20)

Diagnostic Category	n	%
ADHD	5	25.0
Combined type (F90.2)	3	15.0
Predominantly inattentive type (F90.0)	2	10.0
Genetic, congenital, and neurological conditions	5	25.0
Down syndrome/Trisomy 21 (Q90.9)	2	10.0
DiGeorge syndrome (D82.1)	1	5.0
Prune Belly Syndrome and MMIHS (Q87.89, Q43.1)	1	5.0
Idiopathic/generalized epilepsy (G40.909, G40.309)	3	15.0
Congenital cerebral anomalies (cerebral cysts, ventriculomegaly; G93.0, Q04.4)	1	5.0
History of craniosynostosis repair	1	5.0
Additional psychiatric, behavioral, and somatic presentations	7	35.0
Separation anxiety disorder (F93.0)	1	5.0
Unspecified depression (F32.A)	1	5.0
Tourette's syndrome (F95.2)	1	5.0
Conduct/disruptive behavior disorders (F91.8, F91.9)	2	10.0
Sleep disturbance (G47.9)	1	5.0
Sensory food aversion/feeding difficulties (R63.30, R63.39)	2	10.0
Aggressive or oppositional behavior (R46.89)	1	5.0
Sensory processing difficulties and bilateral Eustachian tube dysfunction (R20.8, R20.9, H69.93)	1	5.0
Hypoglycemia (E16.2)	1	5.0
Other somatic/medical conditions (chronic kidney disease, thyroid dysfunction, gastroesophageal reflux, allergic rhinitis, intestinal malrotation)	1	5.0

Note. Categories and subtypes are not mutually exclusive; clients may present with more than one co-occurring diagnosis, so subtype counts within a category can sum to more than the category total.

Recorded treatment duration for parent-mediated intervention services ranged from 3 to 24 months (*M* = 11.55, *SD* = 6.48). Funding sources comprised public Medicaid Managed Care/LME-MCO programs (*n* = 12, 60.0%; Partners: *n* = 6, 30.0%; Vaya: *n* = 6, 30.0%) and commercial insurance plans (*n* = 8, 40.0%; Aetna: *n* = 4, 20.0%; Blue Cross Blue Shield: *n* = 2, 10.0%; Cigna: *n* = 2, 10.0%).

Table 4

Parent-Mediated Intervention Treatment Duration (N = 20)

Measure	M	SD	Min	Max
PMI treatment duration (months)	11.55	6.48	3	24

Note. M = mean; SD = standard deviation.

Table 5

Funding Source by Payer (N = 20)

Payer	n	%
Medicaid Managed Care/LME-MCO	12	60.0
Partners	6	30.0
Vaya	6	30.0
Commercial insurance	8	40.0
Aetna	4	20.0
Blue Cross Blue Shield	2	10.0
Cigna	2	10.0

Note. Percentages are based on N = 20.

Materials

The primary outcome measure used in this study was the Generalization Assessment for Independence and Nurturing (G.A.I.N), a caregiver competency self-assessment designed to evaluate caregivers' perceived confidence and independence in supporting a child's treatment goals in natural environments.

For this analysis, the GAIN-P (Parent Assessment) served as the outcome measure of caregiver competence.

GAIN-P (Parent Assessment): The GAIN-P is a caregiver self-assessment questionnaire that measures caregivers' perceived confidence in their ability to implement behavioral strategies, promote skill generalization, manage challenging behaviors, and support their child's ongoing development outside of direct therapy. The instrument captures caregivers' self-perceived readiness and independence across 15 domains using a 6-point Likert scale ranging from 0 (Not at All) to 5 (Completely Independent). Scores provide insight into caregivers' confidence levels and perceived support needs. Full administration instructions provided to caregivers and clinical staff are reproduced in the Appendix (see Figure A1).

GAIN-P scores range from 0 to 75 and are interpreted using five support-need bands, with higher scores indicating greater caregiver independence and reduced need for professional support: 0-15, High Support (full BST model recommended); 16-30, Moderate Support (continued coaching and modeling recommended); 31-45, Growing Independence (caregiver demonstrates growing independence but benefits from periodic feedback); 46-60, Highly Independent (caregiver is highly independent with limited professional guidance); and 61-75, No Additional Support Needed (caregiver is fully capable of implementing strategies independently). These bands are used in the tier-shift analysis reported in the Results section. To date, no reliability or validity data (e.g., internal consistency, test-retest reliability) for the GAIN-P have been published or independently verified; psychometric evaluation of the instrument represents an important direction for future work (see Limitations).

Procedure

Upon intake, each family completed a baseline GAIN assessment in conjunction with a comprehensive behavior analytic assessment conducted as part of the standard clinical evaluation for ABA services. The behavioral assessment established medical necessity, identified individualized treatment targets, and informed the development of the treatment plan. The GAIN assessment served as a complementary measure of caregiver independence and competency, evaluating the extent to which caregivers could successfully support skill acquisition, generalization, and maintenance within natural environments.

Following the initial assessment, families received individualized ABA intervention consisting of direct clinician-delivered therapy supplemented by structured caregiver guidance sessions; the specific service-hours breakdown for this sample has not yet been verified and is not reported here (see Limitations). Intervention emphasized Behavior Skills Training (BST), including instruction, modeling, rehearsal, and performance feedback, with the goal of transferring treatment implementation from

clinicians to caregivers while promoting generalization across home and community settings.

At the conclusion of each treatment cycle, both the GAIN assessment and the comprehensive behavioral assessment were re-administered. Clinical outcome data — including progress toward individualized treatment goals, reductions in challenging behavior, caregiver implementation fidelity, and changes in GAIN scores — were reviewed to determine the appropriate level of ongoing service. Assessment findings informed whether treatment intensity should be maintained, modified, or reduced according to a structured titration model. This assessment-intervention-reassessment cycle was repeated periodically throughout treatment until families met established graduation criteria and were discharged from ABA services.

The GAIN-P was administered by the case managing professional at intake and again at each reassessment point to address discharge or titration planning. To minimize bias, explicit instructions were provided to caregivers for completing the assessment: caregivers were asked to rate their own confidence and independence directly, and clinicians were limited to providing clarification of question wording without offering examples, hints, or feedback on individual items.

A three-phase titration model was implemented to systematically fade direct clinical services as caregiver independence increased. Progression through each phase required (a) demonstration of generalization and maintenance for at least 75% of individualized treatment goals and (b) a minimum 25% increase in caregiver competence scores relative to the previous assessment period. Upon meeting these criteria, direct therapy hours were reduced by approximately 25%, with caregivers assuming increasing responsibility for intervention implementation.

Data Analysis

All analyses were conducted with the paired sample described above (n = 20) using IBM SPSS Statistics. The normality of the baseline-to-current change scores was evaluated using the Shapiro-Wilk test to confirm the appropriateness of a parametric paired comparison. The primary analysis was a paired-samples t-test comparing GAIN-P scores at baseline and at the current assessment, with Cohen's d calculated as the standardized effect size. A Wilcoxon signed-rank test was conducted as a nonparametric robustness check. Because the GAIN-P is bounded at 75 points, a Pearson correlation between baseline score and the magnitude of change was calculated to evaluate the extent to which caregivers with lower starting scores accounted disproportionately for observed gains (a regression-to-the-mean / ceiling-effect check). Finally, each client's baseline and current score was classified

into one of five support-need tiers (High Support, Moderate Support, Growing Independence, Highly Independent, No Additional Support Needed) using the score bands established for the GAIN Protocol, and a Bowker's test of symmetry was used to evaluate whether tier classification shifted systematically from baseline to the current assessment. To evaluate whether treatment duration accounted for the magnitude of observed gains, a Pearson correlation was calculated between months of intervention and GAIN-P point change; a second correlation between months of intervention and baseline GAIN-P score was calculated to check whether clients with lower baseline scores were systematically retained in treatment longer.

Results

Caregiver Outcomes

Across the 20 clients with complete GAIN-P data, caregiver competence scores increased significantly from baseline (M = 24.00, SD = 14.50) to the current assessment (M = 46.90, SD = 10.51), an average gain of 22.90 points (SD = 13.57) on the 75-point scale. The distribution of change scores did not depart significantly from normality (Shapiro-Wilk W = .943, p = .274), supporting the use of a parametric test. A paired-samples t-test confirmed that this increase was statistically significant, t(19) = 7.55, p < .001, with a large effect size (Cohen's d = 1.69). A Wilcoxon signed-rank test produced a convergent result (p < .001), providing a nonparametric confirmation of the finding.

Table 6

Descriptive Statistics for GAIN-P Scores (n = 20)

Measure	M	SD	Min	Max
Baseline	24.00	14.50	4	59
Current	46.90	10.51	23	63
Change (Current − Baseline)	22.90	13.57	3	47

Note. M = mean; SD = standard deviation.

Table 7

Paired-Samples Test of GAIN-P Scores, Baseline vs. Current

Test	Statistic	df	95% CI of diff.	p	Effect size
Paired t-test	t = 7.55	19	[16.55, 29.25]	< .001	d = 1.69
Wilcoxon signed-rank	—	—	—	< .001	—

Note. The Wilcoxon signed-rank test is reported as a nonparametric robustness check; the change scores did not depart significantly from normality (Shapiro-Wilk W = .943, p = .274), supporting the paired t-test as the primary result.

Baseline score was significantly and negatively correlated with the magnitude of change, r(18) = -.72, p < .001: caregivers who began with lower GAIN-P scores tended to show larger point gains. Given that the instrument has a fixed ceiling of 75 points, this pattern is consistent with a ceiling effect and should be considered when interpreting the magnitude of improvement — gains are not uniform across the sample, and caregivers who started closer to independence had proportionally less room to improve.

Treatment duration ranged from 3 to 24 months across the sample (M = 11.55, SD = 6.48). Months of intervention were not significantly correlated with the magnitude of GAIN-P point change, r(18) = .13, p = .583, indicating that the observed gains are not simply attributable to clients with longer enrollment having more time to improve. Months of intervention were also uncorrelated with baseline GAIN-P score, r(18) = .02, p = .922, indicating that clients with lower baseline scores were not systematically retained in treatment longer than those who started higher. Together, these results suggest the caregiver gains reported above reflect the intervention itself rather than an artifact of variable treatment length.

As a supplementary descriptive measure, the rate of GAIN-P change per month of intervention was calculated for each client (change score ÷ months in treatment). Rates ranged from 0.25 to 7.83 points per month (M = 2.64, SD = 2.10; Table 8). This metric is reported descriptively only — see Discussion for a caveat on its interpretation.

Table 8

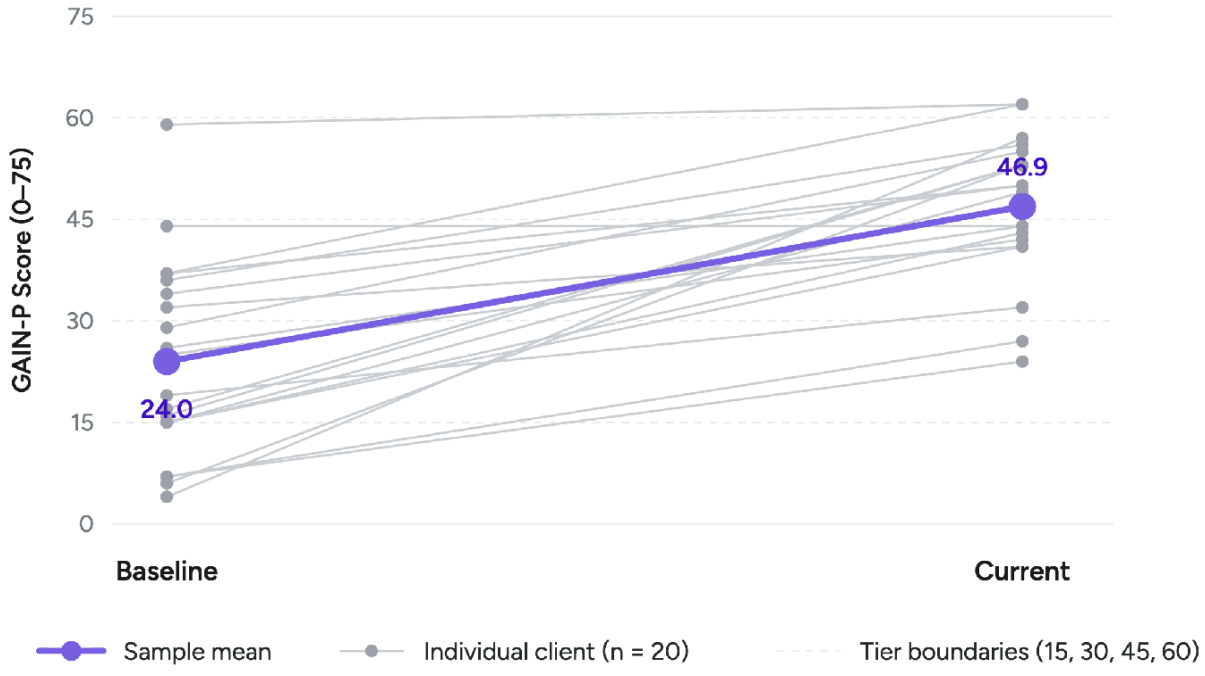
Rate of GAIN-P Change per Month of Intervention (n = 20)

Measure	M	SD	Min	Max
Rate of change (points/month)	2.64	2.10	0.25	7.83

Note. Rate = (Current − Baseline) / months between baseline and current assessment for each client.

Figure 1

Individual GAIN-P Trajectories, Baseline to Current (n = 20)



Note. Each line represents one client's GAIN-P score at baseline and at the current assessment. The bold red line indicates the sample mean. Dashed lines mark the four boundaries separating the five caregiver support tiers (15, 30, 45, and 60).

Shifts in Caregiver Support Tier

Classifying each client's baseline and current score into the five GAIN-P support tiers showed a consistent pattern of upward movement. Eighteen of 20 clients (90.0%) moved into a higher independence tier between baseline and the current assessment, two clients (10.0%) remained within the same tier, and no client regressed to a lower tier. A Bowker's test of symmetry confirmed that this shift in classification was statistically significant, $\chi^2(7) = 18.00$, $p = .012$, indicating that movement across tiers

was systematically toward greater caregiver independence rather than symmetric or random reclassification. At the point of current assessment, 2 of 20 clients (10.0%) had reached the No Additional Support Needed tier, consistent with readiness for discharge from services; the remaining 18 clients continued to receive services at varying levels of support.

Table 9

Caregiver Support Tier at Baseline and Current Assessment (n = 20)

Baseline \ Current	High Support	Moderate Support	Growing Independence	Highly Independent	No Additional Support	Row Total
High Support	0	2	3	4	0	**9**
Moderate Support	0	0	0	3	1	**4**
Growing Independence	0	0	2	4	0	**6**
Highly Independent	0	0	0	0	1	**1**
No Additional Support	0	0	0	0	0	**0**
Column Total	**0**	**2**	**5**	**11**	**2**	**20**

Note. Bowker's test of symmetry, $\chi^2(7) = 18.00$, $p = .012$.

Discussion

The findings of this study are consistent with the feasibility of a caregiver-centered hybrid ABA service model and offer preliminary support for its effectiveness. By operationalizing caregiver competence as a measurable clinical outcome, the GAIN Protocol addresses a longstanding gap in the field by providing an objective framework for assessing caregiver readiness, informing treatment planning, and guiding service titration. This approach aligns intervention practices with the foundational principles of behavior analysis, which emphasize the transfer of stimulus control and the maintenance of behavior under naturally occurring contingencies rather than continued dependence on professional intervention.

The results further support a shift toward value-based, caregiver-centered models of ABA service delivery in which the success of treatment is evaluated not only by improvements in client behavior, but also by the caregiver's ability to independently implement behavioral strategies and sustain meaningful outcomes across everyday environments, consistent with recent evidence that caregiver confidence helps explain the relationship between child functioning and caregiver burden in early intervention (Martínez-Rico et al., 2025). By systematically measuring caregiver competence and incorporating those data into clinical decision-making, the GAIN Protocol positions caregiver independence as a primary indicator of treatment success rather than a secondary objective. The observed improvements in caregiver competence suggest that investing in caregiver capacity may improve both the effectiveness and efficiency of ABA services, though this study did not directly measure service titration outcomes or client-level behavioral maintenance, and these remain important targets for future research. As healthcare systems continue to emphasize value-based care, accountability, and meaningful functional outcomes, caregiver-centered models such as the GAIN Protocol may offer a more scalable, ethical, and sustainable framework for the delivery of high-quality behavior analytic services.

Beyond its clinical implications, this model carries implications for service capacity and workforce sustainability. Structuring treatment around a lower-intensity, caregiver-centered format that supplements reduced direct clinician hours with structured caregiver coaching may allow organizations to expand access to a greater number of families without proportionally increasing staffing demands, while also reducing the sustained burden placed on RBTs and BCBAs under traditional, clinician-dependent frameworks. Because verified service-hours data for this sample were not available for this analysis (see Limitations), the magnitude of any staffing or capacity benefit cannot yet be quantified here. Because clinicians in this model increasingly function as trainers and coaches rather than sole direct interventionists, this approach may also extend behavior-analytic expertise to underserved or

geographically remote families through asynchronous or remote coaching formats.

From a financial perspective, directing resources toward caregiver training and durable skill transfer, rather than prolonged direct service delivery, may improve the return on investment of ABA services by reducing the need for extended, high-cost treatment while maintaining or improving functional outcomes. The GAIN Protocol's tier-based scoring offers funders and providers a quantifiable, if not yet independently validated, indicator of when services can reasonably be reduced (see Limitations regarding the conflict of interest inherent in using organization-generated scores for this purpose).

The three-phase titration model embedded within the GAIN framework provides a structured, data-driven pathway for adjusting service intensity as caregiver competence and client generalization improve, increasing transparency and consistency in clinical decision-making relative to less formalized approaches. However, because titration decisions and GAIN reassessment occur within the same clinical cycle, this model cannot on its own establish whether rising caregiver competence causes reductions in service hours or follows from them (see Limitations). Pairing short-term process indicators, such as caregiver training fidelity and session performance, with longer-term outcome measures like GAIN scores may still offer a more complete picture of whether behavior change is likely to be maintained once formal services end, even as the causal question remains open.

A related caution applies to the rate-of-change metric reported in Table 8 (points of GAIN-P gain per month of intervention). Because this figure is calculated by dividing each client's total point change by their number of months in treatment, it is disproportionately sensitive to clients with short enrollment durations: a modest point gain over a 3-month period yields a much larger rate than the same point gain spread over 24 months, even though the underlying clinical progress may be comparable. The wide range observed (0.25 to 7.83 points per month) should therefore be read as a descriptive summary of variability in this sample rather than a stable, generalizable estimate of how quickly caregiver competence typically improves, and it should not be used to set expectations for individual families or to benchmark clinicians.

Limitations

Several limitations should be considered when interpreting the findings of this retrospective sample. First, although the analytic sample (n = 20) is considerably larger than the six-family case series originally examined in early drafts of this work, it remains modest relative to the field's need for generalizable evidence, and results should still be viewed as preliminary rather than definitive proof of effectiveness. Future research involving larger and more diverse samples across multiple clinical settings, and directly comparing this caregiver-centered hybrid model against other service delivery models, will be necessary to evaluate the reliability and external validity of the GAIN Protocol.

Second, the analytic sample was shaped by substantial attrition from the original pool of 38 client records: 18 records (47.4%) were excluded because they lacked either a baseline or a current/reassessment GAIN-P score, including 13 clients with a baseline score only and no completed current assessment. The reasons these families did not have a completed reassessment—whether disengagement from services, transfer to another provider, early discharge, or simple documentation gaps—were not systematically tracked, and it cannot be ruled out that excluded families differed systematically from the analytic sample on caregiver engagement or outcome trajectory. If families who disengaged from services were also less likely to be building caregiver competence, the reported 90% tier-shift rate would represent an upper bound on the effect within the full referred population rather than an unbiased estimate of it. Future work should track and report reasons for missing follow-up data and, where possible, compare baseline characteristics of included and excluded clients.

Third, the absence of a comparison or control group prevents causal conclusions regarding the relationship between the GAIN Protocol and the observed improvements in caregiver competence. Although participating families demonstrated meaningful gains over time, future studies should compare the protocol against alternative service delivery models to determine whether structured caregiver competency measurement contributes to improved clinical outcomes, accelerated service titration, or greater treatment efficiency beyond what those alternative models would produce on their own.

Relatedly, because titration decisions and GAIN reassessment occurred within the same clinical cycle, this design cannot establish the direction of the relationship between rising caregiver competence and reductions in authorized service hours. It is equally plausible that growing caregiver independence prompted clinicians to reduce hours, that planned reductions in hours prompted caregivers to build independence out of necessity, or that both changes were driven by a third factor such as overall

client progress. Prospective designs that manipulate or independently track hours and GAIN scores on separate timelines would be needed to disentangle this relationship.

An additional, and more significant, limitation concerns the organizational context in which caregiver competence was measured. The GAIN Protocol was developed by, and the GAIN-P was scored within, the same organization that provided the ABA services being evaluated; that organization also uses GAIN scores to authorize reductions in billed service hours through the titration model described above. This represents a direct financial conflict of interest of the kind the field's own ethics code directs behavior analysts to identify and mitigate (Behavior Analyst Certification Board, 2020): the entity with a stake in demonstrating that caregivers are becoming more independent is the same entity generating the data used to make that case, and the same entity whose billing benefits from measured improvement. While standardized administration instructions were used for the GAIN-P (see Procedure), this does not eliminate the underlying incentive structure, and no independent party verified scores, administration fidelity, or the titration decisions made on the basis of those scores. Meaningful independent verification would require, at minimum: scoring or re-scoring by evaluators with no financial or supervisory relationship to the treating organization; reported inter-rater reliability statistics between organization-affiliated and independent raters; and replication of these findings at organizationally unrelated sites using the same protocol. Basic psychometric properties of the GAIN-P, including internal consistency and test-retest reliability, have also not been established (see Materials), further limiting confidence in the precision of individual scores. Future investigations should prioritize these steps before the GAIN Protocol is treated as an objective, bias-free standard for authorizing service reductions.

A related but distinct threat concerns the conditions under which the GAIN-P was completed, separate from who benefits from the result. The same clinician who worked with a caregiver over the course of treatment, and who was aware that higher scores could support a reduction in authorized hours, also typically administered that caregiver's current/reassessment GAIN-P rating. Even with standardized administration instructions and good intentions on all sides, this repeated, non-blinded pairing of rater and respondent creates conditions for instrumentation and demand-characteristic effects that are well documented in behavioral assessment more broadly: caregivers may over-report confidence to a rater they have built a relationship with, and raters may unconsciously interpret ambiguous behavior more favorably at the second assessment. This is a threat to measurement validity distinct from the financial conflict of interest described above, and it would persist even if the organization had no financial stake in the outcome. Blinded or third-party administration of the GAIN-P at reassessment would be the most direct way to rule this out.

The present study captured caregiver competence at a single current/reassessment timepoint rather than specifically at the point of discharge; only 2 of the 20 clients had reached the tier consistent with discharge readiness by the time of this review, and the remaining 18 continued to receive services at varying levels of support. Because long-term follow-up data beyond the current assessment were not available, whether caregiver gains are maintained after formal services end remains unknown. Additional research examining caregiver independence, treatment maintenance, and re-entry into services at 6-, 12-, and 24-month intervals following discharge would provide valuable information regarding the durability of outcomes over time.

Finally, the sample may not fully represent the diversity of families receiving ABA services, particularly with respect to socioeconomic status, cultural background, language, caregiver education, and access to community resources. Because these contextual variables likely influence caregiver learning and implementation, future research should evaluate the performance of the GAIN Protocol across more heterogeneous populations and service delivery settings.

Despite these limitations, the consistency of findings across all participating families suggests that caregiver competence represents a meaningful and measurable clinical outcome that warrants continued investigation. The results provide preliminary support for the feasibility of incorporating caregiver independence as a routine indicator of treatment progress and readiness for service titration.

Future Directions

Building on the limitations described above, the authors are actively pursuing a formal psychometric evaluation of the GAIN-P. Planned next steps include establishing internal consistency (e.g., Cronbach's alpha) using existing item-level data; assessing inter-rater reliability through independent, organizationally unaffiliated scoring of a subset of administrations, directly addressing the conflict-of-interest concern raised above; evaluating test-retest reliability through short-interval repeat administrations; and examining convergent validity against established, independently validated measures of caregiver competence or self-efficacy. Longer-term, the authors intend to pursue confirmatory factor analysis of the GAIN-P's 15-domain structure as sample size accumulates, and to investigate criterion validity by linking GAIN-P scores to independently observed caregiver fidelity and long-term client outcomes. These steps are intended to establish the GAIN Protocol as a psychometrically sound and independently verified tool before it is more broadly adopted as a standard for caregiver competence measurement and service titration decisions.

Concluding Remarks

For decades, success within ABA has been measured primarily by changes in client behavior. That measure remains essential, but it captures only part of what makes treatment gains last: durable change ultimately depends on caregivers' ability to independently implement effective strategies within the routines of everyday life. The findings reported here suggest that caregiver competence, measured systematically via the GAIN Protocol, is one such indicator worth incorporating into routine clinical decision-making alongside client-level outcomes.

These preliminary findings, drawn from a retrospective sample of 20 families, show meaningful increases in caregiver competence over the course of treatment. Whether these gains directly produced the reductions in service intensity observed during titration cannot be established from the present design (see Limitations). Additional research using larger, more diverse samples, independent evaluators, and prospective designs is needed to establish the psychometric properties of the GAIN Protocol and to more directly test its relationship to service intensity. With that caveat, when caregivers gain the knowledge, confidence, and skill to support behavior change independently, it may represent one of the more meaningful measures of treatment success available to clinicians: family capacity that persists beyond the period of formal services.

Disclosure

All data presented in this white paper were retrospectively analyzed from existing clinical records collected during the routine provision of clinical services. No prospective research procedures were implemented, and institutional review board (IRB) approval was not sought. At intake, all clients and their legal caregivers provided informed consent authorizing the use of de-identified, anonymous data for research, program evaluation, and quality improvement purposes. All analyses were conducted using aggregated, de-identified data, and no personally identifiable information is included in this report.

References

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

Autism Commission on Quality. (2022). *Applied behavior analysis accreditation program standards and guide* (Version 1.0). https://autismcommission.org

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

Bearss, K., Burrell, T. L., Stewart, L., & Scahill, L. (2015). Parent training in autism spectrum disorder: What's in a name? *Clinical Child and Family Psychology Review, 18*(2), 170–182. https://doi.org/10.1007/s10567-015-0179-5

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

Brookman-Frazee, L., Stahmer, A., Baker-Ericzén, M. J., & Tsai, K. (2006). Parenting interventions for children with autism spectrum and disruptive behavior disorders: Opportunities for cross-fertilization. *Clinical Child and Family Psychology Review, 9*(3–4), 181–200. https://doi.org/10.1007/s10567-006-0010-4

Conrad, C. E., Rimestad, M. L., Rohde, J. F., Petersen, B. H., Korfitsen, C. B., Tarp, S., Cantio, C., Lauritsen, M. B., & Händel, M. N. (2021). Parent-mediated interventions for children and adolescents with autism spectrum disorders: A systematic review and meta-analysis. *Frontiers in Psychiatry, 12*, 773604. https://doi.org/10.3389/fpsyt.2021.773604

Dixon, D. R., Linstead, E., Granpeesheh, D., Novack, M. N., French, R., Stevens, E., Stevens, L., & Powell, A. (2016). An evaluation of the impact of supervision intensity, supervisor qualifications, and caseload on outcomes in the treatment of autism spectrum disorder. *Behavior Analysis in Practice, 9*(4), 339–348. https://doi.org/10.1007/s40617-016-0132-1

Hume, K., Steinbrenner, J. R., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yücesoy-Özkan, Ş., & Savage, M. N. (2021). Evidence-based practices for children, youth, and young adults with autism: Third generation review. *Journal of Autism and Developmental Disorders, 51*(11), 4013–4032. https://doi.org/10.1007/s10803-020-04844-2

Ingersoll, B., & Dvortcsak, A. (2019). *The Project ImPACT manual for parents.* Guilford Press.

Jónsdóttir, S. L., Brynjarsdóttir, B., Saemundsen, E., & Sigurdsson, J. F. (2018). Long-term outcome of children with autism who received different forms of early intervention during their preschool years: A pilot study of 15 young adults. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 6*(1), 28–39. https://doi.org/10.21307/sjcapp-2018-006

Kasari, C., Gulsrud, A., Paparella, T., Hellemann, G., & Berry, K. (2015). Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. *Journal of Consulting and Clinical Psychology, 83*(3), 554–563. https://doi.org/10.1037/a0039080

Linstead, E., Dixon, D. R., Hong, E., Burns, C. O., French, R., Novack, M. N., & Granpeesheh, D. (2017). An evaluation of the effects of intensity and duration on outcomes across treatment domains for children with autism spectrum disorder. *Translational Psychiatry, 7*(9), e1234. https://doi.org/10.1038/tp.2017.207

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*(1), 3–9. https://doi.org/10.1037/0022-006X.55.1.3

Martínez-Rico, G., García-Grau, P., Cañadas-Pérez, M., & González-García, R. J. (2025). Does family confidence bridge child functioning and caregiver burden in early intervention? *PLOS ONE, 20*(5), e0321997. https://doi.org/10.1371/journal.pone.0321997

Nowell, S. W., Steinbrenner, J. R., Hume, K., Odom, S. L., & Szendrey, S. (2021). *NCAEP null findings summary.* The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice. https://ncaep.fpg.unc.edu/

Ostrovsky, A., Willa, M., Cho, T., Strandberg, M., Howard, S., & Davitian, C. (2022). Data-driven, client-centric applied behavior analysis treatment-dose optimization improves functional outcomes. *World Journal of Pediatrics, 19*, 753–760. https://doi.org/10.1007/s12519-022-00643-0

Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2012). Evidence-based staff training: A guide for practitioners. *Behavior Analysis in Practice, 5*(2), 2–11. https://doi.org/10.1007/BF03391819

Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with

autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*(6), 417–438. https://doi.org/10.1352/0895-8017(2005)110[417:IBTFCW]2.0.CO;2

Smith, T., & Iadarola, S. (2015). Evidence base update for autism spectrum disorder. *Journal of Clinical Child & Adolescent Psychology, 44*(6), 897–922. https://doi.org/10.1080/15374416.2015.1077448

Steinbrenner, J. R., Hume, K., Odom, S. L., Morin, K. L., Nowell, S. W., Tomaszewski, B., Szendrey, S., McIntyre, N. S., Yucesoy-Ozkan, S., & Savage, M. N. (2020). *Evidence-based practices for children, youth, and young adults with autism.* The University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Institute, National Clearinghouse on Autism Evidence and Practice Review Team. https://ncaep.fpg.unc.edu/sites/ncaep.fpg.unc.edu/files/imce/documents/EBP%20Report%202020.pdf

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

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy, 20*(3), 337–355. https://doi.org/10.1016/S0005-7894(89)80054-1

Strauss, K., Vicari, S., Valeri, G., D'Elia, L., Arima, S., & Fava, L. (2012). Parent inclusion in early intensive behavioral intervention: The influence of parental stress, parent treatment fidelity and parent-mediated generalization of behavior targets on child outcomes. *Research in Developmental Disabilities, 33*(2), 688–703. https://doi.org/10.1016/j.ridd.2011.11.008

Tabatabaei, S. H., Shahrokhi, H., Gholipour, K., Iezadi, S., Rezapour, R., Naghibi, D., & Azami-Aghdash, S. (2022). The characteristics and results of parent training interventions in children with autism spectrum disorder: A systematic review. *Iranian Journal of Public Health, 51*(3), 518–530. https://doi.org/10.18502/ijph.v51i3.8927

Turner, K. M., Markie-Dadds, C., & Sanders, M. R. (2010). *Practitioner's manual for primary care Triple P.* Triple P International Pty.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*(7), 1951–1966. https://doi.org/10.1007/s10803-014-2351-z

Appendix
GAIN-P Administration Instructions

Figure A1

GAIN-P Administration Instructions for Caregivers and Staff

Generalization Assessment for Independence and Nurturing (GAIN-P)

Instructions for Caregivers

This questionnaire is designed to help us understand your current skills and your child's current abilities. There are no right or wrong answers. Your responses help us identify what is going well and where we can provide additional support. Please read each statement carefully and select the response that best matches your current experience. Answer each question to the best of your ability based on what typically happens at home and in your daily routines—not based on what happens only when your therapy team is present. If you are unsure about a question, choose the answer that is closest to your current level of confidence or performance. Honest answers help us develop the most appropriate treatment plan for your family.

Instructions for Staff

To maintain the integrity of this assessment, staff must administer the questionnaire consistently across all caregivers.

Staff may:

Read questions aloud if requested.
Repeat a question if needed.
Clarify the meaning of words or phrases used in the question.
Explain how to use the rating scale.

Staff may not:

Provide examples of situations or behaviors.
Offer additional explanations beyond clarifying the wording of the question.
Coach, prompt, or influence the caregiver's response.
Suggest what response should be selected.
Interpret the question in a way that changes its meaning.
Provide feedback indicating whether an answer is "correct" or "incorrect."
The caregiver should answer every question independently, based on their own experiences and current abilities.
Staff should remain neutral throughout administration to ensure the results accurately reflect the caregiver's current level of competence.

How to Rate Each Statement

For each statement, choose the number that best describes how much support you currently need to do the skill on your own. Think about what you can usually do during everyday routines—not just on your best day.

0 – Needs Substantial Support
I cannot do this skill on my own. I need someone to teach me step-by-step, show me how to do it, practice with me, and give me frequent feedback.

1 – Needs More Support
I can do a small part of the skill, but I still need someone to show me what to do and help me correct mistakes most of the time.

2 – Needs Moderate Support
I can usually do the skill, but I still need regular feedback or reminders to do it correctly.

3 – Needs Limited Support
I can do the skill most of the time. I only need occasional reminders or simple prompts to stay on track.

4 – Needs Minimal Support
I can do the skill almost independently. I only need occasional coaching or reminders in new or difficult situations.

5 – No Support Needed
I can do this skill independently and consistently. I do not need help, reminders, or coaching.