Delivering Quality ABA at Scale

Outcomes of an Integrated Instructional Framework Across a Multi-State ABA Provider

Authors:

Brittany Rader, M.Ed., BCBA, LBA, CTP; *President, Clinical Services,* Behavioral Framework

Rebekah Chesko, MS, CCC-SLP, BCBA, LBS, LBA; *Director of Learning and Development* Behavioral Framework

Cate Davis, PhD, BCBA-D, LBA; *Director of Research and Outcomes*, Hi Rasmus;

Geoff Phelps, *Senior Data Analyst*, Hi Rasmus

Behavioral Framework™

Abstract

Maintaining consistent clinical quality across large service organizations has become an increasingly important challenge in Applied Behavior Analysis (ABA). As providers expand across regions, service settings, and funding sources, systems that promote consistent implementation and measurable learner outcomes become increasingly essential. This study examined skill acquisition outcomes among home-based clients of Behavioral Framework, a multi-state ABA provider, comparing two instructional program types: Probe-Teach-Probe (PTP) and standard programming, a form of discrete trial training (DTT). Using a retrospective archival design, data were extracted from the Hi Rasmus clinical data platform across a six-month observation period (January–June 2026). The analytic sample comprised 1,090 clients and 4,022 valid client-month observations. All participants in the sample have a primary diagnosis of Autism Spectrum Disorder and were served in home settings (91.0%). Descriptive statistics confirmed substantial positive skew in the client-level mastery rate (skewness = 3.46) and mastery count (skewness = 3.00) across the full analytic sample, necessitating nonparametric inferential analyses. A Wilcoxon Signed-Rank test comparing monthly program-level mastery rates revealed that PTP significantly outperformed standard programming across all six months of the observation period, $Z = -2.201$, $p = .028$, $r = .90$. PTP program-level rates ranged from 1.171 to 1.366 mastered targets per session activity hour, compared to standard programming rates of 0.200 to 0.754. Over the six-month period, PTP programs produced 130,241 total skill masteries compared to 11,945 for standard programming. These findings provide organizational-scale evidence that an integrated instructional framework combining cold probe data collection, simultaneous prompting, and fluency-based instruction produces significantly higher skill acquisition rates than traditional discrete trial approaches, and that high-quality ABA can be delivered systematically and effectively at scale.

Keywords: applied behavior analysis, skill acquisition, instructional methodology, probe-teach-probe, discrete trial training, organizational scale, outcomes measurement, autism spectrum disorder, home-based services.

Behavioral Framework™

Executive Summary

Delivering Quality ABA at Scale: Outcomes of an Integrated Instructional Framework Across a Multi-State ABA Provider

Behavioral Framework × Hi Rasmus · January–June 2026

By the Numbers

- 1,090 home-based clients across 4 states and Washington, D.C.
- 4,022 individual client-month data points analyzed across the six-month observation period
- 130,241 total skill masteries produced by PTP over 6 months
- 10.9× more total masteries with PTP than Standard Programming
- 85.4% of clients diagnosed with ASD; 69.6% Medicaid-funded

Purpose

This white paper addresses a critical and underexamined challenge in ABA service delivery: whether clinical quality can be maintained as organizations grow. Behavioral Framework and Hi Rasmus present organizational-scale outcomes data comparing two instructional approaches — Probe-Teach-Probe (PTP) and standard programming, a form of discrete trial training (DTT) — across a clinically diverse, Medicaid-predominant home-based population. The paper is designed for ABA practitioners, clinical and organizational leaders, and payers seeking outcome-based evidence of service quality.

PTP is a structured instructional and data collection framework in which each skill progresses through three sequential phases within every session: a pre-instruction probe to assess what the learner can already do independently, a teach phase that concentrates instruction only on unmastered targets using simultaneous prompting, and a post-instruction probe to evaluate the immediate effects of teaching. Standard programming, the comparison approach, follows a traditional discrete trial format in which instruction is delivered across all active targets regardless of prior correct performance.

PTP vs. Standard Programming at a Glance

	PTP	Standard Programming
Monthly program-level mastery rate	1.17–1.37 per session hour	0.20–0.75 per session hour
Month-to-month consistency	Stable (range: 0.20 units)	Variable (range: 0.55 units)
Months leading in performance	6 of 6	0 of 6
Effect size	r = .90 (large)	Not applicable
Statistical significance	p = .028	Not applicable

Note. Effect size (r) and statistical significance (p) reflect the comparison between PTP and Standard Programming program-level mastery rates across the six-month observation period (Wilcoxon Signed-Rank test) and are not properties of either instructional approach independently.

Key Findings

PTP produced significantly higher program-level mastery rates than standard programming in every month of the observation period (1.17–1.37 vs. 0.20–0.75 mastered targets per session activity hour), with a large effect size ($r = .90$) that is clinically *and* statistically meaningful.

PTP program-level mastery rates ranged from 1.17 to 1.37 mastered targets per session activity hour across the six-month observation period, compared to standard programming program-level rates of 0.20 to 0.75 — a consistent and statistically significant advantage confirmed across all six months ($Z = -2.201$, $p = .028$, $r = .90$).

Over six months, PTP programs generated 130,241 total skill masteries compared to 11,945 for standard programming — more than 118,000 additional mastered targets representing a real and compounding advantage in learner progress.

PTP rates were notably stable month-to-month while standard programming rates fluctuated substantially, suggesting PTP produces more predictable and consistent outcomes at scale.

These findings were observed across a large, clinically heterogeneous sample — ASD diagnosis, 69.6% Medicaid-funded, 24.5% Spanish-speaking — strengthening their generalizability to real-world applied settings.

Implications

For clinicians: Integrating assessment and teaching within a single session workflow does not compromise rigor — it enhances the pace of skill acquisition and generates interpretable data every session without sacrificing instructional time.

For organizations: Standardizing instructional methodology is a viable and effective mechanism for maintaining outcome quality across a distributed, multi-state workforce. BCBAs retain full authority over treatment planning and target selection; what PTP standardizes is the procedural infrastructure through which those decisions are implemented and measured.

For payers and policymakers: This paper demonstrates what outcomes-based accountability looks like in practice — more than 4,000 client-month observations analyzed from a single provider's robust data collected through a clinical platform, yielding statistically significant evidence of differential effectiveness between instructional approaches. At a moment when ABA services face increasing scrutiny over both cost and clinical rigor, findings like these offer a concrete model for what it looks like when a provider measures what it does, submits that data to independent analysis, and demonstrates that its approach produces better outcomes.

Introduction

The Challenge of Scaling ABA Quality

The demand for applied behavior analysis (ABA) services continues to outpace the workforce available to deliver them (Behavior Analyst Certification Board, 2026). As rates of autism spectrum disorder (ASD) diagnosis have risen (Maenner et al., 2023), providers have expanded across states and service environments to improve access to care. ABA is among the most extensively researched and empirically supported interventions for individuals with ASD, with decades of evidence demonstrating its effectiveness in improving communication, adaptive, social, and academic skills while reducing behaviors that interfere with learning and daily functioning (Smith & Iadarola, 2015). As a result, ABA has become the standard of care for autism treatment and a primary service covered under Medicaid and private insurance for eligible individuals. However, growth in service delivery has introduced a critical and underexamined challenge: ensuring that treatment quality and learner outcomes are maintained as organizations scale. Increasing emphasis on accountability, outcomes measurement, and value-based care has placed greater expectations on providers to demonstrate clinical effectiveness across diverse populations and service contexts. This paper directly addresses the tension these forces create: whether high-quality, evidence-based ABA can be delivered consistently at organizational scale.

Behavioral Framework is a multi-state ABA provider that offers services across home, clinic, school, and community environments, serving a clinically heterogeneous population of children with ASD ranging from toddlers to adolescents. Clients receive services across a wide range of treatment intensities, from focused intervention programs to comprehensive high-intensity models, reflecting the real-world variability of ABA caseloads in applied settings.

Measurement and Instruction in Applied Practice

ABA practice relies on ongoing measurement to evaluate treatment effectiveness, monitor skill acquisition, and inform clinical decision-making (Cooper et al., 2020). In applied settings, clinicians must simultaneously implement instructional procedures, respond to learner behavior, manage environmental variables, and collect accurate data — competing demands that place considerable procedural burden on frontline staff (Lerman et al., 2011). When data collection systems are complex, time-intensive, or poorly integrated with instructional delivery, implementation fidelity may suffer, reducing both the quality of data obtained and the amount of time available for active teaching (DiGennaro Reed, Reed, Baez, & Maguire, 2011). Research on staff training further underscores that the design and clarity of instructional protocols directly influences the consistency with which treatment is delivered across clinicians and settings (DiGennaro Reed, Hyman, & Hirst, 2011). For large organizations with geographically distributed workforces, these risks are amplified: variation in how individual clinicians approach data collection can produce inconsistencies that undermine outcome tracking and delay clinical response to learner needs. The design of a measurement system is therefore not merely a logistical consideration — it is a clinical one.

The Probe-Teach-Probe Framework

One approach that explicitly integrates measurement and instruction is the probe-teach-probe (PTP) framework. PTP is a structured data collection and instructional methodology in which each skill progresses through its own sequence of pre-instruction probes, targeted teaching, and post-instruction probes. In the pre-test probe phase, the discriminative stimulus (SD) associated with each active target within the skill is presented without prompts or programmed reinforcement to assess independent performance, a procedure commonly referred to as a cold probe (Lerman et al., 2011). Targets on which the learner demonstrates independence during the pre-test are set aside; teaching is concentrated only on unmastered targets. During the teach phase, instruction and a dense reinforcement schedule are provided for each unmastered target. A post-test probe is then administered to assess changes in responding and evaluate the immediate effects of instruction (see Figures 1 and 2). This assessment-embedded instructional framework yields a session-level measure of skill status while seamlessly integrating evaluation and teaching within a single repeatable workflow.

A prerequisite for implementing any standardized instructional methodology at scale is a data infrastructure capable of operationalizing it at the session level. The present study was conducted using Hi Rasmus, a clinical data platform designed specifically for ABA service delivery that provides native support for the PTP framework — enabling technicians to be guided through each phase in real time and allowing session-level data to be extracted in structured formats suitable for organizational-scale analysis. This platform-level operationalization made the present study feasible and represents a meaningful complement to the clinical training infrastructure required for large-scale PTP adoption.

PTP draws from several independently validated instructional and measurement procedures within the ABA literature. Cold probe data collection — in which only the learner's first unprompted response to a target is recorded — has been demonstrated to provide a valid and less labor-intensive alternative to continuous trial-by-trial data collection without compromising instructional decision-making (Lerman et al., 2011; Leaf et al., 2016). Simultaneous prompting, a form of errorless instruction in which the controlling prompt is delivered immediately alongside the SD during teaching trials, has substantial empirical support for promoting rapid skill acquisition and minimizing learner errors across a range of skills and populations (Schuster et al., 1992; Morse & Schuster, 2004; Head et al., 2011). Compared to other response-prompting strategies, simultaneous prompting has consistently demonstrated efficiency advantages in terms of errors during both instructional and maintenance phases (Head et al., 2011). Fluency-based instructional approaches, rooted in the precision teaching literature, emphasize repeated opportunities to respond until skills are performed accurately, quickly, and with stability across varied conditions — outcomes associated with improved maintenance and generalization (Bucklin et al., 2000; Shirley & Pennypacker, 1994; Kubina & Morrison, 2000). Collectively, these procedures form the scientific foundation upon which the PTP framework is built.

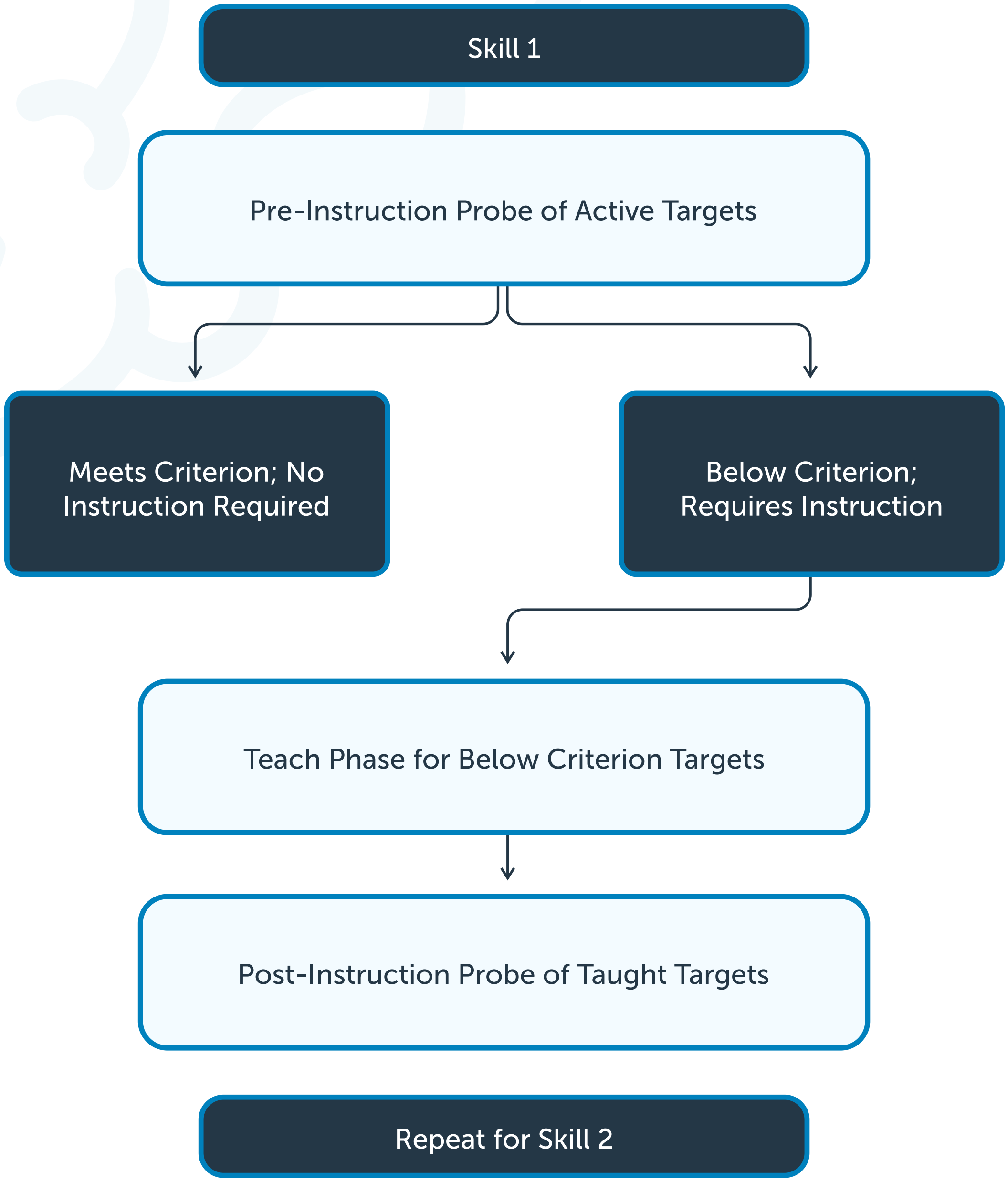

Figure 1.
Probe-Teach-Probe (PTP) instructional framework.

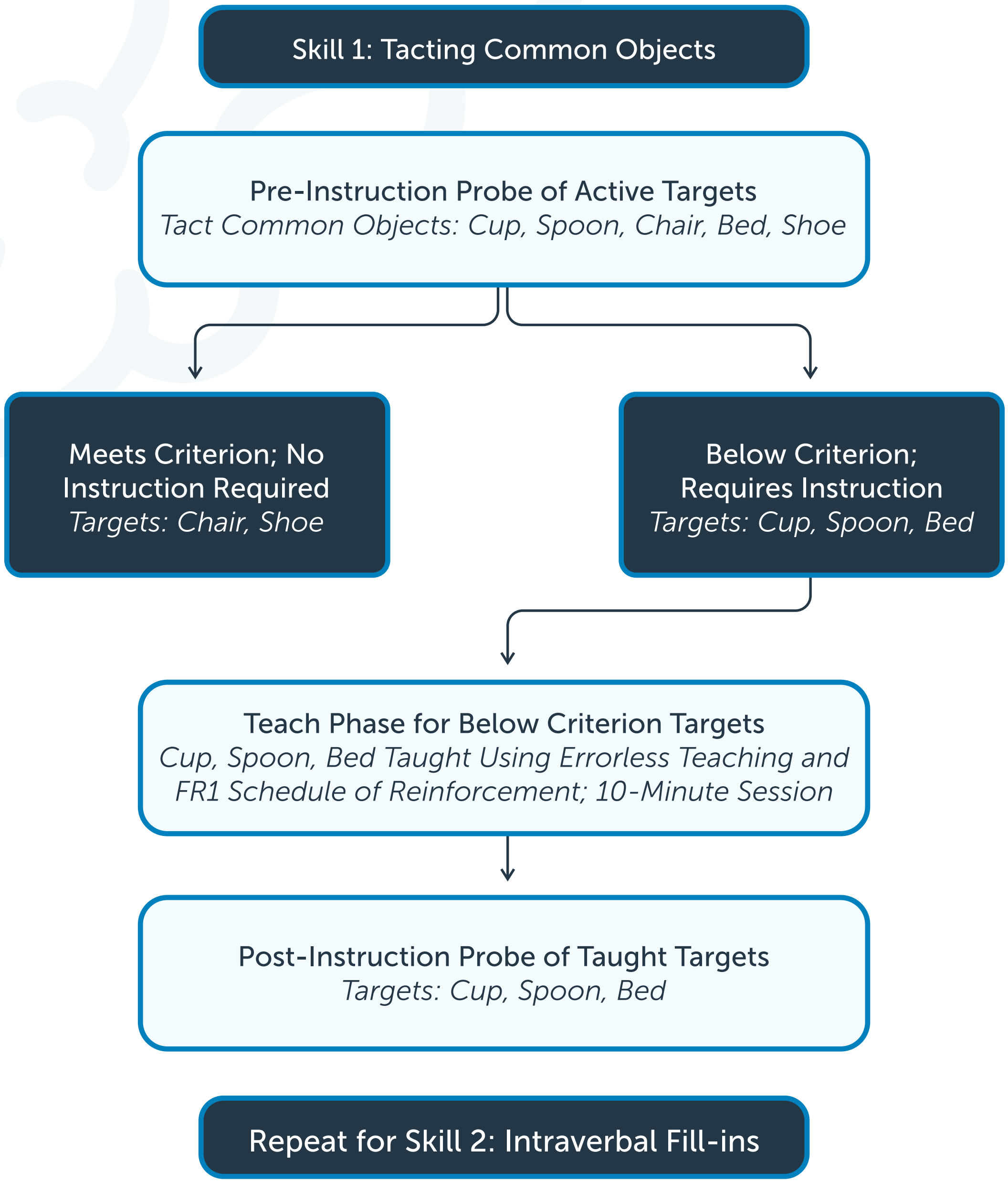

Figure 2.
Example of the Probe-Teach-Probe (PTP) instructional sequence.

PTP and Standard Programming: Instructional Procedures

Probe-Teach-Probe (PTP)

Within each PTP session, three sequential phases are completed for each active skill. During the pre-test probe phase, each active target is presented once without prompts or programmed reinforcement, providing a cold probe measure of independent responding. Targets on which the learner demonstrates independence are recorded as mastered for that session and excluded from the teach phase. During the teach phase, instruction follows a simultaneous prompting procedure: the controlling prompt is delivered immediately alongside the SD, and a dense reinforcement schedule is provided for correct responding. Skills include a minimum of three targets, and teach phase duration is determined by the supervising BCBA, averaging approximately 10 minutes per skill. A post-test probe follows instruction, presenting each target once without prompts to assess immediate changes in responding. A target is considered mastered when performance reaches 80% correct across three consecutive weeks. Hi Rasmus operationalizes this three-phase sequence within its session data collection workflow, guiding technicians through each phase and automatically recording probe and teach trial data at the session level.

Standard Programming and Key Instructional Differences from PTP

Standard programming (DTT) served as the comparison condition, following a discrete trial format in which each teaching opportunity consists of a discriminative stimulus (SD), a learner response, and a programmed consequence (reinforcement or error correction). Unlike PTP, standard programming does not include a pre-session probe phase — instruction begins without first assessing which targets the learner can already perform independently. As a result, teaching trials may be delivered to already-mastered targets, and because prompts can be embedded within or immediately following the SD, the learner may not have an unprompted opportunity to demonstrate the skill before instruction occurs. Data were collected trial-by-trial, and targets continued to receive instruction regardless of prior correct performance. Mastery criteria were consistent with PTP: 80% correct across three consecutive weeks.

It should be noted that the standard programming label reflects the primary program type recorded in Hi Rasmus and does not preclude the use of other instructional procedures during the observation period.

PTP and standard programming differ in two primary ways (see Figure 3). First, PTP begins each instructional cycle with an unprompted probe, ensuring mastery decisions are based on independent performance. Standard programming may incorporate prompting immediately following the instructional cue, which can reduce opportunities to assess — and may obscure — independent skill acquisition. Second, PTP organizes instruction by skill, probing all targets within a skill, teaching only those requiring instruction, and concluding with a post-teaching probe. Standard programming is generally organized target by target, with repeated teaching trials on individual targets before progressing.

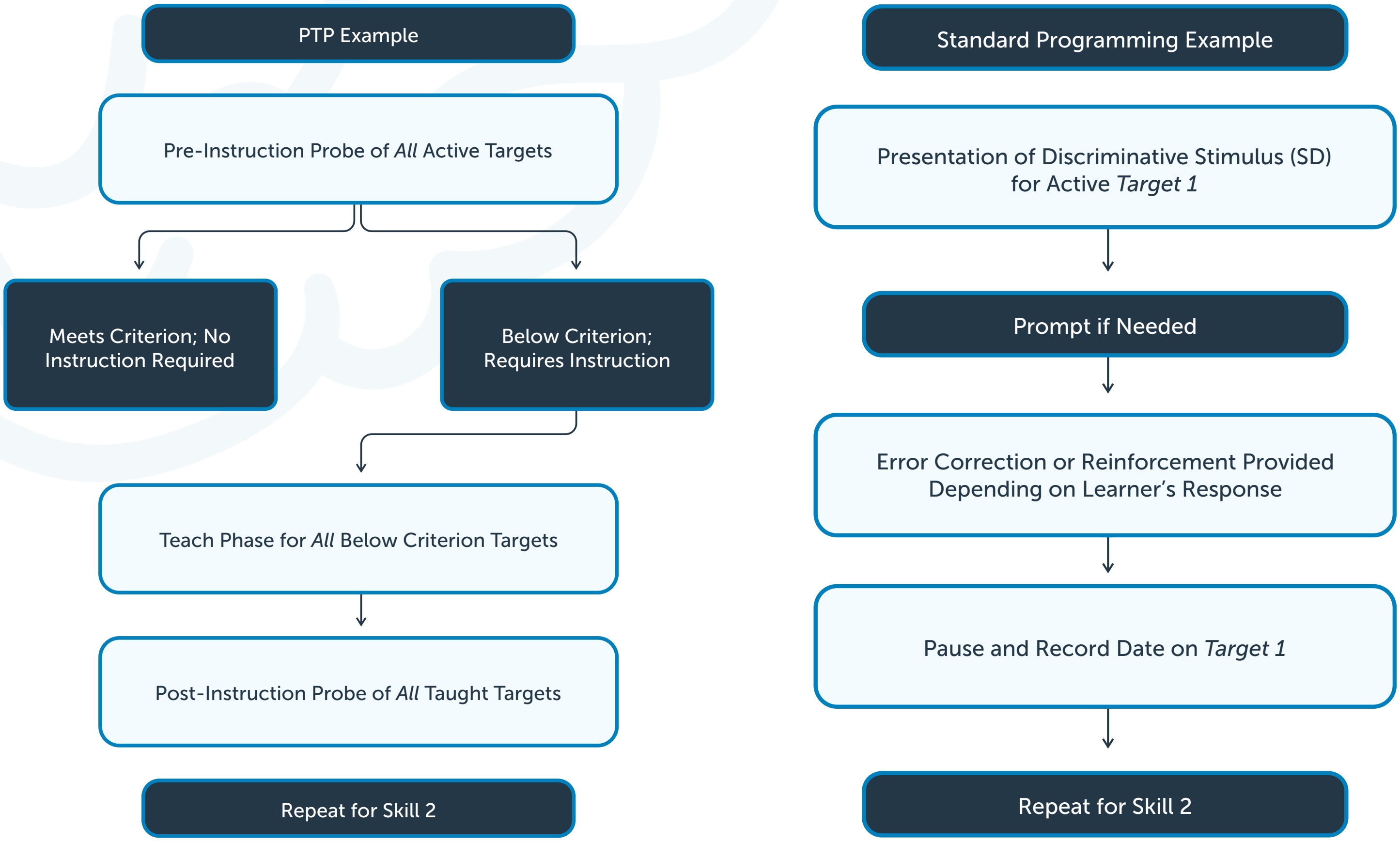

Figure 3.
PTP vs. Standard Programming instructional sequence

Organizational Adoption of PTP

Behavioral Framework adopted PTP as its primary instructional and measurement framework in response to two converging organizational realities. First, as the clinical team grew to encompass more than one hundred BCBAs operating across multiple states and environments, maintaining instructional consistency and confident outcome tracking became increasingly difficult. Second, a persistent technician workforce shortage underscored the need for systems that could more effectively support frontline behavior technicians, who deliver the majority of direct care. Technician retention was recognized not only as an operational priority, but also as a clinical one, given that technicians serve as the primary implementers of treatment and play a critical role in maintaining treatment integrity (Fiske, 2008). PTP — as implemented through the Hi Rasmus platform — addressed both realities by providing a clearly defined, repeatable structure for skill acquisition programming. Within the PTP sequence as operationalized in Hi Rasmus, technicians follow explicit on-screen guidance through each phase: assessing learner performance during the probe, recording responses, and delivering structured instruction during the teach phase. This reduces procedural ambiguity and promotes consistency across clinicians and settings (DiGennaro Reed, Hyman, & Hirst, 2011; Leaf et al., 2017). Critically, PTP functions across both discrete trial teaching and naturalistic environment teaching (NET) contexts, making it adaptable to the full range of service environments within which Behavioral Framework operates.

Although the individual components of PTP are well-supported in the existing literature, no published research has examined PTP as an integrated instructional framework implemented at scale across a large, real-world ABA provider serving a clinically diverse population. Existing comparisons of instructional methodologies have largely been conducted in controlled or small settings that do not reflect the operational complexity of applied service delivery (Lerman et al., 2011; Leaf et al., 2016). This paper addresses that gap. Specifically, we describe the theoretical and practical foundations of PTP, outline its implementation within a large multi-state ABA organization enabled by the Hi Rasmus platform, and present comparative outcomes data examining skill acquisition rates among home-based clients who received both PTP-based instruction programs and traditional DTT instruction programs. Although Behavioral Framework delivers services across a range of environments, the present analysis focuses specifically on the home-based service population, where PTP has been most extensively implemented and where the organization's largest concentration of clinical data is available.

Method

This study employed a retrospective archival design to examine differences in skill acquisition rates between PTP and standard programming instructional program types. Data were drawn from Behavioral Framework's clinical records as exported from the Hi Rasmus platform and analyzed using IBM SPSS Statistics. Given the archival nature of the data and the absence of randomization, this study is observational and descriptive-comparative in design.

Participants and Setting

The sample included 1,090 clients receiving ABA services through Behavioral Framework's Home-Based service unit during a six-month observation period (January–June 2026). Clients ranged from Pre-K through 12th grade (*n* = 945 with documented grade level), with the largest concentrations in 1st Grade (*n* = 156, 14.3%), Kindergarten (*n* = 141, 12.9%), 2nd Grade (*n* = 139, 12.8%), and Pre-K (*n* = 130, 11.9%). Grades 5 through 8 represented a combined 14.8% of the sample (*n* = 159), and grades 9 through 12 accounted for 4.6% (*n* = 50), reflecting a predominantly early school-age population.

All clients carried a primary diagnosis of Autism Spectrum Disorder (ASD; ICD-10: F84.0), with 43.4% classified as Level 2 (requiring substantial support), 29.0% as Level 3 (requiring very substantial support), 12.9% as Level 1 (requiring support), and 14.7% as level unspecified (not recorded). Co-occurring diagnoses included Global Developmental Delay (n = 55, 5.1%), Attention-Deficit/Hyperactivity Disorder (n = 32, 2.9%), Language Disorder (n = 7, 0.6%), and Intellectual Disability across severity levels (n = 9, 0.8%).

Services were delivered across five states and Washington, D.C., with Maryland representing the largest proportion (72.8% of clients with documented location; n = 549), followed by Virginia (14.3%; n = 108) and North Carolina (11.1%; n = 84). The large majority of services were delivered in the home setting (n = 991, 91.0%), with smaller proportions receiving services in home and daycare combined (2.9%), daycare only (2.5%), clinic and/or home (1.4%), home and school (1.3%), school only (0.6%), or clinic only (0.4%).

English was the most frequently reported primary home language (n = 660, 63.4%), followed by Spanish (n = 255, 24.5%), other languages (n = 106, 10.2%), French (n = 10, 1.0%), Mandarin (n = 6, 0.6%), and Arabic (n = 4, 0.4%). The linguistic diversity of the sample underscores the real-world heterogeneity of the clinical population served by Behavioral Framework.

Maryland Medicaid was the primary funding source for the majority of clients (n = 620, 57.2%), followed by CareFirst/BCBS (n = 109, 10.1%), NC Medicaid (n = 68, 6.3%), VA Medicaid (n = 66, 6.1%), Aetna (n = 52, 4.8%), Cigna (n = 48, 4.4%), UnitedHealthcare (n = 42, 3.9%), Anthem (n = 30, 2.8%), and Tricare (n = 22, 2.0%). The predominance of Medicaid funding (approximately 69.6% of clients across three state Medicaid programs) reflects the organization's commitment to serving underinsured and publicly funded populations.

Concurrent therapeutic services were common across the sample. Speech therapy was the most frequently reported additional service (n = 591, 54.2%), followed by occupational therapy (n = 325, 29.8%), physical therapy (n = 27, 2.5%), and feeding therapy (n = 16, 1.5%). The prevalence of concurrent services reflects the complex, multi-system clinical profiles characteristic of this population.

Treatment intensity data obtained from the Treatment_Intensity (Activity Hours) dataset — described in the Data Sources section below — indicated a mean of 16.29 weekly direct service activity hours (Mdn = 16.00 hours per week), consistent with home-based ABA service delivery models. The range of authorized service hours across clients (Min = 10 hours per week, Max = 30 hours per week) reflects the individualized nature of treatment planning and the clinical heterogeneity of the sample. All clients were assigned both instructional program types: PTP and Standard Programming.

Procedures

Program-type assignment was determined at the organizational level rather than by individual supervising BCBAs. However, the decision to implement PTP remained a clinical programming decision based on both the nature of the instructional target and the individual characteristics of the learner. Goals best suited to PTP were assigned that workflow, while goals more appropriately addressed through naturalistic environment teaching or task analysis were assigned corresponding workflows accordingly. Clients receiving PTP-based instruction were supported through Hi Rasmus's in-session data collection interface, which guided technicians through each phase of the PTP sequence in real time.

Regarding implementation conditions, training across the two approaches was broadly comparable during the observation period. PTP was introduced organizationally as a systemic procedural change, and its initial rollout included robust staff training and supporting resources. Since that period, BCBAs receive training on PTP during onboarding — covering both the rationale for its use and the procedural steps — which is similar in scope to the onboarding training provided on standard programming. Ongoing oversight and implementation support did not systematically differ between the two approaches during the observation period, and any observed differences in outcomes reflect practices under routine clinical conditions rather than a controlled experimental context.

Data Sources

Five archival datasets were exported from Hi Rasmus and analyzed. Table X provides an overview of each dataset, its unit of analysis, and the variables it contributed to the present study.

Dataset	Unit of Analysis	Variables/Content
Client_Level (Activity Hours)	Individual client-month observation	Mastery count, session activity hours, client-level mastery rate (*N* = 1,090 clients)
Program_Level_Mastered	Program-month observation	Monthly aggregate mastery counts by program type (Home-Based unit)
Program_Level_Master_Program	Program-month observation	Monthly aggregate program-level mastery rates by program type
Client_Information	Individual Client	Demographic, diagnostic, and service characteristics
Treatment_Intensity (Activity Hours)	Individual client	Aggregate weekly treatment intensity metrics

Hi Rasmus's reporting infrastructure enabled the extraction of all five datasets in structured formats suitable for direct import into SPSS, a capability that is central to the feasibility of organizational-scale outcomes research of this kind. The unit of analysis for client-level descriptive analyses was the individual client-month observation; the unit of analysis for inferential analyses was the program-month observation.

Measures

Mastery Count was defined as the total number of skills mastered by a client within a given calendar month, as recorded in Hi Rasmus.

Session Activity Hours (*session_act_hours*) was defined as the total number of direct service activity hours delivered to a client within a given calendar month, as captured by Hi Rasmus's session logging system. This variable reflects actual time in active therapeutic engagement, distinct from total session or billable time.

Mastery Rate was the primary outcome variable, defined as mastery count divided by session activity hours within a given calendar month. This measure was computed and examined at two levels of analysis. At the client level, mastery rate was derived from the Client_Level dataset and reflects individual client-month performance across all program types; these values are reported as descriptive statistics characterizing the full analytic sample. At the program level, mastery rate was derived from the Program_Level_Master_Program dataset and reflects monthly aggregate skill acquisition rates disaggregated by program type (PTP or standard programming); these values serve as the basis for inferential comparison between instructional approaches. Client-level and program-level mastery rates are not directly comparable, as they reflect different units of analysis and different methods of aggregation.

For clarity, these are referred to throughout as the **client-level mastery rate** (full sample descriptive) and the **program-level mastery rate** (PTP vs. standard programming inferential comparison).

Data Preparation

Client-level data were imported into IBM SPSS Statistics and restructured from wide format (one row per client, 20 variable columns) to long format (one row per client-month observation) using the SPSS Restructure Data Wizard. This yielded a maximum of six observations per client representing January through June 2026 (total possible = 6,540; 1,090 clients × 6 months). Missing data comprised 38.5% of possible client-month observations (n = 2,518 of 6,540), reflecting variable service engagement across the observation window. Missing data were treated as missing at random (MAR). No data exclusions were required; no erroneous or impossible values were identified in the Hi Rasmus export.

Statistical Analysis

Descriptive statistics were computed using the SPSS Explore procedure. Normality was assessed via Shapiro-Wilk tests, Normal Q-Q plots, detrended Q-Q plots, boxplots, stem-and-leaf plots, and skewness and kurtosis statistics. Given significant positive skew confirmed for the primary outcome variables, nonparametric procedures were selected for inferential analyses. A Wilcoxon Signed-Rank test was conducted to compare monthly skill acquisition rates between PTP and standard programming across the six-month observation period ($N = 6$ paired monthly program-level observations). Effect size was calculated as $r = Z / \sqrt{N}$. An alpha level of .05 was used for all tests.

Results

Sample and Descriptive Statistics

Across the six-month observation window, 4,022 valid client-month observations were available for analysis (listwise, 61.5% of 6,540 possible). Descriptive statistics for all three outcome variables are presented in Table 1. The sample demonstrated substantial variability in skill acquisition outcomes consistent with a clinically heterogeneous population. Monthly session activity hours were approximately normally distributed in practice (skewness = 0.26, kurtosis = –.12; $M = 64.24$, $Mdn = 63.26$, $SD = 26.47$), with the 25th and 75th percentiles at 46.10 and 81.05 hours per month respectively, indicating that the central 50% of clients received between approximately 46 and 81 monthly activity hours. The tight correspondence between mean and median for session activity hours suggests a relatively symmetric distribution of service delivery intensity across the sample, lending confidence to the treatment intensity data captured by Hi Rasmus.

Mastery count ranged from 1 to 278 across valid observations ($Mdn = 11.00$, IQR = 18.00), and client-level mastery rate ranged from 0.01 to 6.00 mastered targets per session activity hour ($Mdn = 0.18$, IQR = 0.31). Both variables demonstrated substantial positive skew (mastery count skewness = 3.00; client-level mastery rate skewness = 3.46), consistent with archival ABA service delivery data in which a subset of high-performing clients accumulates disproportionately high skill acquisition counts relative to the broader population. The pronounced mean-to-median discrepancies (mastery count: $M = 18.32$ vs. $Mdn = 11.00$; client-level mastery rate: $M = 0.30$ vs. $Mdn = 0.18$) confirm the right-skewed nature of these distributions and the appropriateness of median-based reporting.

To illustrate what this distribution looks like in practice: a client at the 75th percentile mastered approximately 0.49 targets per session activity hour, while a client at the 25th percentile mastered approximately 0.18 — a more than twofold difference that reflects the genuine heterogeneity of skill acquisition trajectories within a clinically diverse population. This variability is expected and consistent with archival ABA service delivery data; it underscores the importance of median-based reporting as the primary descriptive statistic and the appropriateness of nonparametric inferential procedures for this sample.

Table 1

Descriptive Statistics for Client-Level Outcome Variables (N = 4,022 valid client-month observations)

Variable	N	Mean	Median	SD	Min	Max	IQR	Skewness
mastery_count	4,022	18.32	11.00	21.52	1.00	278.00	18.00	3.00
session_act_ hours	4,022	64.24	63.26	26.47	2.00	169.88	34.96	0.26
mastery_rate	4,022	0.30	0.18	0.36	0.01	6.00	0.31	3.46

Note. Given confirmed non-normality of mastery_rate and mastery_count, medians and interquartile ranges are reported as primary descriptive statistics. Data were extracted from Hi Rasmus client-level reporting modules and reflect performance across all program types. Program-level mastery rates disaggregated by program type are reported separately in Table 3.

Assessment of Normality

Shapiro-Wilk tests indicated statistically significant deviation from normality for all three outcome variables: mastery_count, $W(4022) = .719$, $p < .001$; session_act_hours, $W(4022) = .994$, $p < .001$; mastery_rate, $W(4022) = .701$, $p < .001$ (Table 2). Both the Kolmogorov-Smirnov test with Lilliefors correction and the Shapiro-Wilk test reached significance for all variables ($p < .001$). Consistent with expectations for large samples in which even minor departures from normality are reliably detected, the practical significance of these findings varied substantially across variables.

Session activity hours demonstrated near-perfect normality in practice ($W = .994$, skewness = 0.26, kurtosis = −.12), with mean and median values nearly identical — a finding that validates the internal consistency of Hi Rasmus's session hour logging across a large distributed workforce. In contrast, mastery count and client-level mastery rate exhibited marked distributional skew confirmed by Q-Q plot inspection and pronounced mean-to-median discrepancies. Nonparametric procedures were used for all inferential analyses, and medians and interquartile ranges are reported as primary descriptive statistics throughout.

Table 2

Tests of Normality for Client-Level Outcome Variables (N = 4,022)

Variable	K-S Statistic	K-S df	K-S Sig	S-W Statistic	S-W df	S-W Sig.
mastery_count	.210	4,022	< .001	.719	4,022	< .001
session_act_hours	.023	4,022	< .001	.994	4,022	< .001
mastery_rate	.206	4,022	< .001	.701	4,022	< .001

Note. K-S = Kolmogorov-Smirnov with Lilliefors correction; S-W = Shapiro-Wilk. Despite statistical significance across all variables, session_act_hours demonstrated near-perfect normality in practice (S-W = .994, skewness = 0.26), validating the consistency of session hour data captured by Hi Rasmus.

Program-Type Comparison: PTP vs. Standard Programming

Monthly skill acquisition rates by program type across the six-month observation period are presented in Table 3. PTP rates were consistently higher than standard programming rates in every month, ranging from 1.171 to 1.366 mastered targets per session activity hour compared to standard programming rates of 0.200 to 0.754. The difference between program types was largest in April (PTP — Standard Programming = 1.166), coinciding with the lowest standard programming rate observed across the observation period (0.200), and smallest in June (difference = 0.417), the month in which standard programming rates were highest (0.754). PTP rates remained comparatively stable across the six months (range = 0.195 rate units), while standard programming rates showed greater variability (range = 0.554 rate units), suggesting that PTP-based instruction produced more consistent month-to-month outcomes.

To illustrate the practical magnitude of this difference: a client receiving PTP-based instruction and accumulating 64 monthly session activity hours — the sample median — would be expected to master approximately 75 to 87 targets in a given month. A client receiving standard programming for the same number of hours would be expected to master approximately 13 to 48 targets depending on the month. This translates to a real and compounding difference in learner progress over time that extends well beyond statistical significance into clinical and programmatic relevance.

A Wilcoxon Signed-Rank test confirmed that monthly skill acquisition rates were significantly higher for PTP than standard programming across the six-month observation period, $Z = -2.201$, $p = .028$, $r = .90$. Standard programming rates were lower than PTP rates in all six months without exception (negative ranks = 6, positive ranks = 0, ties = 0), indicating a consistent and unidirectional pattern of superior skill acquisition efficiency associated with PTP. The effect size ($r = .90$, calculated as $Z/\sqrt{N} = -2.201/\sqrt{6}$) is considered large by conventional standards, indicating that the magnitude of the PTP advantage is not merely statistically detectable but clinically substantial. The consistency of PTP's advantage across all six months — with zero months in which standard programming outperformed PTP — provides particularly strong evidence of a systematic performance differential that is unlikely to be attributable to random variation.

Table 3

Monthly Program-Level Mastery Rates by Program Type, January–June 2026

Month	PTP Rate	Standard Programming Rate	Difference (PTP — Standard Programming)
January	1.197	0.418	0.779
February	1.239	0.435	0.804
March	1.305	0.324	0.981
April	1.366	0.200	1.166
May	1.285	0.291	0.994
June	1.171	0.754	0.417
Range	1.171–1.366	0.200–0.754	0.417–1.166

Mastery Count by Program Type

Total aggregate mastery counts by program type across the six-month observation period are presented in Table 4. PTP demonstrated substantially higher mastery counts in every month, with monthly totals ranging from 17,221 to 26,079 compared to standard programming monthly totals of 1,165 to 3,691. PTP mastery counts increased progressively from January through April before declining in May and June. Several factors may contribute to this trajectory. As clients progress through individualized skill sequences, early-phase programs may yield higher mastery counts as foundational targets are acquired; later-phase targets may represent more complex or generalized skills with longer acquisition timelines. Seasonal variation in service engagement — including school calendars and family schedules — may also contribute to fluctuations in monthly activity hours and corresponding mastery counts. These hypotheses warrant examination in future research. Over the full six-month observation period, PTP programs produced 130,241 total skill masteries compared to 11,945 for standard programming, representing a ratio of approximately 10.9:1. This aggregate difference — more than 118,000 additional skill masteries across the observation period — represents a substantial real-world advantage in learner progress that is directly attributable to the instructional efficiency of the PTP framework as operationalized through Hi Rasmus.

Table 4

Monthly Aggregate Mastery Counts by Program Type, January–June 2026

Month	PTP Count	Standard Programming Count	Total
January	17,221	1,627	18,848
February	20,087	2,009	22,096
March	23,331	1,693	25,024
April	26,079	1,165	27,244
May	25,296	1,760	27,056
June	18,227	3,691	21,918
6-Month Total	130,241	11,945	142,186

Discussion

The findings of this study provide organizational-scale evidence that PTP, as an integrated instructional framework combining cold probe data collection, simultaneous prompting, and fluency-based instruction, produces significantly higher skill acquisition program-level mastery rates than standard programming across a large, clinically diverse home-based ABA population. PTP outperformed standard programming in every month of the six-month observation period, with a large effect size (r = .90) and a statistically significant Wilcoxon Signed-Rank result (Z = −2.201, p = .028). Over six months, PTP programs produced more than ten times the total skill masteries of standard programming — a difference of over 118,000 mastered targets that represents a clinically meaningful advantage in learner progress. Comparisons are within-child:

each client received both PTP and standard programming within the same observation window, controlling for individual service hours. PTP is not a high-volume model. Probe trials are assessment-only and do not constitute direct teaching, meaning PTP sessions may involve fewer teaching trials than a continuous DTT block. The efficiency advantage is better explained by targeting precision: clients are taught only what they have not yet independently demonstrated, concentrating instructional effort where acquisition is actually occurring. The result is more masteries per hour, not because more instruction was delivered, but because the instruction delivered was better directed.

Although medians are reported as primary descriptive statistics given the non-normal distributions observed in this sample, the client-level mean mastery rate (M = 0.30) is presented here for comparability with prior studies that reported mean-based acquisition rates. This value reflects performance across all 4,022 client-month observations regardless of program type and should be distinguished from the program-level rates reported in Table 3, which reflect monthly aggregate performance disaggregated by instructional approach. The client-level median mastery rate in the present sample (Mdn = 0.18) falls within or above these benchmarks, and the PTP program-level rates (1.17–1.37 per session activity hour) substantially exceed them.

Clinical Implications

The findings carry direct implications for how ABA providers structure instruction at the program level. The skill acquisition rates observed among clients receiving PTP-based instruction suggest that integrating assessment and teaching within a single structured workflow does not compromise — and may actively enhance — the pace at which learners acquire targeted skills relative to traditional discrete trial approaches. This is consistent with the existing literature on PTP's component procedures: cold probe measurement has been shown to yield data of comparable quality to continuous trial-by-trial recording with substantially less procedural burden (Lerman et al., 2011; Leaf et al., 2016), and simultaneous prompting has repeatedly demonstrated efficiency advantages in promoting rapid acquisition while minimizing errors (Head et al., 2011; Morse & Schuster, 2004). What the present findings add to this evidence base is a demonstration that these components, when integrated into a cohesive framework and deployed through a platform specifically designed to operationalize them — Hi Rasmus — produce outcomes that meet or exceed those associated with more traditional approaches across a large and clinically diverse caseload.

This has practical significance for clinical decision-making: providers selecting instructional methodologies need not view measurement efficiency and instructional rigor as competing priorities. PTP embeds measurement directly into the teaching sequence, which means that every session generates interpretable skill status data without requiring clinicians to sacrifice instructional time for separate assessment procedures. For supervising BCBAs, Hi Rasmus surfaces this data in real time, enabling more timely programmatic adjustments — a critical advantage in home-based settings where session frequency and duration may vary across clients and funding sources.

Beyond measurement, the structure of PTP has implications for how the field thinks about instructional design more broadly. The pre-test probe establishes a clear decision rule: teach only what the learner has not yet demonstrated independently. This concentrates instructional effort where it is most needed within each session, rather than distributing teaching time evenly across mastered and unmastered targets alike. The post-test probe then provides an immediate index of whether instruction produced a change in responding — information that Hi Rasmus aggregates automatically, offering clinicians and supervisors a granular view of both target-level and program-level trajectories over time. Taken together, these features suggest that the field should give greater consideration to instructional frameworks that treat measurement and teaching as inseparable rather than sequential activities, and that clinical data platforms should be evaluated not only for their data collection capabilities but for their ability to operationalize evidence-based instructional frameworks with fidelity.

Organizational Implications

Perhaps the most significant contribution of these findings is the evidence they provide that high-quality, evidence-based ABA can be delivered consistently — and efficiently — at organizational scale. The challenge of maintaining clinical quality as providers grow is well-recognized in the field but rarely examined with outcome data. Behavioral Framework operates across multiple states and service environments with a clinical team of more than one hundred BCBAs — a scale at which variation in clinical practice is not merely possible but expected. The adoption of PTP as a standardized instructional methodology, implemented through Hi Rasmus's purpose-built PTP program type, directly addressed this challenge by providing clinicians and technicians with a common procedural framework for skill acquisition programming.

The outcomes observed in this study suggest that standardization at the methodological level — rather than at the level of individual target selection or clinical judgment — can serve as an effective mechanism for maintaining outcome quality across a distributed workforce. Critically, this was achieved without departing from the evidence-based practices that define comprehensive ABA treatment. PTP does not replace naturalistic environment teaching, task analysis, or other established instructional approaches — it provides a structured framework for discrete skill acquisition that concentrates instructional effort on unmastered material. The result is not a reduction in treatment intensity; comprehensive, consistent hours remain essential. Rather, the findings suggest that when those hours are structured through a methodology verified to accelerate skill acquisition, children may reach their treatment goals faster — potentially reducing the overall length of time in services. For families, that means meaningful progress sooner. For the field, it represents a replicable model for delivering both quality and efficiency at scale.

This distinction — between standardizing structure and constraining judgment — warrants direct attention, because it reflects one of the more persistent cultural tensions within the field. BCBAs are trained as independent clinical decision-makers, and many enter practice with a strong preference for autonomy in selecting instructional procedures, measurement systems, and program designs. That orientation is understandable and, in many respects, a product of quality graduate training. When clinicians within a large organization implement instruction and collect data according to individually preferred methods, aggregating and comparing outcomes across caseloads becomes substantially more difficult. Supervisory feedback may be inconsistent, and technicians working with multiple BCBAs may encounter competing procedural expectations. PTP addresses this not by diminishing the BCBA's clinical role, but by defining a shared procedural infrastructure within which that role is exercised. BCBAs retain full authority over treatment planning, target selection, and programming decisions; what PTP standardizes — and what Hi Rasmus operationalizes within its data platform — is the structure through which those decisions are implemented and measured, creating a shared procedural language that supports training, supervision, and cross-site outcome comparison.

The organizational implications extend to workforce development as well. The persistent shortage of qualified behavior technicians represents one of the most significant threats to service quality in the ABA field. When instructional procedures are ambiguous or vary substantially across supervisors and settings, technicians face a steeper learning curve and greater procedural uncertainty — factors that may affect confidence and consistency in clinical implementation. PTP mitigates these risks by providing a clearly defined and repeatable workflow: probe, teach, probe. The procedural clarity of this sequence reduces the number of in-session decisions a technician must make independently, supporting more consistent implementation and more reliable data — both of which benefit the learner. Critically, this clarity is not merely conceptual: Hi Rasmus operationalizes the PTP sequence within the session data collection interface, guiding technicians through each phase in real time. This platform-level scaffolding represents a meaningful complement to clinical training, reducing procedural ambiguity and supporting fidelity in ways that verbal instruction alone cannot reliably achieve.

As expectations for outcome measurement continue to increase, ABA providers may need to shift from process-based demonstrations of quality to outcome-based demonstrations of effectiveness. Hi Rasmus's reporting infrastructure supports this transition by enabling organizations to extract, aggregate, and analyze outcomes data at a scale and granularity that would be difficult to achieve with traditional data systems. The present study offers one demonstration of what this can look like in practice: a single provider's six-month outcome data analyzed across more than 4,000 client-month observations, yielding statistically significant evidence of differential effectiveness between instructional approaches. Whether this model generalizes across providers and service contexts is an empirical question that future research should address.

Limitations

Several limitations should be considered when interpreting these findings. First, program-level inferential analyses were conducted with N = 6 paired monthly observations from a single business unit, which, while yielding a statistically significant result with a large effect size, represents a methodological constraint on inferential power and generalizability. Client-level program-type linkage data were not available at the time of analysis, precluding individual-level inferential comparisons such as Mann-Whitney U, Friedman Test, or ANCOVA with session hours as a covariate.

Second, the data were derived from routine clinical practice rather than a controlled research environment. This design choice enhances ecological validity and reflects real-world service delivery conditions that controlled studies cannot replicate — and the consistency of PTP's advantage across six consecutive months, with a large effect size and zero exceptions, provides a strong empirical foundation that supports prediction of continued differential performance under comparable conditions. At the same time, the archival design limits the ability to isolate the effects of the instructional framework from other variables, including clinician experience, client severity, and natural variation in service intensity. Future research incorporating prospective designs with matched groups or randomization would strengthen causal inference.

Third, treatment fidelity was not systematically measured as an independent variable, preventing a formal evaluation of the extent to which PTP and standard programming procedures were implemented as designed across clinicians and settings. Although Hi Rasmus's structured data collection interface supports fidelity by guiding technicians through each phase, this does not constitute a formal fidelity assessment. Fourth, the analysis focused exclusively on skill acquisition outcomes and did not evaluate maintenance, generalization, treatment acceptability, social validity, or broader quality-of-life indicators. Finally, program-type assignment was not randomized, and baseline comparability between PTP and standard programming groups could not be formally assessed, introducing the possibility of selection effects that may have contributed to observed differences.

Fifth, the present study was unable to assess the comparability of training provided to staff implementing PTP versus standard programming during the observation period. Given that PTP implementation was operationalized through Hi Rasmus's structured in-session interface and provided real-time procedural guidance to technicians, it is possible that PTP implementers received a higher degree of procedural support than standard programming implementers, independent of any formal training differences. This represents a potential confound that the archival design cannot resolve. Future research should systematically assess training exposure and implementation support across instructional conditions as independent variables. Additionally, the standard programming label reflects the primary program type classification in Hi Rasmus and does not confirm exclusive use of discrete trial procedures; other instructional approaches, including naturalistic environment teaching, may have been used concurrently.

Future Directions

A notable gap emerging from this study is the absence of empirically established benchmarks for rate of skill acquisition in ABA service delivery contexts — a limitation that carries direct implications for defining and operationalizing treatment quality at the organizational level. No universally accepted standard currently exists against which organizational performance can be meaningfully evaluated. The present data offer a starting point: the overall sample client-level median mastery rate of 0.18 targets per session activity hour, the PTP program-level rate range of 1.171–1.366, and the standard program-level rate range of 0.200–0.754 represent empirically observed benchmarks from a large, diverse, real-world population. Future research should prioritize the development of population-stratified reference ranges for rate of skill acquisition, disaggregated by program type, treatment phase, learner verbal behavior level, and service intensity. Hi Rasmus's cross-organizational reporting capabilities position it as a natural platform for this benchmark development, as the same data infrastructure that supported the present study could aggregate outcomes data across multiple provider organizations to establish normative ranges grounded in applied practice.

Future research should also examine the rate of learning opportunities and instructional density as proxies for and predictors of skill acquisition outcomes. Rate of learning opportunities — the number of discrete instructional trials or probe presentations delivered per unit of session time — captures the procedural efficiency of a given instructional approach and may mediate the relationship between program type and mastery rate. Instructional density, defined as the proportion of session time devoted to active instructional engagement relative to non-instructional activities, similarly reflects how efficiently session hours are converted into learning opportunities. PTP's structured three-phase format may produce higher rates of learning opportunities and greater instructional density than standard programming by eliminating the delivery of instructional trials for already-mastered targets — a hypothesis that could be tested with Hi Rasmus's session-level trial data. Examining these variables would help clarify the mechanisms through which PTP produces its observed outcome advantage and identify the specific procedural features most strongly associated with skill acquisition efficiency.

Future research should also replicate these findings across business units and service environments to evaluate generalizability beyond the home-based context. Examining skill acquisition rates across multiple Behavioral Framework locations and service settings would increase statistical power, enable individual-level inferential analyses once client-to-program-type linkage is established, and provide a more complete picture of PTP's effectiveness across the organization's full clinical footprint. Prospective studies conducted under IRB oversight could provide greater control over instructional variables and include measures of treatment integrity, maintenance, generalization, and social validity.

It should be acknowledged that adopting a standardized instructional methodology at this scale is not a straightforward undertaking, and not every organization will be positioned to do so. Successful implementation requires sustained investment in training infrastructure, clinical leadership alignment, a data system capable of operationalizing the methodology at the session level — as Hi Rasmus does for PTP — and an organizational culture willing to prioritize collective outcomes over individual procedural preferences. What the present findings demonstrate is that when these conditions are met, the results are measurable and compelling. PTP moves the conversation from “we do ABA” to “we have a validated instructional framework that produces measurably better outcomes” — a distinction that carries weight with payers, policymakers, and families alike, and one that organizations can now substantiate with real-world data extracted from their own clinical platform.

References

Behavior Analyst Certification Board. (2026). *US employment demand for behavior analysts: 2010–2025.* https://www.bacb.com/wp-content/us_employmentdemand_ba/

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of fluency training and accuracy training using a computer-based simulation task. *Journal of Organizational Behavior Management, 20*(2), 71–94. https://doi.org/10.1300/J075v20n02_05

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

DiGennaro Reed, F. D., Hyman, S. R., & Hirst, J. M. (2011). Applications of staff training procedures to promote treatment integrity. *Behavior Analysis in Practice*, 4(2), 4–18.

DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis, 44(3),* 611–615. https://doi.org/10.1901/jaba.2011.44-611https://doi.org/10.1901/jaba.2011.44-611

Fiske, K. E. (2008). Treatment integrity of school-based behavior analytic interventions: A review of the research. *Behavior Analysis in Practice, 1(2), 19–25.* https://doi.org/10.1007/BF03391724

Head, K. D., Collins, B. C., Schuster, J. W., & Ault, M. J. (2011). A comparison of simultaneous prompting and constant time delay procedures in teaching state capitals. *Journal of Behavioral Education*, 20(3), 182–202. https://doi.org/10.1007/s10864-011-9127-8

Kubina, R. M., & Morrison, R. S. (2000). Fluency in education. *Behavior and Social Issues, 10,* 83–99. https://doi.org/10.5210/bsi.v10i0.133

Leaf, J. B., Cihon, J. H., Ferguson, J. L., Milne, C. M., Leaf, R., & McEachin, J. (2016). A progressive approach to discrete trial teaching: Teaching children with autism. *Behavior Analysis in Practice,* 9(4), 356–375. https://doi.org/10.1007/s40617-016-0143-5

Lerman, D. C., Dittlinger, L. H., Fentress, G., & Lanagan, T. (2011). A comparison of methods of collecting data on performance during discrete trial teaching. *Behavior Analysis in Practice, 4(1),* 53–62. https://doi.org/10.1007/BF03391775

Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A., Grzybowski, A., Hall-Lande, J., Shaw, K. A. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020. *MMWR Surveillance Summaries,* 72(2), 1–14. https://doi.org/10.15585/mmwr.ss7202a1

Morse, T. E., & Schuster, J. W. (2004). Simultaneous prompting: A review of the literature. *Education and Training in Developmental Disabilities,* 39(2), 153–168.

Schuster, J. W., Griffen, A. K., & Wolery, M. (1992). Comparison of simultaneous prompting and constant time delay procedures in teaching sight words to students with moderate mental retardation. *Journal of Behavioral Education, 2(3),* 305–325. https://doi.org/10.1007/BF00947178

Shirley, M. J., & Pennypacker, H. S. (1994). The effects of performance criteria on learning and retention of spelling words. *Journal of Precision Teaching,* 12, 73–86.

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

Behavioral
Framework™