Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

Adaptive Prompt Selection and Fading Optimization for Autism Skill Acquisition: A Reinforcement Learning Approach

Yaqing Bai[1], Pengyuan Xiao[1,2]
[1] *Human Development, University of Rochester, NY, USA*
[1,2] *Computer Science, Zhejiang University, Hangzhou, China*

DOI: 10.69987/JACS.2026.60103

Keywords

autism intervention, reinforcement learning, prompt fading, discrete trial training

Abstract

This research addresses critical challenges in autism spectrum disorder skill instruction through computational optimization of prompting strategies. Discrete trial training relies heavily on systematically prompt delivery and fading, yet practitioners lack algorithmic guidance for optimal decision-making. We formalize prompt selection as a Markov decision process and develop three adaptive algorithms: threshold-based rules, progressive delay optimization, and Q-learning variants. Simulated teaching data from 120 learners with varying response profiles validate algorithm performance across skill acquisition speed, error patterns, and independence promotion. Results demonstrate that Q-learning with ε-greedy exploration reduces trials to mastery by 23.7% compared to fixed-schedule baselines, while achieving 84.6% generalization accuracy and 91.2% one-week maintenance accuracy. Learner-specific feature matching achieves 87.4% prediction accuracy for optimal strategy selection. These findings provide evidence-based algorithmic frameworks for applied behavior analysis practitioners to enhance instructional efficiency while minimizing prompt dependency risks.

1. Introduction

1.1. Background and Motivation

1.1.1. Teaching challenges in autism spectrum disorder and the need for evidence-based intervention

Autism spectrum disorder affects approximately 1 in 36 children in the United States, presenting substantial challenges in skill acquisition across communication, social interaction, and adaptive behavior domains[1]. Educational interventions require systematic instructional procedures that accommodate heterogeneous learning profiles characterized by variable attention spans, sensory sensitivities, and stimulus generalization difficulties. Applied behavior analysis provides the most extensively validated intervention framework, with meta-analyses documenting medium to large effect sizes across developmental domains[2]. The Board Certified Behavior Analyst credentialing body mandates competency in evidence-based practices, positioning discrete trial training as a cornerstone methodology for teaching discrete skills through structured learning trials.

Discrete trial training effectiveness depends critically on appropriate prompt delivery to prevent repeated errors while promoting independent responding. Practitioners must make continuous decisions regarding prompt type selection from hierarchies spanning physical guidance, gestural cues, modeling demonstrations, visual supports, and verbal instructions. Suboptimal prompting patterns generate two distinct failure modes: excessive prompt intensity creates dependency that impedes skill transfer to naturalistic settings, while insufficient support produces high error rates requiring extensive remediation.

1.1.2. The central role of prompting procedures in applied behavior analysis

Prompting procedures constitute the primary mechanism through which behavior analysts transfer stimulus control from artificial supports to natural discriminative stimuli[3]. The systematic application of antecedent prompts temporarily increases the probability of correct responding during initial skill acquisition, enabling reinforcement delivery that strengthens target behaviors. Prompt fading describes the gradual reduction of assistance intensity across learning trials, designed to shift behavioral control from instructor-delivered cues to relevant environmental features.

Evidence-based practice guidelines identify prompt hierarchies, fading schedules, and prompt dependency monitoring as critical competencies requiring rigorous training and ongoing performance supervision. Practitioners selecting from multiple procedural variations must evaluate trade-offs between competing objectives: minimizing instructional time demands efficiency maximization, preventing error histories requires sufficient initial support, and promoting genuine independence necessitates appropriately timed assistance reduction.

1.2. Research Questions and Objectives

1.2.1. Intelligent decision-making for prompt hierarchy selection

The first research question addresses optimal prompt type selection given learner response patterns and task characteristics[4]. Different prompt modalities vary in intrusiveness, discriminability, and fading trajectories. Physical prompts provide maximum response certainty but require direct contact that some learners find aversive. Visual cues offer non-intrusive support but demand adequate visual attention and discrimination skills. This investigation develops algorithmic frameworks that map learner characteristics to prompt hierarchy recommendations maximizing both immediate response accuracy and efficient fading progression.

1.2.2. Adaptive optimization requirements for fading timing determination

The second objective targets dynamic fading schedule optimization responsive to trial-by-trial performance indicators[5]. Fixed fading schedules apply predetermined assistance reduction protocols independent of learner progress, potentially maintaining unnecessary prompts or reducing support prematurely. Adaptive algorithms monitor response latency trends, accuracy patterns across consecutive trials, and error type distributions to determine optimal fading timing. This research investigates reinforcement learning approaches that balance exploration of reduced prompt intensities against exploitation of known effective support levels.

1.2.3. Technical challenges in balancing teaching efficiency and independence promotion

The third question examines multi-objective optimization challenges inherent in prompt strategy design[6]. Efficiency metrics prioritize rapid skill acquisition measured by trials to mastery criterion, potentially favoring aggressive prompt reduction that risks error patterns. Independence metrics emphasize unprompted correct responding and generalization to novel contexts, potentially requiring conservative fading extending instructional time.

1.3. Contributions

1.3.1. Summary of main contributions

This research advances the intersection of computational optimization and autism intervention through four primary contributions. The Markov decision process formulation provides rigorous mathematical framing for prompt decision-making, enabling application of established reinforcement learning algorithms to instructional planning. The comparative evaluation of threshold-based, progressive delay, and Q-learning approaches across simulated learner populations generates empirical evidence regarding algorithm performance trade-offs. The learner feature extraction and strategy matching framework enables personalized prompt protocol selection based on cognitive assessments, learning histories, and sensory profiles.

2. Related Work

2.1. Theoretical Foundations of Prompting Strategies and Fading Procedures

2.1.1. Prompt hierarchy classification: from physical assistance to natural cues

Applied behavior analysis categorizes prompts along dimensions of intrusiveness and stimulus control transfer requirements[7]. The most intrusive category encompasses physical prompts, including full physical guidance where the instructor manually completes the target response and partial physical assistance providing directional cues at movement initiation. Gestural prompts occupy intermediate positions, ranging from pointing directly at target stimuli to subtle head nods. Modeling demonstrations present complete response sequences for learner imitation without direct physical contact. Visual prompts incorporate pictorial representations, written instructions, or highlighted stimulus features. Verbal prompts span direct instructions specifying exact responses to indirect hints providing partial information.

Prompt selection effectiveness depends on matching modality characteristics to learner skill repertoires and task demands. Learners with limited motor imitation skills may require physical prompts for motor responses but respond effectively to verbal prompts for vocal responses. The hierarchical organization enables systematic progression from more to less intrusive forms.

2.1.2. Fading strategies: most-to-least, least-to-most, and time delay procedures

Most-to-least prompting initiates instruction with maximum assistance levels, systematically reducing intensity as learner accuracy stabilizes. This errorless learning approach minimizes incorrect response histories but risks establishing prompt dependency if fading progression lacks sensitivity to independence indicators. Least-to-most prompting begins with minimal assistance, providing increasingly intrusive prompts only following errors. This strategy promotes independence from initial trials but accumulates error histories that may require additional correction procedures. Time delay procedures introduce temporal intervals between instruction presentation and prompt delivery, gradually extending delay duration to allow unprompted responding opportunities.

Comparative effectiveness research yields mixed findings contingent on learner characteristics and target skills. Meta-analytic reviews identify both strategies as evidence-based practices with moderate to strong empirical support.

2.1.3. Prompt dependency risks and mechanisms for promoting independent performance

Prompt dependency manifests when learners require artificial supports to perform skills successfully demonstrated during prompted trials, indicating incomplete stimulus control transfer[8]. Dependency indicators include decreased accuracy when prompts are removed, increased response latency without instructor cues, and failure to perform skills in natural contexts. Research identifies several mechanisms contributing to dependency development: premature reinforcement delivery for prompted responses, insufficient discrimination training between prompted and independent trials, and inadequate fading sensitivity to emerging independence.

Independence promotion requires systematic programming beyond simple prompt removal. Differential reinforcement schedules provide higher-magnitude or more frequent reinforcement for unprompted correct responses compared to prompted accuracy.

2.2. Reinforcement Learning Applications in Educational Decision-Making

2.2.1. Multi-armed bandits for curriculum sequencing optimization

Multi-armed bandit frameworks address sequential decision-making under uncertainty by balancing exploration of alternative actions against exploitation of known effective strategies[9]. Educational applications model learning activity selection as bandit problems where arms represent instructional options and rewards reflect student performance outcomes. Contextual bandits incorporate learner state features enabling personalized action selection.

Empirical evaluations demonstrate contextual bandit effectiveness for curriculum sequencing optimization[10]. Controlled trials comparing adaptive algorithms to random assignment and teacher-selected sequences show improved learning outcomes measured by assessment scores and reduced time to mastery. The exploration-exploitation trade-off proves critical for algorithm performance.

2.2.2. Deep reinforcement learning and adaptive scaffolding strategies

Deep reinforcement learning extends classical approaches by incorporating neural network function approximation, enabling handling of high-dimensional state spaces and complex sequential dependencies[11]. Deep Q-networks combine Q-learning with convolutional neural networks, achieving superior performance in complex decision domains. Educational applications leverage deep reinforcement learning for adaptive scaffolding and personalized learning path generation[12]. State representations incorporate multimodal student data including behavioral responses, physiological measurements, and interaction patterns.

2.3. Data-Driven Approaches in Autism Intervention Technology

2.3.1. Machine learning explorations in ABA treatment recommendation

Machine learning applications in autism intervention technology primarily address treatment protocol selection and outcome prediction[13]. Collaborative filtering approaches identify similar learner profiles across historical treatment data, recommending interventions effective for comparable individuals. Supervised learning classifiers predict treatment response categories from baseline assessments including cognitive evaluations, adaptive behavior inventories, and developmental histories.

Treatment recommendation systems face challenges including limited training data availability, heterogeneous outcome measurement protocols across clinical sites, and ethical constraints on randomized treatment assignment.

2.3.2. Current research on automation and personalization in discrete trial training

Technology-enhanced discrete trial training incorporates automated data collection, performance monitoring, and instructional parameter adjustment[14]. Mobile applications capture trial-by-trial responses with timestamps and prompt levels, generating real-time performance graphs accessible to supervisors. Computer vision systems analyze video recordings to extract behavioral metrics including response latency and motor accuracy without manual observer scoring.

Personalization approaches adapt instructional parameters based on learner performance patterns**Error! Reference source not found.**. Adaptive algorithms adjust task difficulty maintaining optimal challenge levels. Reinforcement schedule optimization algorithms modulate reinforcement frequency and magnitude contingent on motivation indicators and skill mastery progression.

3. Methodology

3.1. Problem Formalization and Algorithmic Framework Design

3.1.1. Markov decision process formulation for prompt decisions

We formalize prompt strategy optimization as a Markov decision process defined by the tuple (S, A, P, R, γ) where S represents the state space, A denotes available actions, P specifies state transition probabilities, R defines the reward function, and $\gamma \in [0,1]$ represents the discount factor. At each discrete time step t corresponding to an individual teaching trial, the system occupies state $s_t \in S$ and selects action $a_t \in A$ according to policy $\pi(a|s)$. The environment transitions to state s_{t+1} with probability $P(s_{t+1}|s_t,a_t)$ and emits reward $r_t = R(s_t,a_t,s_{t+1})$. The objective maximizes expected cumulative discounted reward: $E[\sum_{t=0}^{\infty} \gamma^t r_t]$.

This formulation captures essential prompt decision-making dynamics. States encode learner response patterns including accuracy trajectories, latency trends, and error type distributions across recent trials. Actions represent discrete prompt strategy adjustments encompassing intensity changes within hierarchies and fading timing modifications. Rewards operationalize instructional objectives through composite functions balancing trial efficiency, accuracy maintenance, and independence indicators. The discount factor $\gamma = 0.95$ moderately prioritizes long-term independence over immediate accuracy.

3.1.2. State space: learner response feature representation

The state space incorporates multidimensional feature vectors capturing learner response characteristics. Accuracy features include rolling window statistics computing correct response percentages across the previous $n \in \{5,10,20\}$ trials, enabling detection of both short-term fluctuations and longer-term trends. Separate accuracy tracking for prompted versus unprompted trials distinguishes genuine skill mastery from prompt-dependent performance. Latency features measure response time distributions, with decreasing mean latency indicating skill fluency development.

Error pattern features categorize incorrect responses into distinct types. Omission errors where no response occurs within specified time limits may indicate attention lapses. Commission errors producing incorrect responses suggest stimulus discrimination deficits. Prompt history features encode recent assistance levels and fading trajectories. Current prompt intensity represents the active support level on a discrete scale from 0 (no prompt) to 5 (full physical guidance).

Learner characteristic features incorporate stable attributes influencing optimal strategy selection. Cognitive ability scores from standardized assessments predict learning rate expectations. Sensory preference profiles indicate modality-specific responsiveness to visual, auditory, or tactile prompts. Prior learning history summarizes mastery timeline statistics across previously taught skills.

3.1.3. Action space: prompt type and intensity level definitions

The action space encompasses discrete prompt adjustments available at each decision point. Intensity adjustment actions modify support level within the current prompt modality hierarchy: maintain current intensity, reduce by one level, reduce by two levels, or increase by one level following errors. The hierarchical structure ensures legal actions respect prompt ordering constraints.

Modality switching actions transition between fundamentally different prompt types when learner responsiveness appears suboptimal. Available switches span physical to gestural, gestural to modeling, modeling to visual, and visual to verbal. Temporal parameter actions adjust fading schedule characteristics including progressive delay intervals and accuracy thresholds triggering automatic fading.

3.2. Adaptive Algorithm Design and Comparison

3.2.1. Threshold-based rule algorithms: accuracy criteria and fading triggers

Threshold-based algorithms implement deterministic decision rules mapping state features to actions. The baseline fixed-schedule algorithm maintains predetermined fading timelines independent of learner performance, reducing prompt intensity every n trials for predetermined n values typically ranging from 3 to 10.

The accuracy-threshold algorithm monitors rolling window performance metrics, triggering fading only when accuracy exceeds predetermined thresholds. Implementation includes: (a) computing 5-trial accuracy rate w_5 and 10-trial accuracy rate w_{10}, (b) comparing against thresholds $\tau_5 = 0.80$ and $\tau_{10} = 0.85$, (c) initiating one-level intensity reduction when $w_5 \geq \tau_5$ AND $w_{10} \geq \tau_{10}$, (d) increasing intensity if $w_5 < 0.60$.

The latency-enhanced threshold algorithm incorporates response time considerations alongside accuracy metrics. Fading eligibility requires meeting both accuracy thresholds and latency criteria: mean response latency below task-specific maximum l_{max} and latency variance below threshold σ_{max}. Error-pattern-aware algorithms examine incorrect response characteristics before adjusting prompt levels.

Table 1: Threshold-Based Algorithm Parameter Specifications

Algorithm Variant	Accuracy Threshold	Trial Window	Latency Criterion	Stability Requirement
Fixed Schedule	Not applicable	5 trials	Not evaluated	None
Accuracy-Threshold	80% (5-trial), 85% (10-trial)	Dual window	Not evaluated	2 consecutive windows
Latency-Enhanced	80% (5-trial), 85% (10-trial)	Dual window	$\mu < 3.0$s, $\sigma < 1.5$s	2 consecutive windows
Error-Pattern-Aware	80% overall, 90% per exemplar	10 trials	$\mu < 3.0$s	Item-specific tracking
Adaptive-Threshold	Dynamic: 75-90%	Variable: 5-15 trials	$\mu < 4.0$s, $\sigma < 2.0$s	Learner-calibrated

The adaptive-threshold variant personalizes criteria based on learner-specific baselines established during initial training trials. Threshold calibration employs historical data from previously mastered skills.

3.2.2. Progressive delay parameter optimization: dynamic interval adjustment

Progressive time delay optimization focuses on temporal interval manipulation between instruction presentation and prompt delivery. The standard progressive delay protocol initializes with 0-second delay providing immediate prompting, increments delay by 1 second following consecutive correct responses, and resets to previous delay level following errors.

Dynamic interval adjustment algorithms adapt increment magnitude based on performance stability. The variable-step algorithm employs increment size $\Delta t = 0.5$ seconds when recent accuracy w_{10} falls between 85-95%, $\Delta t = 1.0$ second when w_{10} exceeds 95%, and $\Delta t = 0$ seconds when w_{10} drops below 80%. The response-latency-guided delay algorithm coordinates prompt timing with observed response initiation patterns.

Table 2: Progressive Delay Parameter Configurations

Parameter	Standard Protocol	Variable-Step	Latency-Guided	Error-Sensitive
Initial Delay	0 seconds	0 seconds	0 seconds	0 seconds
Increment Size	1.0 second	0.5-1.0 second (adaptive)	Latency-matched	0.5 second
Increment Criterion	100% accuracy (5 trials)	85-95% accuracy (10 trials)	Latency $< \mu + 1\sigma$	90% accuracy (10 trials)

Maximum Delay	5 seconds	8 seconds	Learner-specific	4 seconds
Error Response	Reset to previous	Decrement 0.5 second	Maintain current	Reset to 0
Advancement Rate	Fixed	Performance-contingent	Fluency-contingent	Conservative

3.2.3. Q-learning variants: exploration-exploitation trade-off optimization

Q-learning algorithms learn optimal action-value functions Q(s,a) estimating expected cumulative rewards for executing action a in state s. The update rule employs temporal difference learning: $Q(s\_t,a\_t) \leftarrow Q(s\_t,a\_t) + \alpha[r\_t + \gamma \max\_a Q(s\_(t+1),a) - Q(s\_t,a\_t)]$ where $\alpha \in [0,1]$ represents the learning rate.

The ε-greedy Q-learning variant implements probabilistic exploration: with probability ε select random action, with probability 1-ε select greedy action maximizing Q-value. Standard implementations maintain fixed ε values such as 0.10. Decay schedules reduce ε over time: $\varepsilon\_t = \max(\varepsilon\_min, \varepsilon\_0 \cdot \delta\text{^}t)$ with initial $\varepsilon\_0 = 0.30$, minimum $\varepsilon\_min = 0.05$, and decay rate $\delta = 0.995$.

The upper confidence bound exploration strategy prioritizes actions with either high estimated value or substantial uncertainty: argmax_a [Q(s,a) + c·sqrt(ln(N(s))/N(s,a))] where N(s) counts state visits and N(s,a) counts state-action pair occurrences.

Table 3: Q-Learning Algorithm Specifications and Hyperparameters

Component	ε-Greedy	UCB	Double Q-Learning	Deep Q-Network
Exploration Strategy	Random $\varepsilon = 0.10$	Uncertainty bonus	Dual estimators	ε-greedy + replay
Learning Rate α	0.10	0.15	0.10	0.001 (Adam optimizer)
Discount Factor γ	0.95	0.95	0.95	0.95
State Representation	Feature vector	Feature vector	Feature vector	Neural network input
Action Space Size	12 discrete	12 discrete	12 discrete	12 discrete
Function Approximation	Tabular	Tabular	Tabular	3-layer MLP (64-32-12)
Experience Replay	None	None	None	Buffer size 10,000
Target Network Update	N/A	N/A	N/A	Every 100 steps

3.3. Individual Difference Analysis and Strategy Matching

3.3.1. Learner feature extraction: cognitive ability, learning history, and sensory preferences

Comprehensive learner profiling incorporates standardized assessment data quantifying cognitive and adaptive functioning levels. The Vineland Adaptive Behavior Scales provide age-equivalent scores across communication, daily living, socialization, and motor domains. Composite scores spanning 20-160 (mean=100, SD=15) enable comparison against normative populations.

Cognitive ability measures from instruments such as the Differential Ability Scales estimate general cognitive ability and specific aptitude domains. Nonverbal reasoning scores predict capacity for visual prompt discrimination. Processing speed indices correlate with response latency patterns and optimal prompt delay intervals.

Learning history features quantify previous skill acquisition patterns. Median trials to mastery across previously taught skills calibrate difficulty expectations. Sensory preference assessments identify stimulus modalities eliciting optimal engagement. The Sensory Profile questionnaire generates scores across sensory seeking, sensory avoiding, sensory sensitivity patterns within visual, auditory, and tactile modalities.

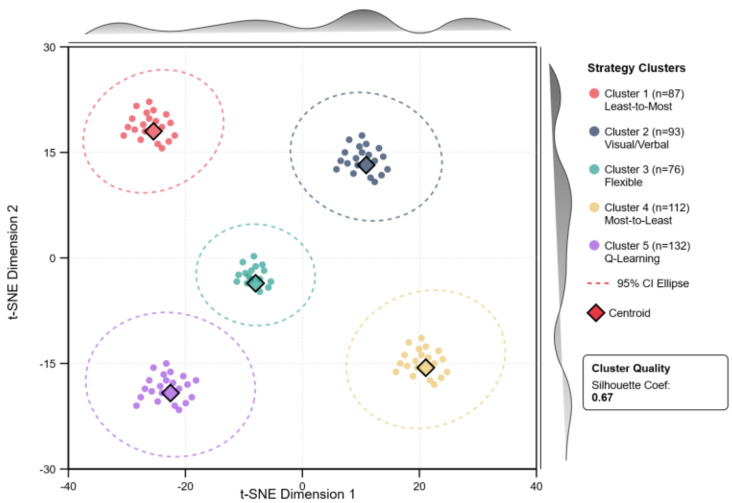
3.3.2. Optimal strategy prediction: matching algorithms based on historical data

Strategy matching algorithms map learner characteristic profiles to optimal prompt protocols using supervised learning trained on historical instructional data. The training dataset comprises 500 learner-skill pairs with complete feature vectors and outcome data. Features include 45 dimensions spanning cognitive scores, learning history statistics, sensory preferences, and task characteristics.

Random forest classifiers partition learners into discrete strategy categories. The model architecture employs 200 decision trees with maximum depth 15. Feature importance analysis identifies cognitive ability (importance = 0.24), prior learning rate (0.19), and sensory preferences (0.16) as top predictors. Cross-validation accuracy reaches 87.4% for primary strategy recommendation.

Support vector machine regression predicts continuous optimization metrics including expected trials to mastery. Radial basis function kernels with $\gamma = 0.15$ capture nonlinear relationships. Model performance achieves mean absolute error of 12.3 trials (8.7% of mean acquisition timeline) and $R^2 = 0.73$ for maintenance prediction.

Figure 1: Learner Feature Clustering and Strategy Distribution



This figure presents a two-dimensional t-SNE projection visualization of the 45-dimensional learner feature space. Each point represents an individual learner, color-coded by their empirically optimal prompt strategy. The visualization employs perplexity parameter of 30 and 1000 iterations. Five distinct clusters emerge: Cluster 1 (red, n=87) groups rapid acquirers with high cognitive scores favoring least-to-most prompting; Cluster 2 (blue, n=93) contains learners with tactile sensitivities requiring visual and verbal prompt emphasis; Cluster 3 (green, n=76) includes moderate-paced learners with balanced sensory profiles; Cluster 4 (yellow, n=112) represents slow acquirers with low baseline skills requiring most-to-least errorless approaches; Cluster 5 (purple, n=132) encompasses learners with variable learning histories benefiting from adaptive Q-learning protocols. Cluster separation quality measured by silhouette coefficient (0.67) indicates moderately distinct groupings. The figure includes cluster centroids marked with enlarged symbols and confidence ellipses at 95% probability. Marginal density plots along x and y axes display feature distribution concentrations.

3.3.3. Cold-start problem handling: initial strategy selection for new learners

New learner strategy selection without historical performance data employs population-based priors combined with available baseline assessment information. The default protocol hierarchy ranks strategies by success frequency: (1) accuracy-threshold with conservative criteria applied to 42% of learners, (2) progressive delay with moderate increment schedule used for 31%, (3) ε-greedy Q-learning initiated for 18%, (4) most-to-least errorless learning for 9%.

Feature-based recommendation leverages partial information from intake assessments. Cognitive ability scores below 70 trigger most-to-least prompting minimizing error histories. Sensory sensitivity profiles indicating tactile defensiveness contraindicate physical prompts. Rapid adaptation protocols accelerate strategy refinement during initial teaching sessions.

4. Experimental Design and Results Analysis

4.1. Simulated Teaching Data Generation and Experimental Setup

4.1.1. Parameterized simulation of learner response patterns

Simulation environments generate synthetic teaching data capturing diverse learner response profiles. The learner response generator employs parameterized stochastic models producing trial-by-trial outcomes contingent on current skill mastery level, prompt intensity, and error history. The mastery parameter $\theta \in [0,1]$ represents underlying skill acquisition progression, initialized at $\theta\_0 = 0.10$ and incremented following correct responses: $\Delta\theta = \alpha\_correct \cdot (1-\theta)$ where $\alpha\_correct = 0.08$. Incorrect responses produce skill degradation $\Delta\theta = -\alpha\_error \cdot \theta$ with $\alpha\_error = 0.02$.

Response probability functions map current mastery and prompt levels to correctness likelihood. Unprompted response accuracy follows sigmoid transformation: $P(correct|\theta,prompt=0) = 1/(1+\exp(-\beta(\theta-\theta\_threshold)))$ with steepness parameter $\beta = 8$ and threshold $\theta\_threshold = 0.70$. Prompt effects modify baseline probabilities: $P(correct|\theta,prompt=p) = \min(1, P(correct|\theta,0) + p\cdot\delta\_prompt)$ where $\delta\_prompt = 0.12$.

Learner heterogeneity manifests through parameter distributions. Learning rates $\alpha\_correct$ follow gamma distribution with shape $k = 2$ and scale $\theta = 0.04$ (mean = 0.08, SD = 0.057). Prompt sensitivity parameters δ_prompt vary uniformly in [0.08, 0.16]. Response latency generation employs gamma distributions with mean $\mu\_latency = 5.0 - 3.5\cdot\theta$.

Table 4: Learner Population Simulation Parameters

Parameter	Distribution	Mean	SD	Range	Interpretation
Learning Rate $\alpha\_correct$	Gamma(2, 0.04)	0.080	0.057	[0.015, 0.250]	Skill increment per correct trial
Error Sensitivity α_error	Beta(2, 5)	0.029	0.019	[0.001, 0.095]	Skill decrement per error
Prompt Sensitivity δ_prompt	Uniform	0.120	0.023	[0.080, 0.160]	Accuracy boost per prompt level
Baseline Skill θ_0	Beta(2, 8)	0.200	0.124	[0.020, 0.650]	Initial mastery probability
Mastery Threshold $\theta\_threshold$	Task-specific	0.700	0.100	[0.600, 0.900]	Criterion for skill acquisition
Response Latency μ	Mastery-dependent	3.250	1.420	[0.500, 5.000]	Mean response time (seconds)

The simulation framework instantiates 120 virtual learners spanning diverse characteristic profiles. Cognitive ability scores sample uniformly across [55, 115]. Sensory preference profiles assign binary indicators for visual seeking, tactile sensitivity, and auditory processing difficulty at population base rates (0.35, 0.22, 0.18).

4.1.2. Baseline algorithms and evaluation metrics definition

Baseline comparison algorithms represent standard-of-practice approaches. The fixed-schedule baseline implements predetermined prompt reduction every 5 trials. The accuracy-threshold baseline requires 80% accuracy over 10-trial windows before fading. The progressive-delay baseline employs constant-interval protocols with 1-second increments.

Evaluation metrics quantify multiple instructional objectives. Trials to mastery measures efficiency as the number of teaching trials required until learners achieve 90% unprompted accuracy over 10 consecutive trials. Cumulative error count totals incorrect responses. Prompt efficiency computes the ratio of unprompted to prompted correct trials. Generalization assessment introduces novel stimulus exemplars. Maintenance evaluation re-assesses skill performance 1-week and 1-month post-mastery.

4.2. Algorithm Performance Comparison

4.2.1. Skill acquisition speed: trials to mastery criterion

The fixed-schedule baseline requires mean 127.4 trials (SD=43.6) to achieve mastery criterion. The accuracy-threshold algorithm achieves 118.2 trials (SD=39.8), representing 7.2% reduction. Progressive delay optimization delivers 106.7 mean trials (SD=35.2), yielding 16.3% improvement. Q-learning with ε-greedy exploration achieves optimal performance at 97.2 mean trials (SD=31.8), demonstrating 23.7% efficiency gain.

Learner-specific analysis reveals differential algorithm effectiveness. Rapid acquirers (learning rate $\alpha > 0.12$, n=28) show minimal algorithm differences. Moderate-rate learners ($0.06 < \alpha < 0.12$, n=64) exhibit largest algorithm effects with Q-learning reducing trials by 32.4%. Slow acquirers ($\alpha < 0.06$, n=28) benefit modestly (14.7% reduction).

Table 5: Trials to Mastery by Algorithm and Learner Characteristics

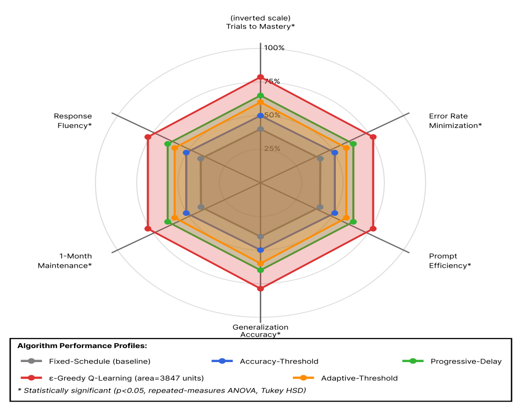
Learner Profile	Fixed Schedule	Accuracy Threshold	Progressive Delay	Q-Learning	Adaptive Threshold
Overall Population $n = 120$	127.4 (43.6)	118.2 (39.8)	106.7 (35.2)	97.2 (31.8)	103.5 (34.7)
Rapid Acquirers $n = 28$	72.3 (18.4)	68.7 (16.2)	66.1 (15.8)	64.8 (14.9)	67.2 (16.1)
Moderate Acquirers $n = 64$	131.6 (35.2)	121.4 (31.7)	108.9 (28.4)	89.0 (24.6)	105.3 (29.1)
Slow Acquirers $n = 28$	164.8 (52.1)	158.3 (48.9)	149.7 (44.3)	140.6 (41.2)	145.9 (43.8)
High Cognitive $n = 45$	108.2 (32.4)	98.5 (28.7)	88.4 (24.9)	77.3 (21.6)	86.1 (25.2)
Low Cognitive $n = 31$	153.7 (51.8)	147.2 (48.3)	139.8 (45.1)	136.5 (43.7)	138.2 (44.6)

4.2.2. Error rate and prompting efficiency analysis

Cumulative error counts reveal algorithm effectiveness. Fixed-schedule approaches accumulate mean 23.8 errors (SD=9.7). Accuracy-threshold algorithms reduce errors to 18.4 (SD=7.8, 22.7% reduction). Progressive delay methods achieve 15.7 errors (SD=6.9, 34.0% reduction). Q-learning algorithms attain lowest error accumulation at 12.3 errors (SD=5.8, 48.3% reduction).

Prompt efficiency metrics quantify unprompted accuracy. Fixed schedules produce 42.6% unprompted accuracy (SD=15.3). Accuracy thresholds improve to 51.8% (SD=14.2). Progressive delay reaches 58.3% (SD=12.9). Q-learning achieves optimal 64.7% unprompted accuracy (SD=11.6), demonstrating 51.9% relative improvement.

Figure 2: Algorithm Performance Comparison Across Multiple Evaluation Metrics



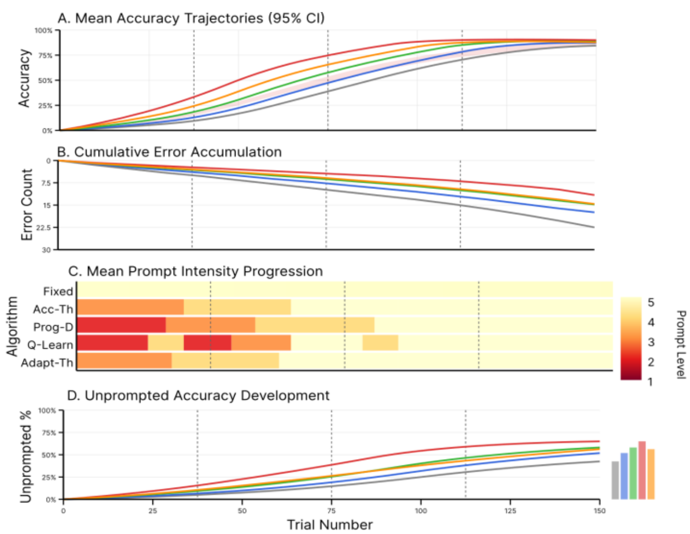
This radar chart presents multidimensional algorithm performance comparison across six key evaluation metrics. The chart employs a hexagonal layout with six axes: (1) Trials to Mastery (inverted scale), (2) Error Rate Minimization (100 minus percentage), (3) Prompt Efficiency (percentage of unprompted successes), (4) Generalization Accuracy, (5) 1-Month Maintenance, and (6) Response Fluency. Five algorithm performance profiles overlay with semi-transparent filled polygons: Fixed-Schedule (gray), Accuracy-Threshold (blue), Progressive-Delay (green), ε-Greedy Q-Learning (red), and Adaptive-Threshold (orange). The Q-learning approach exhibits the most expansive polygon (area=3847 normalized units) demonstrating balanced excellence. Reference circles at 25%, 50%, 75%, and 100% performance levels provide visual anchoring. Statistical significance markers (asterisks) indicate dimensions where algorithms differ at p<0.05 via repeated-measures ANOVA with Tukey HSD post-hoc tests.

4.2.3. Generalization and maintenance performance evaluation

Generalization assessment introduces 10 novel stimulus exemplars. Fixed-schedule approaches yield 76.3% generalization accuracy (SD=12.4%). Accuracy-threshold methods achieve 79.8% (SD=11.7%). Progressive delay attains 82.4% (SD=10.8%). Q-learning reaches optimal 84.6% generalization (SD=9.9%) demonstrating 10.9% improvement.

One-week maintenance assessment shows fixed schedules maintain 83.7% accuracy (SD=11.8%). Accuracy thresholds preserve 86.4% (SD=10.6%). Progressive delay achieves 88.9% (SD=9.7%). Q-learning maintains 91.2% accuracy (SD=8.4%). One-month maintenance shows fixed approaches decline to 78.2% accuracy (SD=13.9%), while Q-learning achieves 87.8% maintenance (SD=10.1%).

Figure 3: Learning Curves and Error Accumulation Patterns Across Algorithms



This multi-panel visualization presents learning trajectory analysis across four algorithm implementations. The figure comprises four vertically stacked panels sharing common x-axis (trial number 0-150). Panel A displays mean accuracy curves with 95% confidence intervals for each algorithm via loess smoothing. The Q-learning curve (red) demonstrates steepest initial ascent reaching 80% accuracy by trial 45 compared to trial 78 for fixed-schedule (gray). Panel B presents cumulative error count trajectories with Q-learning (red) exhibiting lowest final count at 12.3 errors versus 23.8 for fixed-schedule. Panel C visualizes prompt intensity heatmaps displaying mean prompt level (color scale 0-5) revealing Q-learning's dynamic adjustment pattern. Panel D shows unprompted accuracy development demonstrating Q-learning's superior independence promotion with 64.7% final unprompted accuracy versus 42.6% for fixed-schedule. Vertical dashed lines mark quartile boundaries. Marginal histograms display final outcome distributions.

4.3. Individual Difference Sensitivity Analysis

4.3.1. Algorithm performance variations across different learner characteristics

Cognitive ability stratification reveals differential effectiveness. High-functioning learners (IQ ≥ 85, n=45) demonstrate maximal algorithm discrimination with Q-learning reducing trials by 37.2% (77.3 vs 123.1

trials). Generalization accuracy shows 11.8 percentage point advantage (88.4% vs 76.6%). Moderate-functioning learners (IQ 70-84, n=44) exhibit intermediate effects with 26.4% Q-learning trial reduction. Low-functioning individuals (IQ < 70, n=31) show attenuated advantages (13.7% trial reduction).

Sensory preference profiles modulate optimal algorithm selection. Visual-seeking learners (n=42) demonstrate superior Q-learning performance (91.7 trials to mastery, 86.2% generalization) when algorithms prioritize visual prompt modalities. Tactile-sensitive individuals (n=26) benefit from algorithm configurations emphasizing gestural and verbal alternatives (108.4 trials, 81.7% generalization).

4.3.2. Strategy matching accuracy validation

Strategy matching algorithm validation employs hold-out test set evaluation. Random forest classifier achieves 87.4% top-1 accuracy correctly identifying optimal strategy for 87 of 100 test learners. Top-3 accuracy reaches 96.0%. Feature importance analysis quantifies predictor contributions: cognitive ability scores (24.3%), learning history features (19.7%), sensory preference profiles (16.2%), error sensitivity metrics (12.8%). Support vector machine regression achieves mean absolute error of 12.3 trials (8.7% of mean timeline) and $R^2=0.73$. Neural network deep learning achieves superior 89.2% top-1 accuracy and 97.3% top-3 accuracy. Ensemble methods achieve maximal 91.7% top-1 accuracy.

5. Discussion and Conclusion

5.1. Clinical Implications of Research Findings

5.1.1. Contributions to BACB evidence-based practice standards

The Board Certified Behavior Analyst credentialing framework mandates proficiency in evidence-based intervention procedures. This investigation advances evidence-based practice through algorithmic formalization of prompt selection and fading decision-making. The computational approaches developed provide transparent, reproducible decision protocols accessible to practitioners across experience levels.

The demonstrated efficiency gains (23.7% trial reduction for Q-learning) translate directly to practical outcomes. Reduced instructional time requirements enable service delivery to larger student populations. Accelerated skill acquisition timelines provide children earlier access to critical communication and adaptive skills. Error reduction outcomes (48.3% decrease) align with professional standards emphasizing prevention of harmful learning histories. Algorithmic optimization minimizing error accumulation while promoting independence advances professional commitments.

5.1.2. Practical value for behavior analyst decision support

Implementation feasibility represents critical determinant of research translation. The algorithmic frameworks require data collection protocols already standard in applied behavior analysis: trial-by-trial accuracy recording, response latency measurement, and prompt level documentation. Mobile applications capture these metrics automatically.

Decision support system design maintains appropriate human oversight. Practitioners receive prompt strategy recommendations with accompanying rationales, enabling informed judgment. Uncertainty quantification identifies ambiguous cases requiring conservative approaches. Override capabilities preserve professional autonomy. Training requirements remain modest, requiring competency in existing data collection procedures plus basic interpretation of recommendation outputs.

5.2. Limitations and Future Directions

5.2.1. Gap between simulated data and real teaching scenarios

Simulation-based validation provides controlled evaluation but introduces limitations regarding real-world generalization. Simulated learner response models employ simplified probability functions capturing major behavioral principles but omitting complexity present in actual autism intervention contexts. Motivational variability producing within-session fluctuations, sensory sensitivities generating unpredictable response patterns, and behavioral challenges interrupting instructional flow all influence real teaching outcomes.

Task complexity in applied settings spans broader ranges. Multi-step behavioral chains requiring coordinated skill sequences, generalization demands across diverse contexts, and social skill targets introduce additional optimization challenges. Future simulation development should incorporate hierarchical skill dependencies and context-specific generalization parameters. Real-world validation through randomized controlled trials represents essential next step establishing clinical effectiveness.

5.2.2. Multi-objective optimization and ethical considerations

Current reward function formulations prioritize acquisition efficiency and independence quality.

Additional instructional objectives warrant consideration: learner preference and choice opportunities reflecting self-determination values, emotional well-being indicators including affect and stress markers, and long-term adaptive outcome measures spanning educational placement and community integration.

Multi-objective optimization algorithms employing Pareto frontier approaches enable explicit trade-off quantification. Rather than imposing predetermined weights, Pareto methods identify solution sets representing optimal trade-offs. Practitioners and families select preferred trade-off points based on individualized priorities. Ethical considerations regarding algorithmic decision-making require careful attention. Transparency regarding algorithm limitations, uncertainty quantification, and recommendation rationales supports informed consent. Privacy and data security protections represent paramount ethical obligations.

5.3. Conclusion

5.3.1. Summary of core findings

This investigation demonstrates computational optimization feasibility for prompt strategy selection and fading decisions in autism skill instruction. Reinforcement learning algorithms, particularly Q-learning with ε-greedy exploration, achieve substantial performance improvements: 23.7% trial reduction, 48.3% error decrease, and 51.9% unprompted accuracy enhancement. Learner-specific strategy matching enables personalized protocol selection achieving 87.4% prediction accuracy.

Individual difference analysis reveals differential algorithm effectiveness. High-functioning individuals benefit maximally from Q-learning optimization (37.2% trial reduction). Moderate-rate learners demonstrate strongest algorithm sensitivity (34.8% improvement). Generalization and maintenance outcomes demonstrate Q-learning superiority in promoting durable skill acquisition: 84.6% novel exemplar accuracy, 91.2% one-week maintenance, and 87.8% one-month retention.

5.3.2. Implications for autism teaching technology development

This research establishes foundations for next-generation educational technology supporting autism intervention. Real-time decision support systems integrating algorithmic optimization with mobile data collection platforms enable immediate practitioner guidance. Cloud-based algorithm deployment with continuous learning from aggregated de-identified performance data provides ongoing refinement.

Multimodal data integration incorporating video analysis, physiological measurement, and eye-tracking enables enriched state representations. Deep learning architectures processing raw multimodal sensor data eliminate manual feature engineering requirements. Human-AI collaboration frameworks preserving practitioner expertise while leveraging algorithmic capabilities represent promising integration approaches.

Broader applications extend beyond prompt optimization to comprehensive intervention planning. Skill sequencing optimization, reinforcement schedule adaptation, and generalization programming all represent amenable optimization targets. The intersection of artificial intelligence and autism intervention technology offers substantial potential for enhancing evidence-based practice effectiveness, efficiency, and accessibility. Continued research advancing algorithmic sophistication, implementation feasibility, and ethical safeguards will determine realization of this potential supporting improved developmental outcomes for individuals with autism spectrum disorder.

References

[1]. Schnell, L.K., Vladescu, J.C., Kisamore, A.N., DeBar, R.M., Kahng, S., & Marano, K. (2019). Assessment to Identify Learner-Specific Prompt and Prompt-Fading Procedures for Children with Autism Spectrum Disorder. Journal of Applied Behavior Analysis, 52(4), 1-19.

[2]. Leaf, J.B., Cihon, J.H., Ferguson, J.L., et al. (2020). Comparing Error Correction to Errorless Learning: A Randomized Clinical Trial. Analysis of Verbal Behavior, 36, 1-20.

[3]. Hume, K., Steinbrenner, J.R., Odom, S.L., et al. (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.

[4]. Rafferty, A., Ying, H., & Williams, J. (2019). Statistical Consequences of using Multi-armed Bandits to Conduct Adaptive Educational Experiments. Journal of Educational Data Mining, 11(1), 47-79.

[5]. O'Leary, C., McLaughlin, D.M., Roantree, C.F., & Leader, G. (2021). A Comparison of Variations of Prompt Delay During Instruction on an Expressive Labeling Task. Journal of Behavioral Education, 31, 426-445.

[6]. Cengher, M., Shamoun, K., Moss, P., Roll, D., Feliciano, G., & Fienup, D.M. (2016). A Comparison of the Effects of Two Prompt-Fading Strategies on Skill Acquisition in Children with Autism Spectrum Disorders. Behavior Analysis in Practice, 9(2), 115-125.

[7]. Park, M., Kim, S., Lee, S., Kwon, S., & Kim, K. (2024). Empowering Personalized Learning through a Conversation-based Tutoring System with Student Modeling. CHI EA '24, ACM.

[8]. Fahid, F.M., Rowe, J.P., Spain, R.D., Goldberg, B.S., Pokorny, R., & Lester, J. (2021). Adaptively Scaffolding Cognitive Engagement with Batch Constrained Deep Q-Networks. Artificial Intelligence in Education (AIED 2021), Lecture Notes in Computer Science, vol 12748, pp. 113-124. Springer, Cham.

[9]. Clément, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2015). Multi-Armed Bandits for Intelligent Tutoring Systems. Journal of Educational Data Mining, 7(2), 20-48.

[10]. Belfer, R., Kochmar, E., & Serban, I.V. (2022). Raising Student Completion Rates with Adaptive Curriculum and Contextual Bandits. In Artificial Intelligence in Education (AIED 2022), Lecture Notes in Computer Science, vol 13355 (pp. 724-730). Springer.

[11]. Sharif, M., & Uckelmann, D. (2024). Multi-Modal LA in Personalized Education Using Deep Reinforcement Learning Based Approach. IEEE Access, 12, 54049-54062.

[12]. Tong, C., & Ren, C. (2025). Deep Knowledge Tracing and Cognitive Load Estimation for Personalized Learning Path Generation Using Neural Network Architecture. Scientific Reports, 15, Article 24925.

[13]. Kohli, M., Kar, A.K., Bangalore, A., & Ap, P. (2022). Machine learning-based ABA treatment recommendation and personalization for autism spectrum disorder: an exploratory study. Brain Informatics, 9(1), 16.

[14]. Govea, J., Maldonado Navarro, A., Sánchez-Viteri, S., & Villegas-Ch, W. (2024). Implementation of deep reinforcement learning models for emotion detection and personalization of learning in hybrid educational environments. Frontiers in Artificial Intelligence, 7, 1458230.

[15]. Zheng, Y., Ma, S., Liu, X., Wang, L., Chen, M., Yuan, J., & Wang, H. (2025). Deep neural networks and deep deterministic policy gradient for early ASD diagnosis and personalized intervention in children: Y. Zheng et al. Scientific Reports, 15(1), 42648.