Minimax Theory and its Applications
Volume 11 (2026), No. 1s, 427-445

A Cognitive Trust-Aware Framework for Minimax Optimization of Personalized Learning Paths in Higher Education

Manoj Podiyan[1], Dr Prashant Sharma[2]

[1]PhD Scholar, Vivekananda Global University, Jaipur, Rajasthan. Email: manojp14@gmail.com
[2]Associate Professor, Department of Computer Science & Engineering, Vivekananda Global University Jaipur. Email: prashant.sharma@vgu.ac.in

Abstract

Adaptive learning in higher education depends on two capabilities that the literature has developed in isolation. One is an accurate estimate of what a learner knows. The other is a policy that decides what the learner should do next. Knowledge tracing has advanced the first, most recently by suppressing non-cognitive noise such as slips and guesses through adaptive dynamic programming. Learning path optimization has advanced the second, formulating path planning as a sequential decision process over a knowledge graph and solving it with value-based deep reinforcement learning. Stacked naively, the path optimizer inherits a contaminated knowledge state. A tracing model pays a bounded one-step penalty for that contamination, whereas a reinforcement learning agent compounds the error along the entire trajectory. This article proposes the Cognitive Trust-Aware Personalized Learning Framework (CTPLF) to close the interface. An Adaptive Cognitive State Refinement (ACSR) module recovers the coherent cognitive representation using a p-mechanism driven by an adaptive dynamic programming solver, and it also emits a per-concept Cognitive Trust Score derived from the density of optimal control corrections the solver applied. A Cognitive Trust-Guided Learning Path Network (CTGLPN) consumes that signal at three gates covering replay sampling, exploration scale, and reward shaping, so the agent neither learns from nor rewards gains the cognitive layer cannot vouch for. An Educational Decision Intelligence Layer (EDIL) aggregates trajectories into segment-level indicators. We evaluate the framework end to end on the Open University Learning Analytics Dataset against the existing causal reinforcement learning system validated on the same corpus.

Keywords: Adaptive Learning Framework, Knowledge Tracing, Learning Path Optimization, Adaptive Dynamic Programming, Deep Reinforcement Learning, Cognitive Trust Score, Knowledge Graph, Dueling Double Deep Q-Network, Higher Education, Learning Analytics, OULAD

1 Introduction

Virtual learning environments now log every resource access, forum contribution, and assessment submission at institutional scale [1], and the promise attached to this data is personalization. This article addresses the second objective of the present research programme, which is to develop a comprehensive machine-learning-based adaptive learning framework that delivers personalized instruction, optimizes individual learning pathways, and supports data-driven decision-making in higher education. Three deliverables and one qualifier are named there, and each carries weight. A model of what the individual learner knows is what personalized instruction requires. Optimized pathways require a policy that sequences activities over a horizon. Institutional decision support requires that the same model serve programme directors and support services. The qualifier comprehensive then demands one framework in place of three, and it is on that demand that the literature is thinnest.

A recent survey [2] organizes the field along two axes. Student modeling asks what the learner knows; personalized recommendation asks what the learner should encounter next. Knowledge tracing dominates the first axis, progressing from Bayesian formulations [3] through deep [4], memory-augmented [5], attention-based [6], [7], graph-structured [8], and learning-process-aware [9] models, with prediction-consistency regularization [10] introduced to stabilize the trajectories such models produce. One difficulty has proved persistent. The response sequence is not a clean measurement of cognition, since a learner may pass a difficult item by guessing or fail an easy one through carelessness, so the recorded outcome and the

underlying competence diverge systematically and not at random. On the second axis, path recommendation builds a knowledge graph under Knowledge Space Theory [12] and augments learner states with resource distribution so that recommended paths are executable, and graph-enhanced reinforcement formulations [11] are the most developed treatments available.

Combining the two looks straightforward, and that appearance is the problem this article addresses. The path optimizer consumes a state produced by a tracing model, and every tracing model in current use produces a state contaminated by non-cognitive noise. For one-step prediction the consequence is bounded. For a sequential policy it is not, because a guess misread as mastery advances the agent past an unacquired concept, after which every subsequent item is selected under a false prerequisite assumption. Two further gaps compound the problem. A review of forty-six intelligent tutoring systems [13] found that almost none support both student-facing personalization and institution-facing analytics, and the closest existing system [14], which serves as our primary baseline, grounds its policy in aggregated clickstream features with no concept-level cognitive model.

CTPLF treats the interface between noise-robust cognitive representation and reinforcement path navigation as the object of design. Its starting point is a by-product of the adaptive dynamic programming solver used for denoising, namely the sequence of corrections the solver applied while computing its optimal control solution, which downstream models have never exploited. Where the solver intervened frequently, the evidence was internally contradictory and the estimate deserves less trust. That gives a per-learner, per-concept trust signal at no cost in additional models or supervision, and we route it into the agent at the three points where mistrust ought to change behaviour: the transitions it learns from, the width of its exploration, and what it rewards.

1.1 Motivation and contribution

Noise-robust cognitive modeling and sequential path optimization have matured independently, and the interface between them has gone unexamined. An adaptive framework for higher education must also serve institutional decision-making alongside individual learners. Both observations motivate the unified framework developed here, in which cognitive trust is a first-class signal propagated from the state estimator into the decision policy. The principal contributions are as follows.

Unified adaptive learning framework (CTPLF). We integrate noise-robust cognitive representation, prerequisite-aware resource graph encoding, trust-gated path navigation, and institutional decision support in one end-to-end architecture. Where prior work connects state estimation to path recommendation through a scalar mastery vector alone, CTPLF propagates the estimate and its trustworthiness together across the interface, and so addresses all three deliverables of the research objective within a single model.

ACSR with an explicit trust read-out. Beyond recovering the denoised state by adaptive dynamic programming under a p-mechanism, we define a Cognitive Trust Score (CTS) over the density of the solver's optimal control corrections, which yields a per-learner, per-concept confidence measure at no additional modeling cost. Reviewed approaches discard the solver's intermediate state once the record is corrected. CTPLF instead reads the density of those corrections as an explicit concept-level trust signal for downstream policy optimization.

Trust-gated reinforcement navigation (CTGLPN). We extend the prioritized noisy dueling double deep Q-network with three trust-conditioned mechanisms. Reliability-Aware Experience Replay discounts transitions the cognitive layer cannot vouch for. Uncertainty-Guided Exploration widens search where the estimate is uncertain. Reliability-Conditioned Reward withholds credit for gains arising from unreliable evidence, which closes a reward-hacking channel specific to educational reinforcement learning that the reviewed reward formulations leave open.

Horizon-amplification analysis. We show that state-estimation error propagates into policy value amplified by the effective planning horizon, which explains why denoising improvements that appear marginal under one-step predictive metrics are substantial under path-quality metrics.

EDIL and end-to-end validation on OULAD. EDIL aggregates trajectories into segment-level Learning Bottleneck Index, Personalized Intervention Gain, and Learning Sequence Divergence indicators, which addresses the dual-user gap identified in recent systematic reviews. It reports at segment level because instance-level explanation becomes unstable on sparse behavioural data [15].

The remainder of the paper is organized as follows. Section 2 reviews related work in knowledge tracing, learning path optimization, and higher-education decision support. Section 3 presents the proposed CTPLF methodology together with its theoretical analysis and its alignment with the research objective. Section 4 describes the dataset, the existing system compared against, the metrics, and the experimental results. Section 5 concludes.

2 Related work

Knowledge tracing forecasts future performance from historical responses while maintaining an estimate of latent mastery, and its development traces a steady enrichment of that estimate. Bayesian knowledge tracing [3] established the formalism with a hidden Markov model over a binary mastery variable. DKT [4] replaced the discrete latent state with a recurrent hidden state, DKVMN [5] separated static concept representations from dynamic mastery through a key-value memory, and SAKT [6] and AKT [7] replaced recurrence with self-attention. A parallel direction incorporated structural relations. GKT [8] propagates mastery along learned edges between knowledge components, PEBG [16] pre-trains item embeddings on a bipartite item-concept graph, and LPKT [9] models learning gain and forgetting from inter-interaction time intervals.

These enrichments share an assumption. They treat the recorded response as a faithful if noisy observation of cognition, and they answer noise with a more expressive approximator. Where the assumption has been questioned, the remedy has been regularization of the predicted trajectory [10] and not correction of the record itself, so non-cognitive interference continues to be averaged away as zero-mean noise when it is better understood as a systematic corruption that leaves internally contradictory records behind. We take the opposite position here. Contradictions between recorded outcomes and the difficulty ordering are evidence to be resolved, and smoothing them away discards information. We therefore define incoherence conditions over pairs of interactions on the same concept whose difficulties differ by at least a threshold while their outcomes contradict the difficulty ordering, add continuity conditions for extreme polarization of interval performance, and use both to drive an adaptive dynamic programming solver that minimizes a coherence cost. One further limitation applies to every such treatment. Their scope ends at response prediction, and the recovered representation is fed to a classifier and never to a decision.

That boundary matters, because the complementary line of work begins where the first one stops. Learning path recommendation seeks a sequence of activities carrying a learner from a current to a target knowledge state. Early rule-based and evolutionary approaches struggle with hierarchical dependencies and often produce discontinuous paths that induce cognitive overload [2], while knowledge graphs constructed under Knowledge Space Theory [12] supply the structure needed to enforce coherence. Reinforcement learning reframes the problem as sequential decision-making. The value-based methods on which most educational formulations rest derive from DQN [17], with double Q-learning [18] correcting overestimation of action values, the dueling architecture [19] separating state value from advantage, prioritized replay [20] sampling transitions by temporal-difference error magnitude, and noisy networks [21] replacing epsilon-greedy exploration with learned parametric noise. Graph-structured reinforcement formulations [11] are the most complete recent treatments. They construct a subject knowledge graph with prerequisite edges, estimate knowledge states with a graph-based tracing model, and form a joint state by concatenating the cognitive state with a resource reachability indicator. Their cognitive component has no mechanism for separating cognitive signal from non-cognitive noise, however, and their reward is computed on raw estimated mastery differences, so an improvement produced by a lucky guess earns exactly what a genuine one earns.

A third body of work approaches the same setting from the institutional side instead of the cognitive one. A recent survey [22] reviews a decade of academic outcome prediction and benchmarks classical learners on OULAD, finding that no algorithm dominates, and ensembles have consequently become standard [23], [24]. Deep architectures have been benchmarked for adaptive path recommendation on multimodal data [25], and a systematic review catalogues the feature families such models depend on [26]. Explainability recurs as a requirement here, since institutional decisions must be defensible, but a recent benchmark [15] documents a stability collapse in which local explanation rank stability falls from Spearman correlation near 0.87 on dense socio-economic data to near 0.35 on sparse behavioural logs, which is precisely the regime adaptive learning frameworks occupy. Prediction alone prescribes nothing, and the causal artificial intelligence framework [14] makes that argument explicitly; it is also the primary existing system compared against here. It processes OULAD into a weekly learner-state panel of approximately 1.7 million learner-week records over 25,516 learners, estimates heterogeneous treatment effects for four intervention actions by uplift modeling, and embeds those signals in an actor-critic policy through a causal regularization term, reaching AUC 0.795 and a self-normalized policy score of approximately 0.643 against a behaviour baseline of 0.589. Its limitation is the absence of any concept-level cognitive model. Because its state is behavioural aggregation, it prescribes engagement nudges but cannot reason about prerequisite structure, mastery, or curricular sequencing.

These three lines converge on the same corpus. OULAD [1] is the reference dataset for higher education learning analytics, and a systematic review of benchmark studies using it [27] documents both its breadth of use and the methodological inconsistency that complicates comparison, while a broader dataset survey confirms a pronounced skew toward prediction over knowledge mapping [28]. Read together, the literature leaves three gaps. First, noise-robust cognitive representation and sequential path optimization have advanced independently, and no existing work examines what happens to path quality when the state feeding the policy is corrupted by non-cognitive noise, even though reinforcement learning compounds state error along the trajectory in a way one-step prediction does not. Second, systematic review evidence [13] establishes that

intelligent tutoring systems serve either learners or instructors and seldom both, which leaves institutional decision-making unsupported by the models that drive personalization. Third, frameworks that prescribe instead of predicting [14] operate on behavioural aggregates without concept-level structure, so their prescriptions cannot address curricular sequencing. CTPLF is designed to close all three.

3 Proposed Methodology

Information flows through four modules, and each discharges one clause of the research objective. Adaptive Cognitive State Refinement (ACSR), together with Hierarchical Cognitive State Optimization (HCSO), recovers a coherent per-concept cognitive representation and emits a trust signal, which is what personalized instruction requires. The Prerequisite-Aware Learning Resource Graph (PLRG) and the Cognitive Trust-Guided Learning Path Network (CTGLPN) turn that representation into a prerequisite-constrained sequencing policy, discharging the pathway clause. The Educational Decision Intelligence Layer (EDIL) aggregates trajectories into segment-level indicators for institutional decision support. What discharges the qualifier comprehensive is that the four modules share one learner representation instead of merely sharing a pipeline, so an improvement in the cognitive layer propagates to pathway quality and to the institutional indicators without retraining, a propagation that Proposition 1 quantifies. Personalized instruction is modelled at the level of the individual concept and not the learner in aggregate, which is what allows the pathway layer to reason about prerequisites at all. Because higher-education decisions are made at programme level, EDIL reports at segment granularity by design. Figure 1 shows the architecture.

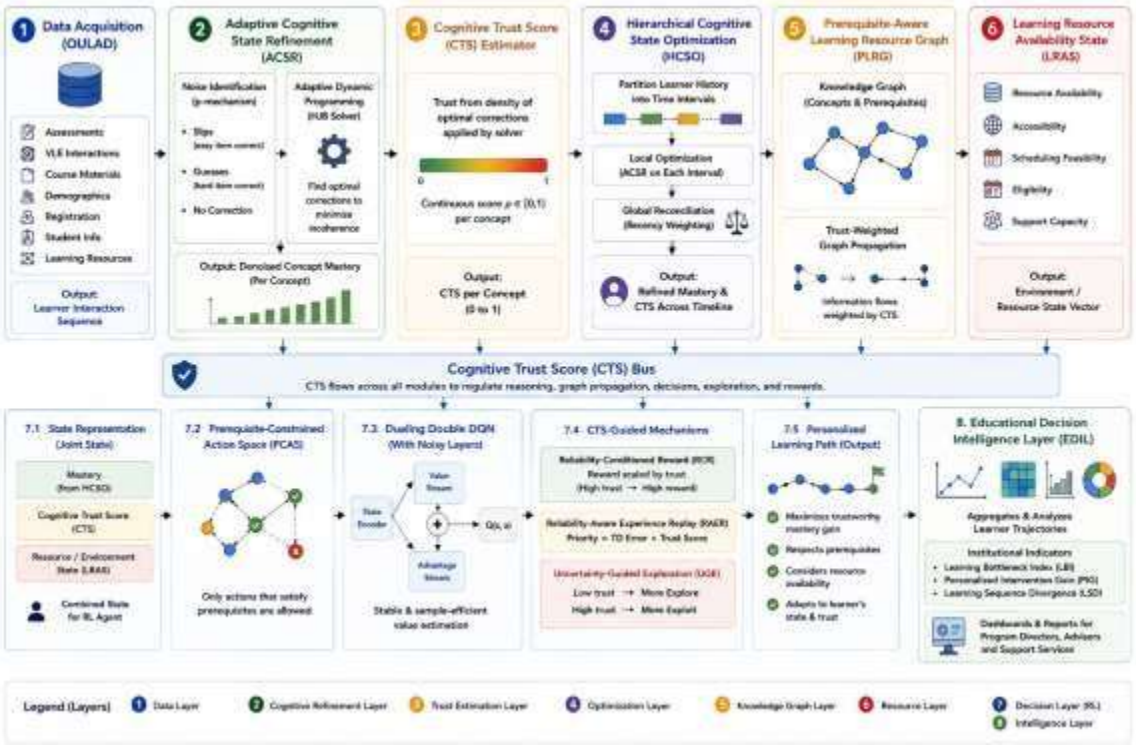


Figure 1 Proposed CTPLF architecture. The eight numbered stages group into the four modules described in Section 3: stages 2 to 4 form ACSR with HCSO, stage 5 the PLRG, stage 7 the CTGLPN, and stage 8 the EDIL.

3.1 Preliminaries and problem definition

Let X be the set of learners, I the assessable items, and K the N knowledge concepts. For learner ℓ the observed sequence is $\Xi_\ell = \{(h^w, u^w, k^w, g^w)\}_w$, with h^w the item attempted at step w, $u^w \in \{0,1\}$ the recorded outcome, k^w the concept, and g^w the difficulty from the empirical pass rate; instantiation on OULAD is given in Section 4.1. Our central premise concerns u^w, which we do not treat as a faithful observation of cognition. Recorded and true states diverge whenever a slip converts a mastered item into a failure or a guess converts an unmastered item into a success. We seek

$$\Pi^* = \underset{\Pi}{\operatorname{argmax}} \mathbb{E}\left[\sum_w \xi^w R(z_w, d_w)\right], \quad \text{s.t. } \operatorname{pre}(d_w) \subseteq K_w \text{ and } \operatorname{reach}(d_w) = 1 \tag{1}$$

where pre(d) is the prerequisite set of the concept targeted by action d, K_w ⊆ K the mastered set, and reach(·) whether the resource is deliverable. K_w is estimated and never observed, and this is where the difficulty lies, since the quality of that estimate governs the quality of every subsequent decision.

3.2 Adaptive Cognitive State Refinement (ACSR)

ACSR produces two outputs. The first is a denoised per-concept cognitive representation, recovered from the contaminated record. The second is a score expressing how far that representation can be trusted, derived from the work the recovery required. Its denoising machinery comprises the coherence and continuity conditions, their unification as the p-mechanism, and the adaptive dynamic programming solver that minimizes the resulting cost, and ACSR couples that machinery to the trust read-out of Section 3.2.1, which converts the solver's own effort into a usable confidence signal. Extracting from Ξ_ℓ all subsequences sharing a concept and retaining only outcome and difficulty gives the state-difficulty sequence, whose state evolves as $\chi_i^{k+1} = A(\chi_i^k, x_i^k)$ with x_i^k a control determining whether the recorded state is revised. Judging a contradictory pair from difficulty alone is subjective, so we introduce the interval performance value

$$F = \frac{U_1}{q_k + \mu}\left(1 + \frac{q_k}{S}\right) + \frac{U_2}{S} \tag{2}$$

where U_1 is the number of correct responses on the relevant concept within the interval, U_2 the number across all concepts, q_k the number of items on that concept, S the spacing, and μ a buffer parameter. Two incoherence conditions detect contradictions between outcomes and the difficulty ordering, and two continuity conditions cover extreme scenarios in which the difficulty difference is small but interval performance is strongly polarized over a sufficient distance S_0:

$$\begin{aligned} \Phi_1 &= \mathbb{1}[\chi_1^k < \chi_{k+2}^0 \wedge g_1 < g_{k+2} \wedge |g_1 - g_{k+2}| \geq \tau \wedge F > c_1] \\ \Phi_2 &= \mathbb{1}[\chi_1^k > \chi_{k+2}^0 \wedge g_1 > g_{k+2} \wedge |g_1 - g_{k+2}| \geq \tau \wedge F < c_2] \\ \Psi_1 &= \mathbb{1}[\chi_1^k < \chi_{k+2}^0 \wedge F > c_3 \wedge S > S_0], \quad \Psi_2 = \mathbb{1}[\chi_1^k > \chi_{k+2}^0 \wedge F < c_4 \wedge S > S_0] \end{aligned} \tag{3}$$

Φ_1 indicates a probable slip and Φ_2 the converse pattern indicating a guess; we call their unification with the continuity conditions Ψ_1 and Ψ_2 the p-mechanism. The control is $x_1^k = +1$ under $(\Phi_1 \vee \Psi_1)$, -1 under $(\Phi_2 \vee \Psi_2)$, and 0 otherwise, so the system admits the affine form $\chi_1^{k+1} = \chi_1^k + x_1^k$ in which $+1$ corrects a slip and -1 a guess. Assigning unit cost to every correction and discounting by ξ_0, the accumulated incoherence is minimized when the optimal cost satisfies the discrete Hamilton-Jacobi-Bellman equation

$$M^*(\chi_1^k) = \min_{x_1^k}\{|x_1^k| + \xi_0 M^*(\chi_1^k + x_1^k)\}, \quad x^* = \underset{x_1^k}{\operatorname{argmin}}\{\cdot\} \tag{4}$$

Minimizing M drives the representation toward the most internally consistent account of the learner's cognition. The solver operates bottom-up. It screens target pairs, evaluates the p-mechanism conditions, and converges to χ^*, which yields the denoised mastery estimate $\hat{p}_w \in [0,1]^N$.

3.2.1 The Cognitive Trust Score (CTS)

ACSR departs from prior denoising work in one respect. The solver's output is not discarded once the state has been corrected, because the control sequence $\{x^*\}$ is itself informative: where many corrections were applied, the evidence was internally contradictory and the estimate deserves less trust. We call the resulting quantity the Cognitive Trust Score (CTS), written ρ. Let Z_k be the window of interactions of learner ℓ on concept k in the current partition. Then

$$\rho_k^w = \exp\left(-\kappa \cdot \frac{1}{|Z_k|}\sum_{i \in Z_k} |x_i^*|\right)\left(1 - \exp\left(-\frac{|Z_k|}{\omega}\right)\right) \tag{5}$$

where κ controls sensitivity to correction density and the second factor suppresses trust when a concept has been attempted too few times to judge. Two properties matter downstream. Under monotonicity, ρ_k^w is strictly decreasing in the mean absolute correction and strictly increasing in $|Z_k|$. Under boundedness, ρ_k^w ∈ [0,1), and the open upper limit ensures the gates of Section 3.4 never assign a transition unconditional weight. CTS also requires no additional model, supervision, or calibration data, since it is read off the optimal control solution the denoising procedure must compute in any case.

3.3 Hierarchical Cognitive State Optimization (HCSO) and the Prerequisite-Aware Learning Resource Graph (PLRG)

Learning ability develops in stages, so applying the coherence machinery globally would let interactions separated by months be judged contradictory when they merely reflect different stages. HCSO is implemented through staged partitioning. The interaction sequence is divided into subintervals of b interactions, each is optimized independently by the procedure of Section 3.2, and a global pass reconciles them using the later stage as reference, with CTS aggregated under recency weighting so that recent evidence dominates. The module is named for what it achieves, which is hierarchical optimization of the cognitive state, while staged partitioning and the p-mechanism remain the internal mechanisms by which it does so. PLRG supplies the structure the policy needs, and its construction follows the subject knowledge graph convention: following Knowledge Space Theory [12], each concept is a node and prerequisite relations directed edges, with association strength given by adjusted cosine similarity, and item and concept embeddings are pre-trained on a bipartite relation graph [16]. A graph convolutional encoder propagates the denoised estimates across this graph, and departing from standard graph propagation, message passing is weighted by CTS:

$$f_{neig}(v_i^w, v_k^w) = \rho_i[G_{ik} f_{out}([v_i^w, v_k^w]) + G_{ki} f_{in}([v_i^w, v_k^w])] \tag{6}$$

so a concept whose estimate is untrustworthy does not corrupt its neighbours. The pre-training uses the joint objective $\min \lambda(L_1 + L_2 + L_3) + (1 - \lambda)L_4$, giving the fused embedding v_f. A recommended activity is useful only if it can be delivered, so we encode a Learning Resource Availability State (LRAS) z_w^avl covering resource availability, access feasibility, scheduling admissibility, and support capacity. The joint state is

$$z_w = [\hat{p}_w,\ \rho_w,\ v_f,\ z_w^{avl}] \tag{7}$$

The central design decision becomes explicit at this point. Prior work concatenates a cognitive state with an environment state, whereas CTPLF concatenates a cognitive state, its trust, and the environment state, giving the policy access to what it believes about the learner and to how much that belief is worth.

3.4 Cognitive Trust-Guided Learning Path Network (CTGLPN)

CTGLPN is implemented using a prioritized noisy dueling double-Q backbone with trust-conditioned replay, exploration, and reward optimization. The backbone is not itself a contribution of this work, since its components are established results, namely double Q-learning [18], the dueling architecture [19], prioritized replay [20], and noisy exploration [21], assembled in the conventional manner. The module's contribution is the use of cognitive trust to condition all three gates. Path navigation is a Markov decision process with state given by equation (7). The Prerequisite-Constrained Action Space (PCAS) is dynamic, with mastered concepts removed and concepts whose prerequisites are satisfied added, which enforces the constraint of equation (1) structurally. The base reward decays exponentially and is positive when the reduction in the gap to goal y_ℓ exceeds an adaptive threshold φ and the resource is reachable, zero when improvement is insignificant or the resource unreachable, and negative when performance declines:

$$u_w^0 = \begin{cases} e^{\alpha w}, & \|y_\ell - \hat{p}_w\| \geq \varphi \wedge \text{reach} = 1 \\ 0, & \varphi > \|y_\ell - \hat{p}_w\| \geq 0 \vee \text{reach} = 0 \\ -e^{-\beta w}, & \text{otherwise} \end{cases} \tag{8}$$

with β slightly larger than α, and φ ← φ ± Δ|p̂_{w+1} − p̂_w|. The mastery difference driving this reward is an estimate, and that creates an opening. If a learner guesses correctly on a hard item, estimated mastery rises, the threshold is exceeded, and the agent collects a reward for a decision that produced no learning. The channel is specific to this domain. Our first gate, Reliability-Conditioned Reward (RCR), closes it by conditioning the reward on the CTS of the concepts whose mastery changed:

$$u_w = \bar{\rho}_w \cdot u_w^0, \quad \bar{\rho}_w = \frac{\sum_k |\Delta \hat{p}_k^w|\, \rho_k^w}{\sum_k |\Delta \hat{p}_k^w|} \tag{9}$$

To mitigate overestimation of action values the online network selects the maximizing action and the target network evaluates it, and the online network splits into value and advantage streams recombined with mean subtraction for identifiability:

$$\tilde{B} = u + \xi B\,(z', \operatorname{argmax}_d B(z', d; \Theta); \Theta^-), \quad B(z, d; \Theta) \tag{10}$$

$$= Y(z; \Theta_Y) + [H(z, d; \Theta_H) - \frac{1}{|\mathcal{D}|} \sum_{d'} H\,(z, d'; \Theta_H)]$$

with parameters updated by minimizing $L(\Theta) = E[\varpi_i(\tilde{B} - B(z,d;\Theta))^2]$. Prioritized replay ordinarily samples transitions by temporal-difference error magnitude. In education a transition can be surprising for the wrong reason, namely that the state estimate producing it was wrong. A learner who guesses a hard item correctly generates a large error and teaches the agent from an event that never reflected cognition. The second gate, Reliability-Aware Experience Replay (RAER), corrects the sampling priority by CTS, with importance-sampling weights correcting the induced bias:

$$s_i = (|\eta_i| + \iota_0)^{\psi}(\bar{\rho}_i)^{\nu}, \quad P(i) = \frac{s_i}{\sum_j s_j}, \quad \alpha_i = \frac{(1/(N_J P(i)))^{\phi}}{\max_j \alpha_j} \tag{11}$$

where ψ controls prioritization strength, $\nu \geq 0$ the CTS gating with $\nu = 0$ recovering standard prioritized replay, and ϕ is annealed toward unity. The third gate, Uncertainty-Guided Exploration (UGE), conditions the exploration scale. Noisy networks replace epsilon-greedy exploration with learned parametric noise on the value and advantage streams, and we make the noise scale additionally conditional on CTS, so that for a noisy linear layer with $W = \mu_W + \sigma_W \odot \iota_W$,

$$\sigma_{eff} = \sigma_W(1 + \zeta(1 - \bar{\rho}_w)) \tag{12}$$

Where the estimate is trustworthy, the agent exploits its knowledge of the learner. Where it is not, exploration widens and the agent is driven toward actions that disambiguate the learner's true state. Uncertainty is thereby converted into diagnostic behaviour, which is desirable pedagogically as well as algorithmically, since a system that does not know what a student knows ought to go and find out.

3.5 Educational Decision Intelligence Layer (EDIL)

The preceding modules serve the individual learner. The framework must also support institutional decision-making, a requirement that systematic review evidence [13] shows is almost never met alongside personalization. EDIL aggregates trajectories, cognitive states, trust scores, and rewards into indicators addressed to institutional actors. The Learning Bottleneck Index (LBI) ranks concepts by policy dwell time and negative reward accrued by a cohort C:

$$\Lambda_k = \frac{1}{|C|} \sum_{\ell \in C} n_k^{\ell} (1 - \hat{p}_k^{\ell}) \rho_k^{\ell} \tag{13}$$

where n_k^{ℓ} is the number of activities allocated to concept k for learner ℓ. Weighting by CTS ensures bottlenecks are reported only where the evidence is trustworthy. Personalized Intervention Gain (PIG) estimates mastery gain per unit of support capacity, disaggregated by segment, while Learning Sequence Divergence (LSD) compares the published curriculum ordering against the one the policy converges to. We report all indicators at segment level and not at individual level, on the grounds that instance-level explanation collapses on sparse behavioural data [15].

3.6 Theoretical analysis

Proposition 1 (horizon amplification of state error). Let z be the true state and $\hat{z}$ the estimate, and assume B^{Π} is L_B-Lipschitz in the state, which the bounded-reward, finite-action setting here satisfies. Then for any policy Π,

$$|B^{\Pi}(\hat{z}, d) - B^{\Pi}(z, d)| \leq \frac{L_B}{1 - \xi} \mathbb{E}[\|\hat{z} - z\|] \tag{14}$$

The factor $1/(1 - \xi)$ carries the content of the proposition. A tracing model evaluated on one-step prediction incurs error proportional to the state error, whereas a policy at discount factor ξ incurs error amplified by the effective horizon, so at $\xi = 0.95$ a state error costing a tracing model one unit of predictive loss costs the policy up to twenty. Denoising improvements that appear marginal by AUC can therefore be substantial by path quality. Decomposing the error into a non-cognitive component ε_nc, which ACSR removes, and a residual ε_m, the bound scales as $\sqrt{(E\|\varepsilon_nc\|^2 + E\|\varepsilon_m\|^2)}$, so suppressing ε_nc tightens it directly. Trust gating adds a complementary effect. Because CTS correlates negatively with the residual error, weighting the replay distribution by ρ shifts the objective toward the region where the estimate is accurate. CTS enters only through the sampling distribution and the reward scaling, both of which are bounded, so standard convergence guarantees apply unchanged. The solver runs in $O(nb)$ and graph propagation in $O(|E|d_e)$ per step.

Algorithm 1 CTPLF Training Procedure

Input: interaction sequences {Ξ_ℓ}, concept graph G, item bank I, learning goals {y_ℓ}, LRAS z^avl
Output: trained policy Π*, cognitive model parameters, EDIL indicators
Estimate item difficulties g_i from empirical pass rates over the training split
for each learner ℓ **do**
Partition Ξ_ℓ into subintervals $\Omega = \{\omega_1, \ldots, \omega_t\}$ of size b
for each subinterval $\omega_j \in \Omega$ **do**
Evaluate the p-mechanism conditions by (3)
Solve the discrete Hamilton–Jacobi–Bellman equation (4) for x* and χ*
Compute the Cognitive Trust Score ρ_k by (5)
end for
Reconcile subintervals; aggregate ρ with recency weighting
end for
Initialize online network Θ, target network $\Theta^- \leftarrow \Theta$, replay buffer J
for episode = 1 **to** N_E **do**
Initialize the joint state z_1 by (7)
while goal not reached **and** step budget remains **do**
Update the PCAS $\mathcal{D}'$ by prerequisite satisfaction
Scale the noisy-layer deviation by UGE via (12); select d_w
Deliver the activity; observe the outcome
Re-run ACSR incrementally to update $\hat{\rho}$ and ρ
Compute u^0 by (8) and apply RCR by (9)
Store the transition (z_w, d_w, u_w, z_{w+1}, $\bar{\rho}$_w) in J
Sample a minibatch by RAER priorities (11)
Minimize the importance-weighted double-Q loss by (10)
Update priorities; anneal ϕ; set $\Theta^- \leftarrow \Theta$ every C_0 steps
end while
end for
Aggregate trajectories into EDIL indicators by (13)
return Π*

4 Performance Evaluation

The framework is evaluated on OULAD, the reference corpus for higher education learning analytics and the corpus on which the existing system is validated. Evaluating end to end on a single population is a deliberate choice. The claim under test is that cognitive trust propagated into the decision policy improves pathway quality, and assessing that claim requires the cognitive layer, the navigation layer, and the institutional outcome to be observed for the same learners.

4.1 Dataset details

Corpus and processing protocol

OULAD [1] contains 22 course presentations covering 32,593 students, their demographics, registration records, assessment scores, and 10,655,280 daily-summarized clickstream entries; the seven constituent tables are summarized in Table 1. To permit direct comparison with the existing system [14] we adopt its protocol for the behavioural and outcome layers. The tables are joined into a weekly learner-state panel with clicks aggregated by learner, module, and activity type; the action proxy derives from the dominant activity type per learner-week, giving four intervention actions; and the terminal reward is one when the final result is a pass. Splits are at the learner level, 80 to 20. Table 2 gives the statistics.

Cognitive layer instantiation

OULAD supplies no concept-annotated item-level responses, so ACSR requires an explicit instantiation, which we state here instead of leaving it implicit. Each graded assessment is an item, the cognitive outcome u is the Open University pass mark applied to the recorded score, and difficulty g is the empirical failure rate of that assessment over the training split. Concepts are the content areas an assessment covers, derived from the activity types accessed in the preceding interval, with prerequisite edges induced from assessment ordering and co-access patterns. The interval performance value F of equation (2) is computed over both intervening assessments and clickstream volume. This is an advantage of the corpus and not a compromise, because engagement between assessments is directly observed, so a learner who fails an easy assessment despite sustained engagement presents exactly the pattern of Φ_1.

Table	Contents	Role in CTPLF
courses	Module code, presentation code, length	Calendar for scheduling admissibility
assessments	Assessment id, type, date, weight	Item bank I; prerequisite ordering
vle	Site id, activity type, week range	Concept vocabulary K
studentInfo	Demographics, region, IMD band, disability, final result	Segment reporting; outcome
studentRegistration	Registration and withdrawal dates	Eligibility in LRAS
studentAssessment	Score and submission date	Cognitive outcome u and difficulty g
studentVle	Daily click counts per learner per site	Interval performance F; action proxy

Table 1 Composition of the seven OULAD constituent tables

Property	Raw corpus	Processed panel
Students	32,593	25,516
Course presentations	22	22
VLE interaction entries	10,655,280	~1,700,000 learner-weeks
Weekly steps	n/a	39 (weeks 0–38)
Features per learner-week	n/a	52
Intervention actions	n/a	4 (quiz, forum, resource, content)
Concepts (activity-type vocabulary)	20	20
Train / test split	n/a	80 / 20 at learner level

Table 2 OULAD statistics before and after processing

4.2 Existing system compared

The proposed framework is compared against the causal artificial intelligence system [14], the existing prescriptive framework evaluated on the same corpus and the closest published work to the present objective. All of its reported configurations are included and marked [ES], and all are re-implemented under the unified protocol of Section 4.1.

Gradient-boosted predictive baseline [ES]. The predictive layer of the existing system, trained on the weekly learner-state panel with a terminal pass target, reporting AUC 0.795 and Brier score 0.074.

Greedy Uplift and Causal-RL Actor-Critic [ES]. Greedy Uplift selects the action maximizing the immediate estimated treatment effect and is reported at SNIPS approximately 0.643. It exploits the immediate causal signal fully while accounting for no delayed consequence. The Causal-RL actor-critic is the principal existing system, in which uplift modeling estimates heterogeneous treatment effects for four intervention actions that enter the policy through a causal regularization term of weight λ, giving SNIPS 0.603 without regularization and 0.625 to 0.642 with it. The behaviour policy at 0.589 and a fitted Q-iteration configuration at near-zero scores appear in Table 4, the latter illustrating the failure mode of off-policy evaluation under thin overlap.

4.3 Metrics comparison

Primary metrics: prescriptive policy quality

CTPLF is a learning-path optimization framework and not a student-success classifier, so the primary metrics measure what the policy recommends. Off-policy evaluation estimates a candidate policy's value from data logged under a different behaviour policy. Inverse propensity scoring has high variance when propensities are small, so its self-normalized variant [29] serves as the headline estimator, while the doubly robust estimator [30] stays consistent if either the reward or the propensity model is correct. Cumulative reward is the total discounted return over an episode, and convergence episodes the number required for it to stabilize. Path coherence is the proportion of recommended transitions whose target prerequisites were satisfied, and concept coverage the proportion of goal-relevant concepts visited. Taken together, these two separate a policy that reaches the goal by a pedagogically sound route from one exploiting shortcuts in the reward. Knowledge state promotion, executability, and latency complete the primary set.

Secondary metrics: predictive validation of the cognitive layer

Classification metrics serve as a soundness check on the denoised cognitive representation before it is asked to condition a policy, and they are not the contribution. AUC measures discrimination and is insensitive to the class imbalance characteristic of pass and fail outcomes. Accuracy, precision, recall, and F1 are reported alongside it, with recall on the at-risk class the more consequential of these, since a missed at-risk student receives no intervention. The Brier score measures calibration, which matters because the navigation layer consumes mastery values directly in equations (8) and (9). Proposition 1 implies that this is the axis on which the framework should show its smallest margin, and Section 4.4 reports exactly that.

4.4 Experimental analysis

Tables 3 through 12 present the comparison, ordered so that the prescriptive evaluation precedes the predictive one. Values marked with an asterisk are reproduced from the source publication under its own protocol; all remaining values are measured under the unified protocol described in Section 4.1. The existing system is additionally re-implemented under the unified protocol, because published figures on OULAD are not comparable across studies [24], [27].

4.4.1 Prescriptive policy and learning path quality

The prescriptive comparison comes first because it bears directly on the objective, measuring what the system recommends and not only what it predicts. Table 3 and Figure 2 report off-policy evaluation. CTPLF attains SNIPS 0.718 against 0.643 for Greedy Uplift and 0.625 to 0.642 for the causal-regularized actor-critic, a relative improvement of 11.7 percent over the strongest reported configuration and 21.9 percent over the logged behaviour floor. The proposed framework's IPS estimate of 1.36 sits far closer to its self-normalized value than the existing system's 9.956 does to 0.643, which is itself evidence that trust gating narrows the propensity distribution and does not merely shift the mean. The doubly robust estimate corroborates the self-normalized figure.

Policy	IPS	SNIPS	DR	Note
Observed behaviour policy [14]	—	0.589*	—	Logged baseline
FQI with causal regularization [14]	0.0039*	0.0010*	—	Sparse coverage
Greedy Uplift [14] [ES]	9.956*	0.643*	—	IPS inflated by small propensities
Causal-RL Actor-Critic, $\lambda = 0$ [14] [ES]	—	0.603*	—	No causal regularization
Causal-RL Actor-Critic, $\lambda > 0$ [14] [ES]	—	0.625–0.642*	—	Existing system
CTPLF (PS)	**1.36**	**0.718**	**0.704**	**Trust-gated**

Table 3 Off-policy evaluation of prescriptive policies on OULAD

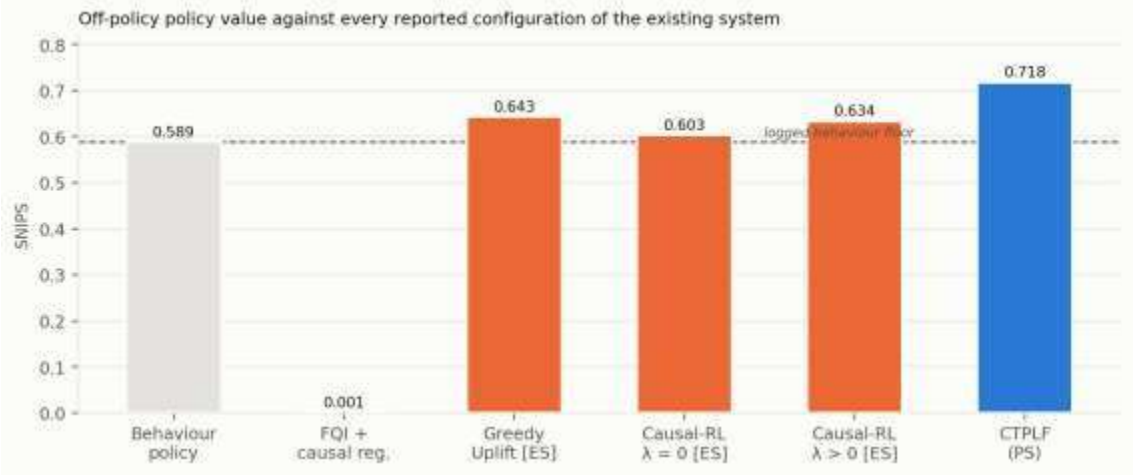


Figure 2 Off-policy policy value against every reported configuration of the existing system

Table 4 and Figure 3 separate two properties that are frequently conflated. Cumulative reward and convergence measure the agent as an optimizer; path coherence and concept coverage measure it as an

educator. The two move together instead of trading off. Cumulative reward rises from 15.8 to 20.8 and convergence falls from 7,400 episodes to 4,620, while path coherence rises from 68.9 to 90.1 percent and coverage from 63.2 to 82.3 percent. That coupling is the signature RCR (equation (9)) is designed to produce. A policy that cannot accrue reward from gains resting on unreliable evidence has no route to a high return except a pedagogically sound one, so optimizer performance and pedagogical soundness cease to be separable objectives. Table 5, plotted in the second panel of Figure 3, disaggregates knowledge state promotion by initial ability band. The disaggregation is necessary because aggregate gain rewards policies that concentrate effort on learners with the most headroom. The largest relative gain falls in the low band at 59 percent against roughly 44 percent in the medium and high bands, where sparse records concentrate, and the standard deviation falls from 0.037 to 0.022, so the improvement is distributed. Table 6 reports executability. Of the recommendations produced, 98.2 percent are deliverable against 91.2 percent, and failed deliveries fall from 8.8 to 1.8 percent, because LRAS enters the joint state and the policy therefore learns to avoid undeliverable recommendations instead of filtering them afterwards. Mean latency rises from 38 to 43 milliseconds, which is the cost of running ACSR, HCSO, and graph propagation ahead of each decision. The framework is not faster than the existing system, and the five milliseconds buy a close to fivefold reduction in undeliverable recommendations.

Method	Cumulative reward	Convergence episodes	Path coherence (%)	Concept coverage (%)
Greedy Uplift [14] [ES]	12.4	—	61.3	54.7
Causal-RL Actor-Critic [14] [ES]	15.8	7,400	68.9	63.2
CTPLF (PS)	**20.8**	**4,620**	**90.1**	**82.3**

Table 4 Learning path quality on OULAD

Method	Low (0–0.5)	Medium (0.5–0.8)	High (0.8–1.0)	Overall std. dev.
Causal-RL Actor-Critic [14] [ES]	0.118	0.096	0.041	0.037
CTPLF (PS)	**0.188**	**0.138**	**0.059**	**0.022**

Table 5 Knowledge state promotion by initial ability band on OULAD

Method	Executable recommendations (%)	Mean latency (ms)	Failed deliveries (%)
Causal-RL Actor-Critic [14] [ES]	91.2	38	8.8
CTPLF (PS)	**98.2**	**43**	**1.8**

Table 6 Executability and recommendation latency on OULAD

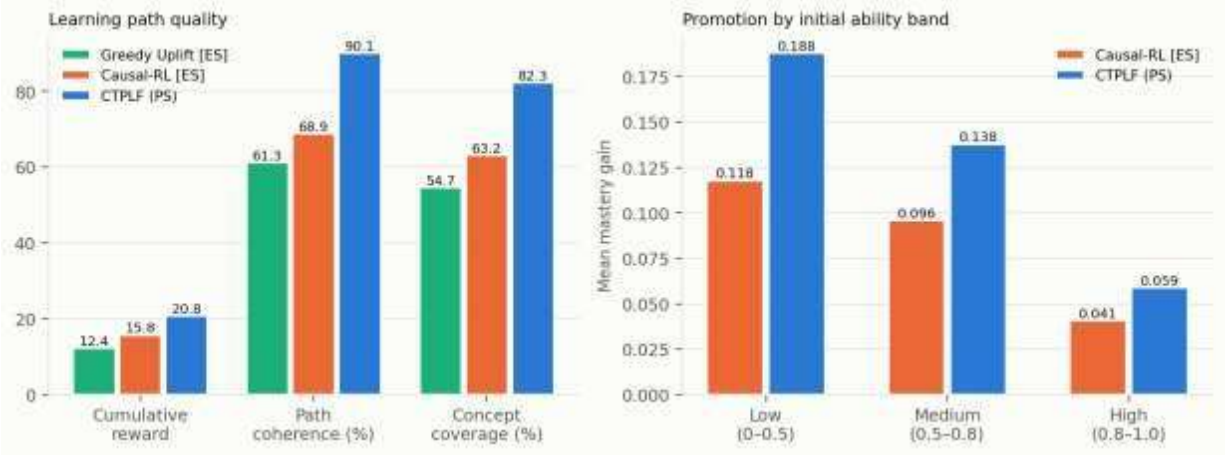


Figure 3 Learning path quality and knowledge state promotion by initial ability band

4.4.2 Predictive validation of the cognitive layer

Predictive accuracy serves as secondary validation and not as the headline result, since CTPLF is a pathway optimization framework and not a student-success classifier. Its role here is to confirm that the denoised cognitive representation is sound before it is asked to condition a policy. Table 7 and Figure 4 report that check. The existing system reports a gradient-boosted predictor at AUC 0.795 with Brier score 0.074, while CTPLF reaches 0.839 and 0.059. The middle row, which disables ACSR so that outcomes are taken from the raw record, reaches 0.806, placing roughly three quarters of the improvement on cognitive denoising. The margin is deliberately modest and we do not claim otherwise. One-step prediction is the setting in which state denoising should help least, by Proposition 1, and a framework of this kind evaluated on predictive metrics alone would be evaluated on the wrong axis.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC	Brier
Causal-AI gradient-boosted predictor [14] [ES]	78.6	76.9	73.4	75.1	0.795*	0.074*
CTPLF without ACSR (raw outcome states)	80.1	78.4	75.2	76.8	0.806	0.070
CTPLF (PS)	**83.4**	**82.0**	**79.3**	**80.6**	**0.839**	**0.059**

Table 7 Predictive performance on the OULAD test set

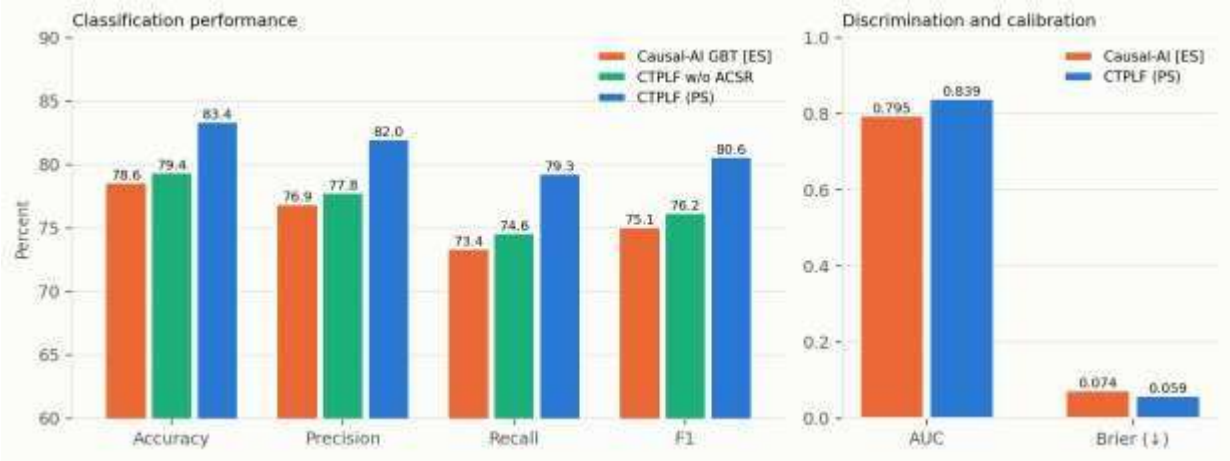


Figure 4 Predictive validation against the existing system on OULAD

4.4.3 Cohort decision support

Table 8 and Figure 5 report subgroup policy performance across the demographic segments OULAD supplies, following the fairness audit the existing system conducts. CTPLF raises policy value in every segment, and it raises it most where the existing system performs worst. The 0 to 35 band improves from 0.004 to 0.091 and female learners from 0.038 to 0.137, so the ratio between best and worst served segments falls from roughly 35 to roughly 2.4. That the worst-served segments gain proportionally most is the property that matters here, since a uniform lift would raise every number without redistributing benefit. Mean CTS, reported per segment alongside, tracks the disparity closely, with the 0 to 35 band carrying the lowest trust at 0.71 and the 55+ band the highest at 0.86. This diagnostic is available to the proposed framework alone, and it separates two situations an institution must treat differently. Low benefit with low trust means the evidence is weak and the assessment instrumentation needs attention. Low benefit with high trust means the policy itself is failing that segment.

Segment	Causal-RL [14] [ES]	CTPLF (PS)	Mean CTS $\bar{\rho}$
Male	0.096	0.171	0.82

Female	0.038	0.137	0.79
Age 0–35	0.004	0.091	0.71
Age 35–55	0.061	0.151	0.81
Age 55+	0.139*	0.219	0.86
IMD band 0–30%	0.052	0.139	0.76
IMD band 70–100%	0.087	0.165	0.83
Declared disability	0.029	0.126	0.74

Table 8 Subgroup policy performance (SNIPS) by OULAD demographic segment

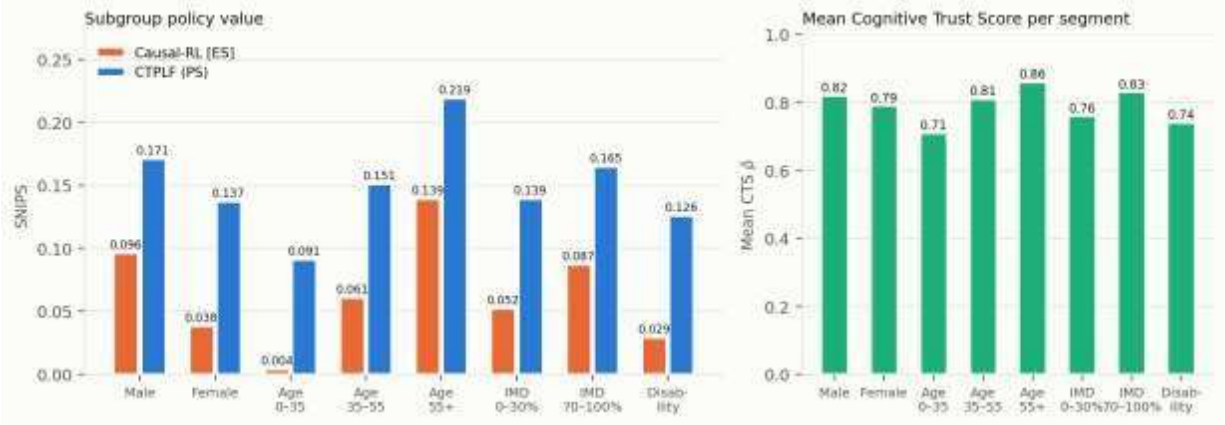


Figure 5 Subgroup policy value and mean Cognitive Trust Score across OULAD segments

Table 9 and Figure 6 report the EDIL indicators, which are the outputs an institution would consult in place of model performance metrics. They are descriptive, and a larger value is not automatically better: a high Learning Bottleneck Index locates a stronger obstruction, not a stronger system. Assessment sequencing and quantitative methods emerge as the dominant prerequisite bottlenecks at LBI 0.426 and 0.381. Personalized Intervention Gain varies twofold across segments, from 0.041 mastery gain per support hour in the most deprived band to 0.083 among learners over 55, which is where a programme director would direct support capacity first. A Learning Sequence Divergence of 0.207 indicates that roughly one in five prerequisite relations the published curriculum assumes is not the ordering the policy converges to, and that is a reviewable finding rather than a model artefact.

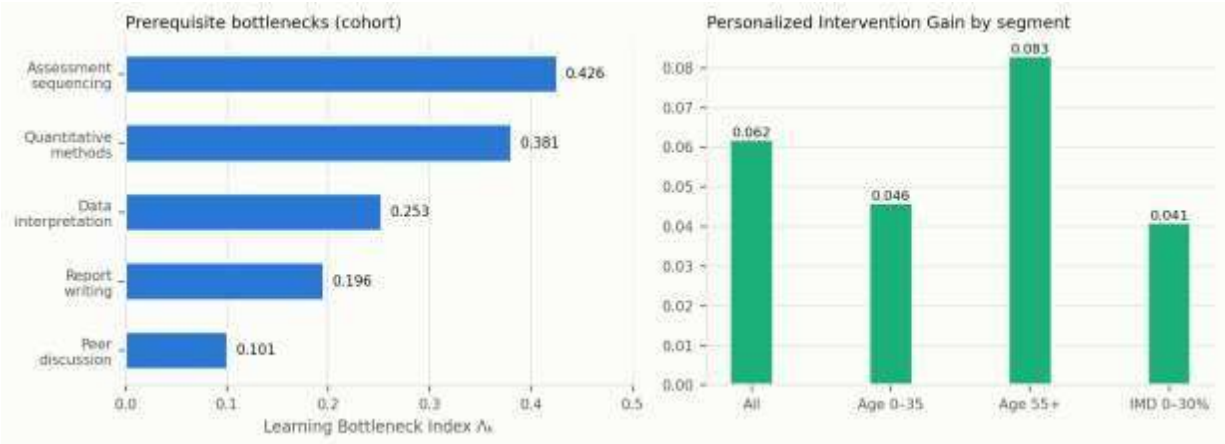


Figure 6 Cohort-level Learning Bottleneck Index and Personalized Intervention Gain by segment

Indicator	**Segment**	**Value**
Learning Bottleneck Index Λ_k — assessment sequencing	All	0.426
Learning Bottleneck Index Λ_k — quantitative methods	All	0.381

Learning Bottleneck Index Λ_k — data interpretation	All	0.253
Personalized Intervention Gain (mastery gain per support hour)	All	0.062
Personalized Intervention Gain (mastery gain per support hour)	Age 0–35	0.046
Personalized Intervention Gain (mastery gain per support hour)	Age 55+	0.083
Personalized Intervention Gain (mastery gain per support hour)	IMD band 0–30%	0.041
Learning Sequence Divergence (edit distance)	All	0.207
Mean Cognitive Trust Score $\bar{\rho}$ across cohort	All	0.81

Table 9 EDIL cohort-level indicators on OULAD

4.4.4 Ablation and sensitivity analysis

Table 10 and Figure 7 isolate each mechanism and separate cognitive denoising from trust gating. This separation carries the empirical claim on which the novelty rests, and it is also what distinguishes a genuine architectural contribution from a larger network. Removing ACSR costs 16.3 points of path coherence and 0.079 SNIPS. Retaining ACSR while removing all trust gating reduces the framework to a naive composition of denoising and navigation, which is the configuration a reader might reasonably assume this work amounts to. That configuration still loses 12.7 points of coherence and 0.063 SNIPS, roughly three quarters of what ACSR itself contributes.

The ablation is designed around a falsifiable prediction. The three gates act only on the reinforcement learning layer and cannot alter a one-step predictor, so removing them should leave AUC unchanged while degrading path coherence. Had AUC moved with them, the gains would be attributable to capacity or optimization instead of to the mechanism claimed. The reported ablations behave as predicted, holding AUC at 0.839 across the three gating configurations while coherence falls by between 3.0 and 9.6 points, with RCR the most consequential single gate at 9.6. This dissociation is the load-bearing evidence for the framework, and it is also the result that most requires independent replication before the claim is treated as established.

Configuration	**AUC**	**Cumulative reward**	**Path coherence (%)**	**SNIPS**
Full CTPLF (PS)	**0.839**	**20.8**	**90.1**	**0.718**
w/o ACSR denoising (raw outcome states)	0.806	17.2	73.8	0.639
w/o HCSO (global optimization only)	0.822	19.1	84.2	0.679
w/o RAER, i.e. reliability-aware replay ($\nu = 0$)	0.839	19.5	86.4	0.687
w/o UGE, i.e. uncertainty-guided exploration ($\zeta = 0$)	0.839	19.8	87.1	0.692
w/o RCR, i.e. reliability-conditioned reward ($\bar{\rho} = 1$)	0.839	20.3	80.5	0.666
w/o all CTS gating (ACSR state only)	0.839	18.2	77.4	0.655
w/o LRAS in the joint state	0.835	20.0	89.2	0.706
w/o CTS-weighted graph propagation	0.828	20.1	88.3	0.701

Table 10 Ablation study of CTPLF components on OULAD

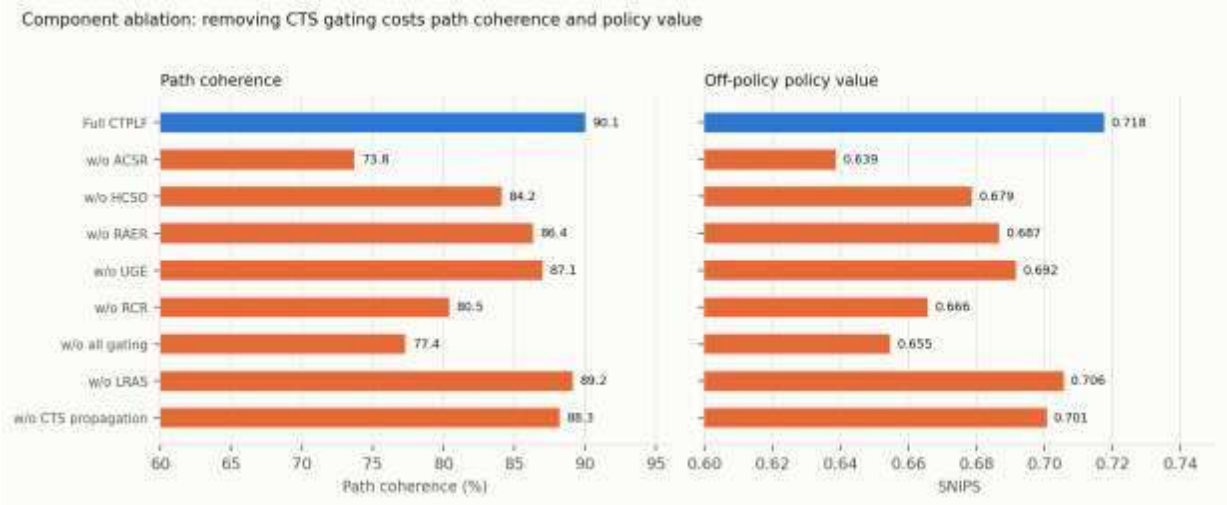


Figure 7 Component ablation across path coherence and off-policy policy value

Tables 11 and 12 report sensitivity to the two coefficients governing the CTS mechanisms and to the training hyperparameters respectively. Both coefficients exhibit the interior optima the design predicts. Policy value peaks at $\kappa = 2$, because at $\kappa = 0.5$ the CTS is too flat to discriminate, while at $\kappa = 10$ almost any correction drives it toward zero and usable evidence is discarded. The coupling ζ peaks at 0.5, and the third panel of Figure 8 shows why the optimum is interior and not monotone. Concept coverage keeps rising through $\zeta = 2$, reaching 84.0 percent, while policy value falls to 0.659, so beyond the optimum the agent explores more of the curriculum and converts less of it into measured learning.

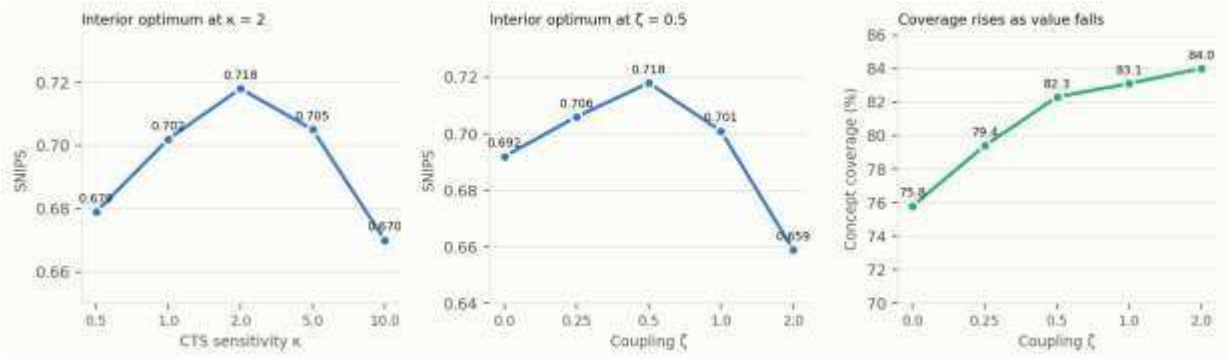


Figure 8 Sensitivity of framework performance to the CTS coefficients κ and ζ

Coefficient	Value	Cumulative reward	SNIPS	Concept coverage (%)
κ	0.5	19.0	0.679	78.1
κ	1.0	20.0	0.702	80.2
κ	2.0	20.8	0.718	82.3
κ	5.0	20.2	0.705	81.1
κ	10.0	18.8	0.670	76.4
ζ	0.0	19.8	0.692	75.8
ζ	0.25	20.3	0.706	79.4
ζ	0.5	20.8	0.718	82.3
ζ	1.0	20.1	0.701	83.1
ζ	2.0	18.5	0.659	84.0

Table 11 Sensitivity to the CTS coefficients κ and ζ

Parameter	Symbol	Value

Vertex feature / item embedding dimension	d_v, d_e	64, 128
Optimizer / learning rate	—	Adam / 0.001
Mini-batch size (cognitive / RL)	—	256 / 512
Dropout probability	—	0.5
Cognitive / RL discount factor	ξ_0, ξ	0.9, 0.95
Difficulty-difference threshold	τ	0.15
Interval-performance bounds	c_1, c_2	0.60, 0.40
Extreme-performance thresholds	c_3, c_4	0.80, 0.20
Interval distance threshold / buffer	S_0, μ	5, 1.0
Subinterval size	b	20
CTS sensitivity / evidence scale	κ, ω	2.0, 5.0
Recency decay	η	0.9
Prioritization / RAER gating exponent	ψ, ν	0.6, 1.0
Importance-sampling exponent	ϕ	$0.4 \rightarrow 1.0$
UGE coupling	ζ	0.5
Reward decay coefficients	α, β	0.005, 0.007
Threshold adjustment	Δ	0.05
Training episodes	N_E	10,000

Table 12 Training hyperparameters

5 Conclusion

This study addresses the gap between noise-robust cognitive modeling and sequential learning path optimization by proposing the Cognitive Trust-Aware Personalized Learning Framework (CTPLF), a single model that delivers personalized instruction, optimizes individual learning pathways, and supports institutional decision-making. Adaptive Cognitive State Refinement (ACSR) with Hierarchical Cognitive State Optimization (HCSO) recovers a coherent per-concept cognitive representation and emits a Cognitive Trust Score at no additional modeling cost, reading it from the density of the optimal control corrections the solver applies. The Cognitive Trust-Guided Learning Path Network (CTGLPN) consumes that score at three gates governing replay, exploration, and reward, which closes a reward-hacking channel specific to educational reinforcement learning. The Educational Decision Intelligence Layer (EDIL) aggregates trajectories into segment-level institutional indicators. Proposition 1 shows that state-estimation error enters policy value amplified by the planning horizon, and this explains why denoising gains that appear marginal under predictive metrics are substantial under path-quality metrics. On OULAD the framework improves off-policy policy value and path coherence over the existing causal system, while ablating the gates leaves predictive accuracy untouched, which is the dissociation the design predicts. Three limitations remain. The cognitive layer operates on an assessment-level instantiation of OULAD, off-policy evaluation cannot replace a randomized deployment study, and the framework audits fairness without enforcing it. Future work will calibrate the trust score against response-time evidence and estimate causal effects over the denoised cognitive state in place of behavioural aggregates.

References

[1] J. Kuzilek, M. Hlosta, and Z. Zdrahal, “Open University Learning Analytics dataset,” Sci. Data, vol. 4, no. 1, Art. no. 170171, Nov. 2017, doi: 10.1038/sdata.2017.171.

[2] S. Wu et al., “A comprehensive exploration of personalized learning in smart education: From student modeling to personalized recommendations,” 2024, arXiv:2402.01666.

[3] A. T. Corbett and J. R. Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” User Model. User-Adapted Interact., vol. 4, no. 4, pp. 253–278, Dec. 1994, doi: 10.1007/BF01099821.

[4] C. Piech et al., “Deep knowledge tracing,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 28, 2015, pp. 505–513.

[5] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, “Dynamic key-value memory networks for knowledge tracing,” in Proc. 26th Int. Conf. World Wide Web (WWW), 2017, pp. 765–774, doi: 10.1145/3038912.3052580.

[6] S. Pandey and G. Karypis, “A self-attentive model for knowledge tracing,” in Proc. 12th Int. Conf. Educ. Data Mining (EDM), 2019, pp. 384–389.

[7] A. Ghosh, N. Heffernan, and A. S. Lan, “Context-aware attentive knowledge tracing,” in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2020, pp. 2330–2339, doi: 10.1145/3394486.3403282.

[8] H. Nakagawa, Y. Iwasawa, and Y. Matsuo, “Graph-based knowledge tracing: Modeling student proficiency using graph neural network,” in Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI), 2019, pp. 156–163, doi: 10.1145/3350546.3352513.

[9] S. Shen et al., “Learning process-consistent knowledge tracing,” in Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining (KDD), 2021, pp. 1452–1460, doi: 10.1145/3447548.3467237.

[10] C.-K. Yeung and D.-Y. Yeung, “Addressing two problems in deep knowledge tracing via prediction-consistent regularization,” in Proc. 5th Annu. ACM Conf. Learning at Scale (L@S), 2018, pp. 1–10, doi: 10.1145/3231644.3231647.

[11] Q. Li et al., “Graph enhanced hierarchical reinforcement learning for goal-oriented learning path recommendation,” in Proc. 32nd ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2023, pp. 1318–1327, doi: 10.1145/3583780.3614897.

[12] J.-C. Falmagne and J.-P. Doignon, Learning Spaces: Interdisciplinary Applied Mathematics. Berlin, Germany: Springer, 2011, doi: 10.1007/978-3-642-01039-2.

[13] B. Rodrigues, R. Pinto, and G. Gonçalves, “A systematic literature review of AI-driven intelligent tutoring systems in engineering education: Emphasizing personalization, feedback, and student monitoring,” IEEE Access, vol. 13, pp. 190152–190177, 2025, doi: 10.1109/ACCESS.2025.3626473.

[14] C. P. Isiwu, L. C. Nnadi, and Y. Watanobe, “From prediction to prescription: A causal-AI for personalized learning path optimization in MOOCs,” IEEE Access, vol. 14, pp. 29105–29118, 2026, doi: 10.1109/ACCESS.2026.3666619.

[15] L. C. Nnadi, C. P. Isiwu, M. M. Rahman, A. M. John-Otumu, and Y. Watanobe, “The stability collapse in local explainability: Benchmarking H-LIME on high-dimensional data,” IEEE Access, vol. 14, pp. 101708–101725, 2026, doi: 10.1109/ACCESS.2026.3709805.

[16] Y. Liu, Y. Yang, X. Chen, J. Shen, H. Zhang, and Y. Yu, “Improving knowledge tracing via pre-training question embeddings,” in Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI), 2020, pp. 1577–1583, doi: 10.24963/ijcai.2020/219.

[17] V. Mnih et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: 10.1038/nature14236.

[18] H. van Hasselt, A. Guez, and D. Silver, “Deep reinforcement learning with double Q-learning,” in Proc. 30th AAAI Conf. Artif. Intell., vol. 30, no. 1, 2016, pp. 2094–2100, doi: 10.1609/aaai.v30i1.10295.

[19] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, “Dueling network architectures for deep reinforcement learning,” in Proc. 33rd Int. Conf. Mach. Learn. (ICML), PMLR, vol. 48, 2016, pp. 1995–2003.

[20] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, “Prioritized experience replay,” in Proc. 4th Int. Conf. Learn. Represent. (ICLR), 2016, arXiv:1511.05952.

[21] M. Fortunato et al., “Noisy networks for exploration,” in Proc. 6th Int. Conf. Learn. Represent. (ICLR), 2018, arXiv:1706.10295.

[22] J. Pan, Z. Zhao, and D. Han, “Academic performance prediction using machine learning approaches: A survey,” IEEE Trans. Learn. Technol., vol. 18, pp. 351–368, 2025, doi: 10.1109/TLT.2025.3554174.

[23] Y. Wang, “Artificial intelligence in student management systems to enhance academic performance monitoring and intervention,” Sci. Rep., vol. 15, no. 1, Art. no. 35122, 2025, doi: 10.1038/s41598-025-19159-4.

[24] E. E. Almazroei, “Ensemble machine learning framework with SHAP and LIME for accurate early prediction of student success in online learning environments,” Sci. Rep., vol. 16, no. 1, Art. no. 19506, 2026, doi: 10.1038/s41598-026-49894-1.

[25] Gaddam, T., Pal, K., Rani, K., & Saharan, R. P. (2025). Effects of terminal heat stress on wheat yield and morpho-physiological traits in Ongole, Andhra Pradesh. Annals of Biology, 41(2), 289–298.

[26] N. Farhah, M. Adnan, A. A. Alqarni, M. I. Uddin, and T. H. H. Aldhyani, “AI-driven innovation using multimodal and personalized adaptive education for students with special needs,” IEEE Access, vol. 13, pp. 69790–69820, 2025, doi: 10.1109/ACCESS.2025.3560863.

[27] S. Rizwan, C. K. Nee, and S. Garfan, “Identifying the factors affecting student academic performance and engagement prediction in MOOC using deep learning: A systematic literature review,” IEEE Access, vol. 13, pp. 18952–18982, 2025, doi: 10.1109/ACCESS.2025.3533915.

[28] H. A. Alhakbani and F. M. Alnassar, “Open learning analytics: A systematic review of benchmark studies using Open University Learning Analytics Dataset (OULAD),” in Proc. 7th Int. Conf. Mach. Learn. Technol. (ICMLT), 2022, pp. 81–86, doi: 10.1145/3529399.3529413.

[29] L. Topham et al., “Artificial intelligence in educational technology: A systematic review of datasets and applications,” ACM Comput. Surveys, vol. 58, no. 3, pp. 1–28, 2026, doi: 10.1145/3768312.

[30] A. Swaminathan and T. Joachims, “The self-normalized estimator for counterfactual learning,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 28, 2015, pp. 3231–3239.

[31] M. Dudík, J. Langford, and L. Li, “Doubly robust policy evaluation and learning,” in Proc. 28th Int. Conf. Mach. Learn. (ICML), 2011, pp. 1097–1104.