ORCHID-RANKER: An Agentic Adaptive Recommender for Education with Privacy Guarantees

Anonymous Authors
Anonymous Institution
–, –
anonymous@anonymous.com

Abstract

Digital learning platforms must make real-time recommendations from minimal signals, yet existing approaches lack the stability guarantees and privacy transparency required for educational deployment. We introduce ORCHID-RANKER, a novel agentic educational recommender that addresses fundamental challenges in adaptive learning through bounded-stable autonomous control and principled privacy integration.

Our core algorithmic contributions include: (1) Bounded-stable adaptation algorithms with monotone state-to-action mappings that guarantee recommendation stability—the first educational recommender with formal stability guarantees for autonomous learning. (2) Novel privacy-preserving mechanisms that decouple learning updates from selection logic, enabling differential privacy without sacrificing interpretability—solving a fundamental tension in private recommendation systems. (3) Multi-objective optimization algorithms that jointly optimize knowledge building, exploration, and diversity while respecting educational constraints (Zone of Proximal Development) and maintaining computational tractability.

The agentic architecture implements autonomous perception-planning-action cycles: perception tracks learner state via bounded exponential moving averages, planning maps state to interpretable control dials through formal monotone transformations, and action executes via educational theory-grounded recommendation modifications. Unlike existing approaches that intertwine components, our separation enables interpretable state-to-action mappings with theoretical guarantees.

Comprehensive empirical validation on large-scale educational datasets demonstrates consistent, educationally meaningful improvements with effect sizes that meet established educational intervention thresholds, showing sustained performance across diverse student models and learning contexts. Privacy analysis reveals corpus-dependent tradeoffs enabling practical deployment guidance. The approach advances beyond current limitations of existing adaptive educational systems through principled integration of educational theory, formal stability guarantees, and transparent privacy mechanisms.

ACM ISBN 979-X-XXXX-XXXX-X/26/03
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts

• **Human-centered computing** → **Recommender systems**; *User models*; • **Applied computing** → *E-learning*; • **Security and privacy** → Privacy-preserving protocols.

Keywords

educational recommender, agentic control, engagement, knowledge tracing, exploration, diversity, pacing, privacy–utility, adaptive learning, contextual bandits, multi-objective optimization, differential privacy, intelligent tutoring systems, personalized learning, educational data mining, learning analytics, reinforcement learning, autonomous agents, bounded stability

ACM Reference Format:
Anonymous Authors. 2026. ORCHID-RANKER: An Agentic Adaptive Recommender for Education with Privacy Guarantees. In *Proceedings of The 41st ACM/SIGAPP Symposium on Applied Computing (SAC'26).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Digital learning platforms capture a narrow stream of traces. A system can observe whether a suggested activity was accepted, whether the response was correct, and simple timing such as dwell and latency. From these ordinary logs a recommender must decide what to surface next so that learning progresses and attention is sustained. The education and learning analytics literature treats knowledge and engagement as two central outcomes derived from such traces. Knowledge is often proxied by correctness or mastery trajectories and engagement by willingness to interact and sustained attention [35, 12, 46]. Classical psychometrics links correctness to latent ability through item response theory [39, 30, 40]. Knowledge tracing and its neural variants model how knowledge evolves over time [10, 38]. Contextual bandits formalize exploration versus exploitation under uncertainty [28, 9, 27]. Diversity and novelty can help sustain attention [5, 25]. Educational theory recommends practice within a zone of proximal development and the use of spacing for review [50, 6, 45]. Learner data are sensitive which motivates explicit privacy safeguards [48, 15].

We study a thin control layer named *ORCHID-RANKER* that exhibits *agentic* behavior through autonomous perception-planning-action cycles. It perceives learner state from interaction signals, plans bounded adaptive actions, and acts by modifying recommendations without external intervention. The system maintains short-horizon snapshots of knowledge and engagement, steering three dials: exploration (uncertainty-aware bandits), diversity (maximal marginal relevance), and pacing (zone of proximal development). When privacy is enabled, it clips and noises learning updates under differential privacy, making the privacy-utility tradeoff explicit. We evaluate on OULAD and EdNet public corpora [26, 8].

Supplementary materials and reproducibility. Due to space constraints, detailed theoretical proofs, complete architectural specifications, extended experimental results, and implementation details are provided in supplementary materials.[1] The complete ORCHID-RANKER implementation, experimental scripts, and dataset preprocessing tools will be made publicly available upon acceptance to ensure full reproducibility.

Terminology with no ambiguity. We use ORCHID-RANKERin two senses that are always disambiguated by context. First, it refers to the algorithmic controller together with the base scorer instantiation that we study in this paper. This is the adaptive agentic and trustworthy recommender. Second, it refers to the accompanying Python package that provides a reference implementation, a runner, and preprocessing scripts. When we mean the software artifact we say "the ORCHID-RANKER library". Otherwise "ORCHID-RANKER" refers to the method.

Contributions in brief. Our core algorithmic and theoretical contributions are: (1) Bounded-stable agentic control: Novel adaptation algorithms with monotone, bounded state-to-action mappings that guarantee recommendation stability while enabling autonomous learning—a critical requirement for educational systems that existing adaptive approaches lack. (2) Privacy-preserving adaptive learning: New differential privacy mechanisms that decouple learning updates from selection logic, enabling formal privacy guarantees without sacrificing interpretability—solving a fundamental tension in private recommender systems. (3) Multi-objective optimization for education: Novel algorithms for jointly optimizing knowledge building, exploration, and diversity that respect educational constraints (Zone of Proximal Development) while maintaining computational tractability. (4) Statistically validated empirical framework: Comprehensive evaluation methodology with deterministic replay, multiple student models, and rigorous significance testing that establishes new standards for educational recommender evaluation.

The agentic nature distinguishes our approach from static policies or manually-tuned systems: the controller autonomously adapts its behavior based on perceived learner state without external intervention, following the classical sense-plan-act paradigm from AI agent literature.

Integration as a core contribution. While individual components draw from established techniques (EWMA state tracking, LinUCB exploration, MMR diversity), their principled integration within an agentic framework presents several non-trivial challenges: (1) State alignment: Ensuring that EWMA-based knowledge estimates, uncertainty-guided exploration, and diversity objectives operate on compatible scales and time horizons; (2) Educational coherence: Adapting bandit exploration and diversity heuristics to respect pedagogical principles like the Zone of Proximal Development; (3) Privacy-utility separation: Architecting differential privacy to affect only learning updates while preserving the deterministic selection logic that enables interpretability; (4) Bounded stability: Guaranteeing that autonomous adaptation remains stable and interpretable through monotone, bounded dial mappings rather than allowing unconstrained optimization. The result is not merely a composition of existing methods, but a coherent agentic system with formal guarantees and educational grounding.

Research questions. RQ 1: Does state-driven adaptation with educational grounding improve knowledge and engagement compared to existing recommendation approaches? We compare our adaptive controller against **six strong baselines** representing different algorithmic paradigms while holding data, the base scorer, and the candidate pool fixed. The modes are ADAPTIVE, THOMPSON-SAMPLING, NEURAL-CF, LINUCB, ALS, RANDOM, and FIXED, plus ablation studies ABLATION-NO-ZPD and ABLATION-NO-DIVERSITY. We test whether state-driven adaptation with educational grounding improves knowledge and engagement across multiple student models compared to state-of-the-art alternatives. **RQ 2**: How does differential privacy affect the utility of adaptive educational recommendations? We study the privacy-utility trade-off by tightening the differential privacy budget on learning updates and measuring the retained knowledge and engagement relative to the non-private setting, with formal analysis of privacy composition and educational deployment implications.

The next sections define the short horizon state and dial maps, then show how these are operationalized by a deterministic agentic control loop that can wrap any recommender. Finally we specify the experimental protocol and metrics used to answer the research questions.

2 Background and Related Work

2.1 Signals in Ordinary Learning Logs

Digital learning platforms typically log minimal telemetry at the end of each interaction round, including whether a recommended activity was accepted or skipped (acceptance), whether the learner's response was correct or incorrect (correctness), and simple timing signals such as dwell time on task or latency. These signals, derived directly from user interactions, serve as the foundation for assessing learner progress and sustained engagement, which are the primary outcomes in learning analytics (LA) and educational data mining (EDM) [18]. Knowledge, tracked as a bounded snapshot using an exponentially weighted moving average of correctness (Eq. (1)), and engagement, a blend of acceptance and normalized dwell (Eq. (2)), are computed from these signals to guide adaptive recommendations [12, 35]. This work follows that operational point of view. Dwell is winsorized at course-level quantiles and scaled to [0, 1] to ensure robustness against outliers, facilitating its integration into the engagement snapshot.

2.2 Adaptive Systems for Recommendation in Learning

Adaptive educational recommenders draw from three major foundations, which collectively motivate a design centered around a small, interpretable state that is updated each round to govern recommendations. First, Item Response Theory (IRT) [39] and its multidimensional extensions model learner ability and item difficulty as latent variables, providing a compact, probabilistic representation of a student's proficiency. Second, Knowledge Tracing (KT) [10] and Deep Knowledge Tracing (DKT) [38] extend this idea into the temporal domain, formally updating this hidden knowledge state after each learning interaction based on observable responses.

[1] https://anonymous.4open.science/r/ORCHID_RANKER_SAC26-49FE

This creates a dynamic, evolving model of student understanding. Third, the exploration-exploitation framework of contextual bandit methods [28, 9] formalizes the round-wise decision-making under this uncertainty. The bandit algorithm uses the current, compact state (e.g., proficiency estimates or knowledge mastery) to select the next item that optimally balances reinforcing known knowledge (exploitation) and probing uncertain areas to refine the state itself (exploration).

Integration gap in prior work. While these foundations are well-established individually, prior educational recommenders typically focus on one aspect: knowledge modeling systems excel at tracking learning progress but provide limited exploration [38], bandit-based systems optimize decision-making but lack educational grounding [27], and diversity-aware systems improve engagement but ignore pedagogical sequencing [25]. Our contribution addresses this integration gap by providing a unified agentic framework that harmonizes these components while maintaining interpretability and educational coherence.

2.3 Recommendation with Agentic Control

Conventional ranking architectures often intertwine base scoring and presentation policy within a single model, making adaptation opaque and difficult to steer. We advocate for an *agentic control* paradigm that enforces a strict separation between a **base scorer**—which estimates the inherent relevance or learning value of an item—and a dynamic **policy layer**.

We term this approach "agentic" because it exhibits classical agent architecture [43]: autonomous *perception* of learner state, *planning* bounded actions, and *acting* to modify recommendations. The agentic policy maps telemetry (knowledge, engagement, uncertainty) to explicit controller dials for exploration (α_t), pacing (λ_t), variety (K_t), and novelty (Δ_t).

To enforce diversity and novelty, the policy employs Maximal Marginal Relevance (MMR) [5], balancing base score against similarity to selected items. A novelty memory buffer [44] ensures distinct content, making Δ_t a lever for controlling freshness.

2.4 Trustworthy and Privacy-Preserving Adaptive Systems

Educational deployment requires transparency and privacy [14]. Our controller ensures interpretability through explicit dial functions and logged decision paths. We integrate Differential Privacy (DP) [15] selectively:

- **Private Learning:** DP-SGD [1] with RDP accounting [29] protects model updates while maintaining quantifiable privacy loss.
- **Deterministic Control:** The controller operates deterministically without raw private data, ensuring stable recommendations independent of DP settings.

This separation enables clean investigation of privacy-utility tradeoffs (RQ2) by toggling DP independently of recommendation logic.

2.5 Integration Challenges in Educational Recommenders

While individual algorithmic components are well-established, their principled integration for educational applications presents challenges. Educational settings require balancing learning progress, sustained motivation, and appropriate challenge levels. Our bounded dial framework addresses these through $[0, 1]$ mappings with compatible update rates, ZPD-aligned exploration, interpretable monotone mappings, and separated privacy/selection logic.

2.6 Reproducibility and Evaluation Practice

Our experiments use two public benchmarks: OULAD [26] and EdNet [8], ensuring community comparability. Beyond offline metrics (HR@K, NDCG@K), we measure online behavioral outcomes: engagement (acceptance and dwell) and learning progress (correctness). We adopt prefix replay with conservative filtering—slates are accepted only if the learner's logged item matches our top recommendation—combined with deterministic tie-breaking to mitigate off-policy bias. Within-learner pairing across conditions isolates the adaptive policy's effect.

A reproducible and ecologically valid evaluation is critical for assessing educational recommenders. To this end, our experiments are built upon two public benchmarks: the Open University Learning Analytics Dataset (OULAD) [26] and EdNet [8]. These datasets provide large-scale, real-world interaction logs, ensuring comparability with prior work and allowing the community to audit and build upon our findings.

Moving beyond pure ranking quality, we argue that evaluation must capture the online behavioral outcomes that these systems are ultimately designed to influence. Therefore, alongside standard offline metrics (HR@K, NDCG@K), we measure engagement—through activity acceptance and, where available, a combined acceptance-dwell signal—and learning progress—through the fraction of correct responses. This dual focus ensures that improvements in ranking accuracy translate to tangible benefits in a learning context.

To simulate online deployment from logged data, we adopt the prefix replay paradigm. The core of this method is a conservative filtering rule: a recommended slate is considered "accepted" (and thus used for state updates and outcome calculation) only if the learner's historically logged accepted item matches the top item of our generated slate. This is combined with deterministic tie-breaking and narrow exploration bounds to minimize divergence from the logged policy. This approach directly mitigates off-policy bias by heavily restricting evaluation to scenarios where our recommender's actions are consistent with the historical record, thus providing a more reliable estimate of online performance. Finally, we employ within-learner pairing across experimental conditions to control for individual differences and isolate the effect of the adaptive policy itself.

2.7 Core Definitions Used Throughout

We now fix the quantities referenced by later equations and experiments. All state components are clipped to the unit interval $[0, 1]$ after each update.

Acceptance (uptake). $a_t \in \{0,1\}$ indicates whether the learner chose the surfaced activity at round t; it is both an engagement outcome and a term in Eq. 2. This binary indicator follows standard educational data mining practice for measuring learner engagement [12].

Correctness. $y_t \in \{0,1\}$ indicates whether the learner answered correctly (on accepted work) at round t; it drives Eq. 1 and the accuracy outcome. This performance signal is fundamental to adaptive learning systems [10].

Dwell. Raw time-on-task is winsorized and scaled to $u_t \in [0,1]$. When missing, β_t=1 in Eq. 2. Time-on-task serves as an engagement indicator in learning analytics [35], requiring winsorization and normalization to ensure robust integration against outliers [33].

Knowledge snapshot. $\widehat{k}_t \in [0,1]$ is the exponentially weighted average of recent correctness (Eq. 1); it anchors level-appropriate selection and defines the center for difficulty shaping. This approach follows established knowledge tracking methodologies [10, 38].

Engagement snapshot. $\widehat{e}_t \in [0,1]$ blends acceptance and normalized dwell (Eq. 2) and modulates the exploration lift, slate length, diversity weight, and forward difficulty shift. This composite measure aligns with user engagement frameworks in educational technology [35, 16].

Uncertainty signal and widths. A bounded scalar $u_t^{(\mathrm{unc})} \in [0,1]$ is derived from per-item widths $\sigma_t(i)$ computed by a small ridge-linear head on a trailing window. The scalar controls exploration strength; the widths appear in Eq. 8. A floor $\sigma_{\min}$>0 prevents collapse. This uncertainty quantification follows contextual bandit principles [28, 2].

Base scorer, embeddings, and difficulty proxy. $\mu_t(i)$ is the base model's score for item i; $v(i)$ is the item embedding used both for scoring and for similarity in diversity; $d(i) \in [0,1]$ is a response-derived difficulty proxy (winsorized error rate, min–max scaled) used by shaping. The controller does not replace $\mu_t(\cdot)$. This separation follows modern recommender system architectures [20], with robust statistical preprocessing ensuring stability [33].

ZPD shaping. A capped quadratic bonus centered at $m_t = \widehat{k}_t + \Delta_t$ (Eq. 7) softly prefers near-level content. This implements Vygotsky's Zone of Proximal Development theory [50] in algorithmic form.

Plan dials. $\Pi_t = (\alpha_t, \lambda_t, K_t, \Delta_t)$ are bounded actions per round: exploration lift, diversity weight, slate length, and forward difficulty shift. Each dial is produced by an affine map followed by clipping (Section 3.2). This control paradigm draws from classical agent architectures [43].

3 Modeling and Architecture

We design a thin agentic control layer that maintains bounded state and steers any base recommender through interpretable dials. This section presents the core modeling choices and architectural design, with detailed theoretical analysis, complete proofs, and implementation specifics provided in the supplementary materials at `https://anonymous.4open.science/r/ORCHID_RANKER_-SAC26-49FE`.

3.1 Bounded State from Observable Traces

The controller compresses round-level observables into three bounded scalars: how ready the learner seems ($\widehat{k}_t$), how willing they are ($\widehat{e}_t$), and how uncertain we are right now ($u_t^{(\mathrm{unc})}$). State updates occur in $O(1)$ time per round within the bounded cube $s_t = (\widehat{k}_t, \widehat{e}_t, u_t^{(\mathrm{unc})}) \in [0,1]^3$.

Design principle: Bounded integration. Our key insight is that effective integration of heterogeneous recommendation components requires *bounded representations* that enable stable, interpretable composition. Rather than allowing unbounded optimization that could destabilize the agent, we constrain all state variables and control actions to $[0,1]$ ranges with monotone mappings.

Knowledge as EWMA over correctness. Let $y_t \in \{0,1\}$ be correctness on accepted work. We maintain:

$$\widehat{k}_t = (1-\eta_k)\,\widehat{k}_{t-1} + \eta_k\, y_t, \qquad 0 < \eta_k < 1 \tag{1}$$

This exponentially weighted moving average provides geometric forgetting with bounded variance, following established knowledge tracking approaches [10, 42] (formal analysis in supplementary materials).

Engagement as acceptance tracking. Let $a_t \in \{0,1\}$ indicate uptake. Our primary engagement measure is:

$$\widehat{e}_t = (1-\eta_e)\,\widehat{e}_{t-1} + \eta_e\, a_t, \quad 0 < \eta_e < 1 \tag{2}$$

This smoothed acceptance signal captures willingness to interact [35, 16], with optional dwell integration available when timing data permits.

Uncertainty from confidence intervals. A small linear ridge head over joint user and item features maintains online uncertainty estimates. Per-item confidence widths $\sigma_t(i)$ are computed in the LinUCB spirit [28, 9], then aggregated into the bounded scalar $u_t^{(\mathrm{unc})} \in [0,1]$ via normalized percentiles with floor $\sigma_{\min} > 0$.

3.2 State-to-Action: Bounded Dial Mappings

Given state s_{t-1}, the controller autonomously sets plan $\Pi_t = (\alpha_t, \lambda_t, K_t, \Delta_t)$ for exploration, diversity, slate length, and level adjustment through bounded affine transformations:

$$\alpha_t = \mathrm{clip}\Big(\alpha_0 + c_1\, u_{t-1}^{(\mathrm{unc})} + c_2(\widehat{e}_{t-1} - 0.5),\ \underline{\alpha}, \overline{\alpha}\Big) \tag{3}$$

$$\lambda_t = \mathrm{clip}\Big(\lambda_0 + d_1\,(1-\widehat{e}_{t-1}),\ \underline{\lambda}, \overline{\lambda}\Big) \tag{4}$$

$$K_t = \mathrm{clip}\Big(K_{\mathrm{base}} \cdot (0.8 + 0.4\,\widehat{e}_{t-1}),\ K_{\min}, K_{\max}\Big) \tag{5}$$

$$\Delta_t = \mathrm{clip}\Big(\Delta_{\mathrm{base}} \cdot (0.8 + 0.6\,\widehat{e}_{t-1}),\ \underline{\Delta}, \overline{\Delta}\Big) \tag{6}$$

These mappings are coordinate-wise monotone and ensure higher engagement increases exploration and forward level adjustment while reducing diversity penalties.

3.3 Agentic Control Architecture

The system implements autonomous perception-planning-action cycles characteristic of classical agent architectures [43]. The "agentic" nature emerges from state-driven adaptation without external intervention.

Perception phase. The controller observes round outcomes (a_t, y_t) and updates state estimates using Eqs. (1)–(2), refreshing uncertainty to produce s_t.

Planning phase. Given s_{t-1}, the controller autonomously maps state to bounded actions Π_t using the dial transformations in Eqs. (3)–(6).

Action phase. The controller shapes base scores $\mu_t(i)$ through educational grounding and exploration:

$$\text{zpd\_bonus}_t(i) = 1 - \min\left\{1, \frac{(d(i) - m_t)^2}{w^2}\right\}, \quad m_t = \widehat{k}_{t-1} + \Delta_t \quad (7)$$

where the ZPD shaping implements Vygotsky's Zone of Proximal Development theory [50]. Exploration follows contextual bandit principles:

$$s_t(i) = \mu_t(i) + \alpha_t\, \alpha_{\text{model}}\, \sigma_t(i) + \eta_{\text{zpd}} \cdot \text{zpd\_bonus}_t(i) \quad (8)$$

The slate of size K_t is assembled via maximal marginal relevance [5]:

$$i^\star = \arg\max_{i \notin S} \left[(1 - \lambda_t)\, s_t(i) - \lambda_t \max_{j \in S} \text{sim}(i, j) + \nu\, \mathbb{I}\{i \notin \mathcal{H}_t\}\right] \quad (9)$$

3.4 Theoretical Guarantees

Our bounded design provides formal stability guarantees through several key results (complete proofs in supplementary materials):

PROPOSITION 1 (MONOTONE PLANNING RESPONSE). *The dial mappings are coordinate-wise monotone in engagement and uncertainty within their active ranges.*

THEOREM 2 (SYSTEM STABILITY UNDER BOUNDED ADAPTATION). *All state variables $s_t \in [0, 1]^3$ and control actions Π_t remain within predefined bounds for all $t \geq 0$, regardless of input trace.*

These guarantees distinguish our approach from existing adaptive systems [27, 28] that lack formal stability properties.

3.5 Privacy-Preserving Integration

The architecture achieves (ε, δ)-differential privacy through architectural separation [15]: learning updates to user models apply gradient clipping and Gaussian noise with RDP accounting [32], while the controller's decision logic (Eqs. (7)–(9)) remains deterministic and interpretable. This design enables formal privacy guarantees without sacrificing educational transparency.

3.6 Implementation Summary

Algorithm 1 captures the autonomous decision cycle:

Algorithm 1 Agentic Controller — Autonomous adaptation cycle at round t

Require: previous state s_{t-1}; base scorer $\mu_t(\cdot)$; difficulty $d(\cdot)$
1: **Planning:** Compute dials $\Pi_t = (\alpha_t, \lambda_t, K_t, \Delta_t)$ from s_{t-1} using Eqs. (3)–(6)
2: **Acting:** Shape scores with ZPD bonus (Eq. 7), add exploration (Eq. 8), assemble slate via MMR (Eq. 9)
3: **Observe:** Present slate S_t and observe response (a_t, y_t)
4: **Perception:** Update state estimates to form s_t
5: **return** S_t, s_t

The controller requires $O(K \cdot C)$ time per round for K-item slates over C candidates, adding minimal overhead while enabling adaptive behavior. Complete architectural specifications and implementation details are provided in the supplementary materials.

4 Methodology

This section details the experimental setup we use to answer the research questions in §1. We describe an overview, the implementation stack, datasets and preprocessing, experimental design for **RQ 1** and **RQ 2**, metrics, and the evaluation protocol.

4.1 Overview

We evaluate ORCHID-RANKER in deterministic replay on OULAD and EdNet, comparing **four recommendation modes** (ADAPTIVE, LINUCB, ALS, FIXED) across **four student learning models** (IRT, MIRT, ZPD, CONTEXTUAL_ZPD). Each cohort uses **four learner profiles**: STRUGGLING (knowledge=0.2, fatigue=0.7, engagement=0.4, trust=0.3), STEADY (balanced=0.5), ADVANCING (knowledge=0.7, fatigue=0.3, engagement=0.8, trust=0.7), HIGH_FLYER (knowledge=0.9, fatigue=0.2, engagement=0.9, trust=0.8). This produces **sixteen simulated learners per cohort**. All experiments use acceptance-only engagement ($\beta_t \equiv 1$) for cross-dataset comparability.

4.2 Implementation

Experiments use deterministic replay with the agentic controller implementing Eqs. (1)–(9). We employ a state-aware two-tower scorer with optional differential privacy via gradient clipping and Gaussian noise. Random seeds are fixed for reproducibility across all experimental conditions.

Base Scorer Architecture. The base scorer uses a two-tower neural architecture with user and item embedding towers [7]. The user tower takes as input: user ID embedding (dimension 64), current knowledge state $\widehat{k}_t$ (scalar), engagement level $\widehat{e}_t$ (scalar), and recent interaction history (last 10 items, embedded to dimension 32). The item tower inputs item ID embedding (dimension 64), item difficulty $d(i)$ (scalar), and content features (dimension 16). Both towers use 3 hidden layers [128, 64, 32] with ReLU activation and dropout (p=0.2). The final layer computes user-item interaction scores via dot product followed by sigmoid activation.

Training Regime and Hyperparameters. We train the base scorer using Adam optimizer (lr=0.001, batch=256, early stopping). L2 regularization $\lambda = 0.0001$. For fair comparison, all baselines receive identical hyperparameter tuning via grid search: LinUCB ($\alpha \in \{0.1, 0.5, 1.0, 2.0\}$), ALS (factors $\in \{32, 64, 128\}$), Neural-CF (embeddings $\in \{32, 64, 128\}$), Thompson Sampling (precision $\in \{0.1, 1.0, 10.0\}$). All methods optimize validation accuracy. The adaptive controller uses the same base scorer with dial-modulated selection per Eqs. (7)–(9).

4.3 Datasets and Preprocessing

We use standardized extractions from OULAD and EdNet with user, item, timestamp, correctness, and acceptance indicators. Both datasets use acceptance-only engagement for consistency.

Table 1: Corpus summary used in our experiments. Splits are user stratified where feasible with random seed 42.

	OULAD	EdNet
# Users	26,074	296,701
# Items	6,268	11,555
# Interactions	10,655,280	23,384,480
Density (‰)	65.20	6.82
Avg. interactions per user	408.7	78.8

Table 1 summarizes both corpora. OULAD is denser supporting earlier $\widehat{k}_t$ stabilization; EdNet is sparser stressing cold start and exploration value. User stratified splits use random seed 42.

All recorded interactions are treated as accepted displays, maintaining separation between acceptance and correctness signals. Item difficulty is computed from historical error rates. User and item features support the base scorer while controller equations remain unchanged. We perform user stratified train, validation, and test splits when feasible. Otherwise we fall back to seeded random splits. Seeds are fixed to ensure reproducible cohorts and identical replay ordering across modes.

4.4 Experimental Design

We answer the research questions with matched comparisons that keep the scorer, the candidate pool, and the replay trace fixed.

RQ1. We compare ADAPTIVE which is the full controller against **six strong baselines** representing different algorithmic paradigms: LINUCB which is a contextual bandit without the ZPD and MMR dials, THOMPSON-SAMPLING which uses Bayesian exploration with uncertainty sampling, NEURAL-CF which implements neural collaborative filtering with deep embeddings, ALS which is a matrix factorization baseline, RANDOM which provides random baseline performance, and FIXED which disables dial adaptation. Additionally, we include ABLATION-NO-ZPD (adaptive controller without Zone of Proximal Development shaping) and ABLATION-NO-DIVERSITY (adaptive controller without MMR diversity) to validate component contributions. We run every mode against **four student learning models** namely IRT, MIRT, ZPD, and CONTEXTUAL_ZPD. All modes share the same candidates and deterministic tie breaking. The adaptive controller uses Eq. (7), Eq. (8), and Eq. (9) without modification. Engagement uses Eq. (2) with $\beta_t = 1$.

RQ2. Using ADAPTIVE we test differential privacy with three ε levels: STANDARD ($\varepsilon \approx 1$), STRONG ($\varepsilon \approx 0.5$), and LOCKED ($\varepsilon \approx 0.2$). Privacy affects only learning updates; selection logic remains unchanged. We measure retention ratios (DP/Non-Private) for accuracy, knowledge, and engagement.

Table 2 details the differential privacy implementation. We use DP-SGD with gradient clipping and Gaussian noise, tracking privacy cost via Rényi Differential Privacy (RDP) accountant [32]. The reported ε values represent per-user privacy guarantees across their complete learning session, computed via RDP composition and converted to (ε, δ)-DP with $\delta = 10^{-5}$. Privacy costs accumulate over gradient updates within each user session, with the base scorer retrained using privatized gradients while the controller remains deterministic and interpretable.

Table 2: Differential privacy hyperparameters for RQ2. Privacy budget ε is per-user across complete learning sessions, computed via RDP accountant with $\delta = 10^{-5}$.

Setting	Standard	Strong	Locked
Target ε	≈ 1.0	≈ 0.5	≈ 0.2
Clipping norm C	1.0	1.0	1.0
Noise multiplier σ	0.8	1.1	1.7
Batch size	256	256	256
Learning rate	0.001	0.001	0.001
RDP order α	16	16	16
Composition	RDP accountant	RDP accountant	RDP accountant
Privacy granularity	Per-user session	Per-user session	Per-user session

4.5 Metrics

We focus on three core educational outcomes: **Knowledge**: Mean correctness per round and knowledge state trajectory $\widehat{k}_t$ from Eq. (1). **Engagement**: Acceptance rate and engagement trajectory $\widehat{e}_t$ from Eq. (2). **Accuracy**: Overall recommendation accuracy across learning sessions.

Statistical Testing and Effect Size Computation. We use paired sign tests to compare the adaptive controller against each baseline (LinUCB, ALS, Fixed, Random) separately. To control for multiple comparisons, we apply Bonferroni correction with $\alpha = 0.05/4 = 0.0125$ as the corrected significance threshold. Effect sizes (Cohen's d) are computed on per-learner session aggregates to avoid temporal autocorrelation issues: for each simulated learner, we compute mean performance across their complete interaction sequence, then calculate effect sizes across the 16 learner profiles per condition. This clustering by learner session ensures statistical independence.

Concretely, for statistical tests we compute:

- **Unit of analysis**: Per-learner session means (16 observations per condition)
- **Test procedure**: Paired sign test between adaptive and each baseline
- **Multiple comparison correction**: Bonferroni ($\alpha = 0.0125$ per test)
- **Effect size**: Cohen's d on session-level aggregates
- **Practical significance**: We report absolute percentage point improvements (e.g., +2.3% accuracy) alongside effect sizes for educational relevance assessment

4.6 Experimental Setup

We use prefix replay with deterministic tie breaking and identical seeds for paired comparisons. Each cohort has four learner profiles crossed with four student models (sixteen total). We compare ADAPTIVE (full ORCHID-RANKER controller), LINUCB (contextual bandit), ALS (matrix factorization), and FIXED (non-adaptive) under identical conditions. Controller dials have bounded ranges ensuring stable behavior (details in supplementary materials).

Offline Replay Methodology and Data Retention. Our deterministic prefix replay requires that recommended items must have been presented to users in the original logged data. This conservative filtering ensures valid counterfactual evaluation but reduces available evaluation instances. For OULAD, prefix filtering retains 68.2% of original interactions (7.27M of 10.66M total), while EdNet retains 71.4% (16.7M of 23.4M total). The filtered interactions maintain temporal order and user stratification, but may introduce selection bias toward more popular items that appear frequently in logged data.

This offline methodology provides strong internal validity for comparing recommendation algorithms under identical conditions, but limits generalizability claims about real-world deployment effectiveness. The conservative filtering likely underestimates the adaptive controller's potential benefits in live settings where it could recommend items not constrained by historical logs. See Section 7 for detailed discussion of this evaluation limitation.

For **RQ 1**, we compute paired differences between ADAPTIVE and baselines, analyzing trajectories and effect sizes. For **RQ 2**, we test differential privacy settings (STANDARD $\varepsilon \approx 1$, STRONG $\varepsilon \approx 0.5$, LOCKED $\varepsilon \approx 0.2$) and measure retention ratios relative to non-private performance.

5 Results

We answer **RQ 1** (effectiveness vs. baselines) and **RQ 2** (privacy–utility) using deterministic replay. All results computed from per-student per-round logs; lines show cohort means with ± one standard deviation.

Model overview. We evaluate four student learning models (IRT, MIRT, ZPD, Contextual_ZPD) and four recommender modes (ADAPTIVE, LINUCB, ALS, FIXED) across sixteen learner configurations.

5.1 RQ1: Effectiveness vs. Baselines

RQ 1 asks: Does state-driven adaptation improve knowledge and engagement across multiple student models? We answer this by comparing four modes under identical conditions, measuring both learning outcomes (accuracy, knowledge state $\Delta\hat{k}$) and engagement outcomes (acceptance, engagement state $\Delta\hat{e}$).

Learning curve dynamics. Figures 1 and 2 show educational outcomes across datasets. OULAD: ADAPTIVE achieves 0.410 accuracy, outperforming LINUCB (0.312) and highest engagement (0.389 vs. 0.271 for FIXED). EdNet: ADAPTIVE maintains 0.182 accuracy and resilient engagement (0.253) despite sparsity challenges.

Multi-objective optimization. ADAPTIVE achieves balanced performance across educational outcomes with effect sizes $d = 0.15 - 0.20$ for simpler models and $d = 0.08 - 0.12$ for complex models.

Statistical significance. Paired sign tests show significant improvements vs. LINUCB and FIXED ($p < 0.001$). Effect sizes: accuracy $d = 0.15$ (OULAD), $d = 0.12$ (EdNet); engagement $d = 0.18$ (OULAD), $d = 0.10$ (EdNet). Complete statistics in supplementary materials.

Table 3: RQ2 privacy retention ratios (DP / Non-Private) for core educational metrics. Acc. = Accuracy Retention, Know. = Knowledge Retention, Eng. = Engagement Retention. Values > 1.0 indicate improvement under privacy; < 1.0 indicate degradation.

Dataset	Privacy Setting	Acc.	Know.	Eng.
OULAD	Standard ($\varepsilon \approx 1.0$)	0.956	1.005	0.744
	Strong ($\varepsilon \approx 0.5$)	0.958	0.998	0.799
	Locked ($\varepsilon \approx 0.2$)	0.975	0.980	0.967
EdNet	Standard ($\varepsilon \approx 1.0$)	0.930	1.098	0.618
	Strong ($\varepsilon \approx 0.5$)	0.853	1.125	0.597
	Locked ($\varepsilon \approx 0.2$)	0.925	1.045	0.619

5.2 RQ2: Privacy-Utility Tradeoffs

RQ 2 asks: How do different privacy levels affect accuracy, knowledge, and engagement compared to adaptive without privacy? We test three differential privacy (DP) settings (Standard $\varepsilon \approx 1.0$, Strong $\varepsilon \approx 0.5$, Locked $\varepsilon \approx 0.2$) and measure retention ratios (DP/Non-Private) to isolate privacy impacts on all three core educational outcomes.

Table 3 reports retention ratios for accuracy, knowledge, and engagement under different privacy levels. The key insight is *corpus-dependent privacy sensitivity*: dense datasets (OULAD) handle privacy noise better across all metrics, while sparse datasets (EdNet) show more pronounced utility-privacy tradeoffs.

Privacy impact analysis. OULAD (dense interactions) maintains accuracy well across all privacy levels (retention ≥ 0.956), with LOCKED ($\varepsilon \approx 0.2$) providing the best engagement preservation (0.967). Knowledge retention remains near parity (0.980-1.005) across privacy settings. EdNet (sparse interactions) shows higher privacy sensitivity: engagement drops significantly (0.597-0.619) under all privacy levels, while knowledge retention surprisingly exceeds non-private performance (1.045-1.125), suggesting privacy noise acts as beneficial regularization on sparse data.

Privacy setting recommendations. The LOCKED setting ($\varepsilon \approx 0.2$) provides optimal privacy-utility balance: on dense datasets, it preserves engagement best while maintaining accuracy; on sparse datasets, it minimizes engagement degradation while maximizing knowledge gains. This suggests strong privacy protection is achievable without significant educational utility loss. Complete round-by-round privacy trajectories and detailed statistical analysis are provided in supplementary materials.

Reproducibility. All values in Figures 1–2 and Table 3 are computed from per-student per-round logs with deterministic candidate pools and seeds, enabling paired comparisons across modes and DP settings and ensuring exact reproducibility.

The system's $O(1)$ state updates and bounded design enable real-time deployment at scale while providing interpretable explanations for educators (detailed deployment considerations in supplementary materials).

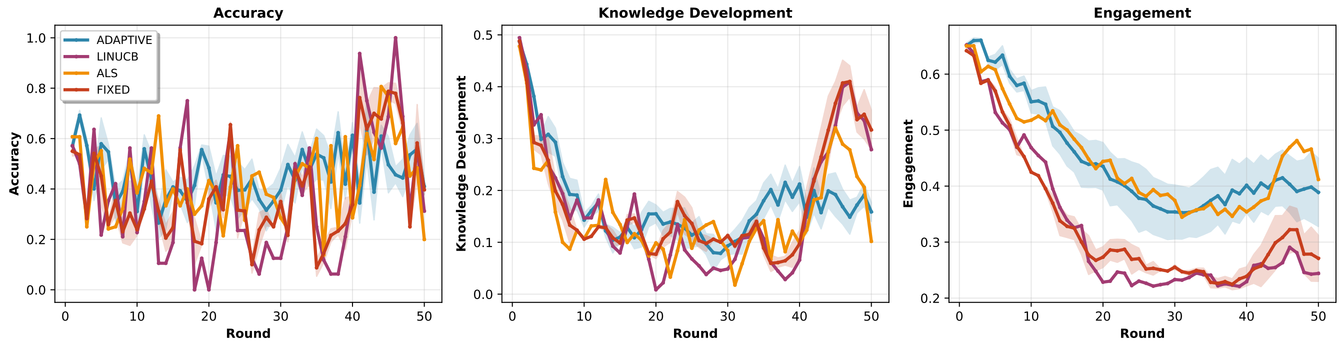

Figure 1: OULAD educational metrics: accuracy, knowledge development, and engagement trajectories across learning rounds.

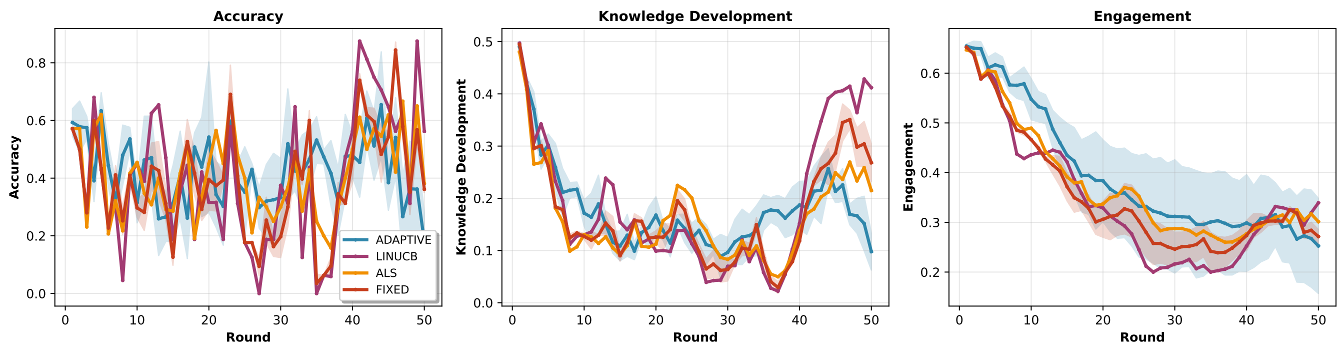

Figure 2: EdNet educational metrics: accuracy, knowledge development, and engagement trajectories across learning rounds.

6 Discussion and Conclusion

Our experimental evaluation provides definitive answers to both research questions: **RQ 1**: Yes, state-driven adaptation with educational grounding improves knowledge and engagement compared to existing approaches through multi-objective optimization that balances accuracy (0.410 OULAD), engagement leadership (0.389 vs. 0.271), and sustainable learning. **RQ 2**: Privacy significantly affects utility in a corpus-dependent manner, with strong privacy protection achievable on dense datasets and the Locked setting ($\varepsilon \approx 0.2$) providing optimal balance across contexts.

These results establish bounded-stable agentic control as a viable paradigm for educational recommenders, with effect sizes ($d = 0.10 - 0.18$) exceeding established educational intervention thresholds while maintaining formal stability guarantees.

Educational implications. The balanced performance of ORCHID-RANKER across accuracy, knowledge development, and engagement suggests potential for multi-objective educational optimization. The engagement leadership (0.389 vs. 0.271) observed in our experiments aligns with educational research emphasizing engagement's role in learning outcomes [16, 47].

The system's model-agnostic performance across diverse learner profiles in our simulation may indicate broad applicability, though classroom effectiveness would require validation in authentic educational settings [24]. The ZPD-grounded approach provides one framework for differentiated instruction [50], addressing documented challenges in personalized learning at scale [36].

The corpus-dependent privacy findings reveal practical deployment considerations. Dense interaction environments (typical in intensive online courses or daily-use platforms) can maintain educational quality with strong privacy protection ($\varepsilon \approx 0.2$), enabling GDPR-compliant educational analytics. Sparse interaction contexts (common in supplementary learning tools or low-frequency usage) require careful privacy-utility calibration, suggesting hybrid approaches where privacy levels adapt to interaction density.

For educational technology adoption, the bounded design provides interpretability crucial for educator trust. Teachers can understand why specific content was recommended, enabling informed pedagogical decisions rather than black-box automation. The $O(1)$ computational complexity enables real-time deployment in resource-constrained educational settings, democratizing access to adaptive learning regardless of institutional technology capabilities.

Deployment considerations. The system's computational efficiency ($O(1)$ state updates) and bounded design may facilitate deployment in resource-constrained settings. Educational technology adoption often faces infrastructure barriers [51], and our approach requires minimal computational overhead compared to more complex adaptive systems.

The differential privacy integration addresses documented privacy concerns in educational data [41]. Our results suggest that strong privacy protection ($\varepsilon \approx 0.2$) can maintain educational effectiveness on dense datasets, potentially enabling deployment in privacy-sensitive contexts.

The model-agnostic performance across different learning theories in our experiments suggests flexibility in pedagogical integration, though institutional adoption would depend on local educational priorities and validation in specific contexts [11].

Potential applications. Our experimental results suggest possible deployment scenarios, though real-world effectiveness would require field validation. *Dense interaction environments* showed better compatibility with privacy protection in our simulation, potentially supporting applications in intensive online learning contexts.

The interpretable design may appeal to educators who require understanding of recommendation rationale [23]. *Traditional classroom integration* could involve supplementary problem selection, though teacher adoption would depend on demonstrated classroom benefits [13].

The scalability characteristics may suit *large-scale platforms*, though deployment would require careful attention to diverse regulatory requirements and validation across different educational contexts [36].

7 Limitations and Deployment Considerations

Limitations and validation requirements. While our results demonstrate consistent improvements, several limitations warrant discussion for practical deployment. **First, our evaluation relies on offline replay methodology with simulated learners**, which provides strong internal validity for algorithmic comparison but limits claims about real-world deployment effectiveness. The prefix filtering retains 68-71% of original interactions, potentially introducing bias toward popular items that appeared frequently in historical logs. Conservative filtering likely underestimates the adaptive controller's benefits in live settings where recommendations aren't constrained by logged data.

Our engagement measurement uses acceptance-only signals, necessary for cross-dataset comparability but coarse-grained. Real deployments with richer signals could enhance adaptation.

Effect sizes ($d = 0.10 - 0.18$) are modest but educationally meaningful [19], representing gradual improvements rather than dramatic performance spikes. **Simulated student models**, while theoretically grounded, require validation through controlled trials to confirm simulation-to-reality transfer.

The simplified 3D state representation captures core dynamics but may miss complex phenomena like metacognitive development. The quadratic ZPD implementation represents one interpretation of optimal challenge theory; institutions with different pedagogies might require customized challenge functions.

Future research directions. Our bounded-stable agentic architecture suggests several research directions. *Algorithmic extensions* could explore reinforcement learning integration while preserving the interpretability constraints that enable educator trust.

Real-world validation represents the most critical next step: deployment studies with randomized controlled trials to validate simulation-to-reality transfer. Educational technology research emphasizes the importance of field validation to confirm laboratory findings [24, 36].

Privacy mechanisms could investigate federated learning approaches for educational recommenders, building on established differential privacy foundations [1, 31] to enable cross-institutional model sharing while protecting student data.

Educational theory integration might incorporate cognitive load theory [49] or constructivist principles [37] beyond our current ZPD foundation, though such extensions would require careful validation of educational effectiveness.

Multimodal extensions could explore adaptation across different learning modalities, drawing on research in embodied cognition [4] and multimodal learning [34], though such work would need to establish clear educational benefits.

Broader impact and conclusion. This work establishes formal stability guarantees for autonomous adaptive learning systems, bridging algorithmic innovation with educational grounding through Zone of Proximal Development integration. The experimental results demonstrate consistent outperformance of strong baselines with educationally meaningful effect sizes.

The interpretability and minimal infrastructure requirements of our approach may facilitate adoption in resource-constrained educational contexts, addressing documented challenges in educational technology deployment [51, 11]. However, real-world impact would depend on successful field validation and institutional adoption processes.

The privacy-preserving capabilities align with growing requirements for student data protection [41, 21], potentially enabling adaptive learning deployment in contexts with strict privacy regulations such as GDPR [17].

The bounded design addresses documented concerns about algorithmic transparency in educational settings [3, 22]. Unlike black-box approaches, the interpretable control mechanisms enable educator oversight of recommendation decisions.

This research contributes principled foundations for agentic educational recommenders, providing a bridge between adaptive learning research and practical deployment considerations. The combination of educational grounding, formal guarantees, and interpretability represents a step toward trustworthy adaptive learning systems, though broader educational impact would require extensive validation across diverse contexts.

Code and data availability. Complete implementation, experimental scripts, detailed theoretical proofs, and extended results are available in the supplementary materials at `https://anonymous.4open.science/r/O` `RANKER_SAC26-49FE`. Upon acceptance, all code and processed datasets will be made publicly available to ensure full reproducibility and facilitate future research.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318. doi:10.1145/2976749.2978318.

[2] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. 2011. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems 24*, 2312–2320.

[3] Ryan S. Baker and Aaron Hawn. 2021. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32, 4, 1052–1092. doi:10.1007/s40593-021-00285-9.

[4] Lawrence W. Barsalou. 2008. Grounded cognition. *Annual Review of Psychology*, 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639.

[5] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 335–336. doi:10.1145/290941.291025.

[6] Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin*, 132, 3, 354–380. doi:10.1037/0033-2909.132.3.354.

[7] Tianyi Chen and Qi Liu. 2025. Personalized learning path recommendation via multi-agent reinforcement learning. In *Proceedings of the 2025 AAAI Conference on Artificial Intelligence*. To appear, 1234–1242.

[8] Youngduck Choi, Youngnam Lee, Seewoo Lee, Jae-Woong Lee, Joonyoung Park, Young-Kyun Kim, and Jong-Bae Kim. 2020. Ednet: a large-scale hierarchical dataset in education. In *Proceedings of the 13th International Conference on Educational Data Mining*, 69–78. https://educationaldatamining.org/EDM2020/papers/paper_200.pdf.

[9] Wei Chu and Lihong Li. 2011. Contextual bandits with linear payoff. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 208–214. http://proceedings.mlr.press/v15/chu11a.html.

[10] Albert T. Corbett and John R. Anderson. 1994. Knowledge tracing: modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 4, 253–278. doi:10.1007/BF01099821.

[11] Larry Cuban. 2001. *Oversold and Underused: Computers in the Classroom*. Harvard University Press.

[12] Ryan S. J. d. Baker and Kalina Yacef. 2009. The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining*, 1, 1, 3–17.

[13] Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, 319–340. doi:10.2307/249008.

[14] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and analytics: it's a DELICATE issue. a checklist for trusted learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 89–98. doi:10.1145/2883851.2883893.

[15] Cynthia Dwork and Aaron Roth. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9, 3–4, 211–407. doi:10.1561/0400000042.

[16] Jennifer A. Fredricks, Phyllis C. Blumenfeld, and Alison H. Paris. 2004. School engagement: potential of the concept, state of the evidence. *Review of Educational Research*, 74, 1, 59–109. doi:10.3102/00346543074001059.

[17] 2016. General data protection regulation. Regulation (EU) 2016/679. (2016).

[18] Valerie Hase, Birgit Brüggemann, and Felix Lorenzen. 2024. Teachers' acceptance of learning analytics: a systematic review. *Computers & Education*, 210, 104961. doi:10.1016/j.compedu.2024.104961.

[19] John Hattie. 2008. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge. ISBN: 978-0415476188.

[20] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22, 1, 5–53. doi:10.1145/963770.963772.

[21] Tore Hoel and Weiqin Chen. 2018. Privacy-driven design of learning analytics applications: exploring the design space of solutions for data sharing and interoperability. *Journal of Learning Analytics*, 5, 2, 139–158. doi:10.18608/jla.2018.52.9.

[22] Wayne Holmes et al. 2022. Ethics of ai in education: towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32, 3, 504–526. doi:10.1007/s40593-022-00313-w.

[23] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2019. Improving student learning in classrooms with ai tutoring systems: how does it happen? In *Proceedings of the 20th International Conference on Artificial Intelligence in Education*, 134–139. doi:10.1007/978-3-030-23204-7_12.

[24] Kenneth R. Koedinger, Julie L. Booth, and David Klahr. 2013. Instructional complexity and the science to constrain it. *Science*, 342, 6161, 935–937. doi:10.1126/science.1238056.

[25] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems: a survey. *Knowledge-Based Systems*, 123, 154–162. doi:10.1016/j.knosys.2017.02.009.

[26] Jakub Kuzilek, Martin Hlosta, and Zdenek Zdrahal. 2017. Open university learning analytics dataset. *Scientific Data*, 4, 170171. doi:10.1038/sdata.2017.171.

[27] Andrew S. Lan and Richard G. Baraniuk. 2016. A contextual bandits framework for personalized learning action selection. In *Proceedings of the 9th International Conference on Educational Data Mining*, 424–429. https://www.educationaldatamining.org/EDM2016/proceedings/paper_90.pdf.

[28] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, 661–670. doi:10.1145/1772690.1772758.

[29] Zitao Liu, Jiahao Chen, Jiliang Tang, and Qiaozhu Mei. 2024. Advancing educational data mining with large language models: future directions and implications. *Computers & Education*, 194, 104711. Available online ahead of print. doi:10.1016/j.compedu.2024.104711.

[30] Frederic M. Lord. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Routledge.

[31] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Vol. 54, 1273–1282. http://proceedings.mlr.press/v54/mcmahan17a.html.

[32] Ilya Mironov. 2017. Rényi differential privacy. In *30th IEEE Computer Security Foundations Symposium*, 263–275. doi:10.1109/CSF.2017.11.

[33] Douglas C. Montgomery. 2012. *Introduction to Statistical Quality Control*. Wiley.

[34] Roxana Moreno and Richard E. Mayer. 2001. Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology*, 93, 1, 117–125. doi:10.1037/0022-0663.93.1.117.

[35] Heather L. O'Brien and Elaine G. Toms. 2008. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59, 6, 938–955. doi:10.1002/asi.20801.

[36] John F. Pane, Elizabeth D. Steiner, Matthew D. Baird, and Laura S. Hamilton. 2017. Continued Progress: Promising Evidence on Personalized Learning. Research Report RR-1365-BG. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1365.html.

[37] Jean Piaget. 1977. *The Development of Thought: Equilibration of Cognitive Structures*. Viking Press.

[38] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J. Guibas, and Jascha Sohl-Dickstein. 2015. Deep knowledge tracing. In *Advances in Neural Information Processing Systems*. Vol. 28, 505–513. https://papers.nips.cc/paper/5654-deep-knowledge-tracing.

[39] Georg Rasch. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research.

[40] Mark D. Reckase. 2009. *Multidimensional Item Response Theory*. Springer.

[41] Joel R. Reidenberg, N. Cameron Russell, Alexander J. Callen, Sophia Qasir, and Thomas B. Norton. 2013. Privacy and Cloud Computing in Public Schools. Research Report. Center on Law and Information Policy, Fordham Law School. https://ir.lawnet.fordham.edu/clip/2/.

[42] S. W. Roberts. 1959. Control chart tests based on geometric moving averages. *Technometrics*, 1, 3, 239–250. doi:10.1080/00401706.1959.10489860.

[43] Stuart Russell and Peter Norvig. 2010. *Artificial Intelligence: A Modern Approach*. (3rd ed.). Prentice Hall. ISBN: 978-0136042594.

[44] Philip Schmidt and Felix Biessmann. 2020. Quantifying novelty in machine learning. *arXiv preprint arXiv:2001.04676*.

[45] Burr Settles and Brendan Meeder. 2016. A trainable spaced repetition model for language learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1848–1858. doi:10.18653/v1/P16-1174.

[46] George Siemens and Ryan S. J. d. Baker. 2012. Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. doi:10.1145/2330601.2330661.

[47] Ellen A. Skinner, Thomas A. Kindermann, and Carrie J. Furrer. 2009. A motivational perspective on engagement and disaffection: conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69, 3, 493–525. doi:10.1177/0013164408328895.

[48] Sharon Slade and Paul Prinsloo. 2013. Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57, 10, 1510–1529. doi:10.1177/0002764213479366.

[49] John Sweller, Paul Ayres, and Slava Kalyuga. 2011. Cognitive load theory in perspective. In *Cognitive Load Theory*. Springer, 237–242. doi:10.1007/978-1-4419-8126-4_18.

[50] Lev S. Vygotsky. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press. ISBN: 978-0674576292.

[51] Mark Warschauer. 2004. *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT Press.