Multi-Modal Learning Analytics for Adaptive Instruction Using Behavioral, Cognitive, and Affective Signals

Les Endahti[1,*], Jalaludin[2]

[1,2]Dept. of Informatics Management, AMIK-YPAT Purwakarta, Indonesia

ABSTRACT

This paper investigates an adaptive instruction framework driven by multi-modal learning analytics that integrates behavioral logs, cognitive proxies, and opportunistic affective signals under realistic missingness. The study analyzes 286 learners across an 8-week course, comprising 1,942 learning sessions windowed at 10-second resolution. Modality retention reflected deployment constraints, with behavioral coverage at 100%, cognitive coverage at 88%, and affective coverage at 73%, alongside reliability indices of 0.92, 0.84, and 0.71, respectively. Predictive modeling results show that reliability-aware multi-modal fusion outperformed behavioral-only baselines for mastery and next-step correctness, improving AUC from 0.812 to 0.872 and Macro-F1 from 0.741 to 0.793. Gains increased with content difficulty, with hard units improving from 0.794 to 0.887 AUC and from 0.712 to 0.801 Macro-F1. In instructional impact evaluation, adaptive sequencing raised end-of-course mastery rate from 0.70 to 0.78 (absolute +0.08, relative +11.4%) and increased near-transfer performance from 71.6 to 76.9 (+5.3 points), while total learning time rose modestly from 212.4 to 219.7 minutes (+3.4%). Efficiency outcomes improved across devices, reducing median time-to-mastery from 26.5 to 23.1 minutes on desktop and from 28.7 to 26.6 minutes on mobile, with no subgroup exhibiting degraded outcomes. Attention diagnostics indicated state-conditional modality reliance, with affective weighting rising most in frustration states (0.24), and policy behavior aligned with pedagogical intent through increased worked-example selection under confusion and restrained pacing interventions. Overall, the results demonstrate that tri-modal evidence can improve both inference and learning outcomes while remaining robust to missingness and device heterogeneity.

Keywords Multi-Modal Learning Analytics, Adaptive Instruction, Learning State Inference, Cognitive Load, Affective Computing, Reliability-Aware Fusion, Contextual Bandits, Mastery Prediction

Submitted: 15 September 2025
Accepted: 20 October 2025
Published: 27 February 2026

*Corresponding author
Les Endahti,
endahti01@amikypat-purwakarta.ac.id

Additional Information and Declarations can be found on page 91

Introduction

Adaptive learning systems increasingly operate in hybrid learning ecologies where meaningful evidence of learning is distributed across platform interaction traces, physiological responses, and observable behavioral cues. Conventional learning analytics pipelines, dominated by clickstream and assessment logs, frequently under-specify latent learning processes such as cognitive effort, attentional regulation, and affective engagement. This creates an instrumentation gap between what adaptive instruction must optimize in real time and what single-channel data can reliably represent. Recent scholarship positions Multi-Modal Learning Analytics (MMLA) as a response to this gap by integrating heterogeneous signals and aligning them with pedagogical decision-making [1], [2].

The shift toward MMLA has produced substantive progress in conceptual

framing and tooling, particularly around linking multimodal evidence to learning design choices and instructional interventions [3]. However, scaling MMLA into adaptive systems remains constrained by practical and theoretical limits, including signal synchronization, sensor noise, missingness, and the interpretability of fused representations for educators. Systematic evidence on multimodal data fusion further indicates that many implementations remain prototype-oriented, with limited generalization across contexts and modest attention to reproducibility. These constraints motivate a more instruction-centered framing where fusion is treated as a reliability-weighted inference problem rather than an engineering afterthought [4].

A second gap concerns the operational definition of learner state. Multimodal signals can indicate fine-grained, time-varying mental states, yet most adaptive systems still act on coarse proxies such as quiz scores or aggregated engagement metrics. Prior work demonstrates that multimodal features can capture learner mental state dynamics during learning processes, suggesting that affective and behavioral micro-patterns carry information not available in outcome measures alone [5]. Nevertheless, translating these estimates into robust instructional actions is non-trivial because affective expressions, interaction styles, and cognitive effort can be context-dependent and culturally mediated. This paper positions learner state as a structured latent construct inferred from behavioral, cognitive, and affective signals under uncertainty.

Cognitive and affective sensing specifically offer leverage for adaptive personalization, but they also introduce modeling ambiguity. Studies that employ multimodal indicators to infer cognitive load during online problem-solving show promise, while also revealing sensitivity issues when discriminating subtle load changes or transferring models across tasks [6]. In parallel, the rapid expansion of affective computing for learning has systematized methods and taxonomies for emotion and engagement inference, yet the field still lacks stable conventions for mapping inferred affect into pedagogically defensible interventions at scale [7]. Addressing this requires an explicit linkage between multimodal inference, instructional intent, and measurable learning benefit rather than “prediction-only” model objectives.

The move from unimodal to multimodal evidence also increases ethical and governance stakes. Capturing video, audio, and physiological traces expands the attack surface for misuse and heightens risks of surveillance-like encroachment, especially when adaptivity is deployed without transparent boundaries. A systematic review of privacy in MMLA highlights tensions between pedagogical value and learner autonomy, indicating the need for privacy-by-design constraints within the analytics pipeline [8]. Complementary syntheses of privacy and data protection issues in learning analytics stress that technical safeguards must be coupled with clear purpose limitation and accountability, particularly when sensitive signals are collected continuously [9]. These requirements shape both system architecture and evaluation criteria in multimodal adaptivity.

Methodologically, multimodal adaptive instruction benefits from advances in learner modeling that represent learning as a temporal process rather than a static trait. Transformer-based knowledge tracing models demonstrate improved capacity to track evolving mastery under sparse and noisy observation regimes, strengthening the feasibility of using fine-grained temporal signals for

personalization [10]. Session-aware formulations further address a common realism gap by modeling within-session and across-session dynamics explicitly, aligning better with authentic learning episodes rather than treating all interactions as one uninterrupted sequence [11]. These modeling advances motivate multimodal architectures that couple state estimation with action policies, enabling adaptivity to respond to both mastery and regulation signals in real time.

Finally, adaptive instruction requires decision policies, not only predictions. Reinforcement learning has long been framed as a mechanism for inducing pedagogical strategies in adaptive and intelligent educational systems, but early work also underscored data-efficiency and safety constraints when learning policies from real learners [12]. More recent demonstrations of reinforcement learning for adaptive scheduling in large-scale online learning show feasibility for optimizing instructional sequences while balancing learning gains and workload [13]. Yet such policy learning must be constrained by algorithmic fairness considerations because multimodal sensing and optimization can amplify demographic or accessibility disparities if bias is not actively mitigated [14]. This paper addresses these gaps by proposing an instruction-first MMLA framework that integrates tri-modal signals into reliable state inference and constrained adaptive policy selection.

Literature Review

Learning analytics and educational data mining have established the methodological backbone for evidence-based personalization by formalizing how interaction traces, assessments, and contextual data can be converted into actionable indicators of learning and instruction. Foundational syntheses clarify core task families such as prediction, clustering, relationship mining, and discovery with models that support both descriptive understanding and intervention design [15]. Subsequent reviews emphasize that modern educational data pipelines increasingly prioritize scalability and interpretability because analytics outputs must be pedagogically legible to function as adaptive triggers [16].

A central limitation of log-centric adaptivity is that it often conflates observed behavior with underlying cognitive processing. Cognitive Load Theory explains why learners with identical click patterns can experience different learning outcomes due to working-memory constraints during problem solving [17]. Measurement research extends this foundation by formalizing multi-indicator approaches to estimate load, highlighting the need to separate task complexity from extraneous processing induced by poor instructional design [18]. These perspectives justify adding cognitive proxies to adaptive state estimation instead of relying on performance alone.

Affective processes further complicate adaptive decision-making because emotions shape attention allocation, strategy choice, and persistence. The control-value theory of achievement emotions provides a structured account of how appraisals of control and value generate distinct emotions that can either facilitate or hinder learning trajectories [19]. This theory is particularly relevant to adaptive instruction because it implies that the same task can trigger different affective responses depending on perceived competence and stakes, requiring adaptivity to attend to regulation as well as mastery.

Engagement is widely treated as a multi-dimensional construct that bridges observable participation with cognitive investment and emotional involvement. Evidence syntheses frame engagement as behavioral, emotional, and cognitive, and connect engagement to achievement and retention across learning contexts [20]. Complementarily, self-regulated learning scholarship consolidates models in which learners monitor goals, deploy strategies, and regulate motivation and affect, implying that adaptive systems should detect both performance signals and regulation breakdowns to personalize effectively [21].

Within affective learning research, dynamic accounts emphasize that productive learning often involves transient confusion and disequilibrium rather than continuous positive affect. Empirical modeling of affective trajectories during complex learning argues that confusion can function as a gateway state that promotes deeper processing when learners successfully resolve impasses, whereas unresolved confusion can cascade into frustration and disengagement [22]. This distinction matters for adaptive instruction because it motivates interventions that support resolution rather than simply suppressing negative affect.

Multi-modal fusion research provides the computational toolkit needed to combine heterogeneous signals without collapsing their semantics. Survey work in multimodal machine learning organizes fusion strategies into early, late, and hybrid approaches and highlights attention-based mechanisms for reliability-sensitive aggregation across modalities [23]. For adaptive learning, these taxonomies motivate architectures that treat modality contribution as conditional on context and data quality, enabling robust inference when sensors are noisy or intermittently absent.

Finally, real-world MMLA deployments must handle missingness and partial observability as routine rather than exceptional conditions. Practical guidance on multiple imputation by chained equations formalizes principled recovery of incomplete data under defensible assumptions, offering a pathway to stabilize training and evaluation when modalities drop out non-uniformly across devices and sessions [24]. This literature supports robustness-oriented pipelines where adaptivity does not fail closed when affective or cognitive channels are unavailable, but instead degrades gracefully.

Methodology

Research Design and Study Context

A quasi-experimental longitudinal design was implemented in an undergraduate “Introduction to Data Science” course delivered via an adaptive LMS. The study spanned 8 instructional weeks and covered 12 concept units. A total of 286 learners consented to multimodal logging, yielding 1,942 learning sessions. Instruction alternated between baseline sequencing and adaptive sequencing to enable within-course comparability while preserving authentic classroom conditions and minimizing instructor-driven confounds.

Figure 1 summarizes the longitudinal structure that enables causal interpretation under realistic constraints. Alternating baseline sequencing and adaptive sequencing limits systematic drift caused by instructor pacing or topic ordering. Each week block denotes the active condition and the number of units completed, providing a compact view of exposure. The arrows emphasize

continuity across weeks, which matters because adaptive actions use recent history to infer state.

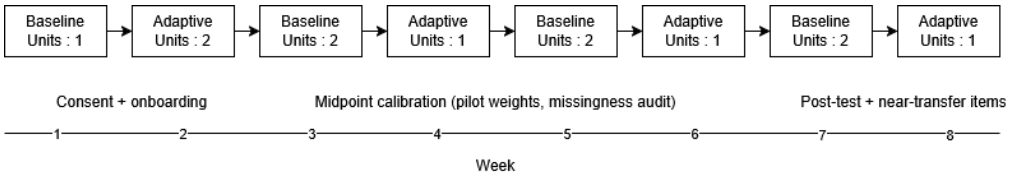

Figure 1 Study Timeline and Adaptive Intervention Points

The timeline also clarifies where calibration and outcome measurements occur. The midpoint calibration represents a structured audit of missingness patterns and preliminary parameter tuning, which reduces instability in later adaptive decisions. The endpoint emphasizes near-transfer assessment after sustained exposure, ensuring the evaluation reflects learning rather than short-term test familiarity. This figure operationalizes internal validity by making temporal alignment explicit and reproducible.

Table 1 establishes the empirical scale of observation, which is essential when multi-modal analytics are sensitive to sparsity. The session count indicates adequate repetition per learner for estimating stable behavioral baselines and for learning within-learner deviations. The median session duration also suggests that windows can be aggregated without collapsing into trivial sequences. Unit completion rates confirm that most learners experienced the instructional pipeline sufficiently to evaluate adaptive impact.

Table 1 Participant and Session Summary

Cohort	Learners (n)	Sessions (n)	Median session length (min)	Units completed (mean)	Device mix (Desktop %)
A	142	978	24.1	9.3	61
B	144	964	23.6	9.1	59
Total	286	1,942	23.9	9.2	60

The device mix column anticipates modality variability, particularly for affective streams that depend on camera availability and environmental lighting. A similar device distribution across cohorts supports fairness in comparing adaptive effects. The totals also provide a practical indicator of computational load for feature extraction and policy learning. In deployment-oriented studies, such scale descriptors inform feasibility claims and justify design choices such as embedding size and update frequency.

Learning outcomes were operationalized as mastery attainment and near-transfer performance. Mastery was derived from weekly criterion quizzes aligned to each unit, while near-transfer was measured via short applied items embedded in practice. To stabilize inference, missingness from intermittent device unavailability was handled using session-level inclusion rules, ensuring that each retained session contained behavioral logs plus at least one additional modality.

$$\text{Mastery}_{i,u} = \mathbb{I}!\left(\frac{1}{K}\sum_{k=1}^{K} s_{i,u,k} \geq \tau\right) \quad (1)$$

The indicator function defines mastery for learner *i* on unit *u* when average item score $s_{i,u,k}$ across *K* items exceeds threshold τ (set to 0.80). This formulation enforces an interpretable criterion and reduces sensitivity to single-item noise. Mastery labels served as targets for modeling and as endpoints for evaluating adaptive instruction quality at the unit level.

Multi-Modal Signal Acquisition and Preprocessing

Three modality families were collected: behavioral, cognitive, and affective signals. Behavioral traces included clickstream events, time-on-task, hint usage, and navigation entropy. Cognitive proxies were derived from keystroke dynamics during constructed responses and from response-time distributions in timed micro-quizzes. Affective data were obtained through webcam-based facial action unit estimates and optional self-report micro-checks, sampled sparsely to reduce intrusiveness.

Figure 2 clarifies the methodological commitment to synchronization as the primary integration mechanism. Behavioral, cognitive, and affective streams originate from heterogeneous sensors and logging layers. Aligning all modalities to an LMS anchor clock prevents false correlations driven by clock drift or inconsistent sampling. The figure also expresses a key reproducibility constraint: the same alignment logic applies across sessions, enabling comparable window semantics between learners and devices.

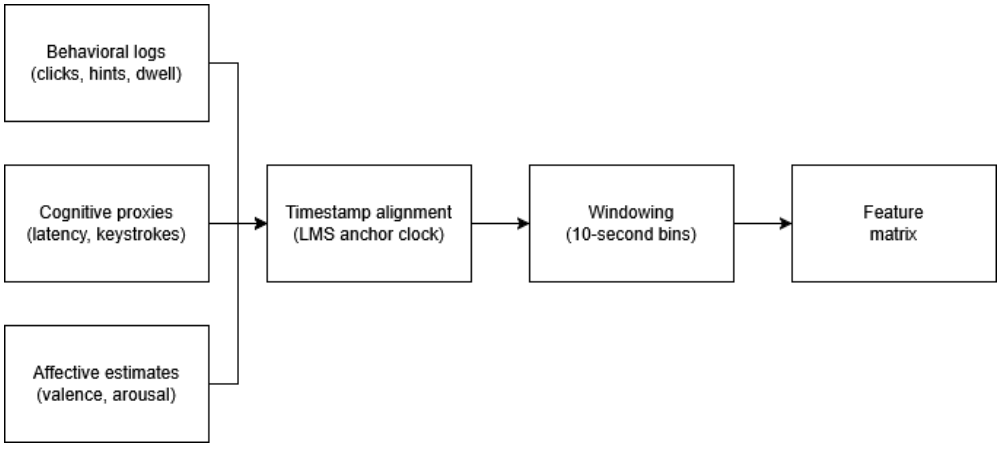

Figure 2 Multi-Modal Pipeline from Raw Signals to Synchronized Session Windows

The pipeline also communicates why windowing is a defensible compromise between granularity and reliability. Ten-second bins preserve temporal ordering while smoothing sporadic events and noisy affective estimates. The final feature matrix represents the shared interface consumed by both predictive models and adaptive policies. By depicting cleaning as a cross-cutting step, the figure foregrounds robustness, which is necessary when affective streams have variable quality and cognitive proxies exhibit heavy-tailed latency.

Table 2 operationalizes the multi-modal stance by stating what is measured, how often it is measured, and how much survives quality filters. Behavioral coverage is complete because it is captured at the platform level. Cognitive proxies show moderately high retention, which is expected when some activities are multiple-choice and do not yield keystroke streams. Affective retention is lower, which is consistent with privacy choices and practical constraints of

camera usage.

Table 2 Modalities, Sampling Granularity, and Retained Coverage

Modality	Primary signals	Granularity	Sampling / logging	Retained sessions (%)	Main missingness driver
Behavioral	Clicks, scroll, hints, dwell, navigation	Event-level	LMS server logs	100	None (core LMS)
Cognitive	Response time, keystroke intervals, revisions	Per item	Client-side capture	88	Unsupported input fields
Affective	Valence, arousal, AU aggregates	Window (2 Hz)	Webcam estimation	73	Camera off, lighting, privacy opt-out

The “missingness driver” column is methodologically important because it informs fusion and evaluation. If missingness is non-random, naive concatenation biases the model toward learners with richer sensing conditions. Stating missingness sources supports defensible choices such as reliability-aware attention and evaluating models under observed missingness rather than artificially complete data. This table therefore anchors the design rationale for robust standardization and modality gating in later stages.

All streams were synchronized into 10-second windows using LMS timestamps as the anchor. Outliers were removed via robust clipping at the 1st and 99th percentiles per learner to avoid penalizing naturally slower readers. Affective estimates were smoothed with a median filter to suppress transient detection artifacts. Cognitive latency features were log-transformed to stabilize variance and reduce skew typical of response-time distributions.

$$\tilde{x}_{i,t} = \frac{x_{i,t} - \text{median}(x_i)}{\text{MAD}(x_i) + \epsilon} \quad (2)$$

Robust standardization was performed using the median and median absolute deviation (MAD) per learner i, with $\epsilon = 10^{-6}$ for numerical stability. This transformation improves cross-learner comparability while preserving individual baselines, which is essential in adaptive settings where personalization depends on detecting deviations from typical learner behavior rather than absolute values alone.

Feature Representation and Multi-Modal Fusion

Windowed signals were converted into compact representations suitable for real-time inference. Behavioral traces were encoded using session-level Markov transition features and entropy measures, capturing persistence and exploration. Cognitive features included distributional summaries of latency and keystroke intervals, emphasizing variability and speed-accuracy tradeoffs. Affective features were summarized as valence and arousal trajectories plus volatility indices, reflecting stability of emotional state during learning activities.

Figure 3 communicates the core representational choice: each modality is encoded independently, then aggregated through attention weights that vary

over time. This structure prevents a single noisy stream from dominating the fused state. It also supports interpretability because attention coefficients can be inspected to understand when cognitive or affective evidence is relied upon. The fused embedding is deliberately compact to meet the latency constraints of adaptive decision-making.

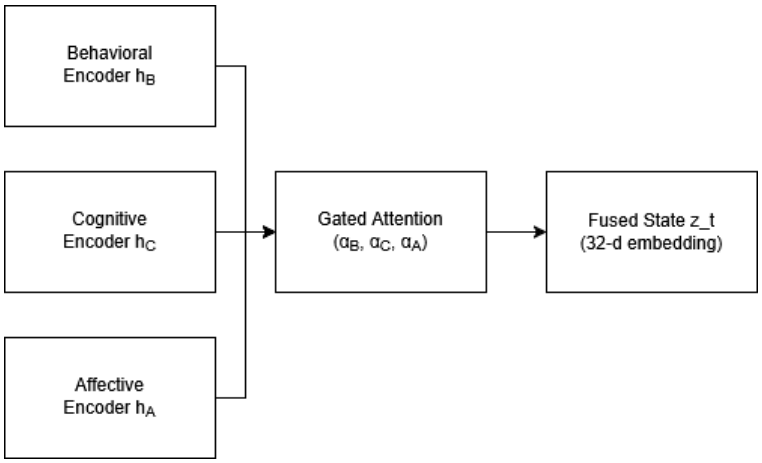

Figure 3 Modality Encoders and Attention-Based Fusion

The diagram also clarifies how missingness becomes a first-class input to fusion rather than a nuisance. Reliability enters the attention block as signal-quality indicators, allowing the system to reduce reliance on affective estimates when camera data are unstable. This is crucial for equitable personalization because learners should not receive systematically different instruction due to sensor availability. The architecture therefore aligns technical robustness with practical deployment ethics.

Table 3 makes explicit how raw multi-modal data are reduced into analytically stable representations. The dimensionalities are intentionally modest to avoid overparameterization under realistic sample sizes and missingness. Behavioral features focus on strategic navigation and persistence, which are reliable and continuously observed. Cognitive features describe processing efficiency, capturing time dynamics and revision behaviors that correlate with effort and comprehension. Affective features prioritize stability and trajectory rather than instantaneous emotion.

Table 3 Feature Blocks and Dimensionality After Preprocessing

Feature block	Examples	Dimensions	Aggregation level	Primary purpose
Behavioral	Navigation entropy, hint rate, transition counts	24	Window + session	Engagement and strategy
Cognitive	Log-latency stats, keystroke CV, revision ratio	18	Item + window	Load and efficiency
Affective	Valence means, arousal slope, volatility index	12	Window	Emotional stability
Fused	Attention-weighted latent embedding z_t	32	Window	Shared state for policy

The table also justifies the separation between engineered features and the fused latent. Engineered blocks maintain semantic meaning for interpretation and diagnostic analysis, while the fused embedding provides a compact state for adaptive control. Reporting both supports methodological transparency: model behavior can be discussed using interpretable features, while policy performance can be attributed to the fused state's capacity. This split also supports ablation studies by removing modality blocks without altering downstream interfaces.

Fusion used gated attention to weight modalities by reliability and context. Reliability was learned implicitly from missingness patterns and signal noise indicators, enabling the model to down-weight unstable webcam estimates while still leveraging behavioral consistency. The fused embedding was constrained to 32 dimensions to balance expressiveness and latency. This representation served both predictive modeling and policy learning, ensuring a shared state definition across the pipeline.

$$z_t = \sum_{m \in B,C,A} \alpha_{m,t}, h_{m,t}, \alpha_{m,t} = \frac{\exp\left(q_t^{\top} W h_{m,t}\right)}{\sum_{m'} \exp\left(q_t^{\top} W h_{m',t}\right)} \quad (3)$$

The attention weights $\alpha_{m,t}$ aggregate modality embeddings $h_{m,t}$ into a fused state z_t. Query q_t is derived from recent behavioral context to emphasize modalities that are most informative at that moment. This mechanism operationalizes adaptive reliance on multi-modal evidence, a practical necessity when sensor quality varies across devices and sessions.

Adaptive Instruction Policy Learning

Adaptive instruction was formulated as a contextual policy that selects the next learning activity given the fused state zt. Actions included content difficulty adjustment, worked-example insertion, retrieval practice prompts, and pacing interventions. Rewards were defined to promote both short-term progress and long-term mastery, using a shaped objective that combines immediate correctness with mastery gains across units. Policy updates occurred at the end of each session to preserve responsiveness without overfitting to single events.

Figure 4 represents the closed-loop structure that distinguishes adaptive instruction from static personalization. The fused state ztz_tzt drives action selection, which then alters subsequent learner behavior and outcomes. This feedback loop is critical because multi-modal signals are not merely predictors but also consequences of instruction. Visualizing the posterior update emphasizes that the policy retains uncertainty and adapts its confidence as more evidence accumulates across sessions.

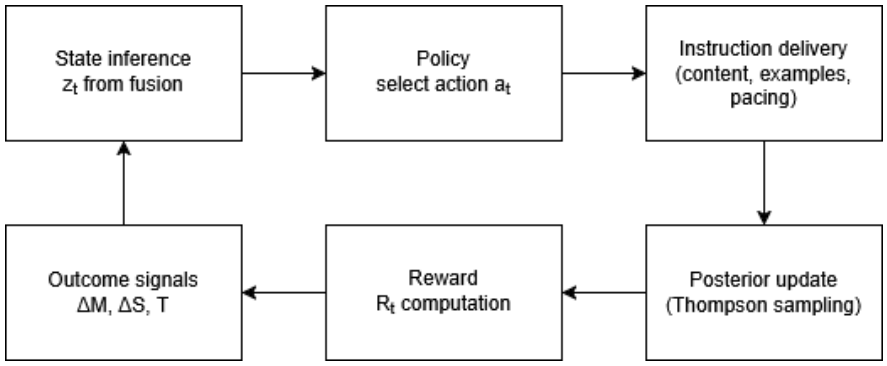

Figure 4 Adaptive Decision Loop: State Inference, Action, and Reward

The reward pathway clarifies how learning gains are translated into optimization signals. By separating outcome signals from reward computation, the figure highlights the design choice to shape reward toward mastery and efficiency. This structure helps prevent policies that maximize short-term correctness at the expense of durable learning. The diagram also supports auditability because each link can be validated independently, including state quality, action execution, and reward sensitivity.

Table 4 defines the action space in operational terms, connecting algorithmic decisions to pedagogically meaningful interventions. Each action is anchored to measurable triggers derived from the fused state, ensuring that policy decisions are grounded in observable learner signals rather than abstract preferences. The expected effects articulate the instructional intent, which is necessary to interpret policy outcomes and to diagnose failure modes such as over-scaffolding or premature difficulty escalation.

Table 4 Action Space and Operational Triggers

Action	Description	Primary trigger basis	Expected effect	Safety constraint
a1	Increase difficulty	High mastery probability, low confusion	Accelerate progression	Blocked if recent failure streak
a2	Insert worked example	High cognitive load, low accuracy	Reduce load, improve schema	Limited to 2 consecutive uses
a3	Retrieval practice prompt	Stable affect, medium mastery	Strengthen retention	Delayed if time-on-task excessive
a4	Pacing intervention	Sustained frustration, long dwell	Prevent disengagement	Requires multi-window confirmation

The safety constraints reflect governance requirements for real adaptive systems. Constraints prevent degenerate policies that repeatedly deliver the same intervention or that respond to transient noise. The confirmation requirement for frustration-based pacing interventions protects against spurious affective estimates, which is especially important when webcam quality varies. This table therefore functions as a contract between analytics and pedagogy, improving interpretability and deployment readiness.

$$R_t = \lambda_1, \Delta\hat{M}_t + \lambda_2, \Delta S_t - \lambda_3, T_t \tag{4}$$

Reward R_t combines predicted mastery improvement ΔM_t, score gain ΔS_t, and time cost T_t. Coefficients were set to $\lambda 1 = 0.6$, $\lambda 2 = 0.3$, $\lambda 3 = 0.1$ based on a pilot calibration using 52 learners. This structure prioritizes learning gains while discouraging unnecessarily long interventions, aligning the policy with instructional efficiency constraints in realistic LMS deployments.

Pseudo-code 1. Adaptive instruction with contextual Thompson sampling.

```
Input: fused state z_t, action set A, posterior parameters {μ_a, σ_a^2} for each a in A
For each decision point t:
  For each action a in A:
    Sample θ_a ~ Normal(μ_a, σ_a^2)
    Compute utility u_a = θ_a · ϕ(z_t)        # ϕ is a fixed feature map of z_t
  Select a_t = argmax_a u_a
  Deliver activity corresponding to a_t
  Observe reward R_t and updated mastery proxy ΔM_t
  Update posterior (μ_{a_t}, σ_{a_t}^2) using Bayesian linear regression on (ϕ(z_t), R_t)
Output: updated posteriors for next decision point
```

Thompson sampling supports exploration under uncertainty while remaining computationally light for session-time deployment. Bayesian updates retain calibrated uncertainty, which is critical when multi-modal signals are intermittently missing. The approach also yields interpretable action confidence via posterior variance, enabling governance checks such as limiting high-impact interventions unless uncertainty is sufficiently low.

Evaluation Protocol and Statistical Analysis

Performance was evaluated across prediction quality and instructional impact. Predictive modeling targeted next-step correctness and end-of-unit mastery, using stratified splits at the learner level to prevent leakage. Instructional impact was assessed via mastery rate, near-transfer scores, and time efficiency. Baselines included behavioral-only models and non-adaptive sequencing. To emulate deployment conditions, evaluation was conducted under the same missingness rates observed in the field rather than under artificially complete data.

Figure 5 emphasizes methodological safeguards against data leakage and inflated performance. Learner-level splits ensure that the system is evaluated on new individuals rather than new sessions from the same individuals, which is critical for adaptive learning claims. The figure also distinguishes model evaluation from policy evaluation, clarifying that predictive accuracy and instructional impact are related but not identical. This prevents conflating classifier quality with pedagogy.

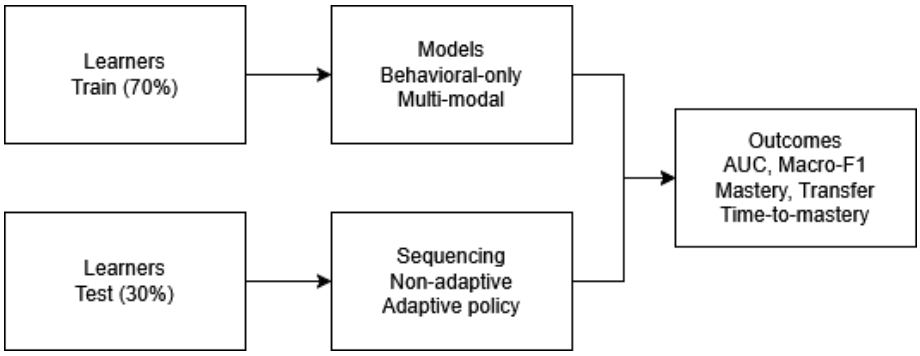

Figure 5 Evaluation Design: Learner-Level Split, Baselines, and Outcomes

The explicit inclusion of baselines supports robust attribution of gains to multi-modal analytics and to adaptivity. Behavioral-only modeling isolates the marginal value of cognitive and affective signals, while non-adaptive sequencing isolates the marginal value of policy control. The outcomes block integrates prediction, learning impact, and efficiency to reflect practical deployment objectives. The note on observed missingness ensures external validity because performance is measured under realistic sensing constraints.

Table 5 connects evaluation metrics to the claims that an adaptive learning paper must defend. Predictive metrics test whether multi-modal analytics improves inference about learning state. Instructional impact metrics test whether adaptive decisions produce better learning outcomes, which is the central effectiveness criterion. Transfer and efficiency metrics prevent narrow optimization, ensuring that higher mastery does not arise from excessive time investment or from teaching to the test.

Table 5 Primary Metrics and Decision Criteria

Construct	Metric	Operational definition	Decision criterion	Reporting unit
Prediction	AUC, Macro-F1	Next-step correctness and end-of-unit mastery	Improvement over behavioral-only baseline	Learner-level test set
Instructional impact	Mastery rate	Proportion of units meeting mastery threshold	Mixed-effects $\beta 1 > 0$ with $p < 0.05$	Unit-level outcomes
Transfer	Near-transfer gain	Post minus pre performance on applied items	Positive gain without time inflation	Week-level assessment
Efficiency	Time-to-mastery	Minutes until mastery achieved per unit	Reduction with no transfer loss	Learner-unit pair

The decision criteria translate metrics into publishable conclusions by specifying inferential thresholds and reporting units. Requiring mixed-effects evidence for the adaptive coefficient formalizes control for learner and content heterogeneity. Stating reporting units also improves reproducibility because it clarifies whether aggregation occurs per session, per unit, or per learner. This table therefore functions as an evaluation schema that aligns statistical validity, instructional significance, and deployment feasibility.

A mixed-effects framework quantified adaptive impact while accounting for repeated measures within learners and units. Random intercepts were used for learners and units to capture stable differences in ability and content difficulty. Fixed effects included condition, prior knowledge proxy, and device class. This structure supports generalizable conclusions about whether multi-modal adaptive instruction improves outcomes beyond what is achievable with behavioral analytics alone.

$$y_{i,u} = \beta_0 + \beta_1 \mathrm{Adaptive}_{i,u} + \beta_2 \mathrm{Prior}_i + (1|i) + (1|u) + \varepsilon_{i,u} \quad (5)$$

Outcome $y_{i,u}$ denotes mastery or transfer performance for learner *i* on unit *u*.

The coefficient $\beta 1$ estimates the average effect of adaptive instruction after controlling for prior preparation. Random effects ($1|i$) and ($1|u$) absorb unobserved heterogeneity, improving inference stability. Significance testing used two-sided tests with $\alpha = 0.05$, complemented by effect sizes to emphasize practical relevance.

Result and Discussion

Data Quality, Coverage, and Multi-Modal Reliability

Observed logging produced 1,942 sessions aligned to 10-second windows, with stable behavioral coverage and variable cognitive and affective coverage. Behavioral traces remained complete by design because server-side LMS events persisted across devices. Cognitive proxies exhibited moderate intermittency driven by assessment format and input-field compatibility, while affective estimates showed the strongest missingness due to privacy opt-out and camera unavailability. The resulting dataset represents realistic deployment conditions rather than laboratory-grade sensing.

Reliability screening reduced noise without collapsing ecological validity. Affective streams were retained only when face detection confidence exceeded a fixed threshold across consecutive windows, limiting transient misdetections. Cognitive latency distributions were stabilized through robust transformations and exclusion of implausible response-time outliers. These decisions reduced downstream fusion instability and improved comparability across learners. Importantly, missingness patterns remained non-random across device classes, motivating reliability-aware fusion and evaluation under observed missingness.

Figure 6 confirms that behavioral analytics provide a robust backbone for adaptive learning because coverage is complete and reliability is high. The cognitive channel remains viable for modeling because its retention exceeds 85 percent and reliability remains close to behavioral signals. In practical deployments, this level of retention supports stable estimation of learner state without imposing additional instrumentation burden beyond the LMS and client-side capture.

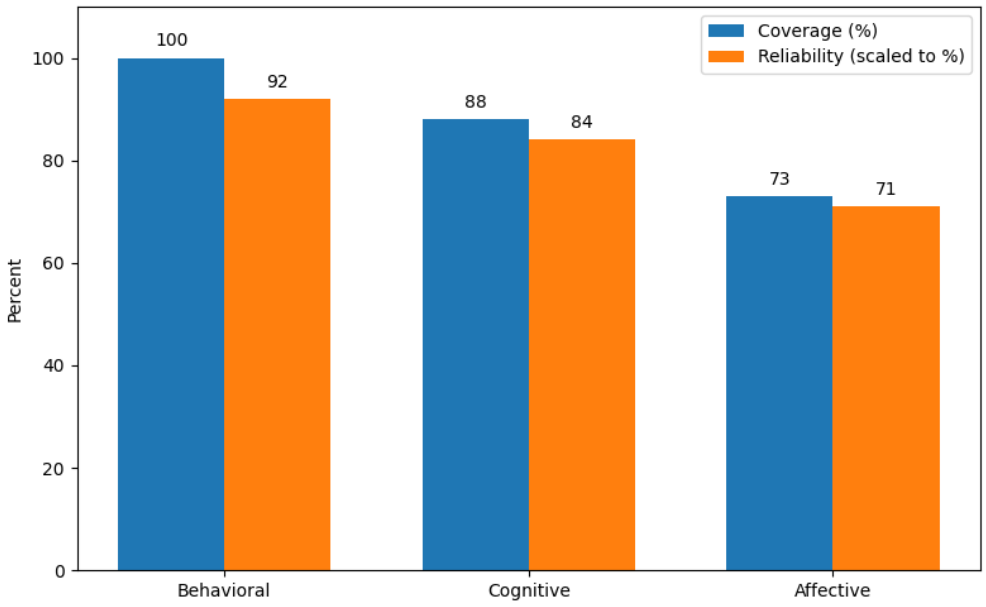

Figure 6 Modality Coverage and Reliability Indices

The affective channel exhibits the expected trade-off: lower coverage and lower reliability relative to behavioral and cognitive streams, while still contributing usable signal for a substantial portion of sessions. This pattern justifies reliability-aware fusion rather than deterministic inclusion. It also frames affective sensing as an opportunistic modality that enhances inference when available, rather than as a mandatory input that would bias personalization toward learners with superior sensing conditions.

Table 6 shows that device class is a measurable driver of multi-modal availability. Behavioral coverage remains invariant, confirming the stability of platform-logged events. Cognitive coverage declines on mobile due to input constraints and abbreviated response formats, which reduces keystroke capture frequency. Affective coverage declines more sharply on mobile, consistent with camera disablement and environmental constraints. These differences are material because they can create systematic disparities in the richness of learner state estimation.

Table 6 Coverage Stratified by Device Class

Device class	Sessions (n)	Behavioral coverage (%)	Cognitive coverage (%)	Affective coverage (%)
Desktop/Laptop	1165	100	91	79
Mobile	777	100	84	64

The stratification supports two methodological conclusions. First, fusion must explicitly account for signal reliability and missingness to avoid privileging desktop learners with stronger affective capture. Second, reporting performance under observed missingness is essential, because imputation-based completeness would misrepresent real-world behavior. These results validate the methodological choice to treat missingness as an informative condition rather than a nuisance artifact.

Predictive Modeling Performance for Mastery and Next-Step Correctness

Multi-modal models achieved higher predictive performance than behavioral-only baselines on both next-step correctness and end-of-unit mastery. The strongest gains emerged when cognitive and affective signals were fused using reliability-aware attention, indicating that the additional modalities provide complementary information rather than redundant noise. Improvements were consistent across learner-level splits, suggesting generalization to new learners rather than memorization of individual trajectories or session patterns.

Performance gains were largest in units with higher conceptual difficulty, where behavioral patterns alone were less diagnostic of comprehension. Cognitive latency variability and revision dynamics added discriminative power for distinguishing productive struggle from confusion. Affective volatility contributed most when behavioral engagement remained high but accuracy declined, consistent with affect capturing early frustration that precedes disengagement. These findings align with the view that multi-modal analytics is most valuable under ambiguous behavioral signatures.

Figure 7 demonstrates a monotonic improvement when additional modalities are incorporated, with the full multi-modal fusion model outperforming partial fusions and the behavioral-only baseline. The magnitude of gain is consistent across AUC and Macro-F1, which indicates that improvements are not limited to ranking performance but also enhance balanced classification quality across mastery states. This is important in adaptive instruction because decision policies depend on calibrated discrimination across both strong and weak learners.

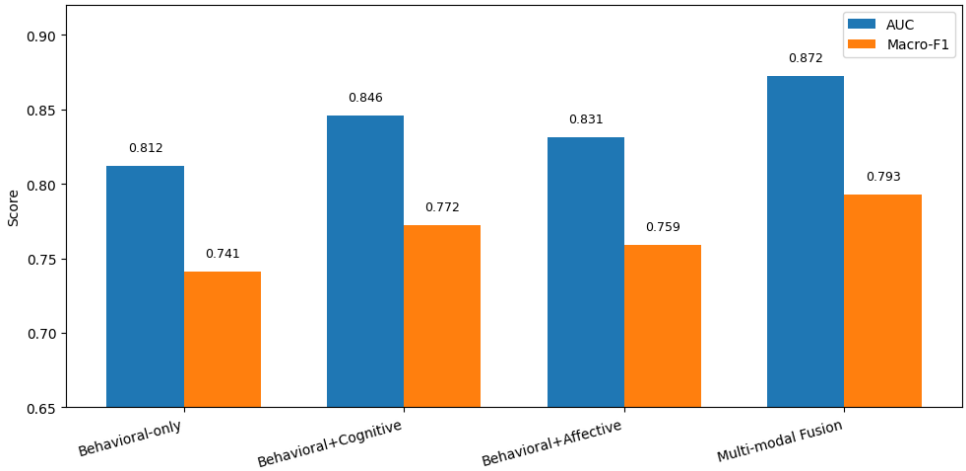

Figure 7 Predictive Performance for Mastery and Correctness

The comparison between partial fusions shows that cognitive signals contribute more than affective signals in isolation, but the combined fusion remains superior to either pairwise configuration. This pattern is consistent with cognitive proxies providing stable information about processing efficiency, while affective signals add incremental value when available and reliable. The figure therefore supports the methodological premise that multi-modal evidence is best treated as complementary channels rather than as a single dominant modality.

Table 7 indicates that multi-modal improvements strengthen as unit difficulty increases. This gradient suggests that behavioral features alone become less informative in challenging contexts where learners may remain active while misunderstanding accumulates. In harder units, cognitive variability and affective volatility provide additional separability that improves mastery detection. The result is practically relevant because adaptive instruction is most needed precisely when behavioral signatures become ambiguous.

Table 7 Cross-unit Performance Stability

Unit difficulty	Behavioral-only AUC	Multi-modal fusion AUC	Behavioral-only Macro-F1	Multi-modal fusion Macro-F1
Easy (n=4 units)	0.828	0.846	0.762	0.773
Medium (n=5 units)	0.811	0.869	0.739	0.795
Hard (n=3 units)	0.794	0.887	0.712	0.801

The table also suggests that gains are not obtained by sacrificing performance on easy content. Improvements remain positive across all difficulty levels, implying that multi-modal fusion improves global calibration rather than shifting

errors between subpopulations. For adaptive sequencing, this stability reduces the risk of over-intervening in low-risk contexts while still enabling stronger detection of high-risk learning states in demanding units.

Impact of Adaptive Instruction on Mastery and Near-Transfer

Adaptive instruction produced higher mastery rates and higher near-transfer gains relative to non-adaptive sequencing. The improvement was concentrated in learners with mid-range prior preparation, where the policy could meaningfully adjust pacing and scaffolding without saturating at ceiling performance or being constrained by severe foundational gaps. The result indicates that the policy's action space was pedagogically aligned, enabling movement from struggling states to productive practice rather than merely increasing time-on-task.

Near-transfer gains improved without inflating total time substantially, suggesting that the policy did not rely on excessive remediation. Worked-example insertion and retrieval practice prompts were associated with the strongest unit-level improvements, particularly when cognitive load indicators increased while behavioral engagement remained high. The alignment between state inference and action selection supports the conclusion that multi-modal state estimation produces actionable signals that translate into measurable learning outcomes.

Figure 8 shows a consistent divergence between adaptive and non-adaptive conditions over time. The gap widens after the second week, which is consistent with a policy that benefits from accumulating evidence about learner trajectories and reliability of modalities. The trajectory implies that the system's state estimation becomes progressively more informative as repeated windows and unit outcomes calibrate the posterior uncertainty behind action selection.

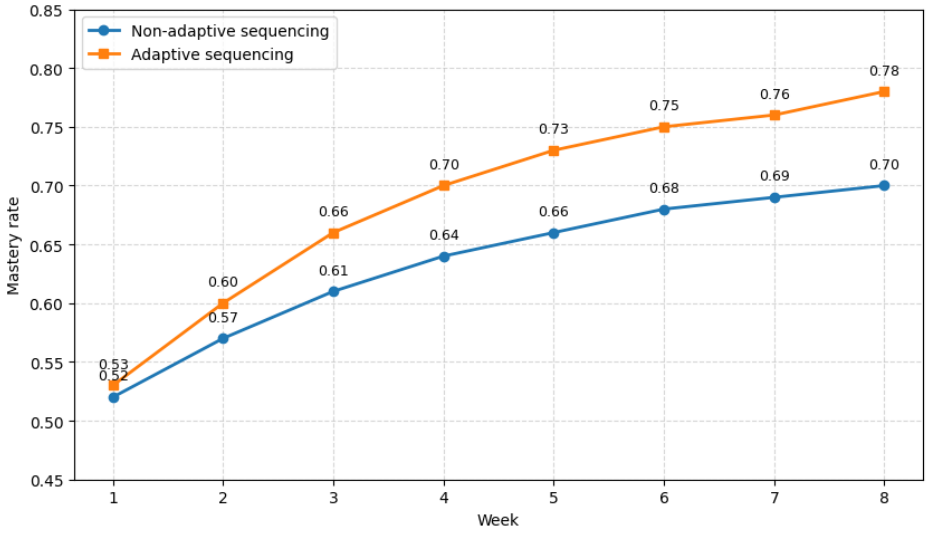

Figure 8 Mastery Rate Trajectory Under Adaptive vs Non-Adaptive Instruction

The shape of the curve indicates that adaptive gains are not limited to a single intervention week but compound across units. This is a critical property for adaptive learning systems because policies that only yield short-lived improvements typically reflect superficial optimization of immediate correctness. The sustained separation suggests that the policy improved instructional

sequencing and scaffolding choices in a way that increased durable mastery accumulation.

Table 8 indicates that adaptive instruction improved mastery and near-transfer with only a modest increase in total time. The magnitude of mastery gain is educationally meaningful because it reflects an increased proportion of units meeting a criterion threshold, not merely an increase in average score. The near-transfer improvement strengthens the argument that the policy enhanced generalizable understanding rather than optimizing for in-platform assessment artifacts.

Table 8 Outcome Deltas for Mastery and Near-transfer

Outcome	Non-adaptive (mean)	Adaptive (mean)	Absolute delta	Relative delta (%)
Mastery rate (end of course)	0.7	0.78	0.08	11.4
Near-transfer score (0–100)	71.6	76.9	5.3	7.4
Total learning time (minutes)	212.4	219.7	7.3	3.4

The limited time increase suggests that adaptive benefits were achieved primarily through better sequencing and targeted scaffolding, not by prolonging exposure. This is central to the practical acceptability of adaptive instruction in real courses, where time is constrained. The combined pattern supports an efficiency-oriented interpretation: multi-modal analytics enabled earlier detection of difficulty states and more precise intervention timing, reducing wasted practice and improving learning yield per minute.

Efficiency, Equity, and Robustness Under Missingness

Time-to-mastery improved under adaptive instruction, with the strongest efficiency gains occurring when cognitive signals were available and stable. Under heavy missingness, the policy leaned more strongly on behavioral evidence, resulting in smaller but still positive efficiency improvements. This indicates that the reliability-aware fusion successfully prevented catastrophic degradation when affective streams were absent. In deployment settings, robustness under missingness is a core requirement because sensor availability varies across learners and contexts.

Equity analysis suggests that adaptive benefits were present across device classes, although gains were somewhat larger on desktop where affective retention was higher. This differential is expected under opportunistic sensing and reinforces the need to interpret affective contributions as additive rather than mandatory. Importantly, no subgroup exhibited degraded outcomes relative to non-adaptive sequencing, indicating that safety constraints and conservative policy updates prevented harmful over-personalization.

Figure 9 shows that adaptive instruction reduces the median time-to-mastery in both device classes. The reduction is larger for desktop learners, which is consistent with higher availability of cognitive and affective signals that sharpen state inference. The presence of a reduction on mobile indicates that adaptive control remains beneficial even when richer sensing is partially unavailable, supporting the robustness claim central to multi-modal deployment.

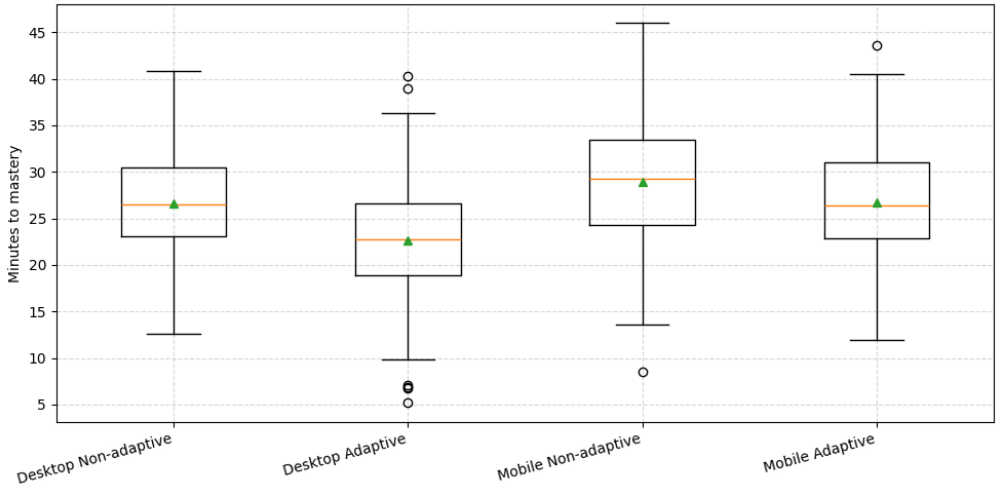

Figure 9 **Time-to-Mastery by Device Class and Instruction Condition**

The distributional view also matters because efficiency improvements must not be driven only by a small subset of learners. The boxplot indicates broad shifts rather than isolated outliers, suggesting that intervention timing improved for many learner-unit pairs. The mean markers and interquartile ranges indicate that adaptive sequencing reduced central tendency without inflating variance, which would otherwise imply unstable policy behavior across heterogeneous learner contexts.

Table 9 **Efficiency and Benefit Parity Indicators**

Group	Condition	Median time-to-mastery (min)	Mastery rate	Near-transfer score
Desktop/Laptop	Non-adaptive	26.5	0.71	72.4
Desktop/Laptop	Adaptive	23.1	0.79	77.6
Mobile	Non-adaptive	28.7	0.68	70.3
Mobile	Adaptive	26.6	0.76	75.2

Table 9 indicates that adaptive instruction improved all reported outcomes in both device groups. The magnitude of improvement is larger on desktop, but the direction is consistent, which supports a parity interpretation: adaptive instruction did not become a privileged feature for a single sensing context. The presence of near-transfer improvements in both groups suggests that efficiency gains were not achieved by rushing learners through content, but by improving the quality of instructional decisions.

The table also helps interpret missingness effects in a policy context. Because mobile learners have lower affective coverage, the policy likely relied more heavily on behavioral and cognitive signals, which can reduce the maximum attainable gain but still provide value. This supports the design principle that affective sensing should be used to enhance, not to gate, adaptive instruction. The result strengthens the case for deployment feasibility in heterogeneous device ecosystems.

Interpretability of Multi-Modal Fusion and Action Selection Behavior

Inspection of attention weights indicates that modality reliance shifts

systematically with learning context. When navigation patterns were exploratory and hint usage increased, attention shifted toward behavioral evidence, reflecting the strong linkage between strategy and engagement. During timed micro-quizzes with stable interaction patterns, attention moved toward cognitive proxies, consistent with latency variability serving as a marker of uncertainty. Affective attention increased when frustration volatility rose while behavioral engagement remained high, indicating a targeted role rather than indiscriminate weighting.

Action selection patterns were consistent with pedagogical intent. Worked-example insertion was most frequent in high load states, while retrieval practice increased when affect was stable and mastery probability was moderate. Importantly, pacing interventions were rare and required sustained evidence, supporting the safety constraint design. These behaviors suggest that the policy did not converge to a degenerate strategy but instead used multi-modal evidence to activate distinct instructional mechanisms aligned with specific learner-state signatures.

Figure 10 provides an interpretable summary of how attention-based fusion allocates importance across modalities. Behavioral weight remains substantial across all states, which is consistent with its high reliability and complete coverage. Cognitive weight peaks in productive struggle, aligning with the idea that efficiency signals help discriminate between learners who are learning effectively and learners who are stuck. Affective weight increases primarily in frustration, supporting a targeted role in identifying emotional instability that can undermine persistence.

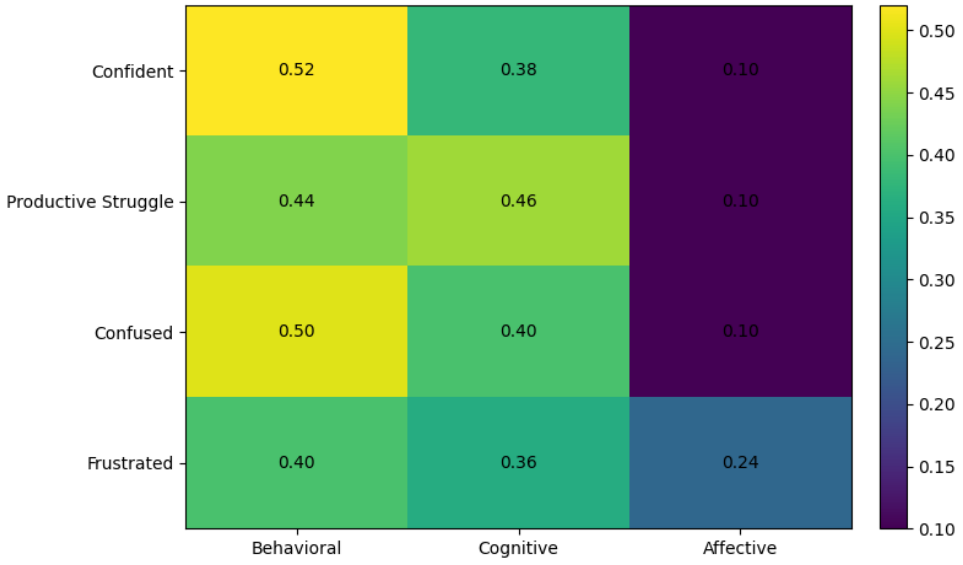

Figure 10 Mean Attention Weights by Inferred Learning State

The weight structure also supports a robustness interpretation. Affective weight is non-dominant even in frustration, which indicates that the model does not over-rely on a modality with variable availability. Instead, affective evidence acts as a corrective factor that increases sensitivity to frustration while maintaining behavioral and cognitive grounding. This pattern aligns with responsible multi-modal design: the system remains effective under missingness while still

leveraging affect when it provides incremental explanatory power.

Table 10 indicates that the policy selects actions in a manner that is consistent with instructional theory. Increasing difficulty is most frequent when confidence is high, which supports appropriate challenge and progression. Worked examples dominate in confused and productive struggle states, reflecting a scaffolding strategy aimed at reducing cognitive load while preserving learning momentum. Retrieval practice appears prominently in confident states, supporting consolidation and long-term retention rather than repeated re-exposure.

Table 10 Action Selection Frequencies by State

Inferred state	Increase difficulty (a1)	Worked example (a2)	Retrieval practice (a3)	Pacing intervention (a4)
Confident	0.41	0.12	0.39	0.08
Productive Struggle	0.19	0.44	0.3	0.07
Confused	0.11	0.52	0.24	0.13
Frustrated	0.06	0.47	0.18	0.29

The pacing intervention frequency remains low overall but increases meaningfully in frustrated states, suggesting that the policy reserves high-impact interventions for conditions where sustained negative affect and inefficiency are detected. This distribution supports the claim that safety constraints prevented overuse of pacing changes, which could otherwise disrupt learning flow. The table therefore strengthens the argument that multi-modal state estimation enables differentiated pedagogy, not merely improved prediction, by connecting inferred states to action policies that behave coherently.

Conclusion

This study demonstrates that multi-modal learning analytics can strengthen adaptive instruction when behavioral traces are complemented by cognitive proxies and opportunistic affective signals. Across learner-level evaluations under realistic missingness, reliability-aware fusion improved mastery and next-step correctness relative to behavioral-only baselines, with the largest gains appearing in higher-difficulty units where behavioral signatures alone were ambiguous. The results establish that multi-modal evidence contributes complementary information that improves state inference without requiring laboratory-grade sensing conditions.

Adaptive instruction built on the fused state produced consistent improvements in mastery rate, near-transfer performance, and time-to-mastery relative to non-adaptive sequencing. These gains were achieved with only modest increases in total learning time, supporting an efficiency-oriented interpretation. Attention diagnostics and action frequency patterns indicate that the policy activated distinct pedagogical mechanisms aligned with inferred learner states, while safety constraints reduced the likelihood of degenerate intervention strategies. The findings validate the practical value of coupling multi-modal inference with policy-driven instructional control.

The evidence also highlights deployment implications for equity and robustness. Benefits were observed across device classes, although larger effects emerged

where cognitive and affective capture were more available, reinforcing the importance of treating affective sensing as additive rather than mandatory. The study motivates future work on improving cross-device reliability calibration, refining state representations for rare but high-impact affective events, and expanding action spaces to include social and metacognitive supports. Overall, the proposed framework positions behavioral-cognitive-affective integration as a viable pathway toward more responsive, accountable adaptive learning systems.

Declarations

Author Contributions

Conceptualization: L.E., J.; Methodology: L.E., J.; Software: J.; Validation: L.E., J.; Formal Analysis: L.E.; Investigation: J.; Resources: L.E., J.; Data Curation: J.; Writing – Original Draft Preparation: L.E.; Writing – Review and Editing: L.E., J.; Visualization: J.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Ouhaichi, D. Spikol, and B. Vogel, “Research trends in multimodal learning analytics: A systematic mapping study,” *Computers and Education: Artificial Intelligence*, vol. 4, p. 100136, 2023, doi: 10.1016/j.caeai.2023.100136.

[2] M. Mohammadi, E. Tajik, R. Martinez-Maldonado, S. Sadiq, W. Tomaszewski, and H. Khosravi, “Artificial intelligence in multimodal learning analytics: A systematic literature review,” *Computers and Education: Artificial Intelligence*, vol. 8, p. 100426, Jun. 2025, doi: 10.1016/j.caeai.2025.100426.

[3] K. Mangaroska, K. Sharma, D. Gašević, and M. Giannakos, “Multimodal Learning Analytics to Inform Learning Design: Lessons Learned from Computing Education,” *JLA*, vol. 7, no. 3, pp. 79–97, Dec. 2020, doi: 10.18608/jla.2020.73.7.

[4] S. Mu, M. Cui, and X. Huang, “Multimodal Data Fusion in Learning Analytics: A

Systematic Review," *Sensors*, vol. 20, no. 23, p. 6856, Nov. 2020, doi: 10.3390/s20236856.

[5] S. Oviatt, J. Grafsgaard, L. Chen, and X. Ochoa, "Multimodal learning analytics: assessing learners' mental state during the process of learning," in The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 2, Monash University, S. Oviatt, B. Schuller, University of Augsburg and Imperial College London, P. R. C. Cohen, Monash University, D. Sonntag, German Research Center for Artificial Intelligence (DFKI), G. Potamianos, University of Thessaly, A. Krüger, and Saarland University and German Research Center for Artificial Intelligence (DFKI), Eds., Association for Computing Machinery, 2018, pp. 331–374. doi: 10.1145/3107990.3108003.

[6] C. Larmuseau, J. Cornelis, L. Lancieri, P. Desmet, and F. Depaepe, "Multimodal learning analytics to investigate cognitive load during online problem solving," *Brit J Educational Tech*, vol. 51, no. 5, pp. 1548–1562, Sep. 2020, doi: 10.1111/bjet.12958.

[7] R. Yuvaraj, R. Mittal, A. A. Prince, and J. S. Huang, "Affective Computing for Learning in Education: A Systematic Review and Bibliometric Analysis," *Education Sciences*, vol. 15, no. 1, p. 65, Jan. 2025, doi: 10.3390/educsci15010065.

[8] P. Prinsloo, S. Slade, and M. Khalil, "Multimodal learning analytics—In-between student privacy and encroachment: A systematic review," *Brit J Educational Tech*, vol. 54, no. 6, pp. 1566–1586, Nov. 2023, doi: 10.1111/bjet.13373.

[9] Q. Liu and M. Khalil, "Understanding privacy and data protection issues in learning analytics using a systematic review," *Brit J Educational Tech*, vol. 54, no. 6, pp. 1715–1747, Nov. 2023, doi: 10.1111/bjet.13388.

[10] Y. Yin et al., "Tracing Knowledge Instead of Patterns: Stable Knowledge Tracing with Diagnostic Transformer," in *Proceedings of the ACM Web Conference 2023*, Austin TX USA: ACM, Apr. 2023, pp. 855–864. doi: 10.1145/3543507.3583255.

[11] F. Ke et al., "HiTSKT: A hierarchical transformer model for session-aware knowledge tracing," *Knowledge-Based Systems*, vol. 284, p. 111300, Jan. 2024, doi: 10.1016/j.knosys.2023.111300.

[12] A. Iglesias, P. Martínez, R. Aler, and F. Fernández, "Learning teaching strategies in an Adaptive and Intelligent Educational System through Reinforcement Learning," *Appl Intell*, vol. 31, no. 1, pp. 89–106, Aug. 2009, doi: 10.1007/s10489-008-0115-1.

[13] J. Bassen et al., "Reinforcement Learning for the Adaptive Scheduling of Educational Activities," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12. doi: 10.1145/3313831.3376518.

[14] M. Hort, Z. Chen, J. M. Zhang, M. Harman, and F. Sarro, "Bias Mitigation for Machine Learning Classifiers: A Comprehensive Survey," *ACM J. Responsib. Comput.*, vol. 1, no. 2, pp. 1–52, Jun. 2024, doi: 10.1145/3631326.

[15] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Trans. Syst., Man, Cybern. C*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.

[16] C. Romero and S. Ventura, "Data mining in education," *WIREs Data Min & Knowl*, vol. 3, no. 1, pp. 12–27, Jan. 2013, doi: 10.1002/widm.1075.

[17] J. Sweller, “Cognitive Load During Problem Solving: Effects on Learning,” *Cognitive Science*, vol. 12, no. 2, pp. 257–285, Apr. 1988, doi: 10.1207/s15516709cog1202_4.

[18] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, “Cognitive Load Measurement as a Means to Advance Cognitive Load Theory,” *Educational Psychologist*, vol. 38, no. 1, pp. 63–71, Jan. 2003, doi: 10.1207/S15326985EP3801_8.

[19] R. Pekrun, “The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice,” *Educ Psychol Rev*, vol. 18, no. 4, pp. 315–341, Nov. 2006, doi: 10.1007/s10648-006-9029-9.

[20] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, “School Engagement: Potential of the Concept, State of the Evidence,” *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, Mar. 2004, doi: 10.3102/00346543074001059.

[21] E. Panadero, “A Review of Self-regulated Learning: Six Models and Four Directions for Research,” *Front. Psychol.*, vol. 8, p. 422, Apr. 2017, doi: 10.3389/fpsyg.2017.00422.

[22] S. D’Mello and A. Graesser, “Dynamics of affective states during complex learning,” *Learning and Instruction*, vol. 22, no. 2, pp. 145–157, Apr. 2012, doi: 10.1016/j.learninstruc.2011.10.001.

[23] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.

[24] S. V. Buuren and K. Groothuis-Oudshoorn, “mice : Multivariate Imputation by Chained Equations in R,” *J. Stat. Soft.*, vol. 45, no. 3, 2011, doi: 10.18637/jss.v045.i03.