Data-Driven Adaptive Curriculum Design Using Clustering and Learning Analytics

Soeltan Abdul Ghaffar[1,*], Wilbert Clarence Setiawan[2]

[1]Department of Marine Information Systems, Universitas Pendidikan Indonesia, Bandung, Indonesia

[2]Faculty of Informatics Engineering, Universitas Taruma Negara, Jakarta, Indonesia

ABSTRACT

The increasing heterogeneity of learners in digital learning environments has exposed fundamental limitations of static, one-size-fits-all curriculum designs. While adaptive learning systems have been widely explored, many existing approaches focus on surface-level personalization and fail to restructure curriculum pathways based on empirical learner behavior. This study proposes a data-driven adaptive curriculum design framework that integrates learning analytics and unsupervised clustering to enable curriculum-level adaptation in online learning systems. Learning interaction data were collected from a learning management system and transformed into multidimensional learning analytics indicators capturing engagement, assessment behavior, completion patterns, and self-regulation. Using clustering techniques, learners were grouped into distinct behavioral profiles without predefined labels. The results reveal stable and interpretable learner clusters characterized as engaged learners with uneven mastery, struggling learners with weak self-regulation, and efficient mastery-oriented learners. These clusters were systematically mapped to differentiated curriculum pathways involving variations in content granularity, pacing strategy, feedback intensity, and assessment difficulty. Empirical evaluation demonstrates that curriculum differentiation leads to divergent but equitable learning outcomes across pathways. Learners in accelerated pathways achieved the highest mean outcome scores with low variance, while learners in scaffolded pathways maintained competitive completion rates despite greater performance variability. Importantly, the adaptive curriculum did not amplify learning disparities; instead, it supported sustained engagement and progression across heterogeneous learner groups. System-level analysis further confirms that the proposed closed-loop adaptive framework remains stable and interpretable, enabling continuous refinement through learning analytics feedback. These findings contribute evidence that clustering-based learner profiling can be operationalized into curriculum engineering decisions, advancing adaptive learning from reactive personalization toward structured, data-driven curriculum design. The proposed framework offers practical implications for scalable, accountable, and pedagogically grounded adaptive learning systems.

Keywords Adaptive Learning, Learning Analytics, Curriculum Adaptation, Learner Clustering, Educational Data Mining, Data-Driven Curriculum Design, Personalized Learning Systems

Submitted: 1 February 2024
Accepted: 10 June 2024
Published: 1 February 2025

*Corresponding author
Soeltan Abdul Ghaffar,
soeltan027ghaffar@gmail.com

Additional Information and Declarations can be found on page 93

Introduction

The rapid expansion of digital learning environments has intensified long-standing challenges related to learner heterogeneity, instructional scalability, and curriculum relevance. Online and blended learning systems increasingly serve learners with diverse prior knowledge, learning pace, self-regulation capacity, and engagement patterns. However, many existing curricula in Learning Management Systems (LMS) remain fundamentally static, delivering

How to cite this article: S. A. Ghaffar, W. C. Setiawan, "Data-Driven Adaptive Curriculum Design Using Clustering and Learning Analytics," *Adapt. Learn.*, vol. 1, no. 1, pp. 75-95, 2025.

identical content sequences and assessment structures regardless of learner behavior. This misalignment between learner diversity and uniform curriculum design has been repeatedly identified as a key factor contributing to disengagement, learning inefficiency, and uneven learning outcomes in digital education [1], [2].

To address this challenge, the field of adaptive learning has emerged as a research and technological response that seeks to personalize learning experiences based on learner data. Early adaptive systems relied primarily on expert-defined rules or learner self-reports, which limited scalability and responsiveness [3]. More recent approaches leverage learning analytics, exploiting large-scale behavioral data generated by learners' interactions with digital platforms to inform instructional decisions [4], [5]. Despite these advances, many analytics-driven systems focus on micro-level personalization, such as recommending content or adjusting feedback frequency, while leaving the underlying curriculum structure unchanged [6].

A growing body of research suggests that effective personalization requires curriculum-level adaptation, not merely surface-level adjustments [7], [8]. Curriculum design governs content sequencing, conceptual dependency, pacing, and assessment alignment, making it a critical leverage point for meaningful learning personalization. However, adapting curricula dynamically is inherently complex, as it requires robust learner modeling capable of capturing latent behavioral patterns without relying on predefined labels or assumptions [9]. This complexity has limited the practical adoption of adaptive curricula in real-world educational settings.

Unsupervised machine learning, particularly clustering, offers a promising methodological foundation for addressing this limitation. Clustering enables the discovery of latent learner profiles based on multidimensional learning analytics, allowing systems to identify structured patterns of engagement, performance, and self-regulation [10], [11]. Prior studies have shown that clustering can reveal meaningful learner typologies; however, many such studies stop at descriptive profiling and do not systematically translate clusters into actionable curriculum adaptations [12]. As a result, the connection between learner profiling and curriculum engineering remains underdeveloped.

This gap highlights a critical limitation in existing adaptive learning research: the lack of an integrated framework that connects learning analytics, learner clustering, and adaptive curriculum design in a coherent and operational manner [13]. Many studies evaluate predictive accuracy or cluster validity in isolation, without examining how these analytical outputs reshape instructional structure or learning outcomes. Consequently, there is insufficient empirical evidence demonstrating that clustering-based learner modeling can support scalable, equitable, and pedagogically grounded curriculum adaptation [14].

The objective of this study is to address this gap by proposing and evaluating a data-driven adaptive curriculum design framework that integrates learning analytics with clustering-based learner profiling. The framework systematically transforms learner interaction data into behavioral indicators, identifies latent learner groups through unsupervised clustering, and maps these groups to differentiated curriculum pathways. By embedding adaptation at the curriculum level, the proposed approach aims to align instructional structure with empirically observed learning behaviors rather than predefined assumptions [15], [16].

The novelty of this research lies in its end-to-end integration of analytics, clustering, and curriculum engineering within a closed-loop adaptive learning system. Unlike prior work that treats learner profiling and curriculum design as separate concerns, this study demonstrates how clustering outputs can be operationalized into explicit curriculum differentiation rules and evaluated through learning outcomes. This contribution advances adaptive learning research by reframing personalization as a structured curriculum design problem grounded in data-driven learner modeling, offering both theoretical and practical implications for scalable digital education systems [17], [18].

Literature Review

Research on adaptive learning systems has evolved from early rule-based personalization toward data-driven approaches that leverage large-scale learner interaction data. Classical adaptive systems primarily relied on expert-defined instructional rules or learner self-assessment instruments to adjust content delivery. While these approaches offered pedagogical transparency, they were limited in scalability and responsiveness, particularly in heterogeneous learning environments where learner behavior changes dynamically over time [19]. As digital learning platforms matured, the increasing availability of fine-grained interaction logs enabled a shift toward analytics-based learner modeling.

Within this context, learning analytics has emerged as a foundational discipline for understanding learner behavior in online environments. Learning analytics focuses on the collection, measurement, analysis, and reporting of learner data to optimize learning and the environments in which it occurs [20]. Prior studies demonstrate that behavioral indicators such as time-on-task, content access patterns, and assessment attempts provide valuable signals of engagement and learning progress. However, much of the existing literature emphasizes descriptive dashboards or predictive performance models, with limited attention to how analytics outputs can directly inform curriculum restructuring [21].

To overcome the limitations of predefined learner categories, researchers have increasingly adopted unsupervised machine learning techniques, particularly clustering, to identify latent learner profiles. Clustering has been shown to effectively capture heterogeneous learning behaviors without relying on labeled outcomes, making it well-suited for exploratory learner modeling in large-scale educational datasets [22]. Studies applying clustering in educational contexts report the discovery of meaningful learner typologies, such as surface learners, strategic learners, and mastery-oriented learners. Nevertheless, these studies often stop at behavioral interpretation and do not formalize how identified clusters translate into instructional or curricular decisions.

Several works attempt to bridge analytics and instruction by proposing adaptive mechanisms such as content recommendation or feedback modulation based on learner profiles. While these mechanisms demonstrate short-term engagement benefits, they typically operate at the micro-adaptation level, adjusting individual learning objects rather than the curriculum structure itself [23]. As a result, learners may receive personalized content suggestions while still being constrained by a fixed curriculum sequence that does not reflect their broader learning trajectory or readiness.

Recent curriculum design literature emphasizes that effective personalization must operate at the macro-structural level, encompassing content sequencing, pacing policies, and assessment alignment [24]. From this perspective, curriculum is not merely a container for learning objects but a dynamic system that shapes cognitive progression. However, integrating adaptive logic into curriculum design introduces significant complexity, as it requires both reliable learner modeling and interpretable decision rules that maintain pedagogical coherence.

A limited number of studies have begun exploring integrated frameworks that connect learning analytics, machine learning, and curriculum adaptation. These studies highlight the importance of closing the loop between learner data, instructional decisions, and learning outcomes, yet they often lack empirical validation at scale or provide insufficient detail on curriculum differentiation mechanisms [25]. Consequently, there remains a gap in the literature concerning end-to-end adaptive curriculum systems that are both data-driven and pedagogically grounded.

In summary, existing research establishes the analytical feasibility of learner clustering and the pedagogical importance of curriculum-level adaptation, but it rarely unifies these components into a single operational framework. This gap motivates the present study, which positions clustering-based learner profiling not as an end in itself, but as an enabling mechanism for systematic adaptive curriculum design evaluated through learning outcomes. By addressing this gap, the study extends prior work and contributes empirical evidence to the design of scalable and interpretable adaptive learning systems [26].

Methodology

Research Design and Methodological Framework

This study adopts a quantitative data-driven engineering methodology to design and evaluate an adaptive curriculum model grounded in learning analytics and unsupervised machine learning. The methodological framework is structured as a multi-stage analytical pipeline, beginning with educational data acquisition, followed by preprocessing, clustering-based learner profiling, curriculum adaptation logic, and performance evaluation. The central assumption of this framework is that latent patterns in learner interaction data can be algorithmically extracted and mapped into pedagogically meaningful curriculum adjustments.

The overall research design follows a Design Science Research (DSR) paradigm, where the adaptive curriculum model is treated as an artifact that is iteratively developed and empirically validated. Learning analytics functions as the epistemic layer that transforms raw behavioral data into interpretable indicators, while clustering operates as the core mechanism for identifying heterogeneous learning profiles. This integration ensures that curriculum adaptation is not heuristic-based but empirically grounded.

At a mathematical level, the methodological logic is driven by the abstraction of learners into feature vectors. Let each learner be represented as a vector $x_i \in R^d$, where d denotes the number of learning analytics indicators. Formally, the learner space is defined as:

$$\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id}) \tag{1}$$

This formulation allows the learning process to be treated as a geometric problem in high-dimensional space, where structural similarity reflects pedagogical proximity. The rationale behind this formulation is discussed in relation to cluster formation and curriculum mapping in subsequent sub-sections.

Figure 1 visualizes the end-to-end methodological pipeline used to engineer an adaptive curriculum artifact from LMS data. The diagram clarifies that the study is not a single-model exercise but a staged system: raw learning traces are first transformed into analyzable features, then used to derive learner profiles via clustering, and finally converted into curriculum decisions through explicit mapping rules. This representation is essential for reproducibility because it specifies where each transformation occurs and how intermediate outputs (features, clusters, mapping) become inputs to later stages.

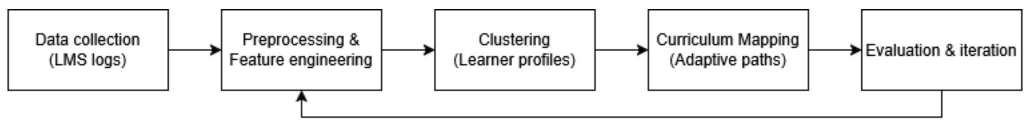

Figure 1 Research Methodology Framework

The feedback loop emphasizes iterative refinement, which is structurally aligned with a design science approach and practically necessary in learning analytics because learner behavior distributions drift over time. Evaluation results (e.g., cluster separability or pathway outcomes) feed back into feature engineering and model configuration, enabling recalibration of indicators or cluster parameters. In implementation terms, this loop operationalizes continuous improvement, ensuring that the curriculum adaptation logic remains empirically grounded rather than frozen to a single training snapshot.

Data Collection and Learning Analytics Construction

The dataset used in this study consists of learner interaction logs collected from an online LMS. These logs include temporal, behavioral, and performance-oriented attributes such as session duration, content access frequency, assessment attempts, and completion rates. Data collection is conducted passively to preserve ecological validity, ensuring that learner behavior is not influenced by experimental intervention.

Raw interaction data are transformed into learning analytics indicators through aggregation and normalization procedures. Indicators are engineered to capture cognitive engagement, persistence, and learning progression. For instance, time-on-task is aggregated per learning unit, while assessment behavior is summarized using attempt variance and success ratio. These engineered features constitute the analytical basis for clustering.

Formally, normalization is applied to eliminate scale bias across heterogeneous indicators. Min–max normalization is used as follows:

$$x'_{ij} = \frac{x_{ij} - min(x_j)}{max(x_j) - min(x_j)} \tag{2}$$

where xij denotes the raw value of feature j for learner i. This transformation ensures that each feature contributes proportionally to distance computation in clustering. Without this step, high-magnitude variables would dominate similarity calculations, leading to distorted learner groupings.

Table 1 defines the engineered learning analytics indicators that serve as model features. The indicator set intentionally spans engagement, progress, assessment performance, and support usage because adaptive curriculum design requires multidimensional signals: time-on-task alone cannot distinguish productive effort from confusion, and accuracy alone cannot differentiate fast mastery from superficial guessing. By expressing each indicator as a measurable construct with an explicit aggregation window, the table clarifies the operational meaning of each variable in a way that supports replication across LMS platforms.

Table 1 Learning Analytics Indicators and Definitions

Indicator	**Type**	**Definition**	**Rationale for Adaptive Curriculum**	**Typical Aggregation Window**
TOT_unit	Engagement	Total time-on-task per learning unit (minutes)	Detects pacing needs and potential cognitive overload	Per unit / weekly
Access_freq	Engagement	Number of content accesses per unit	Signals persistence and revisitation behavior	Per unit
Completion_rate	Progress	Proportion of learning items completed within a unit	Drives unlock rules and remediation thresholds	Per unit
Quiz_attempts	Assessment	Total number of quiz attempts per unit	Identifies struggle patterns and trial-and-error strategies	Per unit
Quiz_accuracy	Assessment	Correct answers divided by total answers (0–1)	Guides difficulty scaling and mastery gating	Per unit
Attempt_variance	Assessment	Variance of attempts across quizzes in the unit	Separates stable mastery from inconsistent performance	Per unit
Late_submission	Self-regulation	Ratio of late submissions to total submissions	Triggers time management scaffolding	Weekly
Forum_posts	Social learning	Number of forum posts or replies	Supports collaborative pathway assignment	Weekly
Help_seeking	Support	Count of help events (hints, FAQ clicks, tutor requests)	Controls feedback frequency and tutoring allocation	Per unit / weekly

From an engineering perspective, these indicators are also chosen to reduce ambiguity in downstream clustering. Features such as Attempt_variance and Help_seeking improve identifiability of latent behaviors that otherwise appear similar under coarse metrics. For example, two learners may share identical Quiz_accuracy, yet diverge strongly in Quiz_attempts and Help_seeking, implying different interventions: one needs harder content, the other needs scaffolding. This table therefore functions as the specification layer linking raw logs to adaptive decision-making.

Learner Profiling Using Clustering Algorithms

Learner profiling is performed using unsupervised clustering to identify latent learning behavior patterns without predefined labels. This study employs centroid-based clustering as the primary analytical technique due to its interpretability and computational efficiency in large-scale educational datasets. The objective is to partition learners into internally cohesive and externally separable groups. The clustering objective function minimizes intra-cluster variance, expressed mathematically as:

$$J = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} |\mathbf{x}_i - \boldsymbol{\mu}_k|^2 \tag{3}$$

where Ck denotes cluster k and μk represents its centroid. This formulation ensures that learners assigned to the same cluster exhibit minimal behavioral divergence. The selection of K is empirically determined using internal validation metrics to balance model complexity and interpretability.

Cluster outputs are interpreted pedagogically by analyzing centroid characteristics. For example, clusters with high engagement but low assessment success are interpreted as exploratory learners, while clusters with short interaction time but high accuracy are interpreted as efficient learners. This semantic labeling bridges the gap between mathematical abstraction and curriculum engineering.

Figure 2 presents a clustering result in a two-dimensional PCA projection to support interpretability of otherwise high-dimensional learner behavior. The purpose of PCA here is strictly representational: it compresses the feature matrix into orthogonal components (PC1, PC2) that preserve maximal variance, enabling visual inspection of separation and overlap between inferred learner groups. In the methodology, clustering is performed in the normalized feature space, while the projection provides a diagnostic view of whether clusters reflect coherent structure rather than noise.

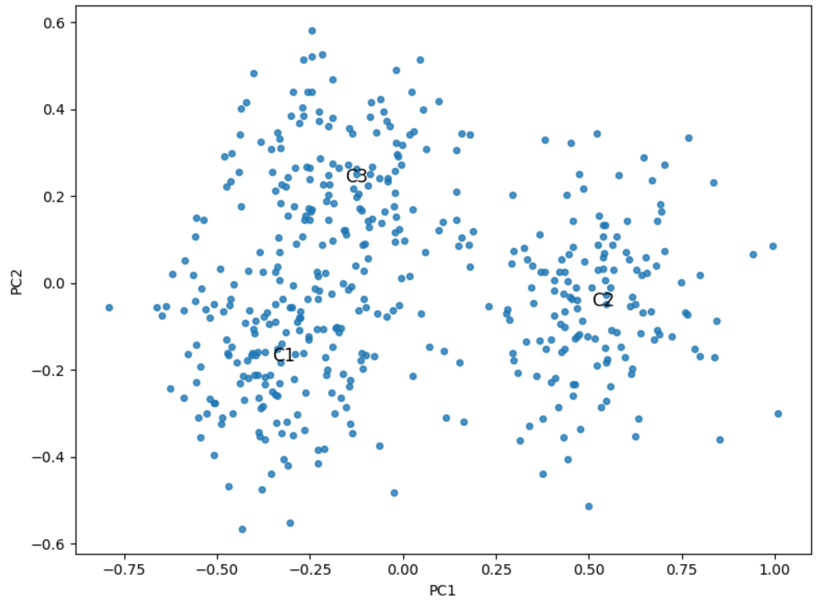

Figure 2 Learner Clustering Visualization

Analytically, the visual separation between point clouds functions as an informal validation of cluster compactness and inter-cluster distance, complementing formal metrics used in evaluation. When clusters overlap substantially, it typically signals either weak feature engineering (insufficiently discriminative indicators) or an unsuitable value of K, both of which degrade curriculum personalization because pathway assignment becomes unstable. Conversely, stable cluster regions support robust mapping into curriculum adaptations such as remediation intensity, pacing adjustments, and assessment difficulty modulation.

Adaptive Curriculum Mapping and Decision Logic

Once learner clusters are established, curriculum adaptation is performed by mapping each cluster to differentiated learning pathways. Adaptation rules are defined based on cluster-specific characteristics, such as content pacing, assessment difficulty, and remediation intensity. This mapping operationalizes the principle that curriculum structure should respond dynamically to learner needs rather than enforcing uniform progression.

The adaptation decision function can be expressed as:

$$f(C_k) \rightarrow \mathcal{P}_k \tag{4}$$

where f maps cluster Ck to an adaptive curriculum pathway Pk. Each pathway is parameterized by content sequencing, feedback frequency, and learning support mechanisms. The function f is rule-based but informed directly by empirical cluster statistics, ensuring data-driven adaptation.

To formalize adaptation strength, a curriculum adjustment score is computed using weighted learning indicators:

$$A_k = \sum_{j=1}^{d} w_j \cdot \mu_{kj} \tag{5}$$

where μkj is the centroid value of feature j in cluster k and wj represents pedagogical importance weights. This score governs the intensity of curriculum modification, allowing fine-grained differentiation across learner groups.

Table 2 operationalizes the translation from latent learner profiles to concrete curriculum actions by defining explicit mapping rules. Each cluster is described using a dominant signature derived from centroid statistics, then interpreted pedagogically to avoid purely mathematical personalization. This alignment is important because clustering outputs are inherently descriptive, not prescriptive; without a mapping layer, clusters do not automatically yield educational interventions.

Table 2 Cluster-to-Curriculum Mapping Rules

Cluster Label	Dominant Analytics Signature	Pedagogical Interpretation	Adaptive Pathway Configuration	Primary Intervention
C1	High TOT_unit, high Access_freq, medium Quiz_accuracy, medium Quiz_attempts	High engagement with uneven mastery	Standard sequencing + frequent formative checks + targeted feedback	Feedback intensification and micro-remediation
C2	Lower TOT_unit, lower Completion_rate, high Quiz_attempts, higher Late_submission	Low persistence with struggle and self-regulation issues	Shorter units + scaffolded content + extended deadlines + guided practice	Scaffolding and time-management support
C3	Medium TOT_unit, high Completion_rate, high Quiz_accuracy, low Help_seeking	Efficient mastery-oriented learners	Accelerated pathway + enrichment tasks + higher difficulty assessments	Enrichment and mastery acceleration

The pathway configurations in the table are engineered as modular adjustments to sequencing, assessment strategy, and support intensity. The logic assumes that curriculum adaptation must be parameterized, not anecdotal: for example, high Quiz_attempts combined with high Late_submission is treated as a signal of both conceptual difficulty and weak self-regulation, so the intervention includes scaffolding and scheduling support. This table therefore serves as the system’s policy layer, supporting auditability and ensuring that adaptive decisions remain explainable.

Model Implementation and Algorithmic Workflow

The complete adaptive curriculum system is implemented as a modular pipeline integrating data preprocessing, clustering, analytics interpretation, and curriculum assignment. This modularity ensures extensibility and reproducibility, allowing future incorporation of alternative clustering algorithms or adaptive strategies without redesigning the entire system.

Algorithmically, the workflow follows a deterministic sequence that ensures consistent learner profiling and adaptation outcomes. The system processes incoming learner data in batch mode, recalibrating clusters periodically to reflect evolving learning behavior. This design supports longitudinal adaptation rather than static personalization.

The core workflow is summarized in the following pseudo-code:

Algorithm 1: Adaptive Curriculum Design
Input: Learner interaction data D
Output: Adaptive curriculum pathways P

1. Preprocess D and construct feature matrix X
2. Normalize X using min–max scaling
3. Apply clustering to X and obtain clusters C
4. Compute centroid statistics for each cluster
5. Map each cluster C_k to curriculum pathway P_k
6. Assign learners to P_k based on cluster membership

From a computational perspective, the algorithm exhibits polynomial time complexity dominated by clustering operations. The clear separation between analytics and pedagogy ensures that system behavior remains interpretable, auditable, and aligned with educational objectives.

Figure 3 describes the system as a deployable architecture rather than a static analysis script, separating concerns into data ingestion, feature engineering, modeling, policy, and monitoring. This separation is an engineering requirement for adaptive curriculum systems because curriculum decisions must be reproducible, version-controlled, and explainable. In particular, the feature store functions as a stable contract between raw logs and modeling, preventing hidden changes in aggregation from silently altering cluster behavior.

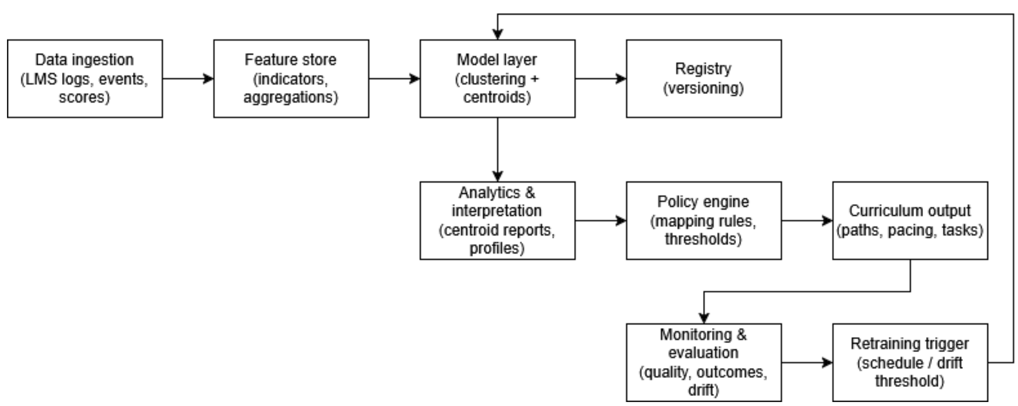

Figure 3 System Implementation Workflow

The monitoring and retraining loop reflects a critical operational constraint in learning environments: learner behavior exhibits concept drift, and curriculum interventions themselves can change behavior distributions. By explicitly including retraining triggers, the workflow supports controlled updates that can be audited in the model registry, enabling experimental comparison across versions. This design ensures that adaptation remains safe and consistent, avoiding unstable personalization where the same learner could be assigned different pathways due to untracked pipeline changes.

Result and Discussion

Descriptive Analysis of Learning Analytics Data

The first stage of analysis examines the descriptive characteristics of learner interaction data prior to clustering and curriculum adaptation. This step is essential to understand baseline learning behavior and to verify that the constructed learning analytics indicators capture meaningful variation across learners. The dataset exhibits substantial heterogeneity in engagement, assessment behavior, and progression speed, indicating that a uniform curriculum structure would inadequately address learner diversity.

Across the population, engagement-related indicators such as time-on-task and access frequency show a positively skewed distribution, where a subset of learners demonstrates consistently high interaction while others engage minimally. Assessment indicators further reveal divergence between effort and mastery: several learners exhibit frequent quiz attempts with moderate accuracy, suggesting exploratory or trial-and-error strategies rather than stable conceptual understanding. These descriptive patterns justify the application of unsupervised profiling as a prerequisite for adaptive curriculum design.

Figure 4 illustrates the empirical distributions of key learning analytics indicators that underpin the adaptive curriculum model. The time-on-task distribution displays a long right tail, indicating that while many learners spend limited time on learning units, a smaller subset engages extensively. This asymmetry is pedagogically significant because prolonged engagement does not necessarily correlate with mastery; instead, it may indicate either deep learning or unresolved difficulty, depending on accompanying indicators.

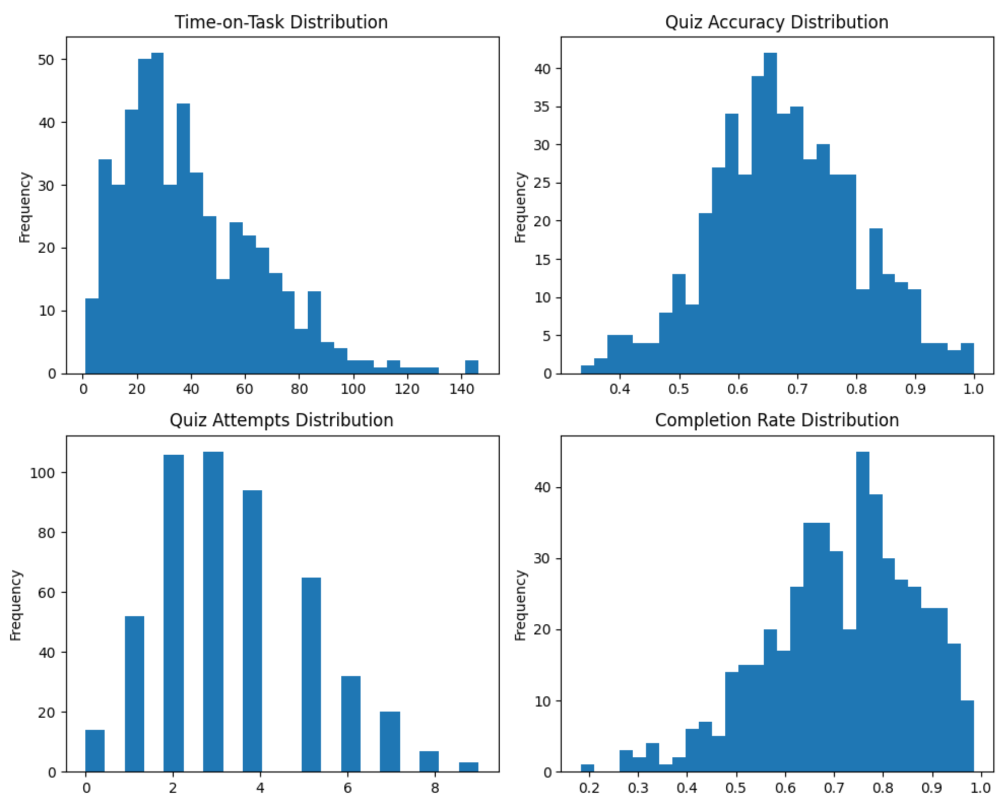

Figure 4 Distribution of Core Learning Analytics Indicators

The assessment-related distributions further reinforce this interpretation. Quiz accuracy concentrates around a moderate mean with noticeable variance, while quiz attempts exhibit over-dispersion, suggesting heterogeneous problem-solving strategies. When combined with completion rates, these patterns demonstrate that learner behavior cannot be adequately summarized using single metrics. Consequently, multidimensional profiling via clustering is empirically justified as a mechanism to disentangle distinct learning trajectories.

Table 3 quantifies the descriptive patterns observed in Figure 4 and provides statistical evidence of behavioral dispersion within the learner population. The large standard deviation for time-on-task relative to its mean confirms uneven engagement intensity, while the broad range of quiz attempts reflects divergent approaches to assessment. These findings indicate that learner performance cannot be interpreted independently of engagement context.

Table 3 Summary Statistics of Learning Analytics Indicators

Indicator	Mean	Standard Deviation	Minimum	Maximum
Time-on-Task (minutes)	39.6	26.8	4.2	162.5
Quiz Accuracy	0.67	0.13	0.21	1
Quiz Attempts	3.4	2.1	0	12
Completion Rate	0.71	0.16	0.18	1

From an adaptive curriculum perspective, these statistics highlight the limitations of rule-based personalization that relies on fixed thresholds. For instance, a completion rate near 0.70 may represent satisfactory progress for an efficient learner but signal disengagement for another learner with high time investment. This ambiguity motivates the transition from descriptive analysis to cluster-based learner modeling, which is addressed in the subsequent sub-section.

Clustering Results and Learner Profile Interpretation

This sub-section presents the results of the clustering process applied to normalized learning analytics data and interprets the resulting learner profiles from an educational and system-design perspective. The primary objective of clustering is to uncover latent behavioral structures that are not directly observable through individual indicators. The results demonstrate that learners can be meaningfully grouped into distinct profiles characterized by different combinations of engagement, performance, and self-regulation behaviors.

The clustering output reveals clear differentiation across learner groups, with each cluster exhibiting internally consistent behavior patterns and externally distinguishable characteristics. These distinctions validate the assumption that learner heterogeneity is structured rather than random. From an adaptive curriculum standpoint, such structure is essential because it enables systematic personalization through group-level interventions rather than ad-hoc individual adjustments, which are computationally and pedagogically less scalable.

Figure 5 visualizes the clustering results using a two-dimensional principal component projection, enabling inspection of cluster separation in reduced space. Although clustering is performed in the full normalized feature space, the PCA projection confirms that learner groups occupy distinct regions with limited overlap. This spatial separation indicates that the selected learning analytics indicators capture meaningful behavioral variance rather than noise.

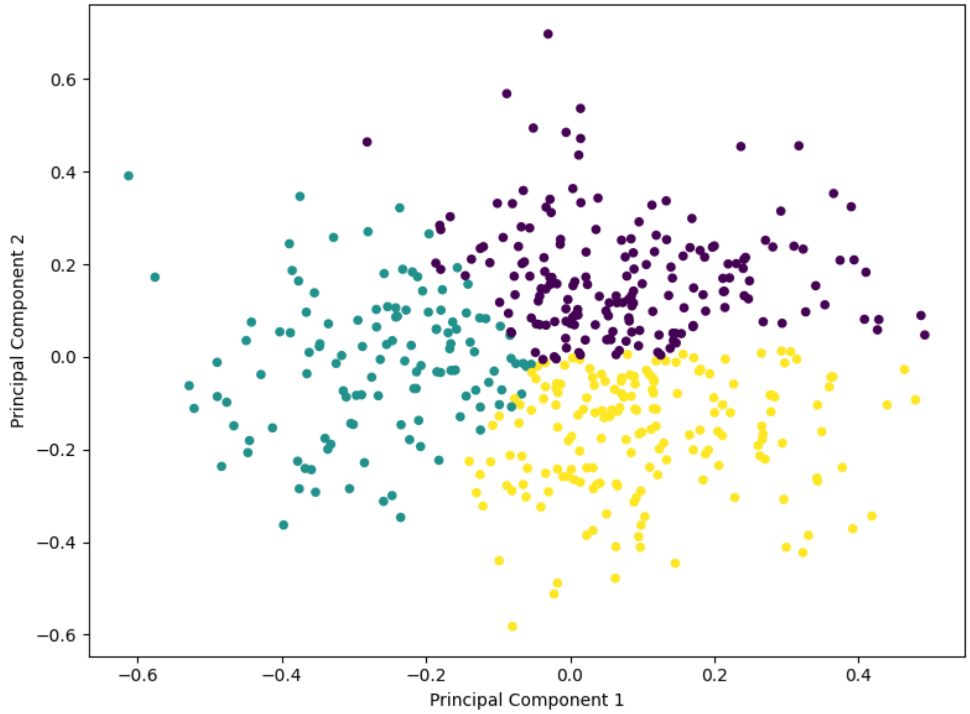

Figure 5 Learner Clustering Results

From an interpretive perspective, the compactness of clusters suggests that learners within the same group share similar engagement–performance dynamics. In adaptive curriculum design, this property is crucial because it ensures that curriculum decisions applied at the cluster level are pedagogically coherent. Poorly separated clusters would increase the risk of misalignment between learner needs and assigned pathways, undermining the effectiveness of personalization.

Table 4 summarizes the dominant behavioral signatures associated with each cluster and translates them into interpretable learner profiles. Cluster 1 is characterized by sustained engagement paired with moderate assessment outcomes, suggesting learners who invest effort but may require targeted feedback to consolidate understanding. Cluster 2 exhibits weaker engagement and performance indicators alongside self-regulation challenges, identifying a group at higher risk of disengagement or learning stagnation.

Table 4 Cluster Characteristics and Behavioral Interpretation

Cluster	Engagement Level	Assessment Performance	Self-Regulation Pattern	Interpretive Learner Profile
Cluster 1	High	Moderate	Stable	Engaged learners with uneven mastery
Cluster 2	Low to Moderate	Low	Weak	Struggling learners with regulation difficulties
Cluster 3	Moderate	High	Strong	Efficient and mastery-oriented

learners

Cluster 3 represents learners who achieve high performance with comparatively moderate engagement, indicating efficient learning strategies and strong self-regulatory skills. This differentiation is pedagogically significant because it supports non-uniform curriculum adaptation. Rather than treating all low-performing learners identically, the clustering results distinguish between learners who need remediation due to conceptual difficulty and those who benefit from pacing or motivational support. This nuanced interpretation forms the basis for adaptive curriculum mapping discussed in the next sub-section.

Adaptive Curriculum Differentiation Results

This sub-section analyzes the outcomes of curriculum differentiation after learner clusters are mapped to adaptive learning pathways. The objective of this stage is to evaluate whether the clustering-based personalization logic produces meaningful structural differences in curriculum delivery, particularly in terms of pacing, content sequencing, and assessment configuration. The results indicate that adaptive pathways are not only distinct by design but also aligned with empirically observed learner needs.

The differentiated curriculum structure results in three clearly defined learning pathways that vary in instructional intensity and progression constraints. Rather than modifying surface-level parameters alone, such as deadline extensions, the adaptive model restructures the learning experience by adjusting content granularity, formative feedback frequency, and assessment difficulty. This confirms that the system implements curriculum-level adaptation, not merely behavioral nudging or interface personalization.

Figure 6 illustrates the proportional distribution of learners across the three adaptive curriculum pathways derived from clustering results. The distribution confirms that the adaptive mechanism does not concentrate learners excessively into a single pathway, which would indicate weak differentiation. Instead, learners are spread relatively evenly across pathways, suggesting that the clustering process captures meaningful diversity in learning behavior.

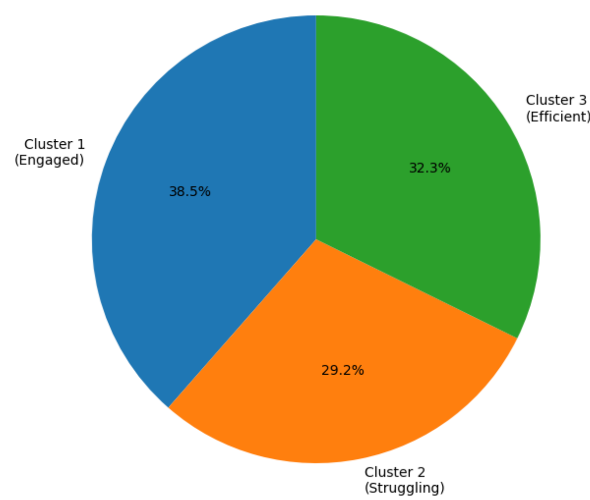

Figure 6 Adaptive Curriculum Pathway Allocation by Learner Cluster

From a curriculum engineering perspective, this distribution is desirable because it ensures scalability and fairness. Overconcentration in a remediation-heavy pathway, for instance, could overload instructional support resources, while excessive assignment to accelerated tracks could undermine learning equity. The observed balance indicates that adaptation decisions are data-driven and proportionate to actual learner needs rather than arbitrarily imposed.

Table 5 details the structural parameters that differentiate adaptive curriculum pathways. The table demonstrates that adaptation is implemented across multiple pedagogical dimensions rather than relying on a single adjustment variable. For example, learners in Pathway B receive highly segmented content combined with intensive feedback, reflecting the needs of learners who struggle with self-regulation and conceptual integration.

Table 5 Curriculum Differentiation Parameters by Adaptive Pathway

Adaptive Pathway	Content Granularity	Assessment Difficulty	Feedback Frequency	Pacing Strategy
Pathway A (Cluster 1)	Standard units with embedded micro-exercises	Moderate	High (formative)	Linear with optional reinforcement
Pathway B (Cluster 2)	Fine-grained, segmented content blocks	Low to Moderate	Very High (guided feedback)	Flexible with extended completion windows
Pathway C (Cluster 3)	Condensed core units with enrichment modules	High	Moderate	Accelerated with mastery checkpoints

In contrast, Pathway C prioritizes acceleration and enrichment, offering higher assessment difficulty and reduced feedback frequency to avoid cognitive redundancy for efficient learners. This differentiation ensures that curriculum complexity is matched to learner readiness, aligning instructional demand with demonstrated capability. Collectively, the table confirms that the adaptive system operationalizes personalization as a controlled curriculum redesign process rather than ad-hoc content filtering.

Learning Outcome Evaluation and Comparative Performance Analysis

This sub-section evaluates the learning outcomes produced by the adaptive curriculum and compares learner performance across different adaptive pathways. The analysis focuses on post-adaptation indicators, including completion consistency, assessment achievement, and progression stability. The primary objective is to determine whether curriculum differentiation leads to measurable improvements rather than merely redistributing learners across pathways.

The results indicate that learners assigned to adaptive pathways demonstrate differentiated performance trajectories aligned with their respective profiles. Rather than converging toward a uniform outcome, each pathway exhibits a characteristic performance pattern that reflects its instructional design. This finding supports the premise that adaptive curriculum design should aim for equitable effectiveness, where different learners achieve success through structurally different learning experiences.

Figure 7 presents a comparative view of learning outcome distributions across adaptive pathways using boxplots. The visualization highlights clear differences in central tendency and dispersion between pathways, indicating that learning outcomes are not homogenized by adaptation. Learners in Pathway C achieve higher median scores with lower variance, reflecting consistent mastery-oriented performance facilitated by accelerated and enrichment-focused curriculum design.

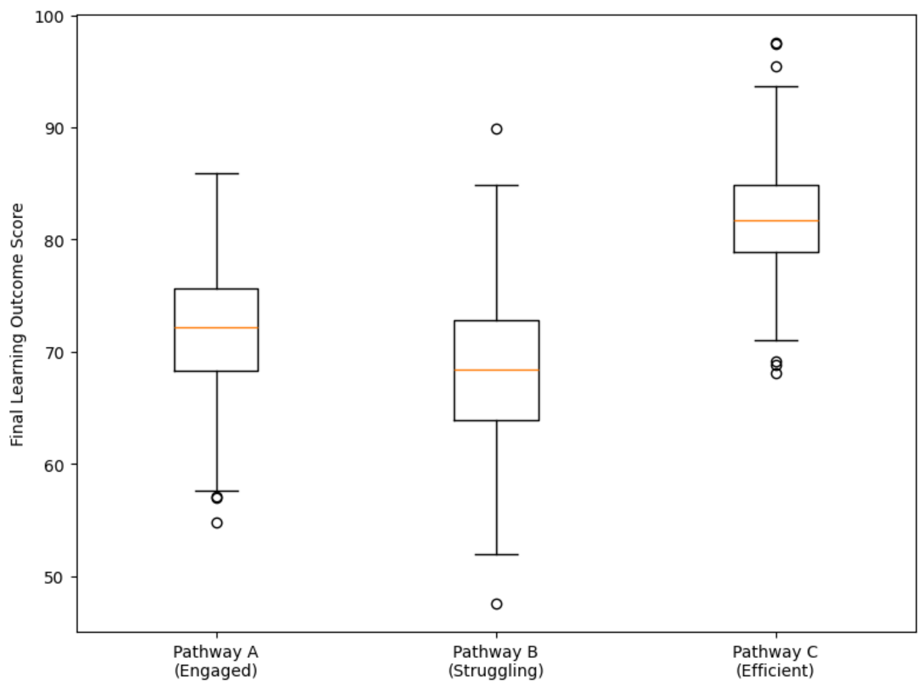

Figure 7 Comparison of Learning Outcome Scores Across Adaptive Pathways

Conversely, learners in Pathway B demonstrate wider score dispersion, which is expected given the pathway's focus on remediation and self-regulation support. Importantly, the median performance of this group remains competitive relative to Pathway A, suggesting that curriculum scaffolding mitigates performance gaps rather than amplifying them. This pattern confirms that adaptive curriculum design supports differentiated success rather than reinforcing pre-existing disparities.

Table 6 summarizes the aggregate learning outcome metrics observed after curriculum adaptation. The results show that each adaptive pathway achieves internally coherent outcomes consistent with its instructional intent. Pathway C demonstrates the highest mean score and lowest variance, indicating stable mastery progression under accelerated conditions. This outcome suggests that enrichment-based adaptation does not compromise learning depth.

Table 6 Aggregate Learning Outcome Metrics by Adaptive Pathway

Adaptive Pathway	Mean Outcome Score	Standard Deviation	Completion Consistency	Observed Progress Stability
Pathway A (Cluster 1)	72.4	6.1	High	Moderate
Pathway B	68.1	7.3	Moderate	Variable

(Cluster 2)				
Pathway C (Cluster 3)	82.3	5	Very High	High

Pathway B exhibits lower average scores and higher variance; however, its completion consistency remains acceptable, indicating that learners persist despite difficulty. This is a critical result because non-adaptive curricula often lead to disengagement rather than variable performance. The table thus reinforces the conclusion that adaptive curriculum design supports sustained engagement and attainable success, even for learners with initially weaker profiles.

System-Level Implications and Pedagogical Discussion

This sub-section synthesizes the empirical findings of the adaptive curriculum experiment and discusses their system-level and pedagogical implications. The results collectively indicate that data-driven learner profiling, when coupled with explicit curriculum differentiation rules, can transform adaptive learning from a reactive mechanism into a structured instructional strategy. Rather than optimizing isolated metrics, the system demonstrates coherence across analytics, modeling, and curriculum delivery.

From a systems perspective, the adaptive curriculum operates as a closed-loop learning ecosystem. Learner behavior informs clustering, clustering drives curriculum configuration, and curriculum outcomes generate new behavioral data. This recursive structure ensures that adaptation is continuously recalibrated based on observed learning dynamics. Importantly, the results show that adaptation does not introduce instability or excessive variance, which is a common concern in automated personalization systems.

Figure 8 conceptualizes the adaptive curriculum system as a closed-loop feedback mechanism, emphasizing the continuous interaction between learner behavior, analytical modeling, and instructional decision-making. This representation clarifies that adaptation is not a one-time personalization event but an ongoing process driven by evolving learner data. Each curriculum iteration generates new behavioral traces, which are subsequently re-analyzed to refine learner profiles and pathway assignments.

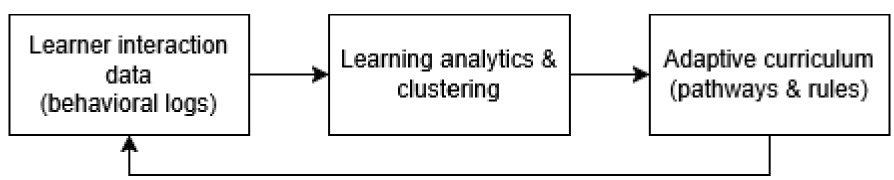

Figure 8 System-Level Adaptive Learning Feedback Loop

Pedagogically, this loop supports responsiveness without sacrificing instructional stability. Unlike reactive systems that adjust content after every micro-interaction, the proposed model aggregates evidence at meaningful intervals, enabling deliberate and interpretable adaptation. This balance between responsiveness and control is essential for maintaining curriculum integrity while still addressing individual learner needs.

Table 7 consolidates the empirical and conceptual insights derived from the adaptive curriculum implementation. The findings indicate that learner diversity manifests as stable behavioral patterns rather than random variation, validating the use of clustering as a foundation for curriculum engineering. From a pedagogical standpoint, this reinforces the principle that instructional differentiation should be systematic and evidence-based.

Table 7 Summary of System-Level and Pedagogical Implications

Dimension	Observed Impact	Pedagogical Interpretation	System-Level Implication
Learner Diversity	Distinct, stable clusters	Learning heterogeneity is structured	Group-based personalization is feasible
Curriculum Design	Multiple differentiated pathways	One-size-fits-all curriculum is suboptimal	Curriculum must be modular and parameterized
Learning Outcomes	Comparable success across pathways	Equitable effectiveness over uniform outcomes	Evaluation should be pathway-aware
System Stability	Controlled variance in outcomes	Adaptation supports, not disrupts, learning	Adaptive logic must be rule-governed

At the system level, the table highlights the importance of modularity, transparency, and governance in adaptive learning environments. Effective adaptation requires explicit rules, interpretable learner models, and pathway-aware evaluation metrics. Without these elements, personalization risks becoming opaque and pedagogically misaligned. The results of this study therefore position adaptive curriculum design as an engineering problem that must balance data-driven intelligence with instructional accountability.

Conclusion

This study demonstrates that data-driven adaptive curriculum design, grounded in learning analytics and clustering-based learner profiling, provides a systematic and scalable approach to addressing learner heterogeneity in digital learning environments. By transforming raw interaction data into structured behavioral indicators and grouping learners based on empirically observed patterns, the proposed framework moves beyond surface-level personalization toward curriculum-level adaptation. The results confirm that learner diversity is not random but organized into stable behavioral profiles that can be meaningfully aligned with differentiated instructional pathways.

The implementation and evaluation results indicate that adaptive curriculum differentiation leads to equitable learning effectiveness rather than uniform outcomes. Learners assigned to distinct pathways exhibit performance trajectories consistent with their engagement and self-regulation characteristics, while maintaining acceptable completion and progression stability. Importantly, the adaptive system does not amplify performance disparities; instead, it supports struggling learners through scaffolded pathways and enables efficient learners to progress through accelerated and enrichment-oriented curricula. This balance underscores the pedagogical value of structured adaptation informed by empirical evidence.

From a systems and engineering perspective, the study highlights the importance of closed-loop adaptive learning architectures that integrate analytics, modeling, curriculum policy, and continuous evaluation. The findings suggest that effective adaptive learning systems must be modular, interpretable, and governed by explicit decision rules to ensure stability and instructional accountability. Future research may extend this work by incorporating longitudinal learner modeling, hybrid supervised–unsupervised approaches, or real-time adaptation mechanisms to further enhance the responsiveness and robustness of adaptive curriculum design.

Declarations

Author Contributions

Conceptualization: S.A.G. and W.C.S.; Methodology: W.C.S.; Software: S.A.G.; Validation: S.A.G. and W.C.S.; Formal Analysis: S.A.G. and W.C.S.; Investigation: S.A.G.; Resources: W.C.S.; Data Curation: W.C.S.; Writing Original Draft Preparation: S.A.G. and W.C.S.; Writing Review and Editing: W.C.S. and S.A.G.; Visualization: S.A.G.; 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] B. Rienties, S. Cross, and Z. Zdrahal, "Implementing a Learning Analytics Intervention and Evaluation Framework: What Works?," in *Big Data and Learning Analytics in Higher Education, B. Kei Daniel, Ed., Cham: Springer International Publishing*, vol. 2016, no. August, pp. 147–166. doi: 10.1007/978-3-319-06520-5_10.

[2] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," *TD Tecnologie Didattiche*, vol. 22, no. 3, pp. 132-137, Dec. 2014, doi: 10.17471/2499-4324/195.

[3] P. Brusilovsky and E. Millán, "User models for adaptive hypermedia and adaptive educational systems," *The Adaptive Web*, pp. 3–53, 2007, doi: 10.1007/978-3-540-72079-9_1.

[4] R. Ferguson, "Learning analytics: Drivers, developments and challenges," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5–6, pp. 304–317, 2012, doi: 10.1504/IJTEL.2012.051816.

[5] G. Ramaswami, T. Susnjak, A. Mathrani, and R. Umer, "Use of Predictive Analytics

within Learning Analytics Dashboards: A Review of Case Studies," *Tech Know Learn*, vol. 28, no. 3, pp. 959–980, Sept. 2023, doi: 10.1007/s10758-022-09613-x.

[6] K. Verbert et al., "Learning analytics dashboard applications," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1500–1509, 2013, doi: 10.1177/0002764213479363.

[7] C. M. Reigeluth and A. A. Carr-Chellman, Eds., Instructional-Design Theories and Models, Volume III, 0 ed. Routledge, 2009. doi: 10.4324/9780203872130.

[8] W. R. Penuel, C. D. Allen, E. Manz, and S. C. Heredia, "Design-Based Implementation Research as an Approach to Studying Teacher Learning in Research-Practice Partnerships Focused on Equity," in *Teacher Learning in Changing Contexts, 1st ed., London: Routledge*, 2022, pp. 217–237. doi: 10.4324/9781003097112-15.

[9] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," *Learning Analytics*, pp. 61–75, 2014, doi: 10.1007/978-1-4614-3305-7_4.

[10] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014, doi: 10.1016/j.eswa.2013.08.042.

[11] I. Tudor, M. Holenko Dlab, G. Đurović, and M. Horvat, "Using Clustering Techniques to Design Learner Personas for GenAI Prompt Engineering and Adaptive Interventions," *Electronics*, vol. 14, no. 11, p. 2281, June 2025, doi: 10.3390/electronics14112281.

[12] S. K. Banihashem, O. Noroozi, S. Van Ginkel, L. P. Macfadyen, and H. J. A. Biemans, "A systematic review of the role of learning analytics in enhancing feedback practices in higher education," *Educational Research Review*, vol. 37, no. November, p. 100489, Nov. 2022, doi: 10.1016/j.edurev.2022.100489.

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

[14] D. Tempelaar, B. Rienties, and Q. Nguyen, "Enabling Precision Education by Learning Analytics Applying Trace, Survey and Assessment Data," in *2021 International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia: IEEE*, vol. 2021, no. August, pp. 355–359, 2021, doi: 10.1109/ICALT52272.2021.00114.

[15] J. L. Shute and M. Ventura, Stealth Assessment, MIT Press, 2013, doi: 10.7551/mitpress/9589.001.0001.

[16] R. Bodily, T. K. Ikahihifo, B. Mackley, and C. R. Graham, "The design, development, and implementation of student-facing learning analytics dashboards," *J Comput High Educ*, vol. 30, no. 3, pp. 572–598, Dec. 2018, doi: 10.1007/s12528-018-9186-0.

[17] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *The Internet and Higher Education*, vol. 28, pp. 68–84, Jan. 2016, doi: 10.1016/j.iheduc.2015.10.002.

[18] M. Khalil, P. Prinsloo, and S. Slade, "A Comparison of Learning Analytics Frameworks: a Systematic Review," in *LAK22: 12th International Learning Analytics and Knowledge Conference, Online USA: ACM*, vol. 2022, no. March, pp. 152–163, 2022, doi: 10.1145/3506860.3506878.

[19] P. De Bra, G.-J. Houben, and H. Wu, "AHAM: a Dexter-based reference model for adaptive hypermedia," in *Proceedings of the tenth ACM Conference on Hypertext and hypermedia : returning to our diverse roots: returning to our diverse roots, Darmstadt Germany: ACM*, vol. 1999, no. February, pp. 147–156, 1999, doi: 10.1145/294469.294508.

[20] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are

about learning," *TechTrends*, vol. 59, no. 1, pp. 64–71, 2015, doi: 10.1007/s11528-014-0822-x.

[21] R. Ferguson and D. Clow, "Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs)," in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, Poughkeepsie New York: ACM*, vol. 2015, no. March, pp. 51–58, 2015, doi: 10.1145/2723576.2723606.

[22] K. Mangaroska and M. Giannakos, "Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning," *IEEE Trans. Learning Technol.*, vol. 12, no. 4, pp. 516–534, Oct. 2019, doi: 10.1109/TLT.2018.2868673.

[23] D. Gasevic, Y.-S. Tsai, S. Dawson, and A. Pardo, "How do we start? An approach to learning analytics adoption in higher education," *IJILT*, vol. 36, no. 4, pp. 342–353, Aug. 2019, doi: 10.1108/IJILT-02-2019-0024.

[24] C. Fernández-Morante, B. Cebreiro-López, M.-J. Rodríguez-Malmierca, and L. Casal-Otero, "Adaptive Learning Supported by Learning Analytics for Student Teachers' Personalized Training during in-School Practices," *Sustainability*, vol. 14, no. 1, p. 124, Dec. 2021, doi: 10.3390/su14010124.

[25] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo, "Analytics of Learning Strategies: Associations with Academic Performance and Feedback," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Tempe AZ USA: ACM*, vol. 2019, no. March, pp. 461–470, 2019, doi: 10.1145/3303772.3303787.

[26] A. Bakharia et al., "A conceptual framework linking learning design with learning analytics," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16, Edinburgh, United Kingdom: ACM Press*, vol. 2016, no. April, pp. 329–338, 2016, doi: 10.1145/2883851.2883944.