

CLUSTERING IN EDUCATIONAL DATA MINING: A CRITICAL REVIEW OF REPRESENTATION, VALIDATION, ANALYTICAL UTILITY AND RESPONSIBLE DECISION-MAKING

Nikhil Thakur

Research Scholar

Department of Computer Science and Engineering, IEC University, Baddi, Solan, Himachal Pradesh, India

Ravinder Singh Madhan

Associate Professor

Department of Computer Science and Engineering, IEC University, Baddi, Solan, Himachal Pradesh, India

Corresponding author: Nikhil Thakur

ABSTRACT

Clustering is widely used in educational data mining (EDM) to identify learner profiles, behavioural trajectories, risk groups and institutional typologies, yet compact groups are often treated as automatically stable, educationally meaningful and suitable for intervention. This critical integrative review synthesises 54 studies published through August 2026 and examines how analytical unit, temporal aggregation, feature representation, scaling, similarity, algorithm choice and missing-data treatment shape educational evidence. Partitioning, hierarchical, density-based, fuzzy, model-based, graph-based and temporal approaches are compared across learner-, course- and institution-level settings. The synthesis shows that internal validity, reproducibility, external usefulness and decision legitimacy answer different questions and cannot be reduced to a single claim of model quality. Longitudinal, cluster-assisted and explainable systems illustrate movement from descriptive profiling towards prediction and recommendation, but require out-of-sample baselines, leakage controls, stakeholder interpretation, fairness assessment and human oversight. The review introduces a Four-Evidence Framework comprising representation validity, structural credibility, analytical utility and decision legitimacy. A consequence-proportionate ladder links the evidence required to the intended use, from exploratory description to consequential support. These contributions provide a practical standard for determining when a cluster solution is informative, should remain exploratory or is sufficiently supported for responsible educational action.

Keywords: educational data mining; clustering; cluster validation; explainable artificial intelligence; responsible decision support.

1. INTRODUCTION

1.1 From Educational Data to Computational Educational Ecosystems

Digital education has shifted from isolated record keeping toward interconnected computational environments. Student Information Systems (SIS) capture enrolment, attendance and administrative records; Learning Management Systems (LMS) record repeated interaction with learning resources; online courses generate temporal and clickstream traces; and multimodal environments add audio, video, gaze, gesture and sensor data. These sources differ in scale, frequency and meaning, so educational analytics depends

as much on integration and representation as on the final learning algorithm [1, 2]. Reviews of higher-education analytics also show that learner-facing applications remain more common than administrative and institutional decision support, despite the growing availability of institution-level data [3]. EDM and Learning Analytics (LA) provide the main computational settings for analysing these data. Contemporary work includes classification, regression, clustering, association and sequence mining, process analysis, deep learning, and explainable AI [4, 5]. Within this broader landscape, clustering is distinctive because it can reveal structure before a target label is defined. This is useful when learners, courses or institutions differ in ways that are not captured by existing categories.

1.2 Why Clustering Matters

Clustering groups observations according to similarity in the selected feature space. Partitioning, hierarchical, density-based, fuzzy, probabilistic and graph-based methods do not search for the same geometry, and their behaviour changes with scaling, dimensionality, outliers and feature construction [6]. The educational value of a partition therefore cannot be judged only by whether an algorithm converges. The groups must also be stable, interpretable, and relevant to the educational question. This distinction is especially important because much of educational machine learning remains prediction-oriented. Clustering can instead expose learner profiles, behavioural trajectories, or institutional typologies and can later be incorporated into supervised models. Recent synthesis of cluster-assisted student-risk prediction illustrates this complementary role while also showing challenges in generalisability and evaluation [7]. Clustering identifies patterns; it does not by itself predict an outcome or justify an intervention.

1.3 Research Gap and Contribution

Existing reviews address EDM and learning analytics broadly, educational big-data architectures, machine-learning methods, higher-education analytics and cluster-assisted prediction, but these strands are often examined separately. The unresolved problem is not a shortage of clustering algorithms. It is the absence of an integrated account of how data representation, algorithmic assumptions, validation evidence, educational interpretation and downstream use jointly determine whether a cluster solution is defensible. This gap is especially important when profiles move from exploratory description into early warning, personalised support, institutional comparison or resource allocation.

This review treats clustering as one stage in a data-to-decision pathway and makes three contributions. First, it integrates data architecture, representation, cluster discovery, validation, predictive integration and governance within one analytical account. Second, it introduces a Four-Evidence Framework that separates representation validity, structural credibility, analytical utility and decision legitimacy. Third, it proposes a consequence ladder: exploratory description may be supported by the first two evidence gates, whereas predictive or consequential use requires progressively stronger out-of-sample, fairness, explainability and oversight evidence. The framework extends beyond learner analytics to course and institutional profiling, where infrastructure, organisational context and resource constraints become central.

1.4 Research Questions

RQ1. How do educational data architecture and feature representation shape the structures recovered by clustering?

RQ2. Which clustering families are used across learner, course and institutional settings, and what assumptions govern their suitability?

RQ3. What evidence is required to establish cluster validity, stability, interpretability and analytical usefulness?

RQ4. How should the evidential threshold change when cluster-derived profiles move from exploration to consequential educational support?

2. REVIEW DESIGN AND LITERATURE SELECTION STRATEGY

2.1 Review Approach

This article adopts a critical integrative review with narrative and framework-building synthesis [8, 9]. The design is appropriate because the evidence spans algorithmic studies, empirical educational applications, review articles, data-architecture research, predictive modelling and scholarship on privacy, fairness and explainability. The objective is analytical integration rather than statistical aggregation: the review asks how methodological choices at one stage of an EDM pipeline constrain the credibility and use of later stages.

2.2 Literature Identification and Selection

Candidate sources were identified iteratively from the four research questions through August 2026. Searches were conducted across SpringerLink, ScienceDirect, IEEE Xplore, the ACM Digital Library, Wiley Online Library, MDPI, the Journal of Educational Data Mining and official EDM proceedings. Query families combined (educational data mining OR learning analytics) with clustering, learner profiling, temporal clustering, cluster validation, stability, reproducibility, cluster-assisted prediction, explainability, fairness, governance and decision support. Backward citation chaining retained foundational algorithms and validity measures; forward citation chaining prioritised recent empirical applications and reviews. Peer-reviewed journal articles and conference papers were eligible when they made a direct technical, empirical or conceptual contribution to the data-to-decision pathway. Non-educational applications, papers that merely mentioned clustering without explaining its analytical role, duplicates and unsupported promotional discussions were excluded. The resulting reference base contains 54 works. Because the purpose is critical framework development rather than exhaustive retrieval or effect-size aggregation, the article does not claim to be a systematic review or meta-analysis and does not present PRISMA-style study-flow counts. The review design and selection framework are summarised in Table 1.

Table 1. Structured Literature Search and Selection Framework

Component	Approach adopted
Review type	Critical integrative review with narrative and framework-building synthesis
Purpose	Connect data architecture, cluster discovery, validation, predictive integration and responsible use
Evidence base	54 recent and foundational peer-reviewed works published through August 2026
Sources searched	SpringerLink, ScienceDirect, IEEE Xplore, ACM DL, Wiley, MDPI, JEDM and official EDM proceedings
Identification	Iterative query families with backward and forward citation chaining
Eligibility	Direct technical, empirical or conceptual relevance to the four research questions
Exclusions	Non-educational applications, incidental mentions of clustering, duplicates and unsupported claims
Extraction frame	Setting, unit, data, representation, method, validation, application, downstream use and governance

Component	Approach adopted
Synthesis logic	Data architecture → representation → clustering → validation → application/prediction → decision/governance

2.3 Evidence Extraction and Synthesis

Each source was examined through a common analytical frame: educational setting and unit of analysis; data source and temporal granularity; feature representation and preprocessing; clustering family and parameter choices; selection of the number of clusters; internal, stability or external validation; educational interpretation; use in downstream prediction or decision support; and privacy, fairness or governance implications. The literature was then organised by its position in the analytical pipeline rather than treated as one homogeneous body of EDM research. The synthesis distinguished cluster formation, cluster credibility and educational usefulness. Greater evidential weight was assigned to studies that defined their analytical unit, justified preprocessing and method selection, reported validation beyond a single favourable index, and explained how profile labels were derived. A partition may be geometrically compact yet unstable, stable yet difficult to interpret, or interpretable yet inappropriate for a consequential decision. Direct comparison of final validity scores across studies was therefore avoided when feature spaces, aggregation levels and validation procedures differed. Representative study-level evidence supporting the synthesis is presented in Table 2.

Table 2. Representative Study-Level Evidence Supporting the Review Synthesis

Study	Setting and data	Analytical role	Contribution and interpretive boundary
Alvarez-Garcia et al. [10]	PISA 2022 student, ICT and performance data	Explainable cluster profiling	Connects ICT conditions with interpretable student profiles; cross-sectional profiles do not establish causal pathways.
Alzahrani et al. [11]	Weekly e-learning engagement records	K-means with label-based validation	Shows that engagement profiles can change over time; patterns remain course- and platform-dependent.
Jin [12]	Vocational-education behavioural data	Behavioural clustering	Differentiates study and participation patterns; activity traces should not be interpreted as direct measures of cognition or motivation.
Song et al. [13]	Multidimensional university-student records	Clustering with subsequent modelling	Produces interpretable group profiles; transfer beyond the originating institutional context requires testing.

Study	Setting and data	Analytical role	Contribution and interpretive boundary
Pecuchova and Drlik [14]	Student digital traces	Clustering within dropout prediction	Uses latent behavioural structure for early-risk modelling; cluster validity and predictive performance require separate evaluation.
Al-Tameemi et al. [15]	Multi-class student-performance datasets	Reduction, K-means and cluster-specific classification	Illustrates hybrid prediction; added value must be tested against a non-clustered baseline with leakage controls.
Hoq et al. [16]	Student programming patterns	Profile clustering with SHAP-based prediction	Links profiles to model explanations; evidence is domain-specific and should be replicated across courses.
Costa et al. [17]	Multi-timepoint Moodle interaction logs	Longitudinal SRL profiling	Demonstrates that profiles vary across course timepoints; fixed, single-time partitions can obscure learner transitions.
Melchor et al. [18]	Administrative and academic dropout data	Comparative clustering processes	Compares alternative processes and interprets original cases near centroids; transformed-space quality does not remove the need for substantive interpretation.
Kisić et al. [19]	Longitudinal project-based learning traces	Clustering, classification and XAI	Links profiles to transparent recommendations; usefulness depends on temporal validation and human-readable rules.
Neffati et al. [20]	Multivariate simulator time series	Temporal-method comparison	Finds that profiles depend on both method and phase selection; temporal context is part of representation validity.
Boccuzzi et al. [21]	Rapid review of AI-based assessment	Ethics and governance	Shows that delegated algorithmic authority raises transparency, fairness, privacy and accountability concerns.

3. EDUCATIONAL BIG DATA ECOSYSTEM AND DATA ARCHITECTURE

Clustering begins with a representation of educational data, not with the educational process itself. Administrative records are comparatively structured and useful for cohort or institution-level analysis; LMS and MOOC logs provide high-frequency behavioural and temporal information; assessment systems provide performance variables; and multimodal environments can add visual, physiological and contextual signals. Each source changes what

similarity means and which patterns an algorithm can discover [2, 3]. The relationship among these sources and analytical stages is mapped in Fig. 1.

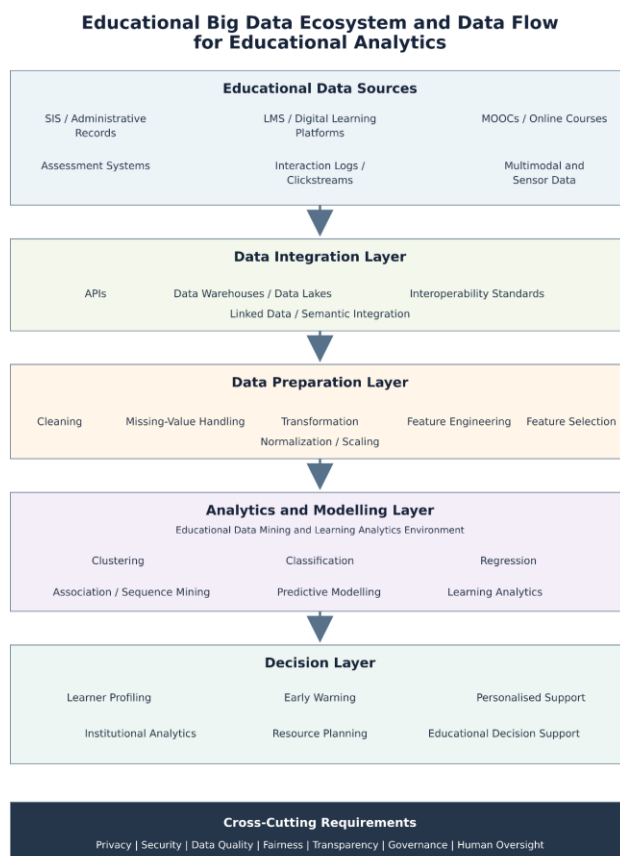


Figure 1. Educational Big Data Ecosystem and Data Flow for Educational Analytics.

Note. Developed by the authors based on the literature synthesised in this review.

3.1 Educational Data Sources and Representation

LMS and online-course data are attractive for clustering because they record repeated activity, but behavioural traces require careful interpretation: clicks, logins and time-on-task indicate observable platform use rather than direct evidence of understanding. Temporal aggregation also matters. Two learners with similar total activity may follow very different weekly trajectories. Multimodal Learning Analytics extends this representation further by integrating channels such as audio, video, gaze, gesture and physiological signals, increasing both analytical richness and dimensionality [22]. AI-based multimodal work has expanded rapidly, but synchronization, missingness, feature extraction and interpretability remain substantial challenges [23]. Administrative and behavioural sources also operate at different levels of granularity. A semester-level attendance measure, a daily LMS count and an event-level clickstream cannot be combined meaningfully without an explicit aggregation strategy. The resulting choice determines whether clustering represents long-term learner characteristics, short-term behaviour, or a mixture of both. For institutional analysis, the same issue appears when school-level resource indicators are combined with learner-level outcomes. Clear definition of the analytical unit is therefore necessary before data from multiple systems are merged.

3.2 Integration, Interoperability and Data Quality

Educational data are frequently distributed across platforms with different identifiers, formats and collection schedules. APIs, data warehouses, learning record stores and semantic integration can connect these systems, but interoperability remains a practical constraint even in mature educational technology environments [24]. Before clustering, inconsistent coding, duplicate identifiers, missing values, platform changes and temporal misalignment must be addressed because unsupervised algorithms lack a known target that would readily expose these errors. Their principal analytical characteristics are summarised in Table 3.

Table 3. Major Educational Data Sources and Their Analytical Characteristics

Data source	Typical characteristics	Clustering / EDM use	Main challenge
Administrative / SIS	Structured, longitudinal records	Cohort, learner and institutional profiling	Coding consistency, linkage, privacy
LMS	High-frequency behavioural events	Engagement and behaviour profiles	Sparse logs, temporal ambiguity
MOOC / clickstream	Large-scale sequential events	Learning-path and dropout profiling	Sequence representation, attrition
Assessment	Structured performance data	Performance and knowledge-state grouping	Validity, missing attempts
Multimodal	High-dimensional, heterogeneous signals	Engagement, affect and collaboration patterns	Synchronisation, feature extraction, privacy
Integrated multi-source	Linked administrative, behavioural and contextual data	Hybrid learner/institution profiles and prediction	Interoperability, missingness, governance

3.3 Architectural Implications for Clustering

The same population can yield different partitions when represented through administrative variables, behavioural sequences, assessments or multimodal features. Integration may improve contextual richness but can also increase dimensionality and allow variables measured on larger scales to dominate distance calculations. Data architecture, feature construction and preprocessing are therefore part of the clustering model itself, not merely preparatory operations.

4. CLUSTERING METHODS IN EDUCATIONAL DATA MINING

No clustering family is universally preferable for educational data. Methods differ in whether they search for compact centres, nested relations, dense regions, probabilistic components or graph structure. Their suitability depends on data type, expected geometry, noise, dimensionality, computational scale and the interpretation required from the final profiles [6]. The principal clustering families considered in this review are organised in Fig. 2.

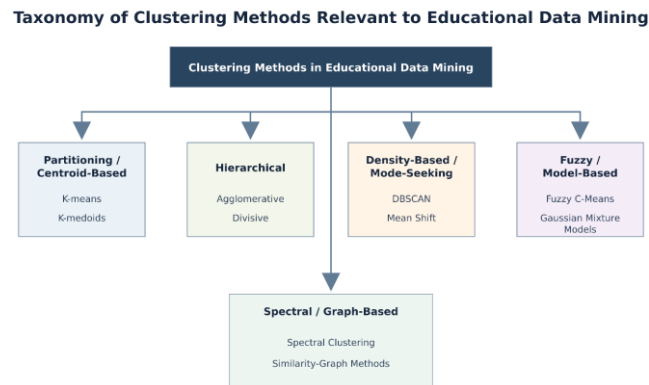


Figure 2. Taxonomy of Clustering Methods Relevant to Educational Data Mining.

Note. Developed by the authors based on the clustering literature synthesised in this review.

4.1 Partitioning and Hierarchical Methods

K-means remains popular because it is simple, scalable and produces interpretable centroids. Its objective is to minimise the within-cluster sum of squared distances:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (1)$$

where C_k is the k th cluster, x_i is an observation assigned to C_k , and μ_k is the centroid of that cluster [25].

The method is sensitive to scale, initialisation, outliers and non-spherical geometry; K-means++ improves centroid initialisation and reduces dependence on arbitrary starting points [26]. K-medoids offers a related alternative using actual observations as representatives, which can be useful when outliers are influential or real cases are easier to interpret than means. Hierarchical clustering builds nested relations rather than one immediate flat partition. Agglomerative methods merge observations or groups, while divisive approaches split larger groups. The dendrogram is useful for exploratory structure, but the result depends on the chosen distance and linkage rule; single, complete, average and Ward-type linkages can produce substantially different hierarchies [27].

4.2 Density, Fuzzy, Model-Based and Graph Methods

DBSCAN forms clusters from locally dense neighbourhoods and can retain sparse observations as noise, making it useful when irregular shapes or atypical cases are expected [28]. Mean Shift instead follows local density modes and does not require a predefined number of clusters, although bandwidth selection and computational cost become important in large or high-dimensional data [29]. Fuzzy C-Means represents overlap by assigning degrees of membership rather than one exclusive label [30]. Gaussian mixture models provide probabilistic membership under a distributional model, typically estimated through the Expectation-Maximization framework [31]. Spectral clustering uses the eigen structure of a similarity graph and is useful for non-linear or relational structure, including collaborative and interaction networks, but it depends heavily on graph construction and can be computationally demanding [32].

4.3 Feature Representation, Scaling and Similarity

The algorithm only sees the features supplied to it. Variables such as grades, attendance, LMS counts and time measures can differ sharply in scale, while redundant variables can unintentionally give one construct disproportionate influence. Standardisation, feature

selection and dimensionality reduction therefore shape the geometry of the solution. Euclidean distance is common for continuous features, cosine similarity can be useful for sparse activity vectors, and mixed data may require measures designed for heterogeneous variable types. These choices should be justified in relation to the educational question rather than applied automatically.

High dimensionality adds another difficulty because pairwise distances can become less discriminative as the number of weakly informative features increases. Feature selection or dimensionality reduction may improve separation and computational efficiency, but transformed components can be harder to explain to educators. The appropriate representation therefore involves a trade-off between statistical compactness and interpretability. Reporting only the clustering algorithm while omitting the representation process conceals one of the main determinants of the resulting partition. Their comparative strengths, limitations and suitable educational uses are consolidated in Table 4.

Table 4. Comparative Characteristics of Clustering Methods for Educational Data

Method	Strength	Main limitation	Suitable educational use	Key consideration
K-means	Efficient; interpretable centroids	Sensitive to scale, initialisation and geometry	Learner or institutional profiling	Scaling, K, stability
K-medoids	Representative real cases; less outlier-sensitive	Higher computational cost	Profiles requiring interpretable exemplars	Distance and sample size
Hierarchical	Nested exploratory structure	Linkage- and distance-sensitive	Exploratory learner/institution similarity	Linkage and dendrogram cut
DBSCAN	Irregular shapes and noise	Sensitive to density parameters	Behavioural anomalies and atypical profiles	Neighbourhood definition
Mean Shift	No predefined K; mode-seeking	Bandwidth and computation	Density-driven behavioural patterns	Bandwidth and dimension
Fuzzy C-Means	Represents overlapping profiles	More complex interpretation	Learners with graded profile membership	Fuzziness and scaling
Gaussian mixture	Probabilistic membership	Distributional assumptions	Latent and uncertainty-aware profiles	Components and covariance
Spectral	Captures complex graph structure	Graph construction and computation	Interaction and collaborative networks	Similarity graph

4.4 Method Selection

Algorithm selection should follow the data and analytical purpose. Compact standardised indicators may suit K-means or medoid-based methods; hierarchical clustering supports nested exploratory structure; density methods are useful when noise or irregular geometry is expected; fuzzy and mixture models represent uncertain boundaries; and spectral methods suit relational structure. Method selection should therefore be reported as an explicit methodological decision rather than as a software default.

5. VALIDATION, STABILITY, AND INTERPRETABILITY OF EDUCATIONAL CLUSTERS

Convergence is not evidence that a partition is credible. Internal geometry, stability under repeated analysis, agreement between alternative partitions, and educational meaning address different questions and should be examined together. Cluster-validity research shows that internal indices can favour different solutions on the same data and should not be treated as ground truth [33].

5.1 Number of Clusters and Internal Validity

For methods requiring K, the elbow in within-cluster variation is a useful exploratory diagnostic but can be subjective. The Gap Statistic compares observed dispersion with a reference distribution and provides a more formal basis for evaluating K [34]. For observation i , the Silhouette coefficient is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

where $a(i)$ is the mean distance from observation i to other observations in its assigned cluster and $b(i)$ is the smallest mean distance to an alternative cluster. This measure evaluates within-cluster cohesion relative to the nearest alternative cluster: values near 1 indicate strong placement, values near zero indicate boundary cases, and negative values suggest possible misassignment [35]. The Davies-Bouldin Index compares within-cluster dispersion with between-cluster separation, with lower values preferred [36]. The Calinski-Harabasz criterion compares between-cluster to within-cluster variation, with larger values indicating stronger separation relative to compactness [37]. Agreement among these measures strengthens evidence, but disagreement should prompt examination of geometry, preprocessing and cluster-size imbalance rather than automatic selection of whichever index is most favourable.

5.2 Stability and Reproducibility

Internal validity describes one fitted partition; stability asks whether comparable structure is recovered after reasonable perturbation. Subsampling approaches compare repeated partitions to determine whether group structure persists beyond one sample [38, 39]. For K-means and related algorithms, different random initialisations test algorithmic stability, while resampling tests sensitivity to the observed data. Temporal stability is a separate requirement. Recent multi-timepoint learner profiling shows that self-regulated-learning patterns can change across course weeks [17], while phase-specific temporal clustering demonstrates that the portion of a sequence selected for analysis can determine whether useful profiles emerge [20]. Stability must therefore be assessed against the kind of variation relevant to the intended claim: seeds, samples, features, timepoints, cohorts or contexts. Agreement between partitions can be quantified with the Adjusted Rand Index (ARI), which corrects pairwise agreement for chance and equals 1 for identical partitions [40]. External validation also requires caution because a reference partition is not necessarily the only defensible grouping of the data; recent work continues to question the assumption that one external measure captures every notion of cluster quality [41].

5.3 Educational Interpretability and Multi-Criterion Validation

A cluster becomes useful only when its defining characteristics can be explained in the educational context. Researchers should report the variables that distinguish groups, cluster sizes, representative centroids or medoids, uncertainty around membership and independent outcomes used to examine external relevance. Comparative dropout research illustrates the value of returning from transformed feature spaces to original observations near centroids so that the partition can be interpreted in educational terms [18]. Cross-national learner profiling

further shows why validation must address contextual sensitivity rather than assuming that one behavioural structure transfers unchanged between settings [42]. A defensible workflow therefore combines justification of K, multiple internal indices, stability, sensitivity analysis, educational interpretation and, where possible, external replication. Validation should also be separated from post-hoc storytelling. Cluster labels ought to be based on reproducible differences in the variables used to form or independently examine the groups, not on convenient names assigned after visual inspection. Where external educational outcomes are available, they can be used to test whether the discovered profiles relate to completion, retention or performance without treating those outcomes as proof that one partition is uniquely correct. This distinction is important in educational settings where more than one grouping can be substantively defensible. The resulting multi-criterion validation approach is summarised in Table 5.

Table 5. Validation, Stability and Interpretation Criteria for Clustering-Based EDM

Criterion	What it assesses	Preferred evidence	Main limitation	Role in EDM
Elbow / WCSS	Change in within-cluster variation	Clear diminishing return as K increases	Subjective when no clear elbow	Initial K screening
Silhouette	Cohesion vs. nearest-cluster separation	Higher values; inspect individual cases	Geometry-sensitive	Partition and case-level assessment
Davies-Bouldin	Compactness relative to separation	Lower values	Representation-sensitive	Compare candidate partitions
Calinski-Harabasz	Between/within dispersion ratio	Higher values	No substantive meaning by itself	Support K selection
Gap Statistic	Observed vs. reference dispersion	Solution exceeding reference expectation	Reference-dependent; extra computation	Formal K assessment
Repeated runs / resampling	Algorithmic and data stability	Recurring membership and structure	Stability ≠ educational validity	Reproducibility
Adjusted Rand Index	Agreement between partitions	Values approaching 1	Agreement, not intrinsic quality	Quantify repeated-solution similarity
Educational interpretation	Substantive meaning of profiles	Coherent feature differences and external relevance	Context-dependent	Connect clusters to educational use

6. APPLICATIONS OF CLUSTERING IN EDUCATIONAL ECOSYSTEMS

Clustering has been used to expose learner heterogeneity, behavioural patterns, academic risk, instructional needs and institutional differences. Its role varies: in some studies the profiles themselves are the final result; in others cluster membership becomes an intermediate representation for prediction or decision support. Keeping these roles separate helps prevent descriptive groups from being presented as causal or predictive findings. The principal application levels, uses and interpretation risks are presented in Table 6.

6.1 Learner, Engagement and Risk Profiling

Learner segmentation can combine academic, demographic, behavioural and technological variables to reveal profiles that are not visible through single measures. Explainable clustering of PISA 2022 data, for example, has been used to relate student profiles to ICT availability and use, performance and expectations [10]. University student profiling has likewise combined clustering with subsequent modelling to organise multidimensional learner information into interpretable groups [13]. Digital learning environments make behavioural clustering especially attractive because activity can be represented over time. Weekly K-means analysis has shown that engagement profiles can change during a course rather than remain fixed at semester level [11]. Multi-timepoint clustering similarly reveals changes in self-regulated-learning profiles across a course [17], and comparative temporal clustering shows that both method and phase selection can alter the learner structure recovered from sequential data [20]. Behavioural clustering in vocational education has also differentiated patterns of study time, participation, discussion and collaboration [12]. These profiles represent observable traces under a chosen temporal representation; they should not be interpreted as direct or permanent measures of cognition, motivation or ability. For academic risk, clustering is most defensible as a way to identify groups with similar trajectories before supervised prediction. Clustering of students' digital traces has been incorporated into early dropout frameworks [14], and recent review evidence confirms increasing use of cluster-assisted student-risk pipelines [7]. The distinction remains important: clustering identifies risk-related structure, whereas supervised models estimate a labelled future outcome.

6.2 Personalised, Course and Institutional Uses

Profiles can support differentiated learning when they identify meaningful combinations of engagement, achievement or support needs, but cluster membership should not become a deterministic intervention rule. Contemporary systems combine clustering with recommendation mechanisms to translate group patterns into student-facing or educator-facing support [43]. Overlapping, uncertain or changing cases still require individual interpretation. The same computational principle can be applied at course, programme, school or institutional level using variables such as infrastructure, access, enrolment, resources and performance. At these higher levels, aggregation, institutional size and contextual differences require careful normalisation and interpretation. Institutional clustering is especially promising where a single ranking obscures different structural pathways to similar outcomes. Two schools may report comparable achievement while differing markedly in connectivity, device access, enrolment pressure or utilisation. A multidimensional typology can preserve those distinctions and support differentiated planning. Yet institutional clusters should not be interpreted as direct measures of quality unless the selected indicators justify that claim; they are profiles of the variables included in the analysis.

6.3 Resource Planning and Decision Support

Decision-support applications use clustering to reduce complex data into a smaller number of interpretable operating conditions or need profiles. Data-mining-based decision-support research demonstrates how pattern discovery can contribute to educational planning when analytical outputs are translated into rules or evidence for decision makers [44]. The appropriate role of clustering is to structure evidence, not to determine an allocation or intervention automatically; cost, equity, policy priorities and local knowledge remain outside many analytical datasets.

Table 6. Applications of Clustering Across Educational Decision Levels

Application	Typical data	Role of clustering	Possible use	Main interpretation risk
Learner profiling	Academic, demographic, survey and ICT variables	Identify multidimensional profiles	Cohort analysis and support	Oversimplified labels
Engagement analysis	LMS logs, clickstreams and submissions	Detect behavioural/temporal patterns	Monitoring and targeted support	Equating activity with learning
Academic risk	Grades, attendance, LMS behaviour	Identify risk-related groups	Input to early-warning models	Confusing clustering with prediction
Adaptive support	Profile and interaction data	Group related instructional needs	Differentiated resources	Uniform intervention within a cluster
Course / institution	Programme, infrastructure and resource indicators	Create structural typologies	Planning and benchmarking	Aggregation and context bias
Hybrid prediction	Cluster features + labelled outcomes	Create latent features or subgroups	Dropout/performance models	Assuming added value without baseline
Decision support	Integrated learner/institutional data	Summarise complex conditions	Advising and resource planning	Treating grouping as autonomous decision

Note. Developed by the authors from the literature synthesised in this review.

7. FROM CLUSTER DISCOVERY TO PREDICTIVE AND DECISION-SUPPORT SYSTEMS

Clustering increasingly functions as an intermediate representation within hybrid EDM systems. Membership labels, distances to centroids, or fuzzy/probabilistic memberships can be added to supervised models, and separate predictive models can be trained within clusters. Such strategies may capture heterogeneity that a single global model overlooks, but they also add parameters, opportunities for overfitting, and potential information leakage. The complete clustering-based data-to-decision pathway is shown in Fig. 3.

7.1 Cluster-Assisted Predictive Modelling

A clear technical distinction should be maintained between unsupervised discovery and supervised estimation. Al-Tameemi et al. [15], for example, combined dimensionality reduction, K-means clustering and cluster-specific classification for student-performance prediction. This kind of design is meaningful only if the added clustering stage is evaluated against an equivalent non-clustered baseline. Improvements should be assessed with metrics appropriate to the outcome, not inferred from cluster quality alone. Hybrid reviews show several recurring designs: cluster labels as features, cluster-specific models, clustering before risk classification and clustering followed by explainable recommendation [7, 19]. Empirical comparisons of clustering processes for early dropout detection further show that results depend on preprocessing, mixed-data handling, method choice and interpretability procedures [18]. These systems should preserve training/evaluation separation across preprocessing, clustering, hyperparameter selection and prediction so that downstream performance is not inflated by leakage.

7.2 Added Value, Explainability and Decision Support

A high Silhouette score does not prove that cluster membership improves prediction, and improved predictive accuracy does not establish that the underlying groups are stable or educationally meaningful. These stages answer different questions. Recent hybrid systems combine clustering with ensemble prediction [45] or with classification and explainable rules that generate personalised recommendations [19]. Such architectures demonstrate technical possibility, not automatic educational value. Explainability becomes more important as model output moves toward action because educators and learners need to understand which profile characteristics and predictive factors contribute to a recommendation [46]. Comparative evaluation should therefore isolate the value added by clustering. A global predictive model, a model augmented with cluster-derived features, and cluster-specific models can be evaluated under the same train-test or cross-validation design. If the hybrid approach yields only a negligible gain, the additional computational and interpretive complexity may not be justified. Conversely, a meaningful improvement should still be examined for stability across cohorts or institutions before the cluster-assisted architecture is treated as generalisable. Decision-support systems should combine model output with interpretable evidence rather than treat predictions as autonomous decisions. Work on educational decision support similarly emphasises extracting patterns and rules that assist decision makers rather than replacing them [44]. This creates a direct transition from computational evaluation to privacy, fairness, accountability and human oversight.

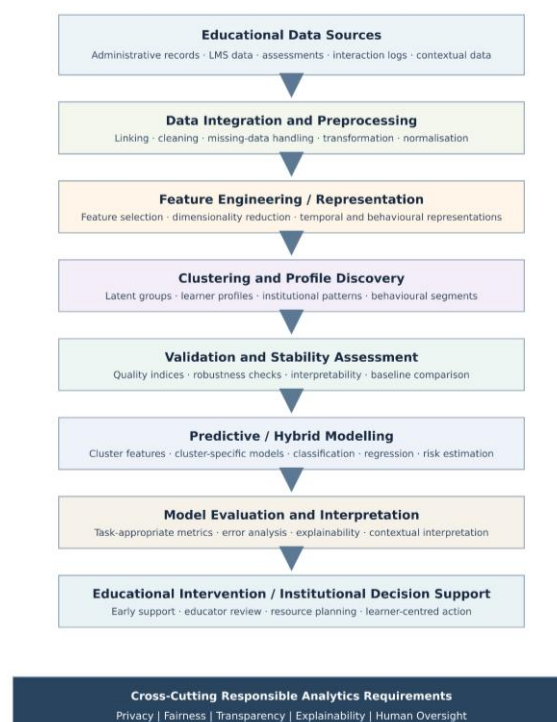


Figure 3. Educational Data-to-Decision Pipeline for Clustering-Based Analytics.

Note. Developed by the authors from the analytical relationships synthesised in this review.

8. RESPONSIBLE EDUCATIONAL ANALYTICS AND GOVERNANCE

Responsible EDM concerns the complete analytical lifecycle. Clustering may be unsupervised, but its results acquire practical consequences once profiles are used for early warning, personalised support, institutional comparison or resource planning. Human-centred learning analytics and AI in education therefore place human judgement, stakeholder

participation and complementary human–AI capabilities alongside model performance [47, 48, 49]. The ethical threshold cannot be detached from the authority delegated to an analytical system.

8.1 Privacy, Bias and Fairness

Linking administrative, behavioural and assessment data can improve analytical richness while increasing privacy risk. Anonymisation is not merely a compliance step because privacy-preserving transformations can change downstream machine-learning utility; educational experiments with k-anonymisation demonstrate this trade-off directly [50]. Data minimisation, controlled access, de-identification and documentation of transformations should therefore precede clustering. Fairness is equally relevant to unsupervised analysis because datasets, selected features, missingness and distance measures can encode contextual disadvantage. Research on demographic features in EDM shows that removing sensitive attributes does not necessarily remove unfairness [51]. Explicit fairness assessment is therefore needed, as illustrated by work on measuring and mitigating unequal model behaviour across student groups [52]. Bias can also differ by data source: institutional, VLE and assessment information have shown different fairness characteristics in student-progress modelling [53]. For clustering, such evidence suggests examining whether disadvantage-related profiles disproportionately contain particular demographic or institutional groups and whether the pattern reflects opportunity, infrastructure or data-generation processes rather than intrinsic learner characteristics.

8.2 Transparency, Explainability and Governance

Transparency requires enough methodological detail to reconstruct the cluster solution: analytical unit, features, preprocessing, scaling, distance measure, parameters, K-selection procedure, validation and stability. Explainability addresses whether the resulting profile or recommendation can be understood by intended users. Work combining student programming profiles with SHAP-based prediction demonstrates how profile information can complement model explanations in educational settings [16]. Governance extends beyond documentation. Auditing perspectives in Learning Analytics argue that fairness, transparency and reliability emerge from interactions among datasets, models, processes and eventual use rather than from one isolated algorithm [54]. Human judgement should therefore remain identifiable throughout model design and deployment. Clusters can prioritise cases for review, organise evidence and support resource planning, but they should not automatically determine learner opportunities or institutional treatment. The level of scrutiny should reflect the consequence of the use. Exploratory clustering used to understand a dataset creates different risks from a profile used to trigger an intervention, assign an opportunity or allocate resources. Rapid-review evidence on AI-based educational assessment identifies transparency, fairness, privacy and accountability as connected concerns when algorithmic systems receive delegated authority [21]. More consequential clustering applications therefore require stronger documentation, subgroup and sensitivity checks, human review, monitoring and mechanisms for contesting or revisiting a recommendation. Responsible analytics is not an ethical appendix to EDM; it changes what counts as adequate technical evidence when model outputs affect people or institutions.

9. FROM CLUSTERS TO CONSEQUENCES: THE FOUR-EVIDENCE FRAMEWORK

9.1 What the Evidence Synthesis Reveals

The synthesis indicates that the central methodological problem in educational clustering is not simply algorithm choice. Feature representation, preprocessing and data quality determine

the geometry that the algorithm receives, while validation determines whether the resulting structure is credible. A method that performs well on standardised school indicators may be inappropriate for sparse clickstreams, mixed administrative records or interaction networks. Recent temporal and comparative studies reinforce this point: profiles can change with the observation window, course phase and preprocessing path [17, 18, 20]. Claims of algorithmic superiority are therefore difficult to interpret when representation, sensitivity and uncertainty are underreported. A second imbalance is the dominance of learner-level analytics over institutional applications. Higher-education reviews continue to report stronger emphasis on learning-oriented uses than administrative decision support [3]. Institution-level clustering could support multidimensional typologies for infrastructure, access and resource planning, but aggregation and context make these applications methodologically distinct from learner segmentation. Computational performance is insufficient: the value of a clustering system depends on whether its profiles are stable, interpretable, analytically useful and appropriate for the decision context. Recent cluster-to-recommendation systems make the translational opportunity visible [19, 43], while also making errors and governance more consequential. The strongest studies keep the stages of the analytical pipeline conceptually separate while testing their dependencies. Data quality is not cluster validation; cluster validation is not predictive evaluation; predictive evaluation is not evidence of intervention effectiveness; and none of these, alone, establishes decision legitimacy. This distinction motivates the Four-Evidence Framework below. The framework is cumulative for consequential use: a study should not compensate for weak representation with strong internal indices, or for unstable clusters with a high downstream accuracy score.

9.2 The Four-Evidence Standard

The framework converts the review synthesis into four evidence gates. Each gate asks a different question, requires different tests and blocks a different form of overclaiming. The standard is intended both as a design tool for researchers and as an appraisal tool for reviewers, editors and institutions. The four evidence gates, minimum requirements and associated decision rules are specified in Table 7.

Table 7. Four-Evidence Standard for Clustering-Based Educational Analytics

Evidence gate	Core question	Minimum evidence	Frequent failure	Decision rule
1. Representation validity	Does the analytical representation match the educational phenomenon?	Defined unit and timeframe; theory-informed features; missingness and scaling reported; sensitivity to aggregation or preprocessing	Treating platform traces or composite scores as direct measures of learning	Do not interpret clusters beyond the constructs and period represented by the data.
2. Structural credibility	Is the partition reproducible, distinguishable and interpretable?	Multiple internal diagnostics; seed/sample/feature stability; uncertainty or boundary cases; educationally grounded labels	Selecting one favourable index or narrating labels after inspection	Call a solution exploratory unless structure survives relevant perturbations and has substantive meaning.

Evidence gate	Core question	Minimum evidence	Frequent failure	Decision rule
3. Analytical utility	Does the cluster information add value outside cluster formation?	External association or out-of-sample evaluation; equivalent non-clustered baseline; leakage control; cohort/context testing	Inferring predictive value from compact geometry or evaluating on data used to create profiles	Retain cluster-assisted complexity only when it adds reliable, practically meaningful value.
4. Decision legitimacy	Is the intended educational use proportionate, fair and accountable?	Purpose specification; subgroup and privacy assessment; understandable rationale; human review; monitoring and contestability	Turning descriptive profiles into fixed labels or automated decisions	Do not use clusters consequentially when affected people cannot understand, review or challenge their use.

Note. The four gates are cumulative when cluster-derived evidence is used for prediction or consequential educational support.

9.3 A Consequence-Proportionate Evidence Ladder

The same cluster solution can be adequate for one purpose and inadequate for another. Exploratory description principally requires credible representation and structure. Diagnostic profiles also require stakeholder interpretation and explicit uncertainty. Predictive augmentation adds out-of-sample evaluation against a non-clustered baseline. When profiles inform interventions, opportunities or resource allocation, all four evidence gates apply, including fairness, privacy, human review and contestability. Fig. 4 translates the cumulative evidence gates into a consequence-proportionate ladder.

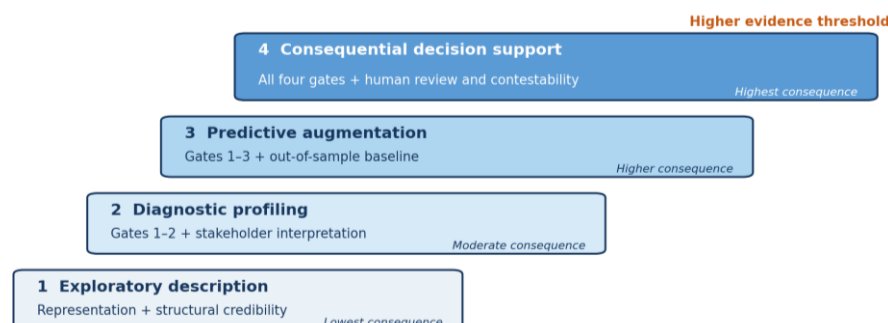


Figure 4. Consequence-Proportionate Evidence Ladder for Cluster-Based Educational Use.

Note. Evidence requirements accumulate as analytical outputs move closer to decisions affecting learners or institutions.

The ladder changes the central question from 'Which clustering algorithm performs best?' to 'What evidence is sufficient for this intended use?' This shift discourages methodological inflation: a partition may be valuable for hypothesis generation without being ready for prediction, and a predictive gain may be useful without authorising automated educational action. It also makes reporting proportional to risk rather than uniform across every clustering study.

10. RESEARCH GAPS AND FUTURE DIRECTIONS

The synthesis undertaken in this review identifies five interconnected research programmes aligned with the Four-Evidence Framework. The first concerns the temporal validity of educational profiles. Much clustering research continues to derive profiles from a single observation period, even though learner behaviour, engagement and self-regulation may change across instructional phases. Longitudinal and multi-timepoint studies should model transitions between profiles, examine the persistence of cluster membership and determine whether the identified trajectories reproduce across cohorts, courses and institutions [17]. This would help distinguish stable educational patterns from temporary responses to a particular course design or observation window.

The second programme concerns representation and methodological benchmarking. Comparative studies should evaluate complete clustering processes not algorithms alone under alternative feature constructions, scaling procedures, similarity measures, missing-data treatments and mixed-type representations [18]. Such comparisons should report sensitivity to preprocessing decisions because an apparently superior algorithm may only reflect the geometry produced by one analytical configuration.

Third, research on cluster-assisted prediction requires stronger evidence of analytical utility. Future studies should use nested evaluation, temporal or external validation, leakage controls and direct comparison with equivalent non-clustered baselines. Ablation analysis should establish whether cluster-derived variables provide information beyond the features from which the clusters were created. When explainable rules or recommendations are generated, their fidelity, stability and usefulness should be evaluated with educators and learners rather than inferred solely from computational measures [19, 45].

Fourth, institutional clustering requires purpose-built theoretical and empirical development. Learner-level assumptions cannot be transferred directly to schools, districts or universities because institutional profiles involve aggregated indicators, infrastructure, digital access, organisational capacity, geographical conditions and resource pressure. Studies at this level should examine whether institutional typologies remain stable across administrative periods and whether they provide information that is useful for planning without obscuring local context [2, 3].

The fifth programme concerns consequential educational use. Systems that influence interventions, opportunities or resource allocation should examine subgroup stability, algorithmic fairness, privacy–utility trade-offs, human override and contestability as components of technical evaluation rather than as considerations introduced after deployment [21, 49]. Shared educational datasets, transparent reporting templates and consequence-specific evaluation standards would improve reproducibility and reveal when apparently strong profiles depend on a single modelling decision. These programmes would shift educational clustering research from algorithm-centred comparison towards evidence-centred assessment of whether identified structures are credible, useful and appropriate for their intended purpose.

CONCLUSION

Clustering occupies an important position in educational data mining because it can reveal patterns among learners, courses and institutions before predefined outcome labels are available. Its educational value, however, cannot be established merely by producing a partition or reporting a favourable internal validity score. This critical integrative review demonstrates that the credibility of educational clusters is shaped throughout the analytical

process. The unit of analysis, observation period, feature construction, scaling procedure, similarity measure, treatment of missing data and selection of the clustering algorithm all influence the structure that is ultimately presented as educational evidence. Cluster interpretation must therefore begin with an examination of how the data represent the educational phenomenon rather than with the labels assigned after the analysis. The evidence examined in this review further shows that cluster formation, cluster validation and educational usefulness are related but analytically distinct. Internal indices assess geometrical properties such as cohesion and separation, yet they do not establish whether a solution is reproducible, substantively meaningful or useful beyond the dataset from which it was produced. Stability analysis, sensitivity testing, external validation and stakeholder interpretation provide different forms of evidence and cannot be reduced to a single preferred metric. Similarly, when cluster membership is incorporated into predictive models, its contribution must be separated from the information already contained in the original variables. An improvement in predictive performance may demonstrate analytical utility, but it does not independently justify an educational intervention or automated decision.

The principal contribution of this review is the Four-Evidence Framework, which organises the requirements for defensible educational clustering into representation validity, structural credibility, analytical utility and decision legitimacy. These evidence gates distinguish between identifying a pattern, demonstrating its reproducibility, establishing its practical value and determining whether its intended use is fair, transparent and educationally appropriate. The consequence-proportionate evidence ladder extends this argument by showing that the adequacy of a cluster solution depends on the purpose for which it is used. Exploratory description requires a lower evidential threshold than predictive augmentation, intervention prioritisation or institutional resource allocation. As the consequences of cluster-derived decisions increase, stronger forms of evidence concerning out-of-sample performance, fairness, privacy, explainability, human review and contestability become necessary. The review therefore positions clustering not as an isolated computational procedure, but as one component of a broader data-to-decision pathway. Its contribution to education depends on the integrity of the chain connecting data construction, methodological choice, validation, interpretation and intended use. The Four-Evidence Framework provides researchers with a structured basis for designing and reporting clustering studies, offers reviewers a standard for evaluating the strength of cluster-derived claims and assists educational institutions in judging whether an analytical output is appropriate for its intended application. It consequently clarifies when a cluster solution is informative, when it should remain exploratory and when it is sufficiently supported for responsible educational action.

REFERENCES

1. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
2. Ang, K. L.-M., Ge, F. L., & Seng, K. P. (2020). Big educational data & analytics: Survey, architecture and challenges. *IEEE Access*, 8, 116392–116414. <https://doi.org/10.1109/ACCESS.2020.2994561>
3. Stojanov, A., & Daniel, B. K. (2024). A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, 29, 5807–5831. <https://doi.org/10.1007/s10639-023-12033-8>
4. Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, 29, 22125–22145. <https://doi.org/10.1007/s10639-024-12704-0>

5. Kalita, E., Oyelere, S. S., Gaftandzhieva, S., et al. (2025). Educational data mining: A 10-year review. *Discover Computing*, 28, 81. <https://doi.org/10.1007/s10791-025-09589-z>
6. Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., et al. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, 104743. <https://doi.org/10.1016/j.engappai.2022.104743>
7. Lu, Y., Yeom, S., Maktoubian, J., et al. (2025). Improve student risk prediction with clustering techniques: A systematic review in education data mining. *Education Sciences*, 15(12), 1695. <https://doi.org/10.3390/educsci15121695>
8. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
9. Sukhera, J. (2022). Narrative reviews: Flexible, rigorous, and practical. *Journal of Graduate Medical Education*, 14(4), 414–417. <https://doi.org/10.4300/JGME-D-22-00480.1>
10. Alvarez-Garcia, M., Arenas-Parra, M., & Ibar-Alonso, R. (2024). Uncovering student profiles: An explainable cluster analysis approach to PISA 2022. *Computers & Education*, 223, 105166. <https://doi.org/10.1016/j.compedu.2024.105166>
11. Alzahrani, N., Meccawy, M., Samra, H., et al. (2025). Identifying weekly student engagement patterns in e-learning via K-means clustering and label-based validation. *Electronics*, 14(15), 3018. <https://doi.org/10.3390/electronics14153018>
12. Jin, J. (2025). Student behavior patterns in vocational education big data based on clustering algorithm. *Discover Artificial Intelligence*, 5, 197. <https://doi.org/10.1007/s44163-025-00433-3>
13. Song, R., Pang, F., Jiang, H., et al. (2024). A machine learning based method for constructing group profiles of university students. *Heliyon*, 10(7), e29181. <https://doi.org/10.1016/j.heliyon.2024.e29181>
14. Pecuchova, J., & Drlik, M. (2024). Enhancing the early student dropout prediction model through clustering analysis of students' digital traces. *IEEE Access*, 12, 159336–159367. <https://doi.org/10.1109/ACCESS.2024.3486762>
15. Al-Tameemi, G., Xue, J., Hadi, I., et al. (2024). A hybrid machine learning approach for predicting student performance using multi-class educational datasets. *Procedia Computer Science*, 238, 888–895. <https://doi.org/10.1016/j.procs.2024.06.108>
16. Hoq, M., Brusilovsky, P., & Akram, B. (2024). Explaining explainability: Early performance prediction with student programming pattern profiling. *Journal of Educational Data Mining*, 16(2), 115–148. <https://doi.org/10.5281/zenodo.14246435>
17. Costa, J., Lima, G., Araújo, R., et al. (2026). Multi-timepoint clustering analysis of self-regulated learning profiles based on user interaction logs in a virtual learning environment. *Discover Artificial Intelligence*, 6, 766. <https://doi.org/10.1007/s44163-026-01267-3>
18. Melchor, F., Conejero, J. M., Fernández-García, A. J., et al. (2026). An empirical evaluation of clustering processes for early detection of university dropout. *International Journal of Data Science and Analytics*, 22, 25. <https://doi.org/10.1007/s41060-025-00965-y>
19. Kisić, E., Raspopović Milić, M., Jović, J., et al. (2026). Tracking student progress and generating personalized recommendations using clustering and explainable artificial intelligence. *Universal Access in the Information Society*, 25, 26. <https://doi.org/10.1007/s10209-025-01295-3>
20. Neffati, C., Tato, A. A. N., Ghazi, G., et al. (2026). Temporal clustering of flight simulator data for learner profiling: A methodological comparison. In A. Botelho, M. M. T. Rodrigo, A. Singla, H. Ogata, H. So, & Y. H. Cho (Eds.), *Proceedings of the 19th International Conference on Educational Data Mining* (pp. 780–785). International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.21040146>
21. Boccuzzi, G., Manganello, F., & Nico, A. (2025). Delegated authority and algorithmic power: A rapid review of ethical issues in AI-based educational assessment. In *Proceedings of the 2025 International Conference on Information Technology for Social Good (GoodIT '25)* (pp. 1–5). Association for Computing Machinery. <https://doi.org/10.1145/3748699.3749766>

22. Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, 4, 100136. <https://doi.org/10.1016/j.caeai.2023.100136>
23. Mohammadi, M., Tajik, E., Martinez-Maldonado, R., et al. (2025). Artificial intelligence in multimodal learning analytics: A systematic literature review. *Computers and Education: Artificial Intelligence*, 8, 100426. <https://doi.org/10.1016/j.caeai.2025.100426>
24. Ogata, H., Liang, C., Toyokawa, Y., et al. (2024). Co-designing data-driven educational technology and practice: Reflections from the Japanese context. *Technology, Knowledge and Learning*, 29, 1711–1732. <https://doi.org/10.1007/s10758-024-09759-w>
25. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). University of California Press.
26. Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035). Society for Industrial and Applied Mathematics.
27. Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *WIREs Data Mining and Knowledge Discovery*, 2(1), 86–97. <https://doi.org/10.1002/widm.53>
28. Ester, M., Kriegl, H.-P., Sander, J., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the second international conference on knowledge discovery and data mining* (pp. 226–231). AAAI Press.
29. Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619. <https://doi.org/10.1109/34.1000236>
30. Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2–3), 191–203. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
31. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
32. Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14* (pp. 849–856). MIT Press.
33. Gagolewski, M., Bartoszek, M., & Cena, A. (2021). Are cluster validity measures (in)valid? *Information Sciences*, 581, 620–636. <https://doi.org/10.1016/j.ins.2021.10.004>
34. Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>
35. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
36. Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
37. Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>
38. Ben-Hur, A., Elisseeff, A., & Guyon, I. (2002). A stability based method for discovering structure in clustered data. *Pacific Symposium on Biocomputing*, 7, 6–17.
39. von Luxburg, U. (2010). Clustering stability: An overview. *Foundations and Trends in Machine Learning*, 2(3), 235–274. <https://doi.org/10.1561/22000000008>
40. Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218. <https://doi.org/10.1007/BF01908075>

41. Gagolewski, M. (2025). Normalised clustering accuracy: An asymmetric external cluster validity measure. *Journal of Classification*, 42, 2–30. <https://doi.org/10.1007/s00357-024-09482-2>
42. Pongsophon, P. (2026). Contrastive modelling of self-regulated inquiry behaviours across nations: Evidence from TIMSS 2023 Earth Science. *British Educational Research Journal*, 52, 1965–1986. <https://doi.org/10.1002/berj.70095>
43. Ouassif, K., Ziani, B., Herrera-Tapia, J., et al. (2025). Empowering education: Leveraging clustering and recommendations for enhanced student insights. *Education Sciences*, 15(7), 819. <https://doi.org/10.3390/educsci15070819>
44. Maniyan, S., Ghousi, R., & Haeri, A. (2024). Data mining-based decision support system for educational decision makers: Extracting rules to enhance academic efficiency. *Computers and Education: Artificial Intelligence*, 6, 100242. <https://doi.org/10.1016/j.caeai.2024.100242>
45. Odoom, S., Osei, E. O., Effah, E. Q., et al. (2026). Intelligent profiling beyond grades using a stacking ensemble framework for student success prediction. *Discover Education*, 5, 242. <https://doi.org/10.1007/s44217-026-01277-4>
46. Gunasekara, S., & Saarela, M. (2024). Explainability in educational data mining and learning analytics: An umbrella review. In B. Paaßen & C. Demmans Epp (Eds.), *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 887–892). International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.12729987>
47. Ifenthaler, D., Gibson, D., Prasse, D., et al. (2021). Putting learning back into learning analytics: Actions for policy makers, researchers, and practitioners. *Educational Technology Research and Development*, 69, 2131–2150. <https://doi.org/10.1007/s11423-020-09909-8>
48. Alfredo, R., Echeverria, V., Jin, Y., et al. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
49. Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, 56(2), 469–488. <https://doi.org/10.1111/bjet.13514>
50. Stinar, F., Xiong, Z., & Bosch, N. (2024). An approach to improve k-anonymization practices in educational data mining. *Journal of Educational Data Mining*, 16(1), 61–83. <https://doi.org/10.5281/zenodo.11056083>
51. Cohausz, L., Tschalzev, A., Bartelt, C., et al. (2024). Investigating demographic features and their connection to performance, predictions, and fairness in EDM models. *Journal of Educational Data Mining*, 16(2), 177–213. <https://doi.org/10.5281/zenodo.14503515>
52. Verger, M., Fan, C., Lallé, S., et al. (2024). A comprehensive study on evaluating and mitigating algorithmic unfairness with the MADD metric. *Journal of Educational Data Mining*, 16(1), 365–409. <https://doi.org/10.5281/zenodo.12180668>
53. Idowu, J. A., Koshiyama, A. S., & Treleaven, P. (2024). Investigating algorithmic bias in student progress monitoring. *Computers and Education: Artificial Intelligence*, 7, 100267. <https://doi.org/10.1016/j.caeai.2024.100267>
54. Simbeck, K. (2024). They shall be fair, transparent, and robust: Auditing learning analytics systems. *AI and Ethics*, 4, 555–571. <https://doi.org/10.1007/s43681-023-00292-7>