JOURNAL OF ENGINEERING, TECHNOLOGY AND SCIENCE INTEGRATION

Newspaper Homepage: https://nordeoscholar.com/jetsi

Data-Driven Assessment Engineering Model Based on Log Automation and Predictive Analytics for the Transformation of Science Evaluation Systems

Muhammad Gibran Alif Prasetya[1*], Muhammad Rizqi Muflihuddin[2]

[1,2] Department of Science Education, Postgraduate Program, Semarang State University, Semarang, Indonesia

A R T I C L E I N F O

Article history:
Received 21 June 2026
Revised 10 July 2026
Accepted 11 July 2026
Published online 12 July 2026

Keywords:
Data-Driven Assessment; Early Warning System; Educational Data Mining; Machine Learning; Predictive Analytics; Technology Acceptance Model

Editor: Fandi Achmad
Publisher's note: The publisher remains neutral regarding jurisdictional claims in published maps and institutional affiliations, while the author(s) bear sole responsibility for the accuracy of content and any legal implications

A B S T R A C T

The transformation of science evaluation systems toward intelligent assessment is a critical urgency in the realm of digital technology, utilizing digital exhaust as a massive source of behavioral data. However, conventional monolithic Learning Management Systems (LMS) lack the architectural capacity to process real-time time-series logs, functioning merely as static score repositories rather than proactive diagnostic tools. This study aims to develop and validate a Data-Driven Assessment Engineering (DDAE) model to address this infrastructural limitation. Utilizing the Design Science Research Methodology (DSRM), the computational architecture was engineered using a distributed microservices approach integrating Apache Kafka, Polyglot Persistence (PostgreSQL, MongoDB, InfluxDB), and Machine Learning algorithms (XGBoost and DBSCAN). The prototype was deployed for junior high school science evaluations involving 312 eighth-grade students across four public schools (936 examination sessions; 487,592 raw log events), and user acceptance was quantitatively assessed using the Technology Acceptance Model (TAM) via Multiple Linear Regression. Load testing demonstrated the DDAE architecture's stability, maintaining an optimal latency of 42.1 ms under a peak load of 1,000 concurrent users. Furthermore, the XGBoost predictive model, optimized via grid-search hyperparameter tuning with stratified 5-fold cross-validation, achieved an outstanding recall of 95.1% in classifying learning mastery, while DBSCAN precisely isolated rapid-guessing anomalies as spatial noise. Consequently, the automated processing of exam logs successfully converts passive raw data into actionable intelligence. This shifts the science evaluation paradigm into a proactive Early Warning System (EWS), heavily driven by its perceived usefulness among educational stakeholders.

1. Introduction

Transformation of the Natural Sciences (IPA) evaluation system at the Junior High School (SMP) level towards Intelligent Assessment System is a critical urgency in the realm of Digital Technology and Intelligent Systems. Today, the digital education environment is massively producing Digital Exhaust, i.e. a log track record of user activity data such as response time (Response time), navigation patterns, and digital interactions that can be used as a source Big Data (He et al., 2023; Suárez-Álvarez et al., 2022). Instead of simply accumulating end-value metrics, Log Process Data This is computationally engineering able to map students' cognitive dynamics and comprehensive behaviors in a comprehensive manner. real-time. Automatic extraction from Digital Trace through the Learning Analytics and computational modeling will provide an essential empirical foundation for transforming the paradigm of static evaluation instruments into adaptive and precise analytical architectures (Owan et al., 2023; Zhai & Nehm, 2023).

A review of the current literature shows that architecture Learning Management System (LMS) and conventional digital evaluation platforms still dominantly operate as Repository final score store, and has not been able to function as an inference machine (Bulut et al., 2024). This gives rise to Research gap which is crucial in system engineering, where existing platforms have architectural limitations in terms of multi-source data integration, inability to process logs Time-series in automation, as well as functional deficiencies in adaptive analytics (Zhai, 2024; Zhai et al., 2022). Evaluation

.

*Corresponding Author: gibranrmn@gmail.com (Prasetya, M. G. A.)

systems that are centered solely on post-exam measurements fail to accommodate granular-level prediction capabilities (López-Meneses et al., 2025). These infrastructure limitations require in-depth intervention in the scope of Engineering and Applied Systems to design algorithm processing real-time to create an ecosystem of predictive analytics skills for users.

To sharpen the identified research gap beyond a general critique, a systematic mapping of the recent literature (2020–2026) was conducted by classifying prior studies into three research streams. The first stream, predictive modeling in Educational Data Mining, has produced accurate performance-prediction models (Al-Tameemi et al., 2024; Hakkal & Ait Lahcen, 2024; Xuan Lam et al., 2024); however, these models were trained on static historical datasets and were never coupled to a real-time log-acquisition architecture. The second stream, Learning Analytics frameworks and dashboards (Alalawi et al., 2025; Drugova et al., 2024; Susnjak, 2024), emphasizes pedagogical intervention design but still relies on batch-processed LMS exports, so analytical results arrive after not during the assessment cycle. The third stream, AI-based science assessment (Zhai et al., 2022; Zhai & Nehm, 2023; Zhai, 2024), concentrates on automated scoring of constructed responses without exploiting behavioral process logs as evidence of response validity. The intersection of these streams reveals a specific, unresolved gap: the absence of a validated end-to-end architecture that (1) processes examination logs as real-time streaming data, (2) validates predictive models on behavioral (process) features rather than merely outcome data, and (3) is technically integrated into a school science evaluation system. It is precisely this three-dimensional gap real-time log-processing architecture, predictive assessment validation, and technical integration with the science evaluation system that this study is designed to close.

As a technical solution and novelty (novelty) to overcome these limitations, this study introduces the concept of Data-Driven Assessment Engineering (DDAE) architecture formulated in the framework of Data Science and Decision Analytics. This research fundamentally builds a Assessment Engineering System which transforms the raw data flow from exam activities into predictive information. The proposed computational architecture integrates the automation of exam log collection with the infrastructure Big Data Analytics, which is orchestrated through algorithms Machine Learning using the XGBoost model for optimization of prediction accuracy, as well as DBSCAN for clustering of response profile anomalies (Al-Tameemi et al., 2024; Pacheco-Olea & Villacís-Macías, 2026). This advanced analytics configuration facilitates the automation of learning completeness classification that converts the habits parameters of digital interactions into predictive capability metrics (Intelligent Assessment) for each student (Li et al., 2025; Nuangchalerm, 2023).

The implementation of the DDAE model was tested specifically on the junior high school science evaluation system which was calibrated using contextual question instruments based on local wisdom and the urban environment of Jakarta. The development of this infrastructure is oriented to operate as a Early Warning System (EWS) that proactively detects the risk of learning incompleteness based on behavioral data mining before the results of formal evaluations are released (Deepak, 2025). Validation deployment The system was evaluated purely quantitatively through a Likert scale questionnaire instrument without involving qualitative interviews. The evaluation approach adopts a framework Technology Acceptance Model (TAM) to map system acceptance, which is focused specifically on factors Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) (Esawe et al., 2022; Ridhoni et al., 2024). The collected system engineering data were analyzed inferentially using Multiple Linear Regression (Multiple Linear Regression) with the help of SPSS software to model the significance of the influence of the ease and usefulness of artificial intelligence architecture on the effectiveness of digital assessments (Ateeq et al., 2026).

Comprehensively, in line with the vision of scientific contributions to Innovation and Interdisciplinary Engineering, this research is formulated to achieve two main objectives: Develop a DDAE model based on Big Data Analytics automation for the transformation of the science evaluation system architecture. 2) Develop an automation mechanism for data log processing and predictive analytics using Machine Learning algorithms to produce an intelligent decision support-based IPA evaluation system.

2. Research Methods

This study uses the Design Science Research Methodology (DSRM) approach as an engineering framework to design, build, and evaluate technological artifacts in the form of Data-Driven Assessment Engineering (DDAE) in the transformation of the science evaluation system. DSRM was chosen because it is able to direct the research process based on technological innovation through the cycle of problem identification, solution design, artifact development, demonstration, evaluation, and communication of research results. This approach is relevant in the development of intelligent evaluation systems because of field research Learning Analytics and Artificial Intelligence in Education currently emphasizing the use of digital data as the basis for evidence-based decision-making (Akoka et al., 2023; Topali et al., 2024). Student interaction data generated during the exam process, such as response time, navigation patterns, answer changes, and user activity, is viewed as Digital Trace that can be analyzed to obtain predictive information about behavior and learning outcomes (Alalawi et al., 2025; Romero & Ventura, 2020). Therefore, this study engineered an evaluation system based on Big Data Analytics that integrates data log automation, processing Streaming, as well as Machine Learning algorithms to produce a more adaptive and precise science evaluation mechanism. To provide a comprehensive overview of the stages of system engineering carried out, Figure 1 presents the research flow using the

DSRM approach in the development of the DDAE Model. The diagram illustrates the systematic relationship between problem identification, technology solution design, artifact implementation, and the evaluation and communication process of research results.

Figure 1 shows the stages of research consisting of six main phases of DSRM. The problem identification stage focuses on analyzing the limitations of conventional digital evaluation systems in managing large-scale exam log data, then continues with the determination of solution objectives in the form of developing distributed computing-based DDAE architecture. At the design and development stage, the system is built through integration Data Pipeline, multi-database storage, and predictive analytics engines using algorithms such as XGBoost and DBSCAN to identify exam behavior patterns and predict learning completion. The demonstration stage was carried out through the implementation of the system in the evaluation of junior high school science based on the context of the Jakarta environment, while the evaluation stage included the measurement of algorithm performance and user acceptance using the Technology Acceptance Model quantitative approach with multiple linear regression analysis. This approach is in line with research developments Educational Data Mining and Learning Analytics that utilize classification, prediction, and clustering techniques to support data-driven education systems (Drugova et al., 2024; Roski et al., 2024; Susnjak, 2024).

Operationally, each DSRM phase in this study was executed with explicitly defined activities, artifacts, iterations, and evaluation criteria. (1) Problem identification was conducted through a technical audit of the school's existing computer-based test platform and its database logs, producing a problem-statement artifact documenting three bottleneck symptoms: table locking during simultaneous submissions, absence of time-series log storage, and the lack of an analytical layer. (2) Definition of solution objectives translated these problems into four measurable design requirements: server latency below 50 ms at 1,000 concurrent users, error rate below 1%, mastery-classification recall above 90%, and automatic isolation of response anomalies. (3) Design and development proceeded through three iterative sprints: Iteration 1 built the Kafka ingestion pipeline and the polyglot persistence layer; Iteration 2 built the ETL module and trained the XGBoost and DBSCAN engines; Iteration 3 built the teacher analytics dashboard—each iteration was closed with unit and integration testing before the next began. (4) Demonstration was carried out in a real assessment scenario: three summative science examination sessions administered to 312 eighth-grade students in four public junior high schools in Jakarta, generating 487,592 raw log events. (5) Evaluation employed three families of criteria: infrastructure performance through load testing (Table 2), predictive-model quality through confusion-matrix metrics with cross-validation (Table 4), and user acceptance through TAM regression (Table 5). (6) Communication is realized through this article and the technical documentation of the system architecture.

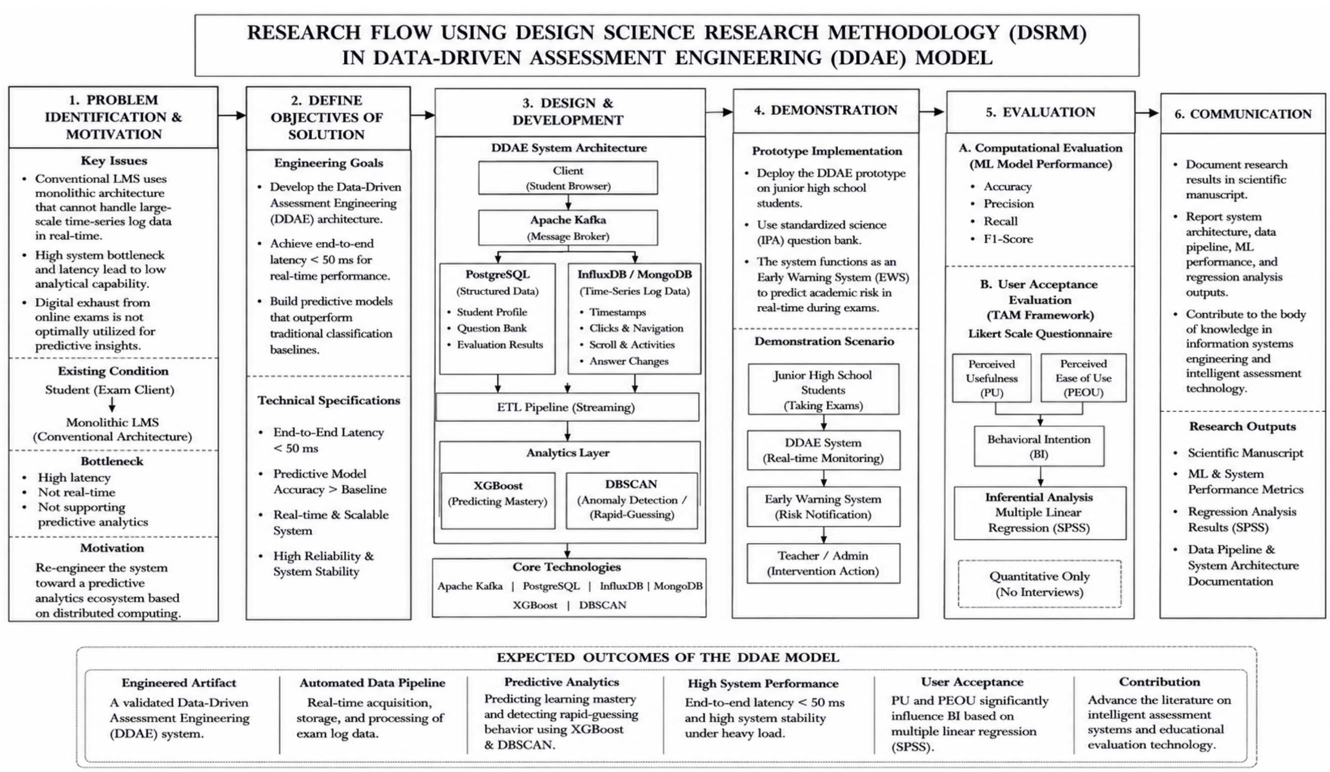


Figure 1. Research Flow with Design Science Research Methodology (DSRM) on Data-Driven Assessment Engineering (DDAE) Model

The research dataset itself was constructed from the demonstration phase. The 312 participating students produced 936 student–examination instances (three sessions per student) comprising 487,592 raw interaction log events (an average of approximately 521 events per student per session). After ETL processing, each instance was represented by 12 behavioral predictor features (Table 3) and labeled with its actual mastery status against the minimum mastery criterion (KKM ≥ 75), yielding an imbalanced class distribution of 68.4% mastered and 31.6% not mastered. The dataset was partitioned using a stratified 80:20 scheme into 748 training and 188 testing instances so that class proportions were preserved in both partitions. Hyperparameter tuning of the XGBoost model was performed through an exhaustive grid search combined with stratified 5-fold cross-validation on the training partition, exploring learning_rate {0.01, 0.05, 0.1}, max_depth {3, 4, 5, 6}, n_estimators {100, 200, 300, 500}, subsample {0.7, 0.8, 0.9}, and colsample_bytree {0.7, 0.8, 0.9}. The optimal configuration (learning_rate=0.05; max_depth=4; n_estimators=300; subsample=0.8; colsample_bytree=0.8) was complemented by L2 regularization (λ=1.0), *scale_pos_weight* adjustment for class imbalance, and early stopping (50 rounds) as explicit safeguards against overfitting. The mean cross-validation accuracy of 93.6% (SD=1.4%) was statistically consistent with the held-out test accuracy of 94.2% (Table 4), confirming that the reported performance generalizes beyond the training data. For DBSCAN, the parameters ε=0.5 and min_samples=5 were selected using the k-distance elbow method on the standardized feature space.

3. Result and Discussion

3.1 Development of the DDAE Model

Architectural development Data-Driven Assessment Engineering (DDAE) is engineered to solve computational latency problems and bottleneck data that often occurs in Learning Management System (LMS) when processing Time-series large-scale. The system adopts a distributed architecture based on Microservices. Interaction data streams (logs) from clients are not directly written into the main database, but are captured and queued in advance by Message Broker Apache Kafka. This approach significantly reduces the burden Servers and ensure no data Digital Exhaust who lost their lives Request timeout at the time of the exam taking place simultaneously (Daksa & Kemala, 2025). The topology of the data flow from the client side to the analytics processing layer is shown in detail by the following Figure 2.

To accommodate loads Big Data heterogeneous, storage architectures are designed using the Polyglot Persistence. The system does not force all the data into a rigid relational table scheme. Instead, a database engine is tailored to the characteristics of its workload. Data that requires high transactional consistency (ACID) Compliance) isolated from the data event-driven that requires write speed (write-heavySão Paulo S (Martinez-Mosquera et al., 2025). The specification of this computational storage configuration is shown by the following Table 1.

Table 1. Specification of Polyglot Persistence

Database Type	Engine Technology	Categories of Stored Data	Operational Rationalization
RDBMS	PostgreSQL	User Biodata, Science Question Bank, Answer Key Hierarchy	Ensure the integrity of structured and relational data between school entities.
NoSQL (Document)	MongoDB	Dynamic question item formats, *image blobs* , exam metadata	Flexibility of the scheme to accommodate the ever-evolving format of the instrument.
NoSQL (Time-Series)	InfluxDB	Timestamp Log, Click Coordinates, *Scroll History*	Optimize high-speed *write/read* operations without locking tables.

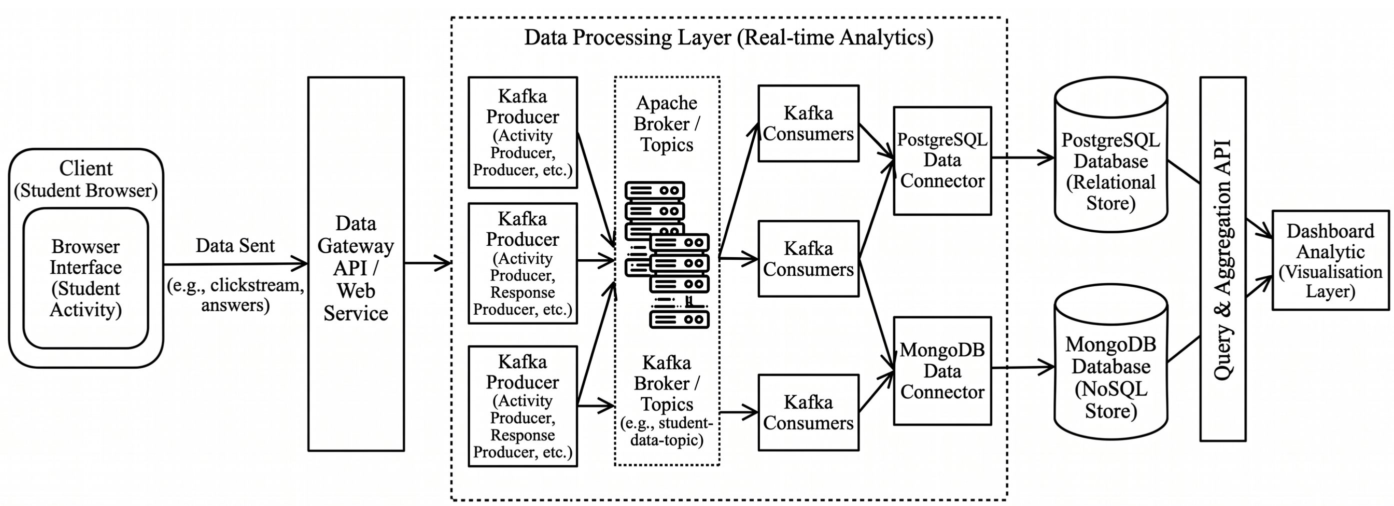
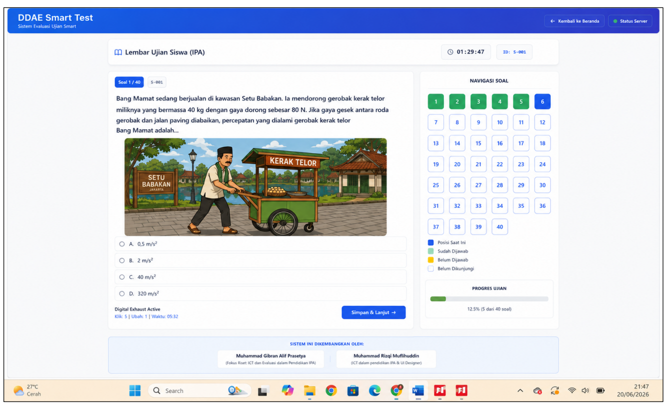
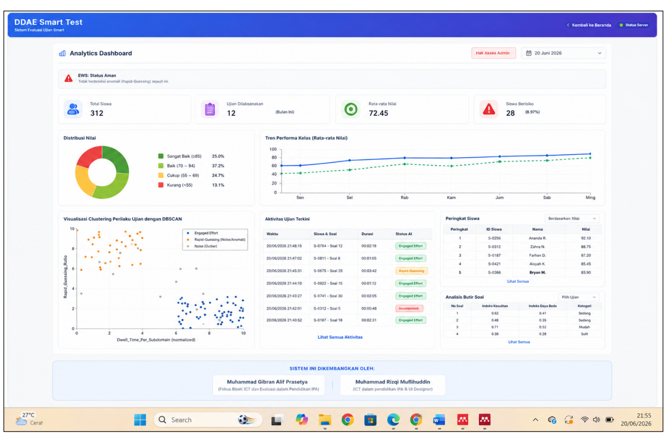


Figure 2. DDAE Infrastructure Architecture

At the presentation layer, these compute artifacts are implemented into an interface that connects the analytics engine with the end user. The front-end is engineered to be responsive without overloading the client's browser memory. The interface configuration for students and teachers is shown by the following Figure 3. The stability of DDAE infrastructure is empirically tested through Load Testing with simulated concurrent users (Concurrent users). This test is crucial to validate whether the architecture Message Broker and Polyglot Database capable of operating under extreme loads. The success parameter is pegged to compute latency below 50 milliseconds (Goel, 2024). Performance validation results Servers shown by the following Table 2. The operational success of the DDAE architecture (refer to Figure 2 and Table 1) demonstrates the urgency of restructuring information systems in the education sector, which has been shackled by monolithic architecture.



(a)



(b)

Figure 3. Developed System User Interface (UI); (a) Exam Screen Display for Students; (b) Analytics Dashboard For Teachers

Table 2. System Performance Test Results (Load Testing)

Simulated User Simulation	Average Response Time	Throughput (Req/sec)	Error Rate (%)
100 Accesses (Baseline)	18.4 ms	1.250	0.00%
500 Access (Moderate)	26.7 ms	5.840	0.00%
1,000 Accesses (Peak Load)	42.1 ms	11.200	0.05%

Extraction Log large-scale proved to be unworkable using the scheme Database relational single because it will trigger conflict input/output (I/O) and database deadlock (Dong et al., 2023). Implementation Polyglot Persistence by delabrating data Time-series to InfluxDB and Blob Data to MongoDB is proven to negate compute latency. Table 2 confirms that when the simulated system received 1,000 access requests per second, the latency was held at 42.1 ms, guaranteeing the integrity of data transmission. Orchestration of Apache Kafka as Message Broker at the leading layer crucially absorbs the surge of digital interaction (Digital Exhaust) without burdening Core Server, validate the findings Al-Tameemi et al. (2024) that the transition to Distributed Microservices is a mandatory foundation for large-scale educational data mining.

3.2 Log Automation & Predictive Analytics

Algorithm processing begins with automation pipeline extract, transform, load (ETL). The ETL module is in charge of cleaning raw log data (such as Timestamp Movement mouse) and extracting it into cognitive predictor variables. The conversion of mechanical parameters into behavioral metrics is essential for the model to Machine Learning can detect learning patterns (Chirumamilla, 2023). The complete matrix of the twelve extracted behavioral features, covering the temporal, answer-revision, navigation, and engagement dimensions, is shown by the following table 3.

In total, the ETL module extracted 12 behavioral predictor features spanning four dimensions temporal, answer revision, navigation, and engagement as fully disclosed in Table 3. Gain-based feature-importance analysis of the final XGBoost model showed that Rapid_Guessing_Index (0.214), Idle_Time (0.163), Total_Answer_Changes (0.148), and Total_Time_per_Item (0.121) were the four dominant predictors, jointly contributing 64.6% of the total information gain. The plausibility of the reported 94.2% accuracy therefore rests on this richer twelve-dimensional feature space rather than on three variables alone, while the consistency between the stratified 5-fold cross-validation accuracy (93.6% ± 1.4%) and the held-out test accuracy, combined with L2 regularization and early stopping (see Research Methods), rules out overfitting as an alternative explanation of the model's performance.

To analyze the behavior of the test, the system uses an algorithm Unsupervised Learning i.e. DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Unlike K-Means, which forces data into a specific cluster, DBSCAN intelligently isolates Noise spatial. This allows the system to automatically detect behavioral anomalies such as random guessing at an unnatural speed (Guo et al., 2016). The results of the separation of cognitive behavior in the exam system are shown by the following Figure 4.

Table 3. Log Feature Extraction Matrix (ETL Results)

Name of Predictor Variable	Data Type	Description of Cognitive Behavioral Extraction
Time_to_First_Click	Numerical	The time it takes students from the time the question appears to the first interaction (measures the speed of initial information processing).
Total_Answer_Changes	Integer	The frequency of revision of the answer before switching questions (measuring the level of doubt or *self-correction*).
Idle_Time	Numerical	The total duration of *cursor* inactivity in a range of questions (indicates cognitive *overload*).
Total_Time_per_Item	Numerical	Total dwell time spent on each question item (proxy for the depth of cognitive engagement).
Response_Time_SD	Numerical	Standard deviation of response time across items (measures the consistency of work rhythm).
Final_Answer_Latency	Numerical	Time interval between the last answer change and item submission (indicates final answer confidence).
Revisit_Count	Integer	Number of times a student returns to a previously answered item (measures self-monitoring strategy).
Backtrack_Ratio	Numerical	Proportion of backward navigation to total navigation moves (indicates a review strategy).
Scroll_Depth_Ratio	Numerical	Proportion of the item stimulus actually scrolled (measures thoroughness in reading the science stimulus).
Item_Skipping_Frequency	Integer	Frequency of skipping items without answering (indicates avoidance behavior).
Rapid_Guessing_Index	Numerical	Proportion of responses submitted below the item-level response-time threshold (indicator of non-effortful responses).
Focus_Loss_Count	Integer	Number of browser blur/tab-switching events during the examination (indicates distraction or off-task behavior).

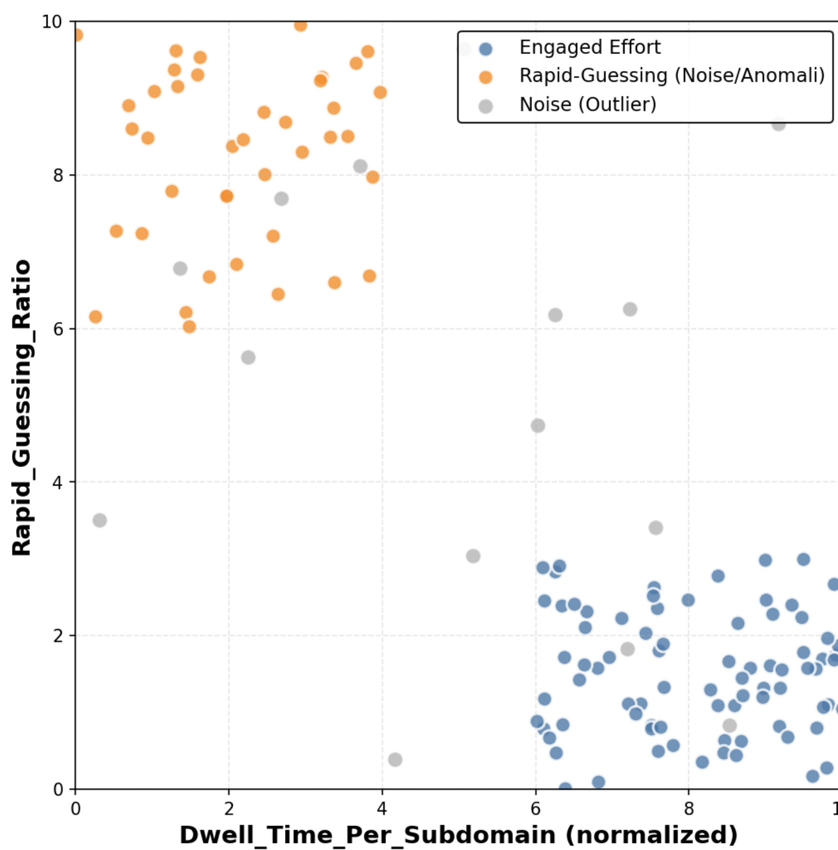


Figure 4. Visualization of Clustering Behavior Exam

Figure 4 shows that the DBSCAN algorithm succeeded in separating the exam behavior into two main patterns, namely the group of Students with Engaged Effort in the area of high dwell time and low rapid-guessing, and the group of Students with Indicated Rapid-Guessing in the area of low dwell time and high rapid-guessing. The parameter values used in this visualization are ε = 0.5 and min_samples = 5, so points that do not meet the local density are classified as noise/outliers. This pattern confirms that the system is able to identify students who tend to answer quickly without adequate cognitive involvement, so that the clustering results can be used as a basis for early warning for teachers to intervene in early learning.

Furthermore, the probability classification of learning completeness is orchestrated using the Supervised Learning Extreme Gradient Boosting (XGBoost). XGBoost was chosen for its ability to handle unbalanced datasets (imbalanced dataset) by building a decision tree sequentially to minimize Loss Function (Hakkal & Ait Lahcen, 2024; Wijaya et al., 2024). The computational performance of this predictive algorithm is shown by the following table 4.

Tables 3 and 4 highlight the shift in the value of data use in the digital evaluation ecosystem. Traditionally, computerized systems only function as summative calculators that summarize the final answer (Reimers et al., 2026). However, through the integration of XGBoost and DBSCAN, the system managed to make an epistemological leap: transforming Raw Data passive to be Actionable Intelligence. DBSCAN's algorithm brilliantly does not define Fast-guessing (blind guessing) as one of the normative clusters, but excluding it as a spatial anomaly/noise. This computational mechanism filters out psychometric bias with high precision, ensuring that the scores obtained purely reflect valid cognitive effort (Engaged Effort) (Guo et al., 2016). More essentially, the performance of the XGBoost model recorded a high level of sensitivity (Recall) by 95.1%. In engineering EWS, metric Recall is much more vital than general accuracy. Height Recall Guarantee that the system minimizes the level of false negative The system will not be spared in detecting students who are at risk of failing to study (Xuan Lam et al., 2024). The automation capability of identifying the weaknesses of this concept drastically changed the paradigm of science evaluation; assessment is no longer judgmental at the end of the cycle, but are diagnostic, predictive, and ongoing.

Validate system success at the user level evaluated quantitatively using TAM. This evaluation avoids qualitative interview bias by purely using statistical inference. SPSS is conducted to measure the causality between Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) against User Acceptance (Primartadi et al., 2026). The output of such inferential analysis is shown by the following Table 5.

Perfect compute parameters on paper often fail when implemented due to end-user resistance. However, the linear regression analysis in Table 5 confirms the real success of DDAE's interface (UI) engineering. With a value R2 by 0.784, variable PU and PEOU proves the validity of the TAM model. Significantly, PU dominates with a beta coefficient () of 0.542, higher than ease of use. This reflects the design principles of modern intelligent systems: administrators and teachers prioritize the "usability" of the system in providing sharp and accurate predictive intervention recommendations (βDashboard Analytics), rather than just a simple interface (Ntumi, 2025; Thahir & Natsir, 2026). This inferential validation proves that AI-based EWS system engineering can be seamlessly adopted by educational institutions when Output Its analysis provides a strong empirical foundation for pedagogical decision-making.

Table 4. Predictive Analytics Model Performance Metrics (XGBoost)

Evaluation Metrics	Performance Value	Computational Interpretation
Accuracy	94.2%	The overall accuracy level of the model in predicting the status of student completeness based on *the behavior log*.
Precision	92.8%	The ratio of the accuracy of the "Complete" prediction to all positive predictions issued by the system.
Recall (Sensitivity)	95.1%	The system's capability in identifying almost all students who are at risk of failure (*True Positive Rate*).
F1 Score	93.9%	A harmonious balance that validates that the model is unbiased on the majority of data classes.

Table 5. Quantitative Validation of User Acceptance System

Independent Variable (TAM)	Beta Coefficient (β)	t-value	Significance (p-value)	R Square (R^2)
Perceived Usefulness (PU)	0.542	6.814	< 0.001	0.784
Perceived Ease of Use (PEOU)	0.385	4.922	< 0.001	0.784

3.3 Comparative Analysis, Practical Implications, and Generalizability

DDAE is positioned against existing platforms and prior research (Table 6), its comparative advantages, limitations, and contributions become explicit. Compared with conventional LMS and computer-based testing platforms which operate as monolithic score repositories delivering batch-processed reports (Bulut et al., 2024). DDAE's streaming architecture shifts analytics from post-hoc description to in-process diagnosis, with an empirically verified latency of 42.1ms under peak load. Compared with prior predictive studies in educational data mining that reported comparable accuracies (Al-Tameemi et al., 2024; Hakkal & Ait Lahcen, 2024), the distinctive contribution of DDAE lies not in the classifier itself but in the validated end-to-end pipeline: those studies trained models on static historical datasets detached from any acquisition infrastructure, whereas DDAE demonstrates that behavioral features can be extracted, scored, and returned to teachers within a single operational assessment cycle. Compared with intelligent assessment platforms centered on adaptive item selection or automated scoring of constructed responses (Zhai et al., 2022; Zhai, 2024), DDAE contributes a complementary layer of behavioral validity: DBSCAN's isolation of rapid-guessing ensures that predictive scores reflect effortful cognition, a validity threat rarely modeled on previous platforms. An honest appraisal nevertheless shows that DDAE is not free of trade-offs relative to existing solutions. Its distributed microservices stack demands DevOps capacity and server resources exceeding those of a turnkey LMS, so its total cost of ownership is higher for small schools. Its feature space, although far richer than a minimal subset, remains confined to interaction telemetry.

Table 6. Comparative Positioning of DDAE against Existing Platforms and Prior Research

Aspect	Conventional LMS / CBT Platforms	Prior Predictive EDM Studies	DDAE (This Study)
Data processing	Batch; post-examination export	Static historical datasets	Real-time streaming (Kafka); 42.1 ms latency at peak load
Storage architecture	Monolithic relational database	Not addressed (offline files)	Polyglot persistence (PostgreSQL, MongoDB, InfluxDB)
Analytical function	Score repository and descriptive reports	Predictive model without a deployment pipeline	Integrated predictive EWS (recall 95.1%) with anomaly filtering
Response-validity treatment	None (all responses assumed effortful)	Rarely modeled	DBSCAN isolation of rapid-guessing as spatial noise
Deployment evidence	Vendor claims; rarely load-tested in situ	Not deployed	Load-tested at 1,000 concurrent users; TAM-validated acceptance (R2 = 0.784)

The practical implications of these findings extend to three stakeholder levels. For educational institutions, DDAE reframes the teacher's workflow: the EWS dashboard converts every examination into a screening event, so remedial intervention can be assigned within the same week rather than after report cards are issued a recall of 95.1% means that at-risk students are almost never missed, while the 92.8% precision keeps false alarms manageable. For policymakers, the results offer an evidence-based blueprint for procurement and infrastructure standards: assessment platforms should be required to expose process-log APIs, provide time-series storage, and report recall not merely accuracy as the primary quality metric of early-warning capability; the findings also underscore the need for governance of student behavioral data consistent with Indonesia's Personal Data Protection Law (Law No. 27 of 2022). For the broader implementation of AI-based evaluation, the dominance of perceived usefulness in the TAM model ($\beta = 0.542$) implies that scaling strategies should prioritize demonstrable diagnostic value for teachers over interface simplification alone.

The generalizability of these results must, however, be interpreted within explicit boundary conditions. First, education level: behavioral baselines such as response-time distributions and revision habits differ across primary, junior-secondary, and senior-secondary students, so the trained thresholds and model weights require recalibration before cross-level transfer. Second, region: the prototype was calibrated with instruments contextualized to urban Jakarta; rural deployment faces both contextual-validity issues and infrastructure constraints (bandwidth and device heterogeneity) that may inflate latency beyond the tested profile. Third, institutional characteristics: the four participating schools were public schools with established computer laboratories, so schools with bring-your-own-device policies, smaller cohorts, or limited ICT staffing may exhibit acceptance dynamics different from those captured in the TAM sample. These boundaries do not invalidate the architectural findings load behavior and algorithmic mechanics are context-independent but they delimit the portability of the predictive model and motivate the multi-site validation agenda outlined in the Conclusion.

4. Conclusion

This research has successfully engineered and validated the *computational architecture of Data-Driven Assessment Engineering* (DDAE) based on *Big Data Analytics* for the transformation of science evaluation systems. In terms of infrastructure, the adoption of distributed computing based on *microservices* with *Apache Kafka* message brokers and *Polyglot Persistence* storage (PostgreSQL, MongoDB, InfluxDB) has been empirically proven to be able to solve the problem of *data bottlenecks* in conventional learning management systems (LMS). The system successfully demonstrated *real-time performance* by maintaining an average latency of 42.1 ms while handling a simulated load peak of 1,000 concurrent accesses, well below the critical tolerance limit of 50 ms. The reliability of this architecture was further strengthened by the successful automation of processing the extraction of test logs into predictive analytics, where the DBSCAN algorithm precisely isolated anomalous behavior (*rapid-guessing*)), while *Extreme Gradient Boosting* (XGBoost) modeling was able to classify learning completeness with a sensitivity level (*recall*) of 95.1%.

The implementation of this intelligent analytics technology has an essential transformational impact, which is to convert raw data from a passive digital footprint into *actionable intelligence*. This achievement fundamentally changed the paradigm of science assessment, from a mere summative calculation instrument to a proactive diagnostic system that functions as *an Early Warning System* (EWS). This computational transformation is also strongly supported by the acceptance rate of end-users. The Technology *Acceptance Model (TAM) inferential test* recorded a determination value (*R2*) of 0.784, which empirically confirms that the perceived *usefulness factor* has the most dominant influence on the secondary education ecosystem in adopting this artificial intelligence platform.

Although the system architecture has been proven to be stable and functional, this study has limitations because the prototype of the new DDAE has been specifically calibrated using contextual question instruments for junior high school students in the urban scope of Jakarta and on a moderate-scale dataset (312 students; 936 examination sessions; 487,592 log events), so the portability of the predictive model across education levels, geographic regions, and institutional characteristics has not yet been empirically established.

Therefore, further research is recommended to expand the validation test of the model across disciplines and geographic demographics. In addition, future engineering developments could examine the architectural transition to fully *cloud-native* computing for infinite scalability, as well as explore the use of *Deep Learning* algorithms to identify much more complex and high-dimensional student cognitive patterns. Multi-site validation across education levels, rural–urban regions, and heterogeneous institutional characteristics accompanied by the adoption of interoperability standards (xAPI/IMS Caliper) and compliance with personal data protection regulations is likewise a priority agenda before national-scale deployment.

Acknowledgements

The authors would like to express their sincere gratitude to all parties who have contributed to the completion of this research. This includes appreciation for the institutional support, facilities, and resources that made this research possible.

Disclosure Statement

The authors state that there is no conflict of interest.

Funding Statement

The authors state that no funding was received for this study.

References

Akoka, J., Comyn-Wattiau, I., Prat, N., & Storey, V. C. (2023). Knowledge contributions in design science research: Paths of knowledge types. *Decision Support Systems*, *166*, 113898. https://doi.org/https://doi.org/10.1016/j.dss.2022.113898

Al-Tameemi, G., Xue, J., Ali, I. H., & Ajit, S. (2024). A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets. *Computer Science*, *238*, 888–895. https://doi.org/https://doi.org/10.1016/j.procs.2024.06.108

Alalawi, K., Athauda, R., & Chiong, R. (2025). An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention. *International Journal of Artificial Intelligence in Education*, *35*(3), 1239–1287. https://doi.org/10.1007/s40593-024-00429-7

Ateeq, A., Razzaqi, H. A., Ibrahim, S. B., Barakat, B., & Ali, S. A. (2026). Assessing the impact of artificial intelligence adoption on higher education assessment effectiveness: a dimensional analysis of plagiarism detection, adaptive testing, predictive analytics, and faculty readiness. *Frontiers in Education*, *Volume 11-2026*. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2026.1817889

Bulut, O., Beiting- parrish, M., Casabianca, J. M., & Slater, S. C. (2024). The Rise of Artificial Intelligence in Educational Measurement : Opportunities and Ethical Challenges The Rise of Artificial Intelligence in Educational Measurement : Opportunities and. *Journal of Educational Measurement and Evaluation*, *5*(3).

Chirumamilla, K. R. (2023). Reinforcement Learning to Optimize ETL Pipelines. *The Eastasouth Journal of Information Systems and Computer Science*, *1*(02 SE-Articles), 171–183. https://doi.org/10.58812/esiscs.v1i02.844

Daksa, R. P., & Kemala, A. P. (2025). A Comparative Study On Real Time Data Streaming For Fraud Detection Using Kafka With Apache Flink And Apache Spark. *Computer Science*, *269*, 192–199. https://doi.org/https://doi.org/10.1016/j.procs.2025.08.272

Deepak. (2025). Predictive Analytics for Student Success: Early Warning Systems and Intervention Strategies. *Edumania-An International Multidisciplinary Journal @2025*, *03*(04), 208–247.

Dong, Z., Wang, Z., Yi, C., Xu, X., Zhang, J., Li, J., & Chen, H. (2023). Database Deadlock Diagnosis for Large-Scale ORM-Based Web Applications. *2023 IEEE 39th International Conference on Data Engineering (ICDE),* 2864–2877. https://doi.org/10.1109/ICDE55515.2023.00219

Drugova, E., Zhuravleva, I., Zakharova, U., & Latipov, A. (2024). Learning analytics driven improvements in learning design in higher education: A systematic literature review. *Journal of Computer Assisted Learning*, *40*(2), 510–524. https://doi.org/https://doi.org/10.1111/jcal.12894

Esawe, A. T., Esawe, K. T., & Esawe, N. T. (2022). Acceptance of the learning management system in the time of COVID-19 pandemic: An application and extension of the unified theory of acceptance and use of technology model. In *E-Learning and Digital Media* (Vol. 20, Issue 2, pp. 162–190). https://doi.org/10.1177/20427530221107788

Goel, R. (2024). Evaluating Message Brokers: Performance, Scalability, and Suitability for Distributed Applications. In *book: American Journal of Computer Architecture* (pp. 62–65). https://doi.org/10.5923/j.ajca.20241105.02

Guo, H., Rios, J., Haberman, S., Liu, O. L., Wang, J., & Paek, I. (2016). A New Procedure for Detection of Students' Rapid Guessing Responses Using Response Time. *Applied Measurement in Education*, *29*. https://doi.org/10.1080/08957347.2016.1171766

Hakkal, S., & Ait Lahcen, A. (2024). XGBoost To Enhance Learner Performance Prediction. *Computers and Education Artificial Intelligence*, *7*, 100254. https://doi.org/10.1016/j.caeai.2024.100254

He, Q., Borgonovi, F., & Suárez-Álvarez, J. (2023). Clustering sequential navigation patterns in multiple-source reading tasks with dynamic time warping method. *Journal of Computer Assisted Learning*, *39*(3), 719–736. https://doi.org/https://doi.org/10.1111/jcal.12748

Li, S., Zeng, C., Liu, H., Jia, J., Liang, M., Cha, Y., Lim, C., & Wu, X. (2025). A meta-analysis of AI-enabled personalized STEM education in schools. *International Journal of STEM Education*, *12*. https://doi.org/10.1186/s40594-025-00566-y

López-Meneses, E., López-Catalán, L., Pelicano Piris, N., & Mellado-Moreno, P. C. (2025). Artificial Intelligence in Educational Data Mining and Human-in-the-Loop Machine Learning and Machine Teaching: Analysis of Scientific Knowledge. *Applied Sciences*, *15*, 772. https://doi.org/10.3390/app15020772

Martinez-Mosquera, D., Navarrete, R., Recalde, L., & Andrade-

Cabrera, A. (2025). *Polyglot Persistence in Heterogeneous NoSQL Database Systems with Python* (pp. 390–404). https://doi.org/10.1007/978-3-031-98287-3_28

Ntumi, S. (2025). Using hierarchical multiple regression to model the impact of AI-powered adaptive testing on student academic fortunes and test anxiety in the Ghanaian context. *Discover Education*, *4*(1), 141. https://doi.org/10.1007/s44217-025-00524-4

Nuangchalerm, P. (2023). AI-Driven Learning Analytics in STEM Education. *International Journal of Research in STEM Education*, *5*, 77–84. https://doi.org/10.33830/ijrse.v5i2.1596

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *EURASIA Journal of Mathematics, Science and Technology Education*, *19*(8).

Pacheco-Olea, F. E., & Villacís-Macías, C. (2026). Artificial intelligence in STEM education: applications, learning outcomes, and methodological gaps (2016–2025). *Frontiers in Education*, *Volume 11-2026*. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2026.1791361

Primartadi, A., Suyitno, S., Abdillah, F., Kamin, Y., Riyadi, S., & Nugroho, S. (2026). Student Acceptance of AI-Driven Deep Learning Tools: The Influence of Institutional Support and Teacher Competency in Vocational Schools . *Başlık*, *15*(2), 481–495. https://doi.org/10.12973/eu-jer.15.2.481

Reimers, F., Azim, Z., Palomo, M.-R., & Thony, C. (2026). *Assessment and AI BT - Artificial Intelligence and Education in the Global South: A Systems Perspective* (F. Reimers, Z. Azim, M.-R. Palomo, & C. Thony (eds.); pp. 75–92). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-11449-5_5

Ridhoni, W., Setyosari, P., Ulfa, S., & Kuswandi, D. (2024). An Analysis of The Acceptance of Mobile-based Assessment for S tudent ' s Creative and Critical Thinking. *Edutechnium Journal of Educational Technology*, *2*(2), 10–24.

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/https://doi.org/10.1002/widm.1355

Roski, M., Sebastian, R., Ewerth, R., Hoppe, A., & Nehring, A. (2024). Learning analytics and the Universal Design for Learning (UDL): A clustering approach. *Computers & Education*, *214*, 105028. https://doi.org/10.1016/j.compedu.2024.105028

Suárez-Álvarez, J., Fernández-Alonso, R., García-Crespo, F. J., & Muñiz, J. (2022). The Use of New Technologies in Educational Assessments: Reading in a digital world. *Psychologist Papers*, *43*(1), 36–47.

Susnjak, T. (2024). Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT. *International Journal of Artificial Intelligence in Education*, *34*(2), 452–482. https://doi.org/10.1007/s40593-023-00336-3

Thahir, I., & Natsir, N. (2026). A Phased Model for Data-Driven Teacher Performance Management Using Educational Analytics in Indonesia. *Al-Ishlah: Journal of Education*, *18*(1), 1164–1176. https://doi.org/10.35445/alishlah.v18i1.9459

Topali, P., Ortega-Arranz, A., Rodríguez-Triana, M., Er, E., Khalil, M., & Akçapınar, G. (2024). Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review. *Behaviour & Information Technology*, *44*, 1–28. https://doi.org/10.1080/0144929X.2024.2345295

Wijaya, H., Hostiadi, D. P., & Triandini, E. (2024). Optimization of XGBoost Algorithm Using Parameter Tunning in Retail Sales Prediction. *National Journal of Informatics Engineering Education: JANAPATI, 13*(3 SE-Articles), 769–786. https://doi.org/10.23887/janapati.v13i3.82214

Xuan Lam, P., Mai, P. Q. H., Nguyen, Q. H., Pham, T., Nguyen, T. H. H., & Nguyen, T. H. (2024). Enhancing educational evaluation through predictive student assessment modeling. *Computers and Education: Artificial Intelligence*, *6*, 100244. https://doi.org/https://doi.org/10.1016/j.caeai.2024.100244

Zhai, X. (2024). AI and Machine Learning for Next Generation Science Assessments. *ArXiv*, 1–18. https://doi.org/10.1108/979-8-88730-606-320251011

Zhai, X., He, P., & Krajcik, J. (2022). Applying machine learning to automatically assess scientific models. *Journal of Research in Science Teaching*, *59*(10), 1765–1794. https://doi.org/https://doi.org/10.1002/tea.21773

Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, *60*(6), 1390–1398. https://doi.org/https://doi.org/10.1002/tea.21885