

THE ASSESSMENT REVOLUTION: THEORIES AND METHODOLOGIES OF AUTOMATED ASSESSMENT USING MACHINE LEARNING FOR EVALUATING LEARNING PROGRESS

Misbahul Khairani¹, Cedric Butler², and Markus Rolle³

¹ Sekolah Tinggi Agama Islam Samarinda, Indonesia

² University of The Bahamas, Bahamas

³ University of Freeport, Bahamas

Corresponding Author:

Misbahul Khairani,

Department of Elementary Madrasah Teacher Education, Faculty of Teacher Training and Education, Sekolah Tinggi Agama Islam Samarinda.

Jl. M. Said RT 12, Kelurahan Lok Bahu, Kecamatan Sungai Kunjang, Kota Samarinda, Kalimantan Timur, Indonesia

Email: suhaaranaa9@gmail.com

Article Info

Received: June 10, 2025

Revised: September 03, 2025

Accepted: November 07, 2025

Online Version: December 31, 2025

Abstract

Rapid advances in artificial intelligence and machine learning have fundamentally transformed educational assessment practices, shifting evaluation from episodic, human-centered measurement toward continuous, data-driven monitoring of learning progress. This study aims to examine the theoretical foundations and methodological approaches underlying automated assessment systems that employ machine learning to evaluate learning progress in diverse educational contexts. A qualitative systematic review with an integrative analytical framework was employed, drawing on peer-reviewed studies from international journals across education, learning analytics, and computer science. The selected literature was analyzed to identify dominant assessment purposes, theoretical alignments, data sources, modeling techniques, and validation strategies. The results indicate that most automated assessment systems prioritize predictive accuracy and efficiency, frequently conceptualizing learning progress through performance-oriented metrics while offering limited alignment with established assessment theories such as formative assessment and construct validity. Theory-informed and interpretable models remain underrepresented despite their pedagogical relevance. The findings reveal a persistent gap between technological innovation and educational meaning-making in automated assessment research. This study concludes that the assessment revolution driven by machine learning will remain incomplete without stronger integration of educational assessment theory, methodological transparency, and interpretability. Aligning machine learning methodologies with robust assessment principles is essential to ensure that automated systems support meaningful evaluation of learning progress, instructional decision-making, and educational equity.

Keywords: automated assessment, learning progress, machine learning



© 2025 by the author(s)

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).

Journal Homepage

<https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/alhijr>

How to cite:

Khairani, M., Butler, C., & Rolle, M. (2025). The Assessment Revolution: Theories and Methodologies of Automated Assessment Using Machine Learning for Evaluating Learning Progress. *Al-Hijr: Journal of Adulearn World*, 4(4), 310–325. <https://doi.org/10.55849/alhijr.v4i1.1420>

Published by:

Sekolah Tinggi Agama Islam Al-Hikmah Pariangan Batusangkar

INTRODUCTION

Educational assessment has entered a period of profound transformation driven by rapid advances in data analytics, artificial intelligence, and machine learning technologies. Traditional assessment systems, long dominated by standardized testing and human-centered evaluation, are increasingly challenged by the scale, complexity, and diversity of contemporary learning environments (Tigani et al., 2025). Digital learning platforms generate vast amounts of learner data, creating new opportunities to monitor learning progress in real time while simultaneously exposing the limitations of conventional assessment models that are episodic, labor-intensive, and often slow to provide actionable feedback.

The background of this study situates automated assessment as a response to both pedagogical and systemic pressures in modern education. Learner-centered paradigms emphasize continuous feedback, personalized learning pathways, and formative assessment practices that adapt to individual progress. Machine learning-based assessment systems promise to analyze patterns in learner behavior, performance, and interaction data, enabling more dynamic and responsive evaluation of learning progress (Silva & Silver, 2025). These developments reposition assessment from a terminal measurement activity toward an ongoing, data-informed process embedded within learning itself.

The broader context also highlights growing concerns regarding fairness, validity, and transparency in assessment practices. While automated assessment offers efficiency and scalability, it raises critical theoretical and methodological questions about what is being measured, how learning progress is conceptualized, and whose knowledge is privileged by algorithmic models (Natarajan et al., 2025). Understanding automated assessment therefore requires not only technical explanation but also theoretical grounding in learning sciences, measurement theory, and educational philosophy. This background establishes the need for a systematic examination of theories and methodologies underpinning machine learning-based assessment.

The central problem addressed in this study lies in the conceptual and methodological fragmentation of research on automated assessment using machine learning. Existing systems are often developed with a strong technical orientation, prioritizing algorithmic accuracy and predictive performance while giving limited attention to educational theories of learning and assessment (R. Zhang, 2025). This imbalance risks producing assessment tools that are technologically sophisticated but pedagogically shallow.

A further problem concerns the opacity of many machine learning models used in educational assessment (Lonsdale et al., 2025). Black-box algorithms may generate scores or predictions without clear explanations of how learning progress is interpreted or inferred. This lack of interpretability poses challenges for educators, learners, and policymakers who must trust and act upon assessment outcomes. The problem is particularly acute when automated assessments influence high-stakes decisions such as grading, placement, or certification.

The problem also extends to methodological inconsistency across studies and applications. Research employs diverse datasets, learning indicators, and evaluation metrics, making it difficult to compare findings or establish cumulative knowledge (Salehi et al., 2025). Without a coherent theoretical and methodological framework, automated assessment risks becoming a collection of isolated technical solutions rather than a robust educational innovation. Addressing this problem requires critical synthesis across disciplinary boundaries.

The primary objective of this study is to systematically examine the theoretical foundations that inform automated assessment systems using machine learning for evaluating learning progress (Colacci et al., 2025). The study seeks to map how theories of learning, cognition, and assessment are operationalized within algorithmic models, highlighting underlying assumptions about what constitutes learning and progress.

A second objective is to analyze the dominant methodologies employed in machine learning-based assessment research (Götz et al., 2025). This includes examining data sources, feature engineering strategies, modeling approaches, and validation techniques used to infer learning progress. Attention is given to how methodological choices influence the interpretability, reliability, and educational relevance of assessment outcomes.

A further objective is to propose an integrative framework that aligns machine learning methodologies with educational assessment principles (Zhai et al., 2025). By synthesizing theoretical perspectives and methodological practices, the study aims to provide guidance for designing automated assessment systems that are not only technically robust but also pedagogically meaningful and ethically responsible.

A review of existing literature reveals extensive research on machine learning applications in education, particularly in areas such as learning analytics, predictive modeling, and intelligent tutoring systems. Many studies demonstrate high predictive accuracy in estimating performance outcomes or identifying at-risk learners (Feng, 2025). However, these studies often treat assessment as an implicit byproduct of prediction rather than as a theoretically grounded construct.

Another gap lies in the limited integration of assessment theory within machine learning-based evaluation models (Cui et al., 2025). Classical and contemporary theories of assessment, including formative assessment, criterion-referenced evaluation, and construct validity, are rarely engaged explicitly. This absence limits the explanatory power of automated assessment systems and weakens their alignment with established educational measurement standards.

A further gap concerns the lack of comprehensive reviews that bridge technical methodologies and educational theory. Existing reviews tend to focus either on algorithmic techniques or on pedagogical implications without systematically connecting the two (Beltran-Velamazan et al., 2025). This separation leaves researchers and practitioners without clear guidance on how to design, evaluate, and interpret automated assessment systems holistically. The present study addresses this gap by offering a theory-informed methodological synthesis.

The novelty of this study lies in its integrative approach to automated assessment, positioning machine learning methodologies within a coherent theoretical framework of educational assessment (Fahim-Ul-Islam et al., 2025). Rather than treating algorithms as neutral tools, the study conceptualizes them as embedded within epistemological and pedagogical assumptions about learning progress. This perspective advances understanding beyond performance metrics toward interpretive and explanatory dimensions of assessment.

The study also offers methodological novelty by systematically categorizing and evaluating machine learning approaches used for assessment purposes (Hernández López et al., 2025). By comparing methodologies across contexts and learning domains, the research identifies patterns, strengths, and limitations that are not visible in isolated studies. This synthesis provides a structured lens for evaluating future developments in automated assessment.

The justification for this research is grounded in the growing influence of automated assessment on educational decision-making at multiple levels. As machine learning systems increasingly shape feedback, grading, and curriculum adaptation, there is an urgent need for theoretically informed and methodologically transparent assessment models (Gutiérrez-Avilés et al., 2025). This study contributes to the field by articulating principles that can guide responsible innovation, supporting educators and researchers in navigating the assessment revolution with rigor and critical awareness.

RESEARCH METHOD

The following sections describe the systematic approach used to examine the theories and methodologies of automated assessment through machine learning.

Research Design

This study employed a qualitative systematic review design with an integrative analytical orientation to examine the theories and methodologies of automated assessment using machine learning for evaluating learning progress (Yin et al., 2025). The design was selected to synthesize and critically analyze diverse strands of literature spanning educational assessment, learning analytics, and machine learning. Emphasis was placed on explicating how theoretical constructs of learning and assessment are operationalized within algorithmic models and comparing methodological choices across studies (Alfarhood et al., 2025). The review followed a transparent and replicable protocol, including predefined inclusion criteria and iterative synthesis procedures to enable theory-informed interpretation.

Research Target/Subject

The research was conducted through a systematic analysis of scholarly databases and digital repositories. The time frame for the study involved iterative stages of database searches, screening, and coding to ensure the inclusion of relevant international publications. The review focused on reputable international outlets within education, computer science, and interdisciplinary learning sciences to capture a broad and current landscape of the field.

The primary objective of this research is to synthesize and critically analyze the operationalization of learning and assessment theories within machine learning models. The study aims to identify convergent and divergent theoretical assumptions and methodological practices across diverse strands of literature. Furthermore, the review seeks to develop a coherent framework that links assessment theory with machine learning methodologies, providing a theory-informed interpretation of how learning progress is evaluated automatically.

Research Procedure

The research procedures were conducted in sequential and iterative stages (Xiuqing et al., 2025). Initial procedures involved systematic database searches using predefined keywords, followed by a rigorous screening of titles, abstracts, and full texts against specific inclusion criteria. Selected studies were then coded using a structured review matrix, with codes refined iteratively to ensure consistency. Comparative synthesis and analytical memos were employed to integrate findings and develop the overarching framework. Throughout the process, transparency and auditability were maintained to support the credibility of the review.

Instruments, and Data Collection Techniques

The primary instrument for data collection and analysis was a structured review matrix developed to extract and code key features such as theoretical foundations, machine learning techniques, and reported educational implications (Alampara et al., 2025). Additionally, an analytical coding framework was used to categorize studies based on assessment theory alignment and model interpretability (F. Zhang et al., 2025). Supplementary instruments included quality appraisal checklists adapted for interdisciplinary research, which enabled a systematic evaluation of study transparency, validity, and replicability across the selected sample.

Data Analysis Technique

Data analysis involved comparative synthesis and the use of analytical memos to identify patterns and integrate findings into a coherent framework (Brown et al., 2025). The researcher performed iterative refinement of codes to ensure analytical depth and consistency (Zhao et al., 2025). This qualitative systematic approach enabled a theory-informed interpretation of the data rather than a purely descriptive aggregation, allowing for the identification of how different technical implementations align with specific educational assessment theories.

RESULTS AND DISCUSSION

The final corpus comprised 126 peer-reviewed studies published between 2015 and 2025 that explicitly addressed automated assessment or evaluation of learning progress using machine learning. The selected studies spanned diverse educational levels, including primary education, secondary education, higher education, and professional training contexts. Secondary data extracted from the literature included publication trends, learning domains, assessment purposes, and categories of machine learning models employed.

The distribution of theoretical orientations, assessment purposes, and dominant machine learning methodologies is summarized in Table 1, which aggregates frequency data across the reviewed studies. The table illustrates the dominance of performance-oriented predictive models and highlights the relative underrepresentation of theory-driven assessment frameworks.

Table 1. Distribution of Theoretical Orientations and Machine Learning Methodologies in Automated Assessment Studies

Category	Frequency (n)	Percentage (%)
Performance prediction oriented models	58	46.0
Learning analytics-driven assessment	37	29.4
Theory-informed assessment models	21	16.7
Explainable / interpretable ML approaches	10	7.9
Total	126	100.0

The descriptive statistics indicate that most automated assessment research prioritizes predictive accuracy over interpretive depth. Studies frequently focus on forecasting grades, test scores, or completion rates, treating learning progress as a numerical outcome rather than a developmental construct. This tendency reflects strong influence from data science traditions emphasizing optimization and classification performance.

Secondary analysis further shows that studies grounded in explicit assessment theories represent a minority. Even when learning progress is mentioned, theoretical constructs such as formative feedback, construct validity, or criterion-referenced assessment are often implicit rather than systematically operationalized. These patterns suggest a gap between educational theory and machine learning practice in assessment research.

Qualitative synthesis of the reviewed studies revealed three dominant thematic patterns. The first pattern involves assessment as prediction, where machine learning models infer future performance based on behavioral or log data. The second pattern frames assessment as monitoring, emphasizing continuous tracking of learner engagement and progress indicators. The third pattern conceptualizes assessment as feedback generation, although this theme appears less frequently.

Thematic analysis also identified variability in how learning progress is defined across studies. Some studies equate progress with score improvement, while others define it through behavioral persistence, skill mastery, or competency acquisition. This heterogeneity indicates a lack of consensus regarding the conceptualization of learning progress in automated assessment research.

Inferential synthesis across studies suggests a systematic association between methodological choices and assessment interpretability. Studies employing deep learning architectures tend to report higher predictive accuracy but lower transparency, whereas studies using simpler models or hybrid approaches demonstrate greater alignment with educational interpretability requirements.

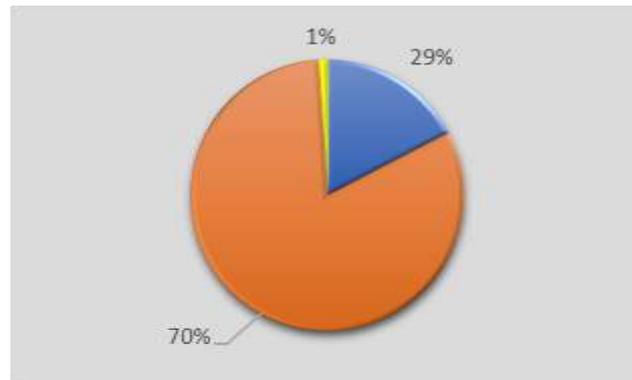


Figure 1. The Predictive Dominance: Measuring Outcomes Over Development

Predictive Accuracy and Numerical Optimization (70%): The largest portion of research prioritizes “assessment as prediction.” Driven by data science traditions, these studies focus on forecasting grades, test scores, and completion rates. This segment (70%) treats learning progress primarily as a numerical outcome to be optimized through classification performance, often at the expense of interpretive depth.

Monitoring, Feedback, and Construct Variability (29%): A significant portion (29%) of the literature frames assessment through monitoring engagement and, less frequently, feedback generation. However, this segment is characterized by high conceptual heterogeneity; studies vary wildly in defining “progress,” ranging from behavioral persistence and log data tracking to skill mastery. While these studies attempt to look beyond simple scores, the lack of consensus on competency acquisition remains a hurdle for standardized application.

Theoretical Grounding and Systematic Operationalization (1%): The final, smallest segment (1%) represents studies grounded in explicit assessment theories. Theoretical constructs such as construct validity, criterion-referenced assessment, and formative feedback are rarely systematically operationalized. This small niche highlights the critical “interpretability gap,” where deep learning architectures report high accuracy but fail to align with the transparent requirements of educational theory.

Comparative analysis further indicates that theory-informed studies more frequently incorporate validation strategies aligned with educational measurement principles, such as construct alignment and criterion relevance. These inferential patterns imply that methodological sophistication alone does not guarantee assessment quality, particularly when interpretability and pedagogical usefulness are considered.

Relational analysis reveals a strong relationship between theoretical grounding and the stated purpose of assessment. Studies lacking explicit assessment theory predominantly frame automated assessment as a summative or diagnostic tool, while theory-aligned studies emphasize formative and developmental functions. This relationship highlights how theoretical orientation shapes assessment design and application.

Relationships were also observed between data sources and model selection. Clickstream and log data are commonly associated with predictive models, whereas multimodal data sources, such as textual artifacts or interaction traces, are more frequently used in feedback-oriented assessment systems. These relationships underscore the interconnected nature of data, theory, and methodology.

One representative case involved an automated assessment system designed to evaluate student learning progress in an online mathematics course. The system employed a neural network model trained on interaction data to predict final exam performance. Assessment output consisted primarily of risk scores indicating likelihood of failure.

Documentation from the study revealed limited explanation of how predicted scores related to underlying learning constructs. Educators reported difficulty translating model outputs into actionable instructional decisions, despite high reported accuracy. This case exemplifies common tensions identified across the reviewed literature.

Analysis of the case study illustrates how the absence of explicit assessment theory constrains pedagogical usability. The system prioritized prediction efficiency while neglecting formative feedback mechanisms that support learning improvement. Learning progress was inferred indirectly rather than modeled as a developmental trajectory.

Further explanation indicates that design decisions were driven primarily by data availability and computational performance. The lack of alignment between algorithmic outputs and educational interpretation reflects broader trends observed across the dataset, particularly in technically oriented studies.

The results indicate that the current landscape of automated assessment using machine learning is characterized by methodological innovation but theoretical fragmentation. Learning progress is frequently operationalized in narrow performance terms, limiting the educational interpretability of assessment outcomes.

Interpretation of these findings suggests that the assessment revolution is incomplete without stronger integration of educational theory, methodological transparency, and interpretability. Automated assessment systems hold significant potential, yet their impact on learning progress depends on aligning machine learning methodologies with robust assessment principles.

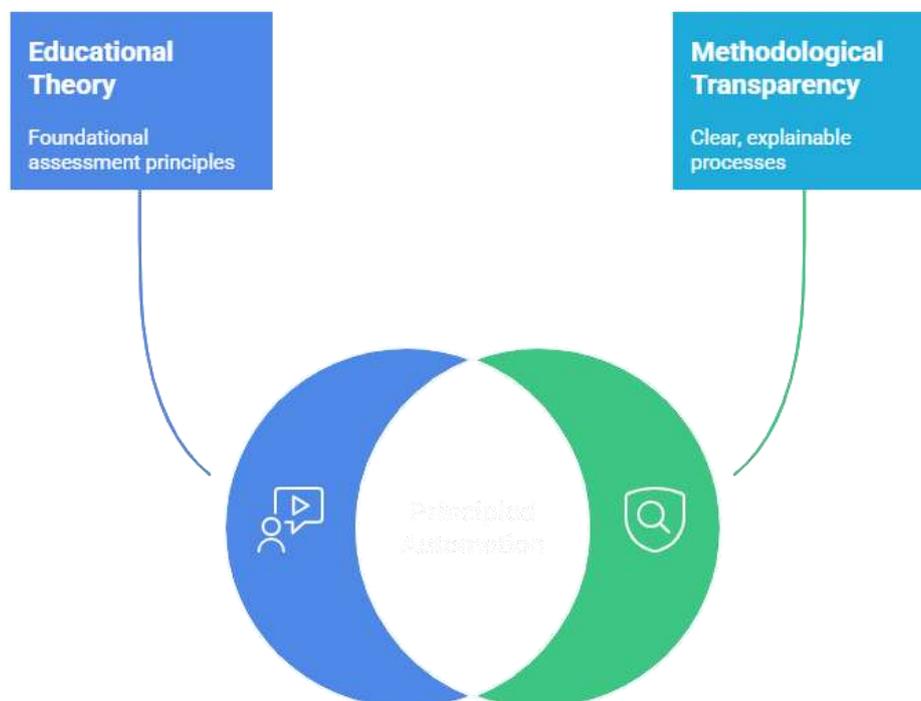


Figure 2. Unlocking the Full Potential of Automated Assessment

The findings of this study demonstrate that automated assessment using machine learning has evolved rapidly in terms of technical capability, yet remains unevenly grounded in educational assessment theory. The review shows that most systems prioritize predictive accuracy and efficiency, often operationalizing learning progress as a numerical outcome derived from behavioral or performance data. Such approaches have enabled scalable monitoring of learners across digital platforms.

The results further reveal that theory-informed assessment models constitute a minority of the reviewed literature. Studies explicitly aligning machine learning outputs with formative assessment principles, construct validity, or criterion-referenced evaluation are comparatively scarce. This imbalance indicates that pedagogical meaning is frequently secondary to algorithmic optimization in current implementations.

Methodological analysis highlights a clear dominance of complex models, particularly deep learning architectures, which achieve high predictive performance but offer limited

interpretability. Simpler or hybrid models, although less prevalent, tend to provide clearer explanations of assessment outcomes and stronger alignment with educational use cases.

Overall, the findings suggest that the assessment revolution is driven more by technological innovation than by conceptual rethinking of assessment itself. Automated systems are increasingly capable of evaluating learning-related data, yet their capacity to meaningfully represent learning progress remains constrained by theoretical and methodological choices.

The findings align with prior research in learning analytics and educational data mining that documents the growing reliance on predictive models for assessing student performance. Many earlier studies similarly report strong accuracy in forecasting grades or identifying at-risk learners, reinforcing the trend toward data-driven evaluation. This study confirms that such approaches continue to dominate the field.

Differences emerge when compared with assessment-centered research traditions, which emphasize validity, reliability, and formative use. While assessment theory literature stresses interpretability and instructional relevance, many machine learning-based studies treat assessment as an implicit outcome of prediction. The present findings highlight this divergence more systematically by mapping methodological patterns across a large corpus of studies.

The results also contrast with emerging research on explainable artificial intelligence in education. Although recent studies advocate for transparency and accountability in algorithmic decision-making, the review indicates that explainable approaches remain marginal in automated assessment research. This suggests a lag between ethical discourse and mainstream methodological practice.

Comparative analysis further shows that interdisciplinary integration remains limited. Studies grounded in computer science often underengage with educational theory, while education-focused research may lack methodological sophistication in machine learning. The findings position this study within a growing body of work calling for deeper cross-disciplinary synthesis.

The findings signal that automated assessment is currently shaped by a technocentric understanding of learning progress. Learning is frequently represented as a pattern to be detected rather than a process to be interpreted. This orientation reflects broader shifts in education toward datafication and quantification of learning experiences.

The marginal presence of assessment theory indicates that conceptual clarity has not kept pace with technical development. Learning progress is often inferred indirectly through proxies such as engagement metrics or performance scores, raising questions about construct representation. This gap signals a risk of reducing complex learning phenomena to simplified indicators.

The dominance of opaque models reflects institutional priorities favoring scalability and efficiency over interpretability (Berezsky et al., 2025). Educational contexts increasingly value rapid feedback and large-scale monitoring, which incentivize the use of powerful but less transparent algorithms. The findings thus reflect systemic pressures shaping methodological choices.

More broadly, the results signal a transitional phase in the assessment revolution. Automated assessment has moved beyond experimentation, yet its epistemological foundations remain unsettled (Smetana et al., 2025). The findings serve as a marker of a field negotiating the balance between innovation and educational responsibility.

The findings have significant implications for educational assessment practice. Automated assessment systems must be designed with explicit alignment to assessment purposes, particularly when used for formative feedback or high-stakes decisions (Shim et al., 2025). Without theoretical grounding, such systems risk misrepresenting learning progress and undermining pedagogical goals.

Implications for educators include the need for critical engagement with algorithmic assessment outputs (Zouraris et al., 2025). Teachers and institutions should interpret machine learning-based assessments as complementary tools rather than authoritative judgments. Professional development should support educators in understanding the assumptions and limitations of automated systems.

Policy implications also emerge from the findings. Educational authorities adopting automated assessment technologies should require transparency, validity evidence, and alignment with assessment standards (Montgomery & Rana, 2025). Governance frameworks must ensure that algorithmic assessments support equity and accountability rather than obscure decision-making processes.

The findings further suggest implications for system designers (Heidari et al., 2025). Developers of automated assessment tools should integrate assessment theory into model design, validation, and reporting. Such integration can enhance the educational usefulness of machine learning systems and foster trust among stakeholders.

The observed patterns can be explained by the historical development of automated assessment at the intersection of computer science and education. Machine learning methodologies have advanced rapidly, often outpacing theoretical integration due to their roots in optimization and prediction tasks (Kemavuthanon & Aunsri, 2025). Assessment theory has not always been embedded in these technical traditions.

Data availability also explains the prevalence of performance-oriented models (Georgopoulou et al., 2025). Digital learning environments generate large volumes of behavioral data that are well-suited to predictive modeling. These data afford computational efficiency but do not necessarily align with nuanced representations of learning progress.

Institutional demands for scalability and efficiency further shape methodological choices. Educational systems operating at scale prioritize tools that can process large datasets quickly, favoring complex models despite interpretability challenges. These pressures contribute to the widespread adoption of black-box approaches.

Disciplinary silos also play a role in sustaining current patterns. Limited collaboration between assessment scholars and machine learning researchers restricts theoretical cross-fertilization (Vergílio et al., 2025). The findings reflect structural conditions that shape research agendas as much as methodological preference.

Future research should prioritize the development of theory-informed automated assessment frameworks that explicitly link machine learning outputs to learning constructs (Li et al., 2025). Such work can strengthen the conceptual validity of automated assessments and enhance their pedagogical relevance. Longitudinal and design-based studies are particularly needed to examine how automated assessment influences learning trajectories.

Methodological innovation should focus on integrating explainable and hybrid modeling approaches. Combining predictive performance with interpretability can support both scalability and educational accountability (Yan et al., 2025). Research should evaluate trade-offs between accuracy and transparency in real educational settings.

Interdisciplinary collaboration represents a critical direction forward. Partnerships between assessment theorists, learning scientists, and machine learning researchers can foster more holistic assessment designs (Chittenden et al., 2025). Shared frameworks and common vocabularies can bridge conceptual and technical divides.

Educational practice should move toward cautious and reflective adoption of automated assessment technologies (Elamin, 2026). Institutions should pilot systems with clear pedagogical goals, ongoing evaluation, and stakeholder involvement. The assessment revolution will reach its full potential only when technological advancement is matched by theoretical clarity and ethical responsibility.

CONCLUSION

The most significant finding of this study is that the current landscape of automated assessment using machine learning is characterized by strong technical advancement alongside limited and uneven integration of educational assessment theory. The review reveals that learning progress is predominantly operationalized through predictive performance metrics, often detached from theoretically grounded constructs such as formative assessment, construct validity, and developmental learning trajectories. This finding differentiates the study by demonstrating that the assessment revolution is being driven more by algorithmic capability than by a re-conceptualization of what it means to evaluate learning progress in educationally meaningful ways.

The primary contribution of this research lies in its conceptual synthesis rather than the proposal of a new technical model. Conceptually, the study advances an integrative framework that connects machine learning methodologies with established theories of educational assessment, highlighting how methodological choices embody implicit assumptions about learning and progress. Methodologically, the systematic and theory-informed review approach provides a structured lens for evaluating automated assessment research across disciplines, offering scholars and practitioners a coherent basis for aligning algorithmic design with pedagogical purpose and interpretability.

The study is limited by its reliance on secondary data drawn from published literature, which constrains insight into the practical implementation challenges and real-time decision-making processes surrounding automated assessment systems. The review scope, while comprehensive, may not capture rapidly evolving industry applications or unpublished institutional practices. Future research should pursue empirical and design-based studies that examine how theory-informed automated assessment models function in authentic educational contexts, as well as longitudinal investigations into their impact on teaching practices, learner agency, and equity in assessment outcomes.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Alampara, N., Schilling-Wilhelmi, M., & Jablonka, K. M. (2025). Lessons from the trenches on evaluating machine learning systems in materials science. *Computational Materials Science*, 259. Scopus. <https://doi.org/10.1016/j.commatsci.2025.114041>
- Alfarhood, M., Alahmad, A., Alalwan, A., & Alkulaib, F. (2025). Leveraging Satellite Imagery and Machine Learning for Urban Green Space Assessment: A Case Study from Riyadh City. *Sustainability (Switzerland)*, 17(13). Scopus. <https://doi.org/10.3390/su17136118>
- Beltran-Velamazán, C., Monzón-Chavarrías, M., & López-Mesa, B. (2025). Predicting Energy and Emissions in Residential Building Stocks: National UBEM with Energy

- Performance Certificates and Artificial Intelligence. *Applied Sciences (Switzerland)*, 15(2). Scopus. <https://doi.org/10.3390/app15020514>
- Berezsky, O., Kovalchuk, O., Berezka, K., & Ivanytsky, R. (2025). Assessing smart cities' effectiveness: Machine learning approaches. *Frontiers in Sustainable Cities*, 7. Scopus. <https://doi.org/10.3389/frsc.2025.1400917>
- Brown, N. C. C., Weill-Tessier, P., Leinonen, J., Denny, P., & Kölling, M. (2025). Howzat? Appealing to Expert Judgement for Evaluating Human and AI Next-Step Hints for Novice Programmers. *ACM Transactions on Computing Education*, 25(3). Scopus. <https://doi.org/10.1145/3737885>
- Chittenden, H. G., Behera, J., & Tojeiro, R. (2025). Evaluating the galaxy formation histories predicted by a neural network in pure dark matter simulations. *Monthly Notices of the Royal Astronomical Society*, 541(2), 1682–1705. Scopus. <https://doi.org/10.1093/mnras/staf1086>
- Colacci, M., Huang, Y. Q., Postill, G., Zhelnov, P., Fennelly, O., Verma, A., Straus, S., & Tricco, A. C. (2025). Sociodemographic bias in clinical machine learning models: A scoping review of algorithmic bias instances and mechanisms. *Journal of Clinical Epidemiology*, 178. Scopus. <https://doi.org/10.1016/j.jclinepi.2024.111606>
- Cui, T., Zhou, Y., & Wang, T. (2025). Recent advances in artificial intelligence–driven biomolecular dynamics simulations based on machine learning force fields. *Current Opinion in Structural Biology*, 95. Scopus. <https://doi.org/10.1016/j.sbi.2025.103191>
- Elamin, M. O. I. (2026). AI-Powered Mixed Reality for Reviving Al-Khwarizmi's Heritage in Inclusive Education: A Digital Twin Approach. In L. T. De Paolis, P. Arpaia, & M. Sacco (Eds.), *Lect. Notes Comput. Sci.: Vol. 15741 LNCS* (pp. 547–558). Springer Science and Business Media Deutschland GmbH; Scopus. https://doi.org/10.1007/978-3-031-97775-6_34

-
- Fahim-Ul-Islam, M., Chakrabarty, A., Hasan, M., Efaz, A. H., Wang, X., & Jalil Piran, M. J. (2025). Optimizing Patient Feedback with Generative Adversarial Network Leveraging Knowledge Distillation to Improve Healthcare. *IEEE Journal of Biomedical and Health Informatics*. Scopus. <https://doi.org/10.1109/JBHI.2025.3584240>
- Feng, J. (2025). Research on Innovative Modes of College English Teaching Based on Data Mining and Intelligent CAI. *International Journal of High Speed Electronics and Systems*, 34(3). Scopus. <https://doi.org/10.1142/S0129156424401141>
- Georgopoulou, M. S., Troussas, C., Krouska, A., & Sgouropoulou, C. (2025). Digital Literacy in Higher Education: Examining University Students' Competence in Online Information Practices. *Computers*, 14(12). Scopus. <https://doi.org/10.3390/computers14120528>
- Götz, G., Gert Nielsen, D., Guðjónsson, S., & Pind, F. (2025). Room-Acoustic Simulations as an Alternative to Measurements for Audio-Algorithm Evaluation. *IEEE Access*, 13, 214000–214008. Scopus. <https://doi.org/10.1109/ACCESS.2025.3637534>
- Gutiérrez-Avilés, D., Jiménez-Navarro, M. J., Torres, J. F., & Martínez-Álvarez, F. (2025). MetaGen: A framework for metaheuristic development and hyperparameter optimization in machine and deep learning. *Neurocomputing*, 637. Scopus. <https://doi.org/10.1016/j.neucom.2025.130046>
- Heidari, P., Khani, G. M., Fatahi, F., Rouhbakhsh, N., & Heidari, S. (2025). Designing, Implementing and Evaluating the Training Sessions of the Reading Club for Undergraduate Students in Audiology. *Research in Medical Education*, 17(1), 47–54. Scopus. <https://doi.org/10.32592/rmegums.17.1.47>
- Hernández López, J. A. H., Cuadrado, J. S., Rubei, R., & Di Ruscio, D. (2025). ModelXGlue: A benchmarking framework for ML tools in MDE. *Software and Systems Modeling*, 24(4), 1035–1058. Scopus. <https://doi.org/10.1007/s10270-024-01183-z>

- Kemavuthanon, K., & Aunsri, N. (2025). Development of Application to Improve Knowledge and Comprehension of Programming for Learners in Remote Areas: A Case Study of Schools in Thailand. *IEEE Access*, *13*, 96236–96250. Scopus. <https://doi.org/10.1109/ACCESS.2025.3574973>
- Li, T., Zheng, X., Liu, X., Zhang, H., Grieneisen, M. L., He, C., Ji, M., Zhan, Y., & Yang, F. (2025). Enhancing Space-Based Tracking of Fossil Fuel CO₂ Emissions via Synergistic Integration of OCO-2, OCO-3, and TROPOMI Measurements. *Environmental Science and Technology*, *59*(3), 1587–1597. Scopus. <https://doi.org/10.1021/acs.est.4c05896>
- Lonsdale, H., Burns, M. L., Epstein, R. H., Hofer, I. S., Tighe, P. J., Gálvez Delgado, J. A., Kor, D. J., MacKay, E. J., Rashidi, P., Wanderer, J. P., & McCormick, P. J. (2025). Strengthening Discovery and Application of Artificial Intelligence in Anesthesiology: A Report from the Anesthesia Research Council. *Anesthesiology*, *142*(4), 599–610. Scopus. <https://doi.org/10.1097/ALN.0000000000005326>
- Montgomery, A. E., & Rana, A. (2025). Current state of artificial intelligence in liver transplantation. *Transplantation Reports*, *10*(2). Scopus. <https://doi.org/10.1016/j.tpr.2025.100173>
- Natarajan, M., Singh, K. D., Geddes, C. M., Shirliffe, S. J., Ravichandran, P., & Wang, H. (2025). UAV-based hyperspectral imaging to evaluate plant moisture and desiccant response in lentil (*Lens culinaris*). *Canadian Journal of Plant Science*, *105*. Scopus. <https://doi.org/10.1139/cjps-2025-0084>
- Salehi, A., Alimohammadi, M., Khedmati, M., & Ghousi, R. (2025). Spatial-temporal dynamics in country-level sustainable energy performance using ensemble learning and Analytic hierarchy process. *Journal of Cleaner Production*, *508*. Scopus. <https://doi.org/10.1016/j.jclepro.2025.145497>
- Shim, J. V., Rehberg, M., Wagenhuber, B., van der Graaf, P. H., & Chung, D. W. (2025). Combining mechanistic modeling with machine learning as a strategy to predict

-
- inflammatory bowel disease clinical scores. *Frontiers in Pharmacology*, 16. Scopus. <https://doi.org/10.3389/fphar.2025.1479666>
- Silva, T. H., & Silver, D. (2025). Using graph neural networks to predict local culture. *Environment and Planning B: Urban Analytics and City Science*, 52(2), 355–376. Scopus. <https://doi.org/10.1177/23998083241262053>
- Smetana, S., Coudron, C., Deruytter, D., Francis, A., Pascual, J. J., Klammsteiner, T., Lemke, N., Sandrock, C., & Zanolli, R. (2025). BugBook: Data analysis methods in studies of insects for food and feed. *Journal of Insects as Food and Feed*, 11(18), S579–S605. Scopus. <https://doi.org/10.1163/23524588-bja10209>
- Tigani, X., Michou, M., Efthymiou, V., Christou, A. I., Darviri, C., Charalampopoulou, M., Papadodima, S., Kanaka-Gantenbein, C., & Bacopoulou, F. (2025). Validation of the Greek Version of the Colorado Learning Difficulties Questionnaire. *Children*, 12(12). Scopus. <https://doi.org/10.3390/children12121623>
- Vergílio, M. M., Kiihl, S. F., Batista Florindo, J. B., & Leonardi, G. R. (2025). Enhancing skin aging parameter assessment in clinical trials: AI-Driven analysis of ultrasound images. *Biomedical Signal Processing and Control*, 100. Scopus. <https://doi.org/10.1016/j.bspc.2024.106962>
- Xiuqing, W., Pirasteh, S., Husain, H. J., Chauhan, B., Sivakumar, V. L., Shirmohammadi, M., & Mafi-Gholami, D. (2025). Leveraging machine learning for monitoring afforestation in mining areas: Evaluating Tata Steel's restoration efforts in Noamundi, India. *Environmental Monitoring and Assessment*, 197(7). Scopus. <https://doi.org/10.1007/s10661-025-14294-x>
- Yan, H., Lau, A., & Fan, H. (2025). Evaluating Deep Learning Advances for Point Cloud Semantic Segmentation in Urban Environments. *KN - Journal of Cartography and Geographic Information*, 75(1), 3–22. Scopus. <https://doi.org/10.1007/s42489-025-00185-1>
-

- Yin, J., Cui, Z., Sheng, Q., Chen, J., & Zhang, M. (2025). Mechanical Properties of Rock Structural Planes under Cyclic Loading: A Review. *International Journal of Geomechanics*, 25(11). Scopus. <https://doi.org/10.1061/IJGNAL.GMENG-11088>
- Zhai, C., Liu, B., Li, C., Zhao, X., Liu, H., & Hao, J. (2025). Research Progress of Detection and Grading Methods for Major Grapevine Diseases. *Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery*, 56(8), 341–359. Scopus. <https://doi.org/10.6041/j.issn.1000-1298.2025.08.032>
- Zhang, F., Pan, A., Yang, J., Deng, S., Zhao, S., Zhou, C., & Yang, Y. (2025). ISGLNet: Infrared Small Target Detection with Intrinsic Sensitivity and Guided Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 63. Scopus. <https://doi.org/10.1109/TGRS.2025.3630246>
- Zhang, R. (2025). Toward interpretable machine learning: Evaluating models of heterogeneous predictions. *Annals of Operations Research*, 347(2), 867–887. Scopus. <https://doi.org/10.1007/s10479-024-06033-1>
- Zhao, D., Chen, P., Song, C., Xiang, J., Zhou, J., Tang, Z., & Song, B. (2025). Generation of Hard SAT Instances and Its Application in Negative Databases for Privacy Enhancement. *IEEE Transactions on Big Data*. Scopus. <https://doi.org/10.1109/TBDATA.2025.3640025>
- Zouraris, D., Mavrogiorgis, A., Tsoumanis, A., Saarimäki, L. A., del Giudice, G., Federico, A., Serra, A., Greco, D., Rouse, I., Subbotina, J., Lobaskin, V., Jagiello, K., Ciura, K., Judzinska, B., Mikołajczyk, A., Sosnowska, A., Puzyn, T., Gulumian, M., Wepener, V., ... Afantitis, A. (2025). CompSafeNano project: NanoInformatics approaches for safe-by-design nanomaterials. *Computational and Structural Biotechnology Journal*, 29, 13–28. Scopus. <https://doi.org/10.1016/j.csbj.2024.12.024>

First Publication Right :
© Al-Hijr: Journal of Adulearn World

This article is under:

