Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

ARTIFICIAL INTELLIGENCE-BASED LEARNING ASSESSMENT: A CONCEPTUAL SYNTHESIS OF ASSESSMENT ACCURACY AND STUDENT LEARNING OUTCOMES

Amroh Umaemah[1*], Hasan Yusuf[2], Inggrid Marlissa[3], Mesrawati Rifai[4], Lilis Suryani[5]

[1] UIN Siber Syekh Nurjati Cirebon
[2] Sekolah Tinggi Ilmu Tarbiyah Az Zahra Tasikmalaya
[3] Universitas Musamus
[4] Universitas Islam DDI Abdurrahman Ambo Dalle
[5] Universitas Panca Sakti Bekasi

[a]*E-mail:* Amrohumaemah@uin.ac.id
[b]*E-mail:* hasan.yusuf@student.unsil.ac.id
[c]*E-mail:* inggrid_fkip@unmus.ac.id
[d]*E-mail:* mesrawarifai@ddipolman.ac.id
[e]*E-mail:* lilisyeyen2019@gmail.com

(*) Corresponding Author
Amrohumaemah@uin.ac.id

ARTICLE HISTORY
Received : 29-08-2026
Revised : 12-11-2026
Accepted : 20-11-2026

KEYWORDS
Artificial Intelligence; learning evaluation; assessment accuracy

ABSTRACT
Artificial Intelligence (AI) offers significant opportunities to transform educational assessment by improving scoring consistency, feedback delivery, and instructional decision-making. However, technological efficiency does not necessarily guarantee assessment accuracy or improved student learning outcomes. This study aims to examine how AI-based learning evaluation can be optimized to enhance assessment accuracy and student learning outcomes in the digital era. A qualitative library research approach was employed to analyze relevant scholarly literature, theoretical perspectives, and institutional publications. The literature was synthesized thematically across four dimensions: assessment accuracy, AI-generated feedback, student learning outcomes, and ethical governance. The analysis indicates that AI can support automated scoring, adaptive assessment, learning analytics, and personalized feedback. Nevertheless, its effectiveness depends on assessment validity, fairness, human oversight, and students' engagement with feedback. This study proposes an integrated optimization framework connecting assessment design, AI-supported evaluation, quality assurance, instructional adaptation, and continuous outcome evaluation. The framework provides conceptual guidance for responsible AI implementation, although further empirical research is required to validate its effectiveness across educational contexts.
This is an open access article under the CC–BY-SA license.

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

INTRODUCTIONS

The rapid development of digital technologies has transformed educational practices, particularly in the areas of teaching, learning, and assessment. Digital learning environments increasingly enable educators to collect, process, and analyze student performance data through learning management systems, online assessment platforms, and intelligent educational applications. Within this transformation, Artificial Intelligence (AI) has emerged as a promising technology for improving the efficiency, adaptability, and analytical capacity of educational assessment. AI-based learning evaluation can support automated scoring, personalized feedback, learning progress monitoring, and the identification of students' academic difficulties. These capabilities offer opportunities to improve assessment practices that have traditionally relied heavily on manual grading and periodic examinations(Zawacki-Richter et al., 2019).

Assessment plays a fundamental role in education because it provides evidence of students' knowledge, skills, competencies, and learning progress. Effective assessment should not merely measure academic achievement but also inform instructional decisions, identify learning gaps, and support continuous improvement. However, conventional assessment practices frequently encounter challenges related to scoring consistency, delayed feedback, limited personalization, and the difficulty of evaluating large volumes of student work. These limitations may reduce the usefulness of assessment results, particularly in digitally mediated learning environments characterized by diverse student needs and increasing demands for timely instructional responses. AI-based assessment has therefore attracted scholarly attention as a means of connecting assessment data with more responsive and adaptive learning processes(Rosyalita et al., 2026).

Recent research has examined the relationship between AI and assessment for learning, emphasizing the potential of intelligent technologies to facilitate feedback and strengthen the connection between evaluation and instructional improvement. A review published in Computers in Human Behavior: Artificial Humans identified a growing body of research on AI-supported assessment for learning while highlighting challenges associated with implementation, assessment design, and the conceptualization of effective feedback (A Review of Assessment for Learning with Artificial Intelligence, 2024). Similarly, research on emerging technologies in higher education assessment has highlighted the increasing use of AI and learning analytics, while identifying human-centered implementation and ethical considerations as important directions for further development (Emerging Technologies in Higher Education Assessment and Feedback Practices, 2024).

Nevertheless, the integration of AI into educational assessment should not automatically be interpreted as evidence of improved assessment accuracy or student learning outcomes. AI systems can process information rapidly, recognize patterns, and generate feedback, but the quality of their outputs depends on the data, algorithms, evaluation criteria, and educational contexts in which they operate. Automated scoring systems, for example, may demonstrate consistency in processing responses while still failing to capture the complexity of students' reasoning, creativity, contextual understanding, or higher-order thinking. Consequently, assessment accuracy should be understood not only in terms of scoring consistency but also in relation to validity, fairness, reliability, and alignment with intended learning objectives.

The distinction between technical efficiency and educational effectiveness can be explained through Assessment for Learning (AfL), which provides the principal pedagogical perspective for this study. AfL conceptualizes assessment as an ongoing process of collecting, interpreting, and using evidence to support instructional decisions and student learning rather than merely measuring achievement. Within this perspective, AI-generated scores, learning analytics, and personalized feedback become educationally meaningful when they inform subsequent teaching and learning activities. Assessment validity further provides a basis for evaluating whether AI-generated interpretations and decisions are supported by appropriate evidence and aligned with the competencies being assessed. Together, these perspectives connect assessment quality, formative feedback, instructional adaptation, and student learning within a coherent conceptual foundation.

This theoretical integration highlights the importance of distinguishing feedback delivery from feedback utilization. An AI system may generate immediate and individualized feedback without necessarily supporting

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

meaningful learning if students cannot interpret or apply its recommendations. From an AfL perspective, feedback becomes pedagogically valuable when learners actively engage with assessment information, identify performance gaps, and undertake appropriate revisions. Similarly, adaptive learning mechanisms should be evaluated according to their responsiveness to demonstrated learning needs rather than their technological sophistication alone. Thus, AI-based assessment is conceptualized as a continuous process involving evidence generation, feedback interpretation, instructional adjustment, and subsequent evaluation.

Ethical and institutional considerations constitute another essential dimension of this conceptual perspective. AI-based assessment may reproduce existing inequalities when training data inadequately represent diverse student populations, languages, educational backgrounds, or learning needs. Student data collection also raises concerns regarding privacy, informed consent, data security, and institutional accountability. UNESCO's guidance emphasizes human-centered AI implementation that protects human agency, promotes inclusion, and addresses ethical risks in educational environments (Miao & Holmes, 2023). These considerations are particularly important when AI-generated scores or recommendations influence students' academic opportunities (Memarian & Doleck, 2024). Accordingly, human oversight and ethical governance are treated as cross-cutting conditions for responsible assessment rather than supplementary technological requirements.

Existing literature on AI in educational assessment can be distinguished according to its primary analytical emphasis. Research addressing automated scoring and AI-supported evaluation primarily examines technological capabilities, scoring consistency, and assessment procedures. Such approaches are useful for understanding operational performance but do not necessarily establish the pedagogical validity of assessment interpretations or their implications for learning. Literature concerning AI-generated feedback and adaptive learning places greater emphasis on personalization, instructional responsiveness, and learner engagement. However, the relationship between feedback availability and actual learning development requires careful consideration because feedback effectiveness depends on students' interpretation, participation, and opportunities for improvement.

A further strand of literature addresses ethical governance, algorithmic fairness, transparency, and institutional responsibility. For example, Miao and Holmes (2023) provide normative guidance for responsible generative AI implementation, whereas Memarian and Doleck (2024) are included in the discussion of AI-related assessment and educational considerations. These contributions represent different analytical orientations: policy-oriented guidance establishes ethical principles, while assessment-related scholarship can help examine their implications for educational practice. Nevertheless, these perspectives cannot be treated as equivalent forms of evidence, and conceptual recommendations should be distinguished from empirically demonstrated educational outcomes. Recent scholarship also highlights the broader challenges surrounding AI integration in education (Sembey et al., 2024), reinforcing the need to examine technological, pedagogical, and institutional dimensions together.

Despite these contributions, the central conceptual gap concerns the limited integration of assessment validity, formative feedback, instructional adaptation, and ethical governance within a unified explanation of AI-supported learning assessment. Technological performance indicators alone cannot establish whether assessment evidence is educationally meaningful, while pedagogical recommendations may remain incomplete without considering algorithmic reliability, data quality, and institutional safeguards. The issue is therefore not simply whether AI can improve assessment procedures, but under what conditions its capabilities can support valid assessment interpretations and meaningful learning processes.

To address this gap, this study employs qualitative library research to synthesize scholarly literature concerning AI-based educational assessment. Guided by Assessment for Learning, assessment validity, and human-centered AI governance, the study examines four interrelated dimensions: assessment accuracy, AI-generated feedback, student learning outcomes, and ethical governance. Its conceptual contribution is an integrated framework explaining how assessment design, AI-supported evaluation, quality assurance, instructional adaptation, and continuous monitoring are connected through pedagogical and institutional conditions. Rather than claiming empirically verified improvements in assessment accuracy or student learning outcomes, the study identifies potential

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

mechanisms, enabling conditions, and limitations that can inform responsible implementation and subsequent empirical investigation (Memarian & Doleck, 2024) (Sembey et al., 2024).

The central problem addressed in this study is the gap between the technological capabilities of AI-based assessment systems and their effective contribution to educational quality. Although AI can facilitate automated scoring, performance prediction, learning analytics, and personalized feedback, these capabilities do not independently guarantee valid assessment results or improved learning outcomes(Holmes et al., 2019).

Three interrelated problems can be identified.

First, assessment accuracy remains a multidimensional challenge. AI-based systems must evaluate student performance consistently and validly while accounting for differences in learning objectives, assessment formats, language use, and student characteristics. Excessive reliance on automated scoring may overlook aspects of learning that require contextual interpretation or professional judgment.

Second, the relationship between assessment feedback and learning improvement requires further conceptual clarification. Timely feedback is educationally valuable when it helps students understand their performance, identify weaknesses, and take corrective action. However, the availability of AI-generated feedback does not necessarily ensure that students engage with the feedback or improve their learning strategies(Luckin et al., 2016).

Third, the optimization of AI-based assessment requires appropriate governance and pedagogical integration. Bias, limited transparency, privacy risks, unequal access to technology, and insufficient teacher preparation can undermine the effectiveness and legitimacy of AI-supported evaluation. These issues indicate that optimization must encompass technical, pedagogical, organizational, and ethical dimensions.

These problems establish the need for a literature-based analysis that explains how AI-based assessment can be designed and implemented to improve measurement quality while supporting meaningful and equitable learning.

METHOD

Research Design

This study employed a qualitative library research design to examine how Artificial Intelligence (AI)-based learning evaluation can be optimized to improve assessment accuracy and student learning outcomes in the digital era. Library research was selected because the study aimed to investigate, interpret, and synthesize existing scholarly knowledge rather than collect primary data through experiments, surveys, interviews, or classroom observations. This approach enabled the researcher to examine theoretical perspectives, empirical findings, conceptual frameworks, and policy documents relevant to the implementation of AI in educational assessment(Shermis & Burstein, 2013).

Qualitative library research emphasizes the systematic examination and critical interpretation of documentary sources to develop an understanding of a particular research problem. In this study, the approach was used to identify recurring themes, compare different scholarly perspectives, and establish conceptual relationships between AI capabilities, assessment accuracy, feedback effectiveness, instructional adaptation, and student learning outcomes. The analysis was therefore interpretive and thematic, with particular attention to the educational implications of AI-supported assessment and the conditions under which its implementation may be effective(Ifenthaler & Yau, 2020).

The research was guided by an integrative perspective that considered AI-based assessment as both a technological process and a pedagogical practice. This perspective allowed the study to examine not only the technical capabilities of AI systems, such as automated scoring, learning analytics, and adaptive assessment, but also their implications for assessment validity, fairness, transparency, and educational decision-making. Accordingly, the study sought to develop a conceptual framework that connects technological optimization with pedagogical effectiveness and responsible AI governance.

The research design was informed by established approaches to qualitative inquiry and literature synthesis, particularly the principles of systematic source identification, critical evaluation, thematic interpretation, and transparent reporting. However, the study was not presented as a systematic review or meta-analysis because it did not apply a registered protocol, a formally exhaustive search strategy, or quantitative aggregation of empirical effect

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

sizes. Instead, it adopted a qualitative and integrative library research approach to generate a theoretically grounded understanding of AI-based learning evaluation.

Sources of Data

The data consisted of secondary sources obtained from published academic literature and authoritative institutional documents relevant to AI-based educational assessment. The sources included peer-reviewed journal articles, scholarly books, research reviews, conceptual papers, and policy publications addressing AI in education, educational measurement, formative assessment, feedback, learning analytics, and ethical AI implementation(Baker & Hawn, 2022).

Academic literature was identified through scholarly databases and search platforms, including Scopus, Web of Science, ERIC, Google Scholar, and ScienceDirect. These sources were selected to provide access to research from educational technology, educational psychology, assessment science, computer science, and related interdisciplinary fields. Institutional publications from organizations such as UNESCO were also considered when they provided authoritative guidance on the ethical and educational implications of AI(Holmes et al., 2022).

The literature search focused primarily on publications from 2019 to 2026 to capture recent developments in AI technologies and their applications in digital learning environments. Earlier publications were considered when they provided foundational theoretical perspectives on formative assessment, feedback, educational measurement, or learning improvement. The inclusion of foundational literature was necessary because contemporary AI applications should be interpreted in relation to established principles of assessment quality and educational effectiveness.

The study prioritized sources that directly addressed at least one of the following areas: AI-supported assessment and automated scoring; assessment accuracy, validity, reliability, and fairness; AI-generated feedback and formative assessment; learning analytics and personalized instruction; student learning outcomes; and ethical, institutional, or pedagogical challenges associated with AI adoption(Miao & Holmes, 2023).

Literature Search Strategy

The literature search was conducted using combinations of keywords related to the principal concepts of the study. Search terms were adapted to the indexing characteristics of each database and combined using Boolean operators, including AND and OR, to identify relevant publications.

The principal search terms included:

- “Artificial Intelligence” AND “educational assessment”
- “AI-based assessment” AND “assessment accuracy”
- “automated scoring” AND “assessment validity”
- “AI-generated feedback” AND “student learning outcomes”
- “learning analytics” AND “formative assessment”
- “adaptive assessment” AND “personalized learning”
- “artificial intelligence” AND “assessment reliability”
- “AI in education” AND “ethical assessment”
- “generative AI” AND “educational evaluation”
- “human oversight” AND “AI-based assessment”

The search process focused on the relationship between AI-supported evaluation and educational improvement rather than on AI technology in isolation. Publications addressing general AI applications were considered only when they offered relevant insights into assessment processes, learning feedback, educational decision-making, or student achievement(Tlili et al., 2023).

The search and selection process involved three broad stages. First, potentially relevant publications were identified through database searches and the examination of reference lists in relevant scholarly works. Second, titles and abstracts were screened to assess their relevance to the research questions. Third, the full texts of potentially

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

eligible publications were examined to determine whether they provided sufficient conceptual, methodological, empirical, or policy-related information for the synthesis(Bearman et al., 2023).

Because this study used qualitative library research rather than a formally registered systematic review protocol, the search process was intended to support a focused and critical synthesis of relevant literature. No claim of exhaustive retrieval or complete coverage of all publications in the field was made.

Inclusion and Exclusion Criteria

To maintain the relevance and analytical quality of the literature, sources were assessed according to predefined inclusion and exclusion criteria.

Criteria	Inclusion	Exclusion
Subject relevance	Studies addressing AI-based educational assessment, learning analytics, feedback, assessment quality, or learning outcomes	Studies focusing exclusively on AI applications unrelated to educational evaluation
Publication type	Peer-reviewed journal articles, scholarly books, research reviews, conceptual studies, and authoritative institutional reports	Unsubstantiated opinion pieces, promotional materials, and sources lacking sufficient methodological or conceptual information
Publication period	Primarily 2019–2026, with earlier foundational works included when theoretically relevant	Older publications without clear relevance to the study's theoretical or conceptual objectives
Language	Publications written in English	Publications in other languages that could not be reliably interpreted for the purposes of this study
Analytical contribution	Sources providing relevant theoretical explanations, empirical findings, methodological insights, or policy guidance	Sources that mentioned AI in education without meaningful relevance to the research questions
Source accessibility	Publications with sufficient accessible information to support critical analysis	Sources for which essential information could not be examined or verified

The inclusion and exclusion criteria were applied to determine whether each source contributed meaningfully to the study's objectives. Greater analytical weight was given to peer-reviewed research and authoritative publications with clearly identifiable arguments, methods, findings, or recommendations. Conceptual publications were also included when they provided important theoretical foundations for understanding assessment accuracy, feedback, or responsible AI implementation.

These criteria were intended to improve the relevance and credibility of the synthesis. Nevertheless, the study recognizes that differences in research design, educational context, assessment format, and AI technology may limit direct comparisons among the selected publications.

Data Collection Procedure

Data collection was conducted through documentary examination of the selected literature. The researcher identified relevant information from each source and organized it according to the study's research questions and analytical categories.

The data collection process consisted of four steps.

First, bibliographic information was recorded, including the author's name, publication year, title, publication venue, and source type. This information supported source identification and reference management.

Second, the researcher examined the purpose, theoretical framework, research design, principal findings, and limitations of each relevant publication. For empirical studies, attention was given to the educational context, type of AI application, assessment method, and reported outcomes. For conceptual and policy publications, the analysis focused on their principal arguments, recommended practices, and implications for educational assessment.

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

Third, the extracted information was organized into a literature matrix. The matrix was designed to facilitate comparison across publications and to identify similarities, differences, recurring concerns, and unresolved questions.

Fourth, the information was reviewed in relation to the research objectives to determine how each source contributed to understanding the optimization of AI-based learning evaluation.

The literature matrix included the following analytical categories:

Analytical category	Information examined
AI assessment applications	Automated scoring, adaptive testing, learning analytics, and AI-generated feedback
Assessment accuracy	Reliability, validity, consistency, and alignment with learning objectives
Feedback and instructional support	Timeliness, specificity, personalization, and usefulness of feedback
Student learning outcomes	Knowledge acquisition, skill development, engagement, self-regulation, and academic performance
Ethical considerations	Algorithmic bias, privacy, transparency, accountability, and fairness
Implementation conditions	Teacher competence, institutional readiness, digital infrastructure, and human oversight
Research limitations	Methodological weaknesses, contextual limitations, and gaps in existing knowledge

This structured approach enabled the researcher to examine the literature systematically while retaining the interpretive flexibility required for qualitative synthesis.

Data Analysis Technique

The literature was analyzed using qualitative thematic analysis and integrative synthesis through coding, thematic categorization, and cross-source comparison. Deductive categories derived from the research objectives were combined with inductively identified patterns. The analysis focused on four interconnected themes: assessment accuracy and AI capabilities, AI-generated feedback and instructional adaptation, student learning outcomes, and ethical governance. Particular attention was given to distinguishing technological efficiency from educational effectiveness and examining the role of human oversight (Nicol, 2021). Convergent and divergent findings were compared across educational contexts, research designs, and assessment approaches to identify recurring patterns, limitations, and conceptual relationships. These findings informed the development of an integrated framework for AI-supported learning assessment.

Trustworthiness of the Research

Research trustworthiness was addressed through source credibility, analytical transparency, and critical comparison. Peer-reviewed publications and authoritative institutional documents were prioritized, while source selection and thematic interpretation followed explicitly defined criteria. Confirmability was supported by distinguishing literature-based evidence from conceptual interpretations (Pardo & Siemens, 2014). Differences in research contexts and contradictory findings were considered to strengthen analytical balance and avoid unsupported generalizations regarding AI effectiveness (Dawson et al., 2019).

Ethical Considerations

As a library-based study, this research involved no human participants or personal educational data. Ethical considerations focused on accurate citation, responsible interpretation, and transparent attribution of published sources. The synthesis also examined algorithmic bias, data privacy, fairness, transparency, and institutional accountability as essential conditions for responsible AI implementation in educational assessment (Crompton & Burke, 2023).

Methodological Limitations

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

This study is limited by its reliance on secondary literature, differences in research designs and educational contexts, and the rapid development of AI technologies. The qualitative synthesis also involves interpretive judgment and does not provide exhaustive systematic-review coverage or establish causal relationships between AI-based assessment and student learning outcomes. Consequently, the proposed conceptual framework requires further empirical validation across diverse educational settings.

RESULT & DISCUSSION

Synthesis of the Reviewed Literature

The reviewed literature identifies a conditional relationship between technical assessment capability and educational value. Three distinctions recur across the cited sources: scoring consistency is not the same as construct validity; feedback generation is not equivalent to feedback uptake; and learning analytics are not direct evidence of improved learning. The literature also differs in evidentiary status: systematic and critical reviews map reported applications and limitations, conceptual work explains feedback mechanisms, and institutional guidance establishes normative safeguards (Zawacki-Richter et al., 2019; Memarian & Doleck, 2024; Sembey et al., 2024; Miao & Holmes, 2023). Table 3 summarizes selected sources represented in the manuscript bibliography; it is an analytical comparison, not a claim that these are all sources screened.

Table 3. Comparative Synthesis of Selected Literature

Source	Scope / approach	AI or assessment focus	Synthesis contribution	Limitation for this study
Zawacki-Richter et al. (2019)	Systematic review; higher education	AI applications broadly	Maps AI research and educator involvement	Broad mapping; cannot demonstrate learning effects from assessment tools
Ifenthaler & Yau (2020)	Systematic review; higher education	Learning analytics and study success	Connects analytics to support for learners	Varied interventions; learning outcomes not uniform
Memarian & Doleck (2024)	Review of assessment for learning with AI	AI-supported formative assessment	Highlights pedagogical feedback mechanisms	Review evidence does not validate one universal model
Sembey et al. (2024)	Systematic literature review; higher education	Emerging assessment and feedback technologies	Shows diversity of practices and design contexts	Technological and contextual heterogeneity
Baker & Hawn (2022)	Scholarly analysis of algorithmic bias	Fairness in education algorithms	Identifies risks of uneven impacts	Not an effectiveness test of this framework
Miao & Holmes (2023)	UNESCO policy guidance	Generative AI governance	Human-centered policy and safeguards	Normative guidance, not empirical impact measurement

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

Assessment Accuracy: Consistency Versus Validity

Automated scoring offers a potential advantage for responses governed by explicit rules or clearly specified rubrics. Yet computational consistency does not demonstrate that a score represents the intended learning construct, particularly when tasks require argumentation, creativity, contextual interpretation, or disciplinary judgment. An assessment validity perspective therefore requires alignment among learning objectives, task specifications, scoring criteria, and review procedures. Technical auditing should examine disagreement with qualified human assessment, variation across learner groups, and the consequences of high-stakes decisions. The synthesis accordingly supports a bounded role for AI in scoring, with teacher moderation where interpretive ambiguity is consequential (Memarian & Doleck, 2024; Baker & Hawn, 2022).

Feedback Uptake and Instructional Adaptation

Assessment for Learning provides the mechanism linking generated evidence to possible educational value: information must be interpreted, acted upon, and reassessed. The literature indicates that rapid or personalized feedback is useful only insofar as learners can understand it, compare it with their work, and revise their learning strategies. Educators must also use patterns in student responses to adapt teaching, rather than treating dashboard indicators as sufficient educational outcomes. Nicol (2021) emphasizes learners' internal comparison and feedback processes, while Ifenthaler and Yau (2020) review learning analytics in relation to study success. Their different emphases suggest that data provision and pedagogical action are distinct steps in a feedback cycle.

Learning Outcomes and Limits of Inference

A recurring limitation is the substitution of proxy metrics—time saved, consistency across automatic scores, volume of comments, or prediction quality—for actual evidence of learning. Reported AI capabilities may support instructional responsiveness, but the present library study contains no pre-test/post-test results, classroom comparison, or pooled effect estimate. Learning improvement remains a plausible downstream outcome rather than an observed result of this research. Future evaluations should examine student revisions, transfer of knowledge, achievement against defined criteria, and differential impacts across learning contexts (Zawacki-Richter et al., 2019; Sembey et al., 2024).

Ethical Governance as a Boundary Condition

Fairness, transparency, privacy, inclusion, teacher preparedness, and institutional accountability condition the responsible use of AI-generated assessment evidence. Algorithmic bias can arise through unrepresentative data or uneven model performance; consequently, apparent efficiency must be evaluated alongside the distribution of errors and the right to contest consequential decisions (Baker & Hawn, 2022). UNESCO's human-centered guidance supports consent, oversight, and meaningful human agency rather than uncritical automation (Miao & Holmes, 2023). These governance conditions intersect with assessment validity: a technically consistent score that systematically disadvantages a subgroup remains educationally indefensible.

Proposed Integrated Conceptual Framework

Synthesizing the four themes yields a conditional sequence: assessment objectives and rubrics establish the intended constructs; AI generates provisional scores or feedback; quality assurance tests the alignment and fairness of outputs; teachers and learners interpret the feedback; instructional decisions and revision opportunities follow; and subsequent assessment checks learning against the original objectives. Ethical governance and human oversight operate across every stage. The sequence is a conceptual hypothesis, not an experimentally validated optimization algorithm. Table 4 identifies its components, theoretical rationale, and limiting conditions.

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

Table 4. Proposed Integrated Conceptual Framework for AI-Based Learning Assessment

Stage	Function and relationship	Evidence / conceptual anchor	Boundary / safeguard
1. Assessment design	Define constructs, rubrics, and acceptable interpretations	Validity-oriented assessment; Memarian & Doleck (2024)	Avoid construct underrepresentation
2. AI-supported evaluation	Generate preliminary scores, feedback, and analytics	Zawacki-Richter et al, (2019); Sembey et al. (2024)	Check data suitability and model errors
3. Quality assurance	Moderate scores, audit bias, review ambiguous answers	Baker & Hawn (2022)	Human oversight and contestability
4. Feedback uptake	Learners interpret, compare, and act on feedback	Assessment for Learning; Nicol (2021)	Comprehensibility and opportunity to revise
5. Instructional adaptation	Teachers target support using assessment evidence	Ifenthaler & Yau (2020)	Professional judgment and access equity
6. Follow-up evaluation	Reassess competencies to examine possible learning change	AfL feedback cycle	No assumption of demonstrated gain
Cross-cutting governance	Privacy, fairness, transparency, accountability	Miao & Holmes (2023)	Monitor throughout all stages

Discussion and Research Implications

The framework integrates Assessment for Learning with validity-oriented assessment and human-centered governance. Its contribution is to locate educational value in the relationship among evidence quality, feedback uptake, pedagogical action, and follow-up assessment, rather than in the technical sophistication of AI alone. The synthesis also clarifies why results from different publications cannot be treated as interchangeable: an application survey, a systematic review, an argument about feedback, and policy guidance support different kinds of claims. Comparative empirical studies are needed to test the proposed links across disciplines, grade levels, learner groups, and assessment stakes (Memarian & Doleck, 2024; Nicol, 2021; Miao & Holmes, 2023).

CONCLUSION

This study concludes that optimizing Artificial Intelligence (AI)-based learning evaluation requires an integrated approach that combines technological capabilities, assessment quality, pedagogical practices, and ethical governance. Based on the qualitative library research, AI has the potential to support educational assessment through automated scoring, adaptive assessment, learning analytics, and personalized feedback. These applications may improve scoring consistency, accelerate feedback delivery, identify learning difficulties, and support more responsive instructional decisions. However, technological efficiency does not automatically guarantee assessment validity or improved student learning outcomes.

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

The analysis indicates that assessment accuracy depends on reliability, validity, fairness, transparency, and alignment with intended learning objectives. Furthermore, the contribution of AI-based evaluation to learning outcomes depends on whether students understand and apply feedback and whether educators use assessment evidence to implement appropriate instructional interventions. Human oversight remains essential to address algorithmic bias, inaccurate feedback, privacy risks, and inappropriate automated decisions.

Accordingly, this study proposes an integrated optimization framework encompassing assessment design, AI-supported evaluation, quality assurance, feedback utilization, instructional adaptation, and continuous outcome evaluation. Educational institutions should implement AI responsibly, strengthen educators' AI literacy, protect student data, and evaluate both technical performance and educational effectiveness. Future empirical research is needed to validate the proposed framework and examine its effectiveness across educational contexts.

REFERENCES

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, *86*, 369–385. https://doi.org/10.1007/s10734-022-00937-2

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*, 22. https://doi.org/10.1186/s41239-023-00392-8

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 446–455. https://doi.org/10.1145/3303772.3303784

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. In *Center for Curriculum Redesign*.

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., & al., et. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*, 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. In *Pearson*.

Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, *2*, 100040. https://doi.org/10.1016/j.chbah.2023.100040

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. In *UNESCO*. https://doi.org/10.54675/EWZM9535

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*, 756–778. https://doi.org/10.1080/02602938.2020.1823314

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*, 438–450. https://doi.org/10.1111/bjet.12152

Rosyalita, D., Nabiilah, S. H., Prastiyanto, E., & Al Basri, B. A. (2026). Analyzing the Integration of Sharia Based Business Management Principles in Responding to Contemporary Global Market Challenges. *The Journal of Academic Science*, *3*(2), 60–69.

Sembey, R., Hoda, R., & Grundy, J. (2024). Emerging technologies in higher education assessment and feedback practices: A systematic literature review. *Journal of Systems and Software*, *211*, 111988. https://doi.org/10.1016/j.jss.2024.111988

Shermis, M. D., & Burstein, J. (eds. . (2013). Handbook of automated essay evaluation: Current applications and new

Sosioedukasi
JURNAL ILMIAH ILMU PENDIDIKAN DAN SOSIAL
ISSN: 2086-6135 E-ISSN: 2541-612x
Volume 15 No. 4 Desember 2026 https://ejournal.unibabwi.ac.id/index.php/sosioedukasi/index

directions. In *Routledge*.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*, 39. https://doi.org/10.1186/s41239-019-0171-0