Curricula: Journal of Curriculum Development *- e-ISSN 2830-7917 & p-ISSN 2964-7339*
Volume 4 No 2 (2025) 1833-1846

Curricula:
Journal of Curriculum Development
https://ejournal.upi.edu/index.php/CURRICULA/

Assessment and evaluation of an artificial intelligence-based deep learning curriculum

Aam Hamdani[1], Amay Suherman[2], Bambang Darmawan[3], Enda Permana[4]
[1,2,3,4]Universitas Pendidikan Indonesia, Kota Bandung, Indonesia
aam_hamdani@upi.edu[1], a_suherman@upi.edu[2], bambang_darmawan@upi.edu[3], enda_per@upi.edu[4]

ABSTRACT

The rapid development of Artificial Intelligence (AI) and the implementation of deep learning curricula in education have necessitated reforms to learning assessment and evaluation. This article examines the conceptual framework and assessment practices appropriate for AI-based deep learning curricula: how assessments can measure critical thinking, collaboration, knowledge transfer, and data literacy competencies; how AI-based platforms enable personalization, automated assessment, and learning analytics; and the ethical and validity challenges of using AI for evaluation. The method used is a systematic literature review and a comparative analysis of current practices in school and higher education contexts. The results indicate that AI-based assessments can improve learning effectiveness by providing real-time feedback and progress tracking; however, their success depends heavily on authentic task design, teacher readiness, infrastructure, and ethical/data security arrangements. Recommendations are provided for educational institutions to adopt a combination of authentic assessments (portfolios, projects, observations) and AI metrics, and to conduct regular algorithm audits to ensure the validity, reliability, and fairness of assessments.

ARTICLE INFO

Article History:
Received: 12 Aug 2025
Revised: 16 Nov 2025
Accepted: 23 Nov 2025
Publish online: 27 Dec 2025

Keywords:
artificial intelligence; authentic assessment; deep learning curriculum; learning evaluation; personalized learning

Open access
Curricula: Journal of Curriculum Development is a peer-reviewed open-access journal.

ABSTRAK

Perkembangan pesat di bidang Artificial Intelligence (AI) dan penerapan kurikulum pembelajaran mendalam (deep learning) dalam ranah pendidikan memunculkan kebutuhan untuk mereformasi asesmen dan evaluasi pembelajaran. Artikel ini mengkaji kerangka konseptual dan praktik asesmen yang sesuai untuk kurikulum deep-learning berbasis AI: bagaimana asesmen dapat mengukur kompetensi berpikir kritis, kolaborasi, transfer pengetahuan, dan literasi data; bagaimana platform berbasis AI memungkinkan personalisasi, penilaian otomatis, serta analitik pembelajaran; serta tantangan-etika dan validitas dalam penggunaan AI untuk evaluasi. Metode yang digunakan adalah studi literatur sistematis dan analisis komparatif terhadap praktik terkini dalam konteks sekolah dan perguruan tinggi. Hasil menunjukkan bahwa asesmen berbasis AI dapat meningkatkan efektivitas pembelajaran dengan menyediakan umpan balik real-time dan pelacakan kemajuan, namun keberhasilannya sangat bergantung pada desain tugas otentik, kesiapan guru, infrastruktur, dan pengaturan etika/keamanan data. Rekomendasi diberikan agar institusi pendidikan mengadopsi kombinasi penilaian otentik (portofolio, proyek, observasi) dengan metrik AI, serta melakukan audit algoritma secara berkala untuk memelihara validitas, reliabilitas, dan keadilan penilaian.

Kata Kunci: *asesmen otentik; evaluasi pembelajaran; kecerdasan buatan; kurikulum deep learning; personalisasi pembelajaran*

How to cite (APA 7)
Hamdani, A., Suherman, A., Darmawan, B., & Permana, E. (2025). Assessment and evaluation of an artificial intelligence-based deep learning curriculum. *Curricula: Journal of Curriculum Development, 4*(2), 1833-1846

Peer review
This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymised during review.

https://doi.org/10.17509/curricula.v4i2.95294

INTRODUCTION

In the 21st-century skills era, education emphasizes not only the mastery of facts and procedures but also students' ability to think critically, solve complex problems, work in teams, and apply knowledge to new situations. The term "deep learning" emerged as a concept that was subsequently implemented in the form of a curriculum and operationalized as a pedagogical approach. This study positions deep learning as a pedagogical approach aimed at fostering more meaningful and critical learning, as well as generating new solutions through creative insights into environmental problems. This concept focuses not only on mastery of subject matter but also on deep conceptual understanding, connections between ideas, application in various contexts, and self-reflection. This concept is becoming increasingly relevant for students to think more critically and deeply, collaborate actively, and solve problems in a meaningful way (Halim, 2025; Kasi et al., 2025; Nafi'ah & Faruq, 2025).

This challenge is becoming increasingly pronounced because the learning process in schools tends to be superficial (surface learning), written tests dominate assessment, and teachers do not yet have adequate evaluation tools to map higher-order competencies. This situation has a direct impact on the quality of student learning: critical thinking skills and problem-solving abilities are not developed to their full potential, the transfer of knowledge to new contexts is limited, motivation to learn declines, and graduates are less prepared to meet the demands of the workplace and industry. Therefore, ideal learning and assessment should be personalized, adaptive, and authentic to comprehensively capture higher-order thinking skills (Putri et al., 2025).

The integration of Artificial Intelligence (AI) technology into the learning process also presents significant opportunities to bridge this gap. AI technology can support personalized learning, dynamic content adaptation, and automated feedback for pupils and teachers (Yuliansyah & Saidah, 2025). For example, an AI-based learning management system can track pupils' learning progress, provide recommendations for further study materials, or identify areas where pupils are struggling. The integration of an AI-based management system into the implementation of a deep learning curriculum in primary schools resulted in an average increase in N-Gain scores of 58.13%, compared with 37.48% in the control group (Ali et al., 2025). The N-Gain score indicates the effectiveness of AI-based interventions in improving pupils' conceptual understanding, including their analytical and problem-solving skills.

However, implementing assessment within a curriculum framework that adopts AI-based deep learning principles presents significant challenges. Traditional written examinations are insufficient for measuring higher-order competencies such as knowledge transfer, collaboration, and creativity. However, the use of AI in assessment can also introduce additional layers of complexity, including issues of validity and reliability, algorithmic bias, fairness, and data privacy and ethics (Bulut et al., 2024). These issues have been highlighted in various studies, including the integration of assessment validity theory with the responsible AI framework, as well as reliability factors in data collection that influence the performance and fairness of AI systems (Burstein & LaFlair, 2024; Inel et al., 2023). It is therefore important to design a holistic assessment framework—one that combines authentic

https://doi.org/10.17509/curricula.v4i2.95294

tasks, AI-based learning analytics, and expert supervision—so that assessment is not merely a final measurement tool, but also an integral part of deep learning.

In light of these issues, this article offers a solution through a comprehensive analysis of the design and implementation of learning assessment and evaluation within an AI-based deep-learning curriculum. In the context of this literature review, the key variables discussed include assessment, learning evaluation, deep-learning curricula, AI, and authentic assessment and personalized learning. The literature review indicates that many current educational studies and practices still face challenges, such as assessments that are overly focused on written tests, the limited availability of instruments capable of measuring higher-order thinking skills, and the sub-optimal use of AI in supporting the learning evaluation process (Bulut et al., 2024; Chinta et al., 2025; Yan et al., 2023). In fact, ideally, a deep learning curriculum requires authentic, adaptive assessments that are integrated with AI-based learning analytics in order to provide a comprehensive picture of learning outcomes.

This study develops a conceptual framework that integrates the principles of deep learning, authentic assessment, and the use of AI in formative and summative assessment, whilst highlighting aspects of validity, reliability, fairness, and the role of teacher supervision in maintaining the quality of evaluation. Furthermore, this research examines the gap between deep learning curriculum theory and AI-based assessment practices in various previous studies. Specifically, this article will: 1) Identify the characteristics of assessment suitable for deep learning; 2) Explore the role of AI in supporting formative and summative assessment; 3) Discuss the challenges of validity, reliability, and ethics in AI-based assessment; and 4) Provide recommendations for education practitioners to design evaluation systems that are effective, ethical, and relevant to the learning needs of the 21st century.

LITERATURE REVIEW

Deep Learning Pedagogy

Deep learning pedagogy emphasizes conceptual understanding, higher-order thinking, and knowledge transfer through authentic tasks and metacognitive reflection, with its effectiveness depending on task design and the teacher's preparedness (Sadykova et al., 2024; Suherman et al., 2025). In this study, deep learning is understood as a pedagogical approach rooted in a concept and implemented through a curriculum that adopts the principles of deep learning, namely an emphasis on strong conceptual understanding, higher-order thinking skills, and the transfer of knowledge to various new contexts. Unlike surface learning, which focuses on rote memorization or the procedural completion of tasks, deep learning requires students to develop skills in analysis, synthesis, evaluation, and reflection on their learning process.

The international literature explains that deep learning involves integrating ideas, elaborating meaning, and establishing connections between concepts, thereby making knowledge more flexible and adaptable (Chen & Singh, 2024). The national study also confirms that deep learning is relevant to the demands of a modern curriculum that emphasizes competencies, reasoning, and problem-solving skills in a real-world context (Nafi'ah & Faruq, 2025).

https://doi.org/10.17509/curricula.v4i2.95294

From a pedagogical perspective, there are a number of key characteristics that define this approach, including the use of authentic tasks, context-based problem-solving, project-based or design-based learning, and the stimulation of metacognition to enhance learning awareness. The design-based learning approach can foster deep learning through exploration, collaboration, and reflection on the design process undertaken by students (Weng et al., 2023).

Assessment

Assessment as pedagogy positions authentic tasks, portfolios, and formative assessment as an integral part of teaching to facilitate deep learning, with a focus on continuous feedback and process-based assessment that improves student learning (Sappaile, 2025). Furthermore, assessment plays a central role as part of pedagogy, rather than merely serving as a tool for measuring final outcomes. The concept of 'assessment as pedagogy' emphasizes assessment that provides ongoing feedback, demands authentic performance, and allows pupils the opportunity to demonstrate their thought processes (Timperley & Schick, 2025). In line with this, the literature emphasizes that deep learning can be achieved only if assessments, learning activities, and learning objectives are designed coherently.

Assessment that supports deep learning emphasizes the authenticity of tasks, the evaluation of both process and product, and ongoing formative assessment. Authentic tasks (e.g., projects, case studies, portfolios, performances) that replicate real-world contexts compel students to integrate, elaborate on, and transfer concepts, thereby fostering higher-order thinking and metacognitive reflection. Analytical rubric design, iterative feedback, and instruments that measure the process (rather than just the final outcome) are key features, as they enable educators to assess higher-order cognitive aspects such as analysis, synthesis, and problem-solving (Baines et al. 2024; Banihashem et al., 2025; Cahyani et al., 2024). Furthermore, assessments that support deep learning require alignment among curriculum, instruction, and assessment, as well as educators' ability to design authentic tasks and to provide technological support for personalization and real-time feedback.

Artificial Intelligence (AI)

Artificial Intelligence (AI) in assessment (automated scoring, large language model feedback, learning analytics) offers efficiency and personalization, but raises issues regarding the validity of inferences, bias and fairness, and privacy, thereby necessitating human audits and responsible AI governance (Aloisi, 2023). One of the most significant potential uses of AI in assessment is to speed up the marking process and provide automated feedback at scale, particularly for open-ended tasks such as essays and programming questions. Automated scoring systems and AI-powered grading can reduce assessors' workload whilst producing consistent, rapid scores, thereby shortening the time required for pedagogical intervention. Recent research evidence indicates that AI-powered grading enhances the efficiency of marking without compromising consistency, particularly when combined with human audit mechanisms to ensure the validity of the assessment (Deepshikha, 2025; Tan et al., 2025).

https://doi.org/10.17509/curricula.v4i2.95294

Furthermore, AI opens up opportunities for real-time feedback, personalization, and big data processing (learning analytics) that support continuous formative assessment. Through learning analytics, AI-based platforms can provide immediate and actionable feedback, identify patterns of learner difficulty, and adapt learning pathways in an adaptive manner, thereby making interventions more targeted. Systematic studies and experiments show that a combination of adaptive testing, Intelligent Tutoring Systems (ITS), and LLM-based support can enhance personalization and the effectiveness of learning, provided there is robust data governance and oversight of algorithm quality. Local implementation examples also demonstrate increased learning gains following the adoption of AI systems that provide automatic feedback and content adaptation (Campos, 2025; Weidlich et al., 2025).

The application of AI in assessment raises serious challenges regarding the validity, reliability, and integrity of assessment inferences: AI models sometimes assess aspects that differ from the intended learning objectives, meaning that arguments regarding validity must be rebuilt for each system; furthermore, the results of automated scoring need to be tested for consistency and robustness against manipulation. This has been widely discussed in the literature, which links classical assessment validation principles with the principles of responsible AI—suggesting strict standards, audits, and documentation to ensure that AI scores are trustworthy (Burstein & LaFlair, 2024).

METHODS

His study employs a qualitative approach using a systematic literature review method to conduct an in-depth analysis of the concepts of assessment and learning evaluation within an artificial intelligence-based deep learning curriculum. The research subjects consist of various literature sources in the form of journal articles, whilst the research objects are the concepts, models, and practices of assessment, learning evaluation, AI in personalized learning, and the implementation of the deep learning curriculum discussed in these sources. The selection of literature was based on criteria of relevance, topicality, suitability to the educational context, and its contribution to the study of AI-based assessment and curricula, ensuring that the analyzed literature truly represents current theoretical and practical developments. The success of this research is determined by its ability to synthesize and construct a comprehensive understanding of various literature, thereby producing a clear picture of effective assessment and evaluation models within AI-based deep learning curricula. Meanwhile, the research tools consist of content analysis and systematic analysis, which are used to classify literature findings and identify conceptual patterns. The following are the detailed methodological steps.

Literature Selection

The literature search was conducted using open-access databases and journals in the fields of education and technology, focusing on publications from 2023 to 2025 to ensure the research remains up to date. The search terms included: “deep learning curriculum”, “artificial intelligence in educational assessment”, “AI in educational measurement”, “deep learning pedagogy assessment”, and “AI-based assessment tools”. From the initial results, several articles were reviewed, such as 1) Measurement and ethics in AI (Bulut et al., 2024);

https://doi.org/10.17509/curricula.v4i2.95294

2) The use of deep learning in primary education (Nafi'ah & Faruq, 2025); as well as 3) the effectiveness of AI-based management systems, and other topics as a basis for up-to-date reference (Ali et al., 2025).

Inclusion and Exclusion Criteria

The inclusion criteria are as follows: 1) Studies that explicitly address learning assessment or evaluation within the context of deep learning or AI curricula; 2) Empirical studies or systematic reviews; 3) Publications from 2023 to 2025; 4) Formal learning contexts (primary, secondary, or higher education). The exclusion criteria are as follows: 1) Research that focuses solely on technical AI algorithms without any connection to learning assessment; 2) Publications prior to 2023; 3) Non-formal contexts (e.g., corporate training) that are not directly relevant to school or university curricula.

Data Selection and Extraction Process

Following the search, the selection process involved screening titles and abstracts, followed by a review of the full text where relevant. Each article that passed was analysed for data extraction: type of assessment (formative, summative, portfolio, project, collaborative), the role of AI (automatic feedback, learning path adaptation, student analytics), learning outcomes (e.g. improved scores, student engagement, knowledge transfer), as well as issues of validity, reliability, ethics, teacher readiness and infrastructure. Data was extracted into a spreadsheet for comparative analysis.

Comparative Analysis and Synthesis

Following data extraction, the research involved analyzing patterns and findings across studies, specifically examining whether AI-based assessments were successful, what factors contributed to their success, the challenges that arose, how authentic task design relates to AI, and how validity, reliability, and fairness were addressed. The analysis results were synthesized into a conceptual framework for assessment within an AI-based deep learning curriculum, encompassing key elements such as authentic tasks, data analytics, personalization, and algorithm auditing.

Limitations of the Method

There are several limitations to this study, including: 1) It is literature-based; 2) No primary field data were collected; 3) Methodological details in the articles reviewed are sometimes limited; and 4) As AI technology and assessment methods continue to evolve, the findings presented should be viewed as a snapshot of the current situation and are subject to change. Nevertheless, this study can serve as a foundation for further empirical research.

https://doi.org/10.17509/curricula.v4i2.95294

RESULTS AND DISCUSSION

Assessment methods suitable for a deep learning curriculum have several key characteristics, namely: 1) Authentic tasks (real-world, project-based, collaborative); 2) Opportunities for knowledge transfer to new contexts; 3) Opportunities for student reflection; and 4) In-depth feedback. Authentic tasks are crucial because only through real-world situations or complex simulations can students internalize concepts and make connections between ideas. In the context of AI, tasks can be designed so that students use AI tools as part of the process (for example, data exploration using AI and peer collaboration via AI platforms). However, summative assessment still measures students' cognitive and reflective contributions. AI plays a key role, particularly in the formative stage, enabling real-time monitoring of student progress, providing immediate feedback, and recommending personalized learning pathways. However, for summative assessment, the use of AI still requires caution, as automated or analytical assessments must be validated against traditional assessment rubrics and supported by human assessment.

The implementation of AI-based assessments presents significant challenges in maintaining the validity and reliability of assessments. Assessment validity remains a key issue, particularly regarding the extent to which AI-based instruments can accurately measure the intended competencies, whilst reliability relates to the consistency of assessment results across students and over time. Furthermore, the use of automated assessment algorithms has the potential to introduce bias, particularly when the dataset is not representative or when the algorithm tends to prioritize certain types of answers. To minimize these risks, researchers have found that it is essential to design assessments that combine validated assessment rubrics, human assessor involvement, and transparent algorithms. Preliminary assessment trials are required to evaluate the system's reliability, whilst analyzing achievement gaps between student groups, such as those based on socio-economic background, is a crucial step in ensuring fairness in AI-based assessment.

The implementation of AI-based assessment is influenced by several challenges and supporting interrelated factors, as follows:

1. Teachers' readiness is a crucial factor, as they are required to possess data literacy, an understanding of AI analytics, and the ability to interpret the feedback generated. Various studies indicate that the lack of training and support for teachers remains a significant barrier to the optimal use of AI;
2. Technological infrastructure factors, particularly relating to the availability of devices, internet connectivity, and AI-based learning management systems;
3. Assignments should also be designed to allow AI to be used as a tool, rather than a substitute for students' own thinking, whilst assessment rubrics must be explicitly defined and aligned with the competencies to be assessed;
4. Ethical, privacy, and data security considerations must also be a top priority, as the use of AI involves the collection and analysis of student data; therefore, policies are needed regarding privacy protection, student/parent consent, and algorithm audits to ensure the responsible use of AI.

https://doi.org/10.17509/curricula.v4i2.95294

Discussion

Effective Assessment for the Deep Learning Curriculum

Effective assessment for a deep learning curriculum must place authentic tasks (real-world projects, case studies, simulations, and collaborative activities) at the heart of its design, as such tasks facilitate the internalization of concepts, the integration of concepts, and the transfer of knowledge to new contexts. Current literature indicates that task authenticity is characterized by contextual relevance, situational complexity, and the need to apply knowledge, enabling assessment that captures students' deep thinking processes rather than merely quantifiable product outputs. Furthermore, assessments that focus on both products and processes (e.g., portfolios, analytical rubrics, peer/collaborative assessment) provide opportunities for students to reflect on their cognitive strategies and demonstrate evidence of skill transfer to other situations, which lies at the heart of deep learning. Systematic reviews and recent empirical studies support a shift towards authentic assessment as a key strategy for achieving deep learning outcomes (Baines et al., 2025; Vlachopoulos & Makri, 2024).

In the context of AI integration, task design should allow for the use of AI tools as part of the learning process, for example, data exploration using AI tools, ITS support for scaffolding, or collaboration facilitated by AI platforms. However, final assessments must be designed to explicitly evaluate students' cognitive and reflective contributions. The literature from 2023–2025 emphasizes models in which AI acts as a scaffolding or augmenting tool, rather than a substitute for students' thinking. Within this framework, assessment instruments need to incorporate indicators that evaluate students' reasoning, decision-making justifications, design processes, and metacognitive reflection. Furthermore, case studies on the use of AI in Problem-Based Learning (PBL) demonstrate that establishing rules for AI usage, rubrics that distinguish between human contributions and tool outputs, and verification mechanisms (e.g., process artifacts, oral interviews, or direct observation) are crucial for maintaining the validity and integrity of assessment in the GenAI era. Research on AI-based assessment for learning also emphasizes the need for audit mechanisms, algorithmic transparency, and ethical policies to ensure that AI feedback supports rather than undermines the objectives of deep learning (Gonsalves, 2025; Memarian & Doleck, 2024).

The Role of AI in Supporting Formative and Summative Assessment

AI plays the most transformative role at the formative stage, when progress monitoring, interventions, and regular feedback are most needed. Through learning analytics and intelligent tutoring systems, AI can monitor students' learning trajectories in real time, provide immediate diagnostic feedback, and recommend personalized learning pathways based on individual patterns of difficulty; these features are considered to strengthen the processes of assessment for learning and self-regulated learning. Empirical studies have found that analytics-driven feedback enhances reflective engagement and helps teachers target formative interventions. Furthermore, research on GenAI also highlights the potential

https://doi.org/10.17509/curricula.v4i2.95294

of real-time, dialogic feedback to support continuous learning (Arslan et al., 2024; Banihashem et al., 2025; Manganello & Fante, 2025).

However, the use of AI for summative assessment requires caution and rigorous validation. Automated or analytical assessment must be aligned with traditional assessment rubrics and tested against validity standards, as without evidence that the algorithm's output reflects the intended constructs, score inferences may be inaccurate. Recent literature emphasizes 'human-in-the-loop' practices: verification/review by educators, algorithmic audits, and transparent documentation are necessary to ensure reliability, fairness, and accountability in the use of AI for summative assessment. Implementation studies also indicate that whilst teachers value the speed of AI feedback in a formative context, they tend to have less confidence in automated scores for summative purposes without human supervision and clear validation mechanisms (Berg & Papadopoulos, 2025; Kaldaras et al., 2024).

Challenges of Validity, Reliability, and Ethics in AI-Based Assessment

One of the main challenges in implementing AI-based assessment is ensuring validity, that is, whether the scores or indicators produced truly reflect the intended competency constructs and reliability, namely, the consistency of results across time, across items, and across groups of respondents. Recent literature confirms that automated models often capture surface-level linguistic patterns or technical features that correlate with scores, but do not always measure the deep thinking processes targeted by deep learning curricula; therefore, evidence of validity must be re-collected for each AI system, including content evidence, response evidence, the relationship between scores and external indicators, and the consequences of using the scores. New studies and guidelines recommend a combination of empirical evidence (e.g., correlations with analytical rubrics and expert assessments), model development documentation, and human-in-the-loop mechanisms to address issues of score interpretation (Kaldaras et al., 2024; Kinnear et al., 2024).

The risks of bias and unfairness are also a major concern: unrepresentative training data, linguistic differences, and sampling phenomena can lead to algorithms treating diverse groups unequally (for example, second-language learners or certain socio-economic groups). Therefore, the design of AI-based assessments must incorporate mitigation measures to test the reliability and equity of scores, conduct item-by-item analysis across sub-groups, utilize fairness metrics (e.g. disparate impact, ABROCA), carry out periodic algorithm audits, and incorporate validated rubrics and expert assessment as a final verification. International studies also emphasise the importance of model transparency, reporting of bias metrics, and the involvement of human evaluators in the decision-making loop to ensure that assessment inferences remain valid and fair; local studies highlight the additional need for improved AI literacy among educators and the adaptation of local data contexts prior to large-scale implementation (Ali et al., 2025; Andersen et al., 2025; Chinta et al., 2025).

The results of the literature review reveal several practical implications that can be applied in this context, namely: 1) Using clear analytical rubrics as construct standards; 2) Conducting reliability tests and analysing inter-group differences prior to summative assessment; 3) Implementing a human-in-the-loop (educator verification) process for high-stakes cases; and 4) Conducting regular algorithmic audits and transparent reporting of

https://doi.org/10.17509/curricula.v4i2.95294

validity and fairness metrics. This integrated approach, combining empirical evidence, human verification, and technical governance, makes it more likely that AI-based assessment will produce valid, reliable, and fair inferences within the context of a deep learning curriculum (Johnson & McCaffrey, 2023; Kinnear et al., 2024; Zheng et al., 2024).

Challenges and Enabling Factors in AI-Based Assessment

Teachers' readiness heavily influences the implementation of AI-based assessment. Data literacy, an understanding of AI analytics, and the ability to interpret analytical feedback are prerequisites for AI to be used pedagogically rather than merely technically. Many studies highlight that a lack of professional training and capacity-building programs is a major obstacle, as if teachers are not properly trained, AI outputs can easily be misinterpreted or even undermine the reflective assessment practices that should be in place. Therefore, training programs for AI literacy and Professional Development (PD) that focus on data interpretation, task design, and human-in-the-loop practices are essential (Tan et al., 2025).

Technology infrastructure is a critical enabling factor. The availability of adequate devices, stable internet connectivity, and the integration of a learning management system (LMS) with AI modules are operational prerequisites. Without reliable infrastructure, analytics systems cannot process real-time data or provide adaptive feedback, meaning that the potential of AI for assessment-for-learning is not fully realized. Implementation research and literature reviews also indicate that infrastructure gaps (including reliance on commercial LMS vendors) frequently lead to issues with interoperability, costs, and the scale of adoption in schools with limited resources (Pardosi et al., 2024; Vergara et al., 2024).

When designing tasks and rubrics, the challenge lies in devising tasks that position AI as a scaffold or aid to the thinking process, rather than a substitute for students' own thinking, whilst also developing explicit rubrics that distinguish and assess human cognitive contributions from AI-generated outputs. Recent literature emphasizes two key practices: 1) validated analytical rubrics to capture process-related aspects (argumentation, justification, reflection); and 2) verification mechanisms (process artifacts, oral interviews, or portfolios) to ensure the validity of assessment inferences. Several experimental studies involving LLMs and chatbots also recommend procedures for adapting rubrics when AI is integrated, to ensure that construct reliability and validity are maintained (Fernandez-Sanchez et al., 2024; Hmoud et al., 2024).

Ethical, privacy, and data security issues are crucial non-technical challenges. The use of AI requires the collection and processing of sensitive student data; therefore, data protection policies, parental/student consent, encryption, and algorithmic audits must be implemented. Ethical reviews also emphasize the need for model transparency, bias documentation, and accountability mechanisms when decisions have high-stakes implications (such as determining graduation or rankings). In practice, institutions need to implement responsible AI policies, including regular audits and reporting on fairness metrics, and build trust through open communication with stakeholders (Akgun & Greenhow, 2022; Lim et al., 2025).

The literature review identified several relevant practical implications to be considered in the implementation of AI-based assessment, namely: 1) Delivering practical, case-based professional development programmes for teachers' AI literacy; 2) Prioritising investment in

https://doi.org/10.17509/curricula.v4i2.95294

the interoperability and connectivity of LMS infrastructure; 3) Designing analytical rubrics and process verification procedures when using AI; 4) Implementing privacy policies, explicit consent, and periodic algorithm audits. An integrated approach combining human capabilities, technological readiness, well-designed assessments, and ethical governance forms the foundation for AI-based assessment to function effectively and fairly within the deep learning curriculum (Hmoud et al., 2024; Lim et al., 2025; Vergara et al., 2024).

CONCLUSION

Based on the literature review conducted, this study proposes the conceptual hypothesis that assessment within an AI-based deep learning curriculum will be more effective if designed as an integrated system combining authentic assessment, AI-assisted learning analytics, and expert evaluation as central components. This integration is expected to enhance the accuracy of high-level competency assessment whilst supporting personalized learning.

The research objectives have been achieved through mapping assessment characteristics aligned with deep learning, explaining the role of AI in supporting both formative and summative assessment, identifying validity challenges, and formulating assessment design recommendations consistent with 21st-century learning principles. This study demonstrates that whilst AI holds great potential to assist the evaluation process, the quality of assessment remains heavily influenced by educators' active involvement, the readiness of the technological ecosystem, and the alignment of assessment design with curriculum objectives.

As a follow-up, sustained efforts are required to strengthen educational capacity in AI literacy and assessment design, build a secure and transparent data infrastructure, and develop algorithmic audit mechanisms to ensure the fairness and accountability of the system. Further research is also recommended to test this integrated assessment model in real-world implementation contexts, thereby providing empirical evidence regarding its effectiveness, practical challenges, and the most appropriate adaptation strategies for educational institutions.

AUTHOR'S NOTE

The author declares that there are no conflicts of interest regarding the publication of this article. The author confirms that the data and content of the article are free from plagiarism. We would like to thank the Directorate of Research and Community Service at the Indonesia University of Education for the 2025 financial year.

REFERENCES

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics, 2,* 431-440.

Ali, E. Y., Nugraha, R. G., & Bagja, N. (2025). The effectiveness of AI-Based education management systems in implementing deep learning curriculum in elementary schools. *Journal of Integrated Elementary Education*, *5*(2), 339-352.

https://doi.org/10.17509/curricula.v4i2.95294

Aloisi, C. (2023). The future of standardised assessment: Validity and trust in algorithms for assessment and scoring. *European Journal of Education, 58*(1), 98-110.

Andersen, N., Mang, J., Goldhammer, F., & Zehner, F. (2025). Algorithmic fairness in automatic short answer scoring. *International Journal of Artificial Intelligence in Education, 38,* 1-38.

Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Frontiers in Artificial Intelligence*, *7*(1), 1-8.

Baines, S., Otermans, P., Tree, D., & Worsfold, N. (2025). Measuring and mapping authentic assessment with a novel quantitative typology. *Teacher in Higher Education, 30*(3), 663-682.

Banihashem, S. K., Gasevic, D., Noroozi, O., Jarodzka, H., Brinke, D. J., & Draschler, H. (2025). Optimizing formative assessment with learning analytics. *Review of Educational Research, 1*(1), 1-49.

Berg, S. V. D., & Papadopoulos, P. M. (2025). Summative assessment with artificial intelligence: Qualitative analysis and comparison of technology acceptance in student and teacher populations. *Innovations in Education and Teaching International, 62*(5), 1529-1544.

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., & Walsh, C. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges. *Chinese/English Journal of Educational Measurment and Evaluation, 5*(3), 1-34.

Burstein, J., & LaFlair, G. T. (2024). Where assessment validation and responsible AI meet. *Europe Knowledge Development Institute, 1*(1), 1-32.

Cahyani, V. P., Ahmad, F., Soekarman, & Nawi, M. Z. M. (2024). Authentic assessment as a solution to enhance deep learning and maintain academic integrity in higher education. *Journal of Educational Analytics (JEDA), 3*(3), 531-544.

Campos, C. M. (2025). The impact of artificial intelligence on personalized learning in higher education: A systematic review. *Trends in Higher Education, 4*(2), 1-15.

Chen, J., & Singh, C. K. S. (2024). A systematic review on deep learning in education: Concepts, factors, models and measurements. *Journal of Education and Educational Research*, *7*(1), 125-129.

Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Quy, T. L., & Zhang, W. (2024). FairAIED: Navigating fairness, bias, and ethics in educational AI applications. CORR, 1*(1)*, 1-59.

Deepshikha, D. (2025). A comprehensive review of AI-Powered grading and tailored feedback in universities. *Discover Artificial Intelligence, 5*(1)*,* 1-18.

Fernandez-Sanchez, A., Lorenzo-Castineiras, J., & Sanchez-Bello, A. (2024). Navigating the future of pedagogy: The integration of AI tools in developing educational assessment rubrics. *European Journal of Education, 60*(1), 1-13.

https://doi.org/10.17509/curricula.v4i2.95294

Gonsalves, C. (2025). Contextual assessment design in the age of generative AI. *Journal of Learning Development in Higher Education, 34*(1), 1-14.

Halim, M. F. (2025). Transforming educational leadership: The role of school principals in implementing deep learning-based curricula. *Journal of Educational Management Research, 4*(1), 308-321.

Hmoud, M., Swaity, H., Anjass, E., & Aguaded-Ramirez, E. (2024). Rubric development and validation for assessing tasks' solving via AI chatbots. *The Electronic Journal of E-Learning, 22*(6), 1-17.

Inel, O., Draws, T., & Aroyo, L. (2023). Collect, measure, repeat: Reliability factors for responsible AI data collection. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *11*(1), 51-64.

Johnson, M. S., & McCaffrey, D. F. (2023). Evaluating fairness of automated scoring in educational measurement. *Advancing Natural Language Processing in Educational Assessment*, *142*(1), 1-10.

Kaldaras, L., Akaeze, H. O., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Sec. Assessment, Testing and Applied Measurement, 9*(1)*,* 1-10.

Kasi, Y. F., Bai, D. V., Novia, N., Deporos, S. R. C., & Mabbaya, A. D. (2025). Implementation of deep learning in school curriculum: Perspective of teachers in Nagekeo Regency. *Jurnal Penelitian Pendidikan Paedagogia, 28*(2), 320-328.

Kinnear, B., St-Onge, C., Schumacher, D. J., Marceau, M., & Naidu, T. (2024). Validity in the next era of assessment: Consequences, social impact, and equity. *Perspectives on Medical Education, 13*(1), 452-459.

Lim, T., Gottipati, S., & Cheong, M. (2025). What students really think: Unpacking AI ethics in educational assessments through a triadic framework. *International Journal of Educational Technology in Higher Education, 22*(56), 1-32.

Manganello, F., & Fante, C. (2025). Scoping review on the role of learning analytics in assessing and fostering creativity in educational contexts. *Journal of Learning Analytics, 12*(2), 5-18.

Memarian, B., & Doleck, T. (2024). A Review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans, 2*(1)*,* 1-11.

Nafi'ah, J., & Faruq, D. J. (2025). Conceptualizing deep learning approach in primary education: Integrating mindful, meaningful, and joyful. *Journal of Educational Research and Practice, 6*(2), 78-89.

Pardosi, V. B. A., Xu, S., Umurohmi, U. Nurdiana, & Sabur, F. (2024). Implementation of an artificial intelligence based learning management system for adaptive learning. *Al-Fikrah: Jurnal Manajemen Pendidikan, 12*(1), 149-161.

Putri, D. N., Mayasril, N., & Gusman, R. (2025). Asesmen abad 21: Menakar kompetensi, bukan sekedar nilai. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, *11*(2), 221-228.

https://doi.org/10.17509/curricula.v4i2.95294

Sadykova, A., Iskakova, M., Ismailova, G., Ishmukhametova, A., Sovetova, A., & Mukasheva, K. (2024). The impact of a metacognition-based course on school students' metacognitive skills and biology comprehension. *Frontiers in Education, 9*(1), 1-9.

Sappaile, B. I. (2025). A comparative analysis of the merdeka belajar framework as a means to achieve deep learning in elementary school students: A literature review. *Cosmos: Jurnal Ilmu Pendidikan, Ekonomi dan Teknologi*, *2*(4), 944-963.

Suherman, A., Supriyadi, T., Safari, I., Saptani, E., Fauzi, R. A., Sudirjo, E., Komiljon O'g'li, T. S., & Abdisamiyevich. (2025). Bridging teacher readiness and deep learning-based teaching practice: Assessing the effectiveness of the active model for enhancing teacher pedagogy. *International Research Journal of Multidisciplinary Scope, 6*(4), 284-298.

Tan, L. Y., Hu, S., Yeo, D. J., & Cheong, K. H. (2025). A comprehensive review on automated grading systems in STEM using AI techniques. *Mathematics, 13*(1)*,* 1-10.

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence, 8*(1)*,* 1-19.

Timperley, C., & Schick K. (2024). Assessment as pedagogy: Inviting authenticity through relationality, vulnerability and wonder. *Teaching in Higher Education, 30*(3), 592-607.

Vergara, D., Lampropoulos, G., Antón-Sancho, Á., & Fernández-Arias, P. (2024). Impact of artificial intelligence on learning management systems: A bibliometric review. *Multimodal Technologies and Interaction*, *8*(9), 1-19.

Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation, 83*(1)*,* 1-13.

Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., Giorgashvili, T., Yau, J., & Drachsler, H. (2025). Highly informative feedback using learning analytics: How feedback literacy moderates students perceptions of feedback. *International Journal of Education Technology in Higher Eduucation, 22,* 43(1), 1-25*.*

Weng, C., Chen, C., & Ai, X. (2022). A pedagogical study on promoting students' deep learning through design-based learning. *International Journal of Technology and Design Education, 33*(1)*,* 1653-1647.

Yan, L., Sha, L., Zhao, L., Li, Y., Maldano, R. M., Chen, G., Li, X., Jin, Y., & Gasevic, D. (2023). Pracatical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90-112.

Yuliansyah, H., & Saidah, N. (2025). Transformasi kurikulum berbasis kecerdasan buatan sebagai strategi pendidikan abad 21. *Jurnal Pendidikan Islam Educan, 9*(1), 53-66.

Zheng, E. L., Jin, W., Hamarneh, G., & Lee, S. S. J. (2024). From human-in-the-loop to human-in-power. *The American Journal of Bioethics*, *24*(9), 84-86.

https://doi.org/10.17509/curricula.v4i2.95294