Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence Driven Grading and Personalised Feedback in Higher Education Assessment

Mziwendoda Cyprian Madwe

madwem@unizulu.ac.za

University of Zululand

Systematic Review

Keywords: Artificial intelligence, Automated grading, Personalised feedback, Higher education institutions

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10810076/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Traditional assessment practices in higher education are constrained by time-intensive grading. Inconsistencies in evaluation, limited scalability and delays in providing meaningful feedback. Artificial intelligence has emerged as a promising solution capable of transforming assessment through automated grading and personalized feedback. This systematic literature review synthesizes empirical evidence from studies published between 2020 and 2026 to examine the applications, benefits, challenges and future directions of AI-powered grading and tailored feedback in higher education institutions. The review reveals that AI technologies, particularly machine learning, natural processing and computer vision, significantly enhance grading efficiency, consistency and scalability while enabling timely, personalized feedback that supports student learning and engagement. These technologies also reduce academic workload, allowing educators to devote greater attention to higher-order teaching activities. Despite these advantages, the review identified persistent challenges, including algorithmic bias, data privacy and security concerns, limited transparency AI decision-making and the need to preserve academic integrity through meaningful human oversight. The findings suggest that AI should augment rather than replace academic judgement, requiring robust governance frameworks, ethical safeguards and continuous validation to ensure fairness, accountability and educational effectiveness. This review contributes to the growing scholarship on AI-enabled assessment by providing a comprehensive synthesis of current evidence and identifying key priorities for researchers, institutional leaders and policymakers seeking to implement responsible, equitable and learner-centred AI assessment systems in higher education.

1. Introduction

Assessment is a fundamental component in teaching and learning in higher education, strengthening the assessment of student learning and providing feedback that supports academic development (Levy-Feldman, 2025). However, traditional assessment practices remain heavily dependent on manual grading and feedback processes that are labour-intensive, time-consuming, difficult to scale, and prone to human biased (Saez, Garcia, Mochon & Isasi, 2026). As university class sizes increase and learning environments become more diverse, it becomes more difficult for educators to offer consistent and timely grading, as well as detailed and personalised feedback (Swiecki et al, 2022). Generative artificial intelligence (Gen AI) has emerged as potential powerful technology to mitigate this problem through automated grading and timely individualised feedback. Gen AI can support student engagement with learning tasks and enhance the efficiency and scalability of assessments in higher education (Kizilcec et al., 2024; Deepshikha, 2025; Zhan & Yan, 2025).

Deepshikha (2025) stresses the potential of Gen AI-powered tools into education to transform current practices through assessment and feedback. Gen AI-driven automated grading systems (AGS) have emerged a solution for reducing the grading time and offer timely feedback to a larger university class (Tan et al., 2025). The AGS employs large language models (LLMs) to assess and score student work, mitigating human bias inherent in conventional grading systems (Ansor et al., 2023). Gen AI-enabled

AGS can provide personalised and immediate student feedback, fostering more responsive learning environment (Alghamdi, Alanezi & Khan, 2020). Gen AI-powered AGS may reduce the teachers' marking/grading burdens and free up valuable time that can be redirected to support students' learning (Papageorgiou and Pierrakeas, 2026).

Gen AI has emerged powerful tool to provide a personalised student feedback that identify strengths and weaknesses, supporting student to understand their learning achievement and guide them towards areas requiring more attention and development (Khosravi et al., 2022; Cooper, Ilich & Sturrock, 2026). Gen AI-powered feedback provides students with real-time, personalised insights into their performance, enhancing the feeling of competence (Gao, 2025; Zhan et al., 2025). This individualised student feedback facilitates a sense of mastery over students' studies, improving academic self-efficacy in learning tasks (Xu, Liu & Li, 2025).

Although GenAI provides transformative potential in educational assessment, it also presents significant challenges and ethical concerns (Madwe, Chonco & Zungu, 2025). These challenges and concerns include algorithmic bias, data privacy and security, academic integrity, potential risk of over-reliance on AI-driven judgements and reduced human oversight in assessment (Singh& Hiran, 2022; Zhao, Chapman & Sabet, 2024; Dai & Chan, 2026). AI-driven assessment systems often required detailed and structured student data to generate assessment, grades/scores and personalised feedback (Chiu et al., 2023). This raises concerns about the privacy, security, ownership and responsible use of student information (Chiu et al., 2023). Furthermore, some studies reported some teachers may have insufficient knowledge and competence in understanding and application of AI technologies (Ilma & Rohmah, 2025; Ren & Wu, 2025). This limited AI knowledge may constrain teachers' ability to critically assess AI-generated grades and feedback, identify bias or errors, and exercise appropriate human oversight over AI-generated outputs.

These challenges indicate that effective AI-driven scoring/grading and feedback depends on not merely technical abilities of the AI systems and its capability to support fair, trustworthy and pedagogically sound assessment practices. Therefore, AI-driven assessment should be considered as an instrument that complements rather than replaces academic judgement, with teachers retaining an important role in assessing, interpreting and evaluating AI-generated assessment grades and feedback output. Prior research indicates that human oversight is important because over-reliance on AI may lead to the uncritical acceptance of GenAI-generated outcomes, weakening students' intellectual development and learning growth (Singh & Hiran, 2022; Dai & Chan, 2026).

The emergence of GenAI technologies, especially LLMs such as ChatGPT and similar systems, has increasingly expanded the use of AI in higher education assessment (Wang et al., 2024). Compared with earlier rule-based or machine learning approaches, GenAI technologies provide exceptional capabilities for generating feedback, evaluating and grading complex written responses, and supporting adaptive learning (Nongkhai et al., 2026).

Despite increasing studies on AI-driven assessment in higher education (Myint, Lo & Zhang, 2024; Deepshikha, 2025; Yldirim-Erbasli et al., 2026), the empirical evidence remains fragmented across different AI systems, assessment activities, educational context and outcomes. Previous studies have often investigated automated grading and AI-driven feedback separately, making it challenging to establish how these collectively influence assessment processes in higher education. Furthermore, limited studies focus on accuracy, personalisation, efficiency, fairness, data privacy and human oversight. Therefore, a study is needed to review and synthesise how different AI systems are employed to grade students' assessments and to provide feedback. This study should identify the benefits and challenges reported across higher education contexts, examining the condition required for responsible implementation. Such study is especially relevant as HEIs seeks evidence-based approaches to integrating AI into assessment while maintaining educational quality, academic integrity, fairness and meaningful human involvement.

Against this background, this SLR examine studies published between 2015 and 2026 on AI-powered grading and personalised feedback in higher education. The study is guided by three objectives:

1. To examine the application of artificial intelligence technologies in higher education assessment, with particular emphasis on automated grading and personalised feedback systems.
2. To synthesise the reported benefits and challenges of AI-driven grading and personalised feedback in enhancing assessment quality in higher education institutions.
3. To identify emerging trends and propose research priorities for the responsible implementation of AI-enabled assessment, with particular attention to fairness, transparency, academic integrity, and effective human oversight in higher education.

By addressing these research objectives, the present study combines AI tools employed for grading/marking and feedback, their reported benefits and challenges, priorities for future research. The study's findings may provide insights for educators, researchers, institutional leaders and policymakers seeking to develop responsible, equitable and learner-centred approaches to AI-supported assessment in higher education.

2. Research methodology

In consistent with Hamilton et al. (2026), this study adopted a systematic literature review (SLR) following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines. The PRISMA 2020 framework provides a transparent and rigorous process for identifying, screening, assessing, and selecting relevant studies, thereby improving the reliability and reproducibility of systematic reviews (Kolaski, Logan & Ioannidis, 2023).

Braun and Clarke (2026) thematic analysis model was employed to synthesis the findings of studies included in final review. The researcher selected this model due to its ability to identify, organise, analyse, and interpret recurring patterns across qualitative and quantitative studies (Ahmed et al.,2025). This

framework has enabled the research to develop themes that explain how artificial intelligence is transforming assessment practices in higher education.

2.1 Search Strategy

The research conducted a literature search on 07 August 2026 using three major academic databases, namely Web of Science (WoS), Scopus, and ScienceDirect. These databases are widely used because they offer extensive coverage of peer-reviews studied across education, technology, AI, assessment and interdisciplinary research (Dai, Lui & Zhang, 2026; Madwe, 2026; Mbandlwa, Ndlovu & Mpofu, 2026). In consistent with previous studies, the researcher adopted Population, Concept and Context (PCC) framework as the search strategy. This framework was considered appropriate for the current SLR (Yousefiazar et al., 2025; Joshi et al., 2026). The population and context in this SLR consist of students, teachers/educators and HEIs, whereas the central concept focused on the use/application of AI in educational assessment, especially GenAI-driven grading, automated assessment and tailored personalised feedback.

The present study operationalized the PCC model using Boolean operators, synonyms and related search terms (Adugbire et al., 2025). The OR operator was used to combine terms within each concept, while AND operator was adopted to combine the main concepts. The search terms included: “artificial intelligence,” “generative intelligence,” “generative AI,” “GenAI,” “large language mode,” “machine learning,” “automated grading,” “automated assessment,” “automated scoring, “AI grading,” “AI-powered assessment,” “AI-driven assessment,” “personalised feedback,” “personalized feedback,” “AI-generated feedback,” “AI-supported feedback,” “automated feedback,” “higher education,” “university”, “tertiary education,” * and “university student.” The researcher used database-specific search syntax and truncation where applicable.

This search strategy was used to get literature exploring AI-driven grading and AI-powered or personalised feedback, specifically AI application in higher education assessment. The present SLR limited to articles published between 2015 and 2026, written in English, and peer-reviewed studies. Open-access publications were used as a database filter where applicable. Table 1 presents the database-specific filtering and screening process used in this SLR.

2.2 Study Identification and Screening

The identification and screening of studies were undertaken using the PRISMA 2020 framework. The search was performed in WoS, Scopus and ScienceDirect using the inclusion criteria and predefined search strategy. The researcher applied database-specific filters to remove studies that did not meet the inclusion criteria. These filters included publication year, document type, subject area (where applicable), open-access availability, and publication language. Table 1 presents the database search process and the number of articles retained after each screening stage. Following PRISMA 2020 guideline, the initial search returned 98 articles from WoS, 6 from Scopus and 298 from ScienceDirect. After filtering the search by the publication period (2015–2026), document type, subject area, open-access and language,

42 articles from WoS, 2 from Scopus and 42 from ScienceDirect were retrieved. Total of 86 articles were retained for subsequent screening. The titles of these retained articles (86) were screened against the predefined inclusion criteria. The title screening focuses on articles on the application of GenAI or AI in higher education assessment, AGS, AI-driven assessment or personalized feedback, education measurement and related AI-powered assessment practices. Based on the title screening, 34 papers from WoS and 32 from ScienceDirect were excluded because titles did not focus on research objectives. Therefore, 10 papers from ScienceDirect, 8 from WoS and 2 from, yielding a total of 20 papers were retained for abstract and full-text screening.

Table 1
PRISMA 2020 Identification of Studies

Screening Stage	**Web of Science**	**Scopus**	**ScienceDirect**
Initial search results	98	6	298
Publication period (2015–2026)	93	6	296
Document type (Articles/Review Articles)	63	3	197
Subject area (Social Sciences)	—	—	54
Open-access articles	42	2	42
English language	42	2	42
Studies retained for screening	42	2	42
Articles excluded based on title screening	(34)	(0)	(32)
Articles retained for abstract and full-text screening	8	2	10
Articles excluded after abstract and full-text screening	(4)	(0)	(4)
Total articles included in final review	4	2	6
Total articles included for final review			**12**
Source: Author’s compilation			

2.3 Data Extraction

Relevant information was extracted from each eligible study using a structured data extraction form. The extracted information included author(s), publication year, country, research objectives, study design, educational context, employed AI technology, assessment application, principal findings, reported benefits, implementation challenges, and recommendations for future research. The structured extraction ensured consistency during data synthesis and facilitated comparison across studies.

2.4 Data Analysis

The selected studies were analysed using thematic analysis. Following Braun and Clarke's (2006) six-phase framework, the analysis involved: (1) familiarisation with the selected studies, (2) generating initial codes, (3) searching for recurring themes, (4) reviewing and refining themes, (5) defining and naming themes, and (6) producing the final narrative synthesis.

Coding focused on recurring concepts relating to the use of artificial intelligence in higher education assessment. Similar codes were grouped into broader themes that represented the collective evidence reported across the literature. This approach enabled the review to move beyond describing individual studies towards identifying broader patterns, similarities, differences, and emerging trends regarding AI-powered grading and personalised feedback.

The thematic synthesis was guided by the review objectives and emphasised both the opportunities and challenges associated with AI-enabled assessment. The findings were subsequently organised into major themes that collectively explain the current state of knowledge, implementation considerations, and future research priorities.

Following the coding and theme development process, the findings were organised into five (5) overarching themes that address the objectives of this SLR. These themes were developed through an iterative analysis of the selected studies and represent the dominant patterns identified across the literature on artificial intelligence in higher education assessment.

To analyse and synthesise the results of the 12 studies included in the final review, the evidence was arranged into five themes that align with research objectives. The thematic analysis was performed and focused on recurring patterns about the AI technologies and their application in higher education assessment, automated grading, personalised feedback, institutional and educational benefits and the ethical, pedagogical and governance challenges associated with AI-driven assessment. Table 2 presents thematic framework that provide a structured analysis and synthesising the findings across the included studies.

Table 2
Thematic Analysis Framework

Theme	Description	Key dimension explored	Supporting studies
Theme 1: Artificial Intelligence Technologies Used in Higher Education Assessment	Examine the range of AI technologies and their use in higher education assessment.	Large language models (LLMs), natural language processing (NLP), machine learning, automated assessment systems, AI-assisted grading, AI-supported feedback and AI-generated quizzes.	Maier, Seibold and Klotz (2026); Myint et al. (2024); Ramnarain-Seetohul, Rosunally and Bassoo (2025); Schneider, Richner and Riser (2023); Pozdniakov et al. (2026); Xiong and Li (2026); Deepshikha (2025); Durgungoz and Durgungoz (2025)
Theme 2: AI-Driven Automated Grading and assessment efficiency	Examines the capacity of AI to automate grading and assessment processes while improving efficiency, consistency, scalability and reliability.	Automated scoring, short-answer assessment, grading accuracy, consistency, assessment time, workload reduction, scalability and human-AI collaborative grading.	Myint et al. (2024); Ramnarain-Seetohul et al. (2025); Schneider et al. (2023); Papageorgiou and Pierrakeas (2026); Yildirim-Erbasli et al. (2026); Deepshikha (2025); Durgungoz and Durgungoz (2025);
Theme 3: AI-Powered Personalised Feedback and student learning	Examines the role of AI-generated and AI-supported feedback in providing timely, personalised and formative information to support student learning.	Personalised feedback, formative feedback, feedback quality, peer feedback, student engagement, learning support and feedback responsiveness.	Maier et al. (2026); Pozdniakov et al. (2026); Yildirim-Erbasli et al. (2026); Durgungoz and Durgungoz (2025); Deepshikha (2025); Hooda et al. (2022); Durgungoz and Durgungoz (2025)
Theme 4: Educational and institutional Benefits of AI in Higher Education Assessment	Synthesises the reported educational and institutional value of AI-enabled grading and feedback for students, educators and HEIs	Reduced educator workload, faster assessment, improved consistency, scalability, timely feedback, learning support and enhanced assessment capacity	Maier et al. (2026); Schneider et al. (2023); Pozdniakov et al. (2026); Papageorgiou and Pierrakeas (2026); Yildirim-Erbasli et al. (2026); Deepshikha (2025); Durgungoz and Durgungoz (2025)
Theme 5: Ethical, pedagogical and governance challenges of AI-powered assessment	Examines the limitations and risks associated with AI-enabled assessment and the conditions required for responsible implementation.	Algorithmic bias, fairness, transparency, explainability, data privacy, academic integrity, limitations in evaluating complex responses, human oversight, AI literacy and institutional governance.	Maier et al. (2026); Xiong and Li (2026); Papageorgiou and Pierrakeas (2026); Yildirim-Erbasli et al. (2026); Durgungoz and Durgungoz (2025); Deepshikha (2025); Suazo-Galdames and Chaple-Gil (2025)

Source: Author's own compilation

3. Results and Discussion

Following the screening and eligibility assessment, 12 articles were included for final review. Six articles were from ScienceDirect, four from WoS and two from Scopus. Although, the search strategy covered articles published between 2015 and 2026, after the application of inclusion criteria and screening process, the 12 papers included in the final review were published between 2022 and 2026. This indicates the recent expansion of literature on AI in higher education assessment, especially in relation to automated grading, AI-powered personalised feedback and AI-enabled assessment. The included articles were published across journals focusing on AI in higher education, educational technology, assessment and higher education.

Table 3 show that studies on AI-driven assessment are international dispersed. The studies are from Europe, Asia, Africa, Northen Americal and South America. Germany, Singapore, Mauritius, Liechtenstein/Switzerland, the Netherlands, China, India/Saidi Arabia, Greece, Canada/ Türkiye, Türkiye/United Kingdom and Chile were represented among the final reviewed papers. This shows that AI-powered assessment is being explored across different institutions contexts. However, Table 3 indicates that distribution is geographically uneven, with only Ramnarain-Seetohul et al. (2025) provide evidence from African higher education setting, namely Mauritius. This suggests that the evidence regarding AI integration, teacher preparedness, HIEs governance and student responses may not generalised to the environments characterised by different technological resources, institutional capacities and educational traditions.

Table 3
Geographical Distribution of Reviewed Studies

No	Author(s) and initials, year of publication	Journal	No. of citations*	Country/region of study	Continent
1	Maier, Seibold and Klotz (2026)	*Computers and Education Open*	0	Germany	Europe
2	Myint, Lo and Zhang (2024)	*Computers and Education: Artificial Intelligence*	3	Singapore	Asia
3	Ramnarain-Seetohul, Rosunally and Bassoo (2025)	*International Journal of Artificial Intelligence in Education*	1	Mauritius	Africa
4	Schneider, Richner and Riser (2023)	*International Journal of Artificial Intelligence in Education*	43	Liechtenstein/Switzerland	Europe
5	Pozdniakov, Brazil, Banihashem, Noroozi, Gašević, D., Sadiq and Khosravi (2026)	*Computers & Education*	2	Netherlands	Europe
6	Xiong and Li (2026)	*Educational Research Review*	0	China	Asia
7	Hooda, Rana, Dahiya, Rizwan and Hossain (2022)	*Mathematical Problems in Engineering*	396**	India/Saudi Arabia	Asia
8	Papageorgiou and Pierrakeas (2026)	*Education Sciences*	0	Greece	Europe
9	Yildirim-Erbasli, Dibek, Thomas and Lesoway (2026)	*Behavioral Sciences*	0	Canada and Türkiye	North America / Asia
10	Durgungoz and Durgungoz (2025)	*Education and Information Technologies*	2	Türkiye and United Kingdom	Asia/Europe

Source: Author's own compilation

No	Author(s) and initials, year of publication	Journal	No. of citations*	Country/region of study	Continent
11	Deepshikha (2025)	*Discover Artificial Intelligence*	11	India	Asia
12	Suazo-Galdames and Chaple-Gil (2025)	*Ingénierie des Systèmes d'Information*	2	Chile	South America
Source: Author's own compilation					

3.1 Theme 1: AI Technologies and Applications in Higher Education Assessment

The first theme is the AI technologies and assessment applications through which AI is being infused in higher education assessment. The 12 reviewed studies indicate that AI or GenAI-driven assessment is not based on a single technological approach. Instead, prior research reveals an expanding technological landscape involving LLMs, NLP, AGS systems and AI-supported grading collaborative assessment, AI-driven feedback, peer-feedback and general AI use in educational measurement.

The reviewed literature reports that different AI technologies are being used to different assessment activities. For example, Maier. (2026) explored LLM-generated formative feedback and reported the growing role of GenAI in facilitating feedback processes. Myint et al. (2024) explored an AI-Instructor collaborative grading approach for semi-open ended multipart question. This study illustrates a model which supports rather than completely replace the educator. Ramnarain-Seetohul et al. (2025) proposed a hybrid framework for automated assessment of short answers, while Schneider et al. (2023) examined trustworthy automated grading of multilingual and multi-type shorts questions. These researchers show a shift from traditional automated scoring towards AI systems that can process more sophisticated and diverse responses from students.

Prior research also revealed that AI applications are extending beyond direct automated grading. Pozdniakov et al. (2026) investigated AI assistance in peer-feedback provision, while Xiong and Li (2026) reviewed the broader use of AI in educational measurement. Deepshikha (2025) also synthesised evidence on AI-driven grading and tailored feedback in universities. Futhermore, Durgungoz and Durgungoz (2025) show the expansion of AI-driven assessment through AI-powered quizzes. Collectively, these researchers report that AI is being applied across multiple assessment functions, including scoring, formative feedback, peer feedback and assessment support. They also suggest that AI is positioned as both instrument for generating a grade and as infrastructure that support many phases of the assessment process, staring from generation and marking/scoring to personalised feedback and learner support. Therefore, the evidence highlights a transition from narrowly defined

automated grading towards more integrated AI-powered assessment ecosystems. Prior literature opposes treating AI-enabled assessment as a single technological intervention with a uniform pedagogical function. Therefore, reviewed studies that evaluation of AI-assessment should differentiate between the technology application, the assessment activity conducted, the extent of automation and the level of human involvement.

3.2 Theme 2: AI-Driven Automated Grading and Assessment Efficiency

This second theme concentrates on the ability of AI to foster grading/scoring efficiency, consistency and scalability, focusing on the environments where teachers must assess large number of students. Automation has emerged as one of the clearest rationales for integrating AI in higher education assessment. Myint et al. (2024) reports the potential of AI-instructor collaborative grading to support the assessment of semi-open-ended questions. These researchers do not defined efficiency as complete automation, instead acknowledging that AI can inadequately perform grading process without retaining evaluative role of educator. In the same vein, Ramnarain-Seetohul et al. (2025) designed a hybrid model for automated short -answer assessment, suggesting the potential of automated grading while recognising the continuing importance of human intervention. Scheider et al. (2023) focused on trustworthy automated grading of short, multilingual and multi-type answers. Their work shows the importance of developing grading systems that can produce reliable assessment across different responses format and languages. This is especially important because automated grading is only educationally useful when resulting scores demonstrates acceptable levels of consistency and accuracy.

Furthermore, Papageorgiou and Pierrakeas (2026) reveals that AI-assisted grading can reduce the time required to assess student-written responses and reduce aspects of teachers' grading workload. Similarly, Deepshikha (2025) identifies efficiency, scalability and workload reduction as essential values of AI-driven grading and personalised feedback. However, Yildirim-Erbasli et al. (2026) introduce significant qualification by demonstrating that students' preferences for AI grading and feedback may depend on the assessment context. They report that AI-driven grading is more appropriate for lower-stakes assessment when combined with expert human judgement. These findings show that the acceptability of AI-driven assessment is shaped by the assessment context and the role of human expertise.

Overall, the reviewed studied suggests that automation is one of the principal advantages of AI-driven assessment, especially in large class where manual grading can impose significant demands on educators. However, prior research does not support the interpretation that automation should result in the complete removal of educator from grading. Rather, the stronger evidence points towards AI-assisted or human-AI collaborative grading, where technological efficiently is combined with academic judgement. The efficiency benefits are best defined and understood as an opportunity to reallocate teachers' time rather than reducing human involvement in higher education assessment.

3.3 Theme 3: AI-Generated Personalised Feedback and Student Learning

The third theme reported in this study is the capacity of AI to provide timely, personalised and formative feedback. Traditional automated grading/scoring predominantly focuses on producing assessment scores, whereas AI-driven feedback can provide students with information aimed to explain their performance and identify areas for improvement.

Maier et al. (2026) offer especially relevant evidence through their investigation of LLM-driven formative feedback. These authors highlight the potential of GenAI to support learning through feedback that is produced promptly and adapted to students' responses. This expands the role of AI beyond grading towards a more formative function in which feedback becomes part of the learning process. The prominence of this application lies in its ability to make feedback more personalised, immediate and scalable, especially in contexts where teachers provide inadequate extensive and individualised feedback to every student.

Pozdniakov et al. (2026) investigated AI assistance in peer- feedback provision and found that AI can support students in producing higher-quality feedback. This indicates that AI may shape both the feedback received by students and students' capacity to participate in feedback processes. This perspective positions AI-driven feedback as "feedback generator", "facilitator of feedback engagement" and "feedback literacy". Similarly, Yildirim-Erbasli et al. (2026) explored student preferences about AI grading and feedback and reported that students' acceptance of AI-driven assessment depends partly on the assessment context and the involvement of human experts. These authors emphasise the importance of human expert in determining when AI-generated assessment is appropriate. Durgungoz and Durgungoz (2025) provide harmonising evidence from AI-powered quizzes. Their study reveal that AI-enabled assessment can improve student engagement, provide immediate personalised feedback, enable students identify areas for improvements. They also found that AI-generated assessment can create opportunity for identifying students who may be at-risk and allowing timely intervention. Their study also reports that AI-driven assessment can reduce educator workload while retaining the important of human oversight.

Together, these researchers reveal that AI-enabled feedback has the potential to address one of the persistent limitations of higher education assessment, namely the difficulty of providing timely and personalised feedback at scale. However, they emphasise that individualised feedback should not be equated automatically with effective and constructive feedback. The educational value of feedback depends on relevance, accuracy, depth and contextual appropriateness of the feedback, as well as students' ability to critically engage with it.

3.4 Theme 4: Educational and Institutional Benefits of AI-Powered Assessment

The fourth theme synthesises the educational and institutional benefits reported in final reviewed articles. A consistent results show that AI can reduce the time and effort associated with assessment while increasing the capacity of educators to offer feedback to large classes. Some reviewed studies found workload-related benefits. For example, Papageorgiou and Pierrakeas (2026) reported that AI-driven grading can reduce the time required for assessment and reduce educators and marking burden. In the same vein, Deepshikha (2025) found broader benefits including, improved efficiency, personalised support, reduced educator workload and enhanced learning experiences.

At the student level, reviewed studies report that potential gains extending beyond receiving grade or score. Maier et al. (2026) report that AI -driven feedback improves assessment and support learning processes. Pozdniakov et al. (2026) reveal how AI can support student in providing more effective and constructive feedback. Moreover, Durgundoz and Durgungoz (2025) add that that AI-driven feedback enhances student engagement, immediate personalised feedback, and help identifying areas for improvement and potential identifying at-risk student for timely intervention.

AI-driven assessment ability to identify student who may be at-risk is especially important because it indicates that AI systems may contribute to a movement from retrospective assessment to more proactive learning support.

The prior research indicates that the principal value of AI-powered assessment lies not simple replacing manual assessment but in expanding the capacity of educators and institutions to provide timely, consistent and scalable assessment support. AI can potentially free educators from repetitive assessment tasks and allow greater attention to be directed towards higher-order educational activities. However, these gains depend on the reliability and pedagogical appropriateness of underlying AI systems.

3.5 Theme 5: Ethical, Pedagogical and Governance Challenges of AI-Powered Assessment

The fifth theme is about ethical pedagogical and governance challenges associates with integrating AI-driven assessment. Although the final reviewed studies show significant potential, they also report important challenges that prevent AI from being treated as an autonomous replacement for academic judgement. One of the key concerns is fairness and algorithmic bias, consistent and valid across different students and response ty. Xiong and Li (2026) report algorithmic bias as an important issue in AI-powered educational measurement. Their study explore difference between AI and human grading, raising questions regarding bias and consistency. Similarly, Papageorgiou and Pierrakeas (2026) raise concern regarding fairness, transparency and the interpretation of complex student answers. These findings suggest that automated assessment systems may not consistently reproduce the judgment of expert educators, especially where responses involve ambiguity, creativity or complex reasoning. Scheneider et al, (2023) demonstrate why assessment validity cannot be reduced to whether an algorithm can generate a grade, suggesting a scoring system must be able to respond appropriately to different kind and format and complexity of student’s responses.

Prior research also report transparency and explainability as a concern. Students and educators need to understand how assessment decisions are produced, especially when AI-enabled scores shape academic outcomes. Therefore, previous studies show that technological performance alone is inadequate. AI-driven assessment must also be sufficiently transparent and trustworthy to support meaningful educational decision-making.

Human oversight has emerged as another important requirement (Myint et al., 2024; Deepshikha, 2025; Durgungoz and Durgungoz, 2025; Yildirim-Erbasli et al., 2026). Yildirim-Erbasli et al. (2026) indicate that AI-driven grading and feedback may be especially appropriate in lower-stakes contexts when combines with human expertise. Similarly, Durgungoz and Durgungoz (2025) report limitations in AI-generated feedback and the continuing importance of human oversight. Deepshikha (2025) argues that the successful implementation of AI-driven grading and feedback requires careful integration of human expertise, continuous validation and appropriate ethical safeguards. This emerging evidence is essential because it questions a simplistic technology-substitution narrative, suggesting how responsibility should be allocated between AI systems and teachers. Therefore, prior research supports a human-AI partneship model, instead of full automation.

The literature also indicates concerns regarding data privacy and responsible use of student information, especially because AI-driven assessment systems can require substantial amounts of student data. These concerns are closely connected to the need for institutional governance and clear policies governing how assessment data are collected, processed and used.

Overall, the findings indicate that the central challenge is not whether AI can perform assessment tasks, bit how AI can be integrated without compromising fairness, academic integrity, educational quality and human judgement. Therefore, the evidence supports a huma-AI partnership model in which AI provides computational support while educators retain responsibility for interpreting and conceptualising assessment outcomes.

Across the five themes, the prior research shows a consistent pattern. AI-driven assessment is advancing from relatively narrow automated scoring systems towards broader systems incorporating automated grading, formative feedback, peer-feedback support and human-AI collaborative assessment. The reviewed studies show significant potential for enhancing assessment efficiency, reducing teacher workload and increasing access to timely and individualised feedback.

At the same time, the previous studies included in the final review shows that technological capability alone does not guarantee effective assessment. Challenges regarding algorithmic bias, transparency, privacy, academic integrity, the assessment of complex questions/responses and human oversight remain significant. The strongest rising position across the prior research is not replacement of teachers by AI, but the advancement of responsible human-AI assessment models in which AI facilitates repetitive and scalable activities as teachers retain responsibility for professional judgement and interpretation of AI-generated responses.

This evidence from literature is especially essential for HEIs because integration of AI-driven assessment needs more than technology implementation, it demands appropriate governance, teachers AI knowledge and competence, continuous validation, transparent assessment methods and clear boundaries regarding the role of AI in academic decision making.

The benefits emerging from reviewed studies depend on the reliability of AI technologies, the competence of teachers, institutional governance and meaningful human oversight.

4. Conclusion

This study reviewed existing literature to investigate the application, benefits and challenges of AI-driven grading and personalised feedback higher education. Employing PRISMA 2020 guidelines, the search and screening were performed across WoS, Scopus and ScienceDirect. The 12 articles published between 2022 and 2026 were included for final review. This SLR reports that AI-enabled assessment is transforming beyond traditional automated grading towards a broader ecosystem involving automated grading, LMM-driven feedback, AI-assisted peer feedback, AI-generated quizzes, hybrid assessment approaches and AI-supported educational measurement.

The SLR's findings show that the principal contribution of AI technologies to higher education assessment lies in its ability to augment assessment processes instead of simply automating them. Prior research reports that AI can increase the scalability and efficiency of grading, reduce repetitive student assessment scoring/marking demands, offer more immediate and personalised feedback, and facilitate forms of student engagement and provide intervention that are difficult to main through manual assessment. The literature on AI-driven quizzes also suggest potential for identifying student who may require additional support, creating opportunities for more timely intervention. These gains are especially relevant to HEIs handling large and diverse student population.

However, the review also shows that technological ability should not be equated with pedagogical effectiveness. AI-enabled scores and feedback remain subject to concern surrounding bias, consistency, transparency, privacy, assessment validity and the interpretation of complex student's answers. Educational benefits of AI depend on the quality, contextual appropriateness and accountability of assessment processes in which it is embedded.

oversight has emerged as one of the important factors in the reviewed literature. The existing studies support that AI benefits include, computational speed, consistency and scalability, while educators retain responsibility for professional evaluation, contextual interpretation, moderation and consequential academic decisions. This suggests that the future of AI-enabled assessment should not be conceptualised primarily as a choice between human assessment and machine assessment. These studies suggest the combination of human and artificial intelligence to produce assessment that is transparent, efficient, valid, and pedagogically sound. the Reviewed studies report that the most promising future for AI in higher education assessment is not the replacement of academic judgement,

but its augmentation, suggesting that pedagogical legitimacy of AI-driven assessment depends on retaining human expert to interpret, question and contextualised AI-generated responses.

For HEIs, this indicates that integrating AI-driven assessment demands more than acquiring or implementing AI technologies. Clear policies regarding responsible uses of AI, instruments for validating AI-powered assessment responses, educator training in AI tools, protecting student data. HEIs also need guidelines for addressing inaccurate or biased AI outputs.

The review also provides important avenues for future research. Further empirical studies are needed to determine the long-term effects of AI-driven grading and personalised feedback on student learning, feedback engagement, student academic development and educational equity. The future researcher should conduct longitudinal and comparative studies to establish whether reported benefits of AI-driven assessment in efficiency and engagement contribute to long-term and sustainable improvements in learning and assessment quality.

Declarations

Ethics, Consent to Participate, and Consent to Publication declaration

Not Applicable

Consent for publication

Not Applicable

Competing interests

The author declares no competing interests.

Funding

Declaration

The research received no specific funding.

Author Contribution

M.C wrote the main manuscript, prepare tables, proofread and reviewed the final manuscript

Data Availability Statement

Not Applicable

References

1. Adugbire BA, Jordan PJ, Young C, Iwu-Jaja C. Strategies for implementation of patient and family-centered care in the perioperative context: A scoping review. Health Sci Rev. 2025;14. https://doi.org/10.1016/j.hsr.2025.100220.
2. Ahmed SK, Mohammed RA, Nashwan AJ, Ibrahim RH, Abdalla AQ, Ameen BMM, Khdhir RM. Using thematic analysis in qualitative research. J Med Surg Public Health. 2025;6. https://doi.org/10.1016/j.glmedi.2025.100198.
3. Alghamdi AA, Alanezi MA, Khan ZF. (2020). Design and implementation of a computer aided intelligent examination system. *International Journal of Emerging Technologies in Learning, 15(1)*. https://doi.org/10.3991/ijet.v15i01.11102
4. Ansor F, Zulkifli NA, Jannad DSM, Krisnaresanti A. (2023). Adaptive learning based on artificial intelligence to overcome student academic inequalities. *Journal of Social Science Utilising Technology, 1(4*). https://doi.org/10.70177/jssut.v1i4.663
5. Braun V, Clarke V. (2026). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3(2).* https://doi.org/10.1191/1478088706qp063oa
6. Chiu TFF, Xia Q, Zhou X, Chai CS, Cheng M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers Education: Artif Intell. 2023;4. https://doi.org/10.1016/j.caeai.2022.100118.
7. Cooper G, llich K, Sturrock K. (2026). Using generative artificial intelligence to reimagine feedback in higher education: a collaborative autoethnography. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2026.2653889
8. Dai W, Chan CKY. Shaping responsible GenAI use in research through AI literacy-oriented guidelines: insights from postgraduate students. Int J Educational Technol High Educ. 2026. https://doi.org/10.1186/s41239-026-00609-6. 23.
9. Dai K, Liu Y, Zhang X. Generative AI in higher education: A bibliometric review of emerging trends, power dynamics, and global research landscapes. Computers Education: Artif Intell. 2026;10. https://doi.org/10.1016/j.caeai.2026.100544.
10. Deepshikha D. A comprehensive review of AI-powered grading and tailored feedback in universities. Discover Artif Intell. 2025. https://doi.org/10.1007/s44163-025-00517-0. *5*.
11. Durgungozi A, Durgungoz FC. Exploring effortless AI-generated gamified quizzes in an online special education module: evaluating question quality, student engagement, and its potential to identify at-risk students. Educ Inform Technol. 2025;30. https://doi.org/10.1007/s10639-025-13765-5.

12. Gao Y. Deep learning-based strategies for evaluating and enhancing university teaching quality. Computers Education: Artif Intell. 2025;8. https://doi.org/10.1016/j.caeai.2025.100362.
13. Hamilton D, McKenzie G, Nejstgaard JE, Brennan CH, Moher SE, D., Page MJ. J Clin Epidemiol. 2026;192. https://doi.org/10.1016/j.jclinepi.2026.112133.
14. Hooda M, Rana C, Dahiya O, Rizwan A, Hossain S, Kumar V. (2022). Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education. *Mathematical Problems in Engineering, 2022(1).* https://doi.org/10.1155/2022/5215722
15. Ilma A, Rohmah Z. (2025). AI in EFL education: teachers' competence and the roadblocks to teaching material development. *Cogent Education, 12(1).* https://doi.org/10.1080/2331186X.2025.2588471
16. Joshi M, Linnansaari A, Ngowi J, Mwaniki E, Kahshamba K, Salla A. Understanding context in the implementation of school health promotion: a scoping review and conceptual framework. Syst Reviews. 2026;15. https://doi.org/10.1186/s13643-026-03232-8.
17. Khosravi H, Shum SB, Chen G, Conati C, Tsai Y, Knight S, Martinez-Maldonado R, Sadiq S, Gasevic D. Explainable Artificial Intelligence in education. Computers Education: Artif Intell. 2022;3. https://doi.org/10.1016/j.caeai.2022.100074.
18. Kizilcec RF, Ruber E, Papanastasiou EC, Cram A, Makridis CA, Smolansky A, Zeivots S, Raduescu C. Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. Computers Education: Artif Intell. 2024;7. https://doi.org/10.1016/j.caeai.2024.100269.
19. Kolaski K, Logan LR, Ioannidis JPA. Guidance to best tools and practices for systematic reviews. Syst Reviews. 2023;12. https://doi.org/10.1186/s13643-023-02255-9.
20. Levy-Feldman I. The Role of Assessment in Improving Education and Promoting Educational Equity. Educ Sci. 2025;15(2). https://doi.org/10.3390/educsci15020224.
21. Madwe MC, Chonco CM, Zungu A. (2025). Artificial Intelligence in Higher Education Assessment: Opportunities, Challenges and Pedagogical Considerations. *International Journal of Applied Research in Business and Management, 6(2).* https://doi.org/10.51137/wrp.ijarbm.2025.mmaa.45846
22. Madwe MC. (2026). Generative Artificial Intelligence Student Feedback in Large University Accounting Classes: Prompting Techniques. *International Journal of Sustainability in Business and Economics, 2(1).* https://doi.org/10.51137/wrp.ijsbe.575
23. Maier U, Seibold M, Klotz C. LLM-generated formative feedback in education: A qualitative systematic literature review. Computers Educ Open. 2026;10. https://doi.org/10.1016/j.caeo.2026.100374.
24. Mbandlwa Z, Ndlovu J, Mpofu L. (2026). Generative Artificial Intelligence and the Crisis of Research Integrity: Implications for Knowledge Production in Higher Education. *Journal of Public Administration and Development Alternative, 11(si1).*
25. Myint PYW, Lo SL, Zhang Y. Harnessing the power of AI-instructor collaborative grading approach: Topic-based effective grading for semi open-ended multipart questions. Computers Education: Artif

Intell. 2024;7. https://doi.org/10.1016/j.caeai.2024.100339.

26. Nongkhai LN, Wang J, Wynn A, Mendori T. Evaluating adaptive and generative AI-based feedback and recommendations in a knowledge-graph-integrated programming learning system. Computers Education: Artif Intell. 2026;10. https://doi.org/10.1016/j.caeai.2025.100526.
27. Papageorgiou K, Pierrakeas C. (2026). Do AI Grading Systems Systematically Differ from Human Teachers' Grading? Evidence of Bias and Consistency in Educational Assessment. *Education Sciences, 16(7).* https://doi.org/10.3390/educsci16071147
28. Pozdniakov S, Brazil J, Banihashem SK, Noroozi O, Gasevic D, Sadiq S, Khosravi H. AI assistance in peer feedback provision: Pedagogically sound, but minimally adopted. Comput Educ. 2026;248. https://doi.org/10.1016/j.compedu.2026.105591.
29. Ramnarain-Seetohul V, Rosunally Y, Bassoo V. A Unified Conceptual Hybrid Framework for the Automated Assessment of Short Answers. Artif Intell Educ. 2025;35. https://doi.org/10.1007/s40593-025-00487-5.
30. Ren x, Wu ML. Examining Teaching Competencies and Challenges While Integrating Artificial Intelligence in Higher Education. Tech Trends. 2025;69. https://doi.org/10.1007/s11528-025-01055-3.
31. Saez Y, Garcia LM, Mochon A, Isasi P. Evaluating large language models for AI-assisted grading: a framework and case study in higher education. Sci Rep. 2026;18. https://doi.org/10.1038/s41598-026-48656-3.
32. Schneider J, Richner R, Riser M. Towards Trustworthy AutoGrading of Short, Multi-lingual, Multi-type Answers. Artif Intell Educ. 2023;33. https://doi.org/10.1007/s40593-022-00289-z.
33. Singh SV, Hiran KK. (2022). The impact of AI on teaching and learning in higher education technology. *Journal of Higher Education Theory and Practice, 22(13).* https://doi.org/10.33423/jhetp.v22i13.5514
34. Suazo-Galdames IC, Chaple-Gil AM. (2025). AI-Driven Assessment Systems in Higher Education: Effectiveness for Enhancing Critical Thinking and Creativity. *Ingénierie des Systèmes d'Information,30(6).* https://doi.org/10.18280/isi.300624
35. Swiecki Z, Khosravi H, Chen G, Martinez-Maldonado R, Lodge JM, Milligan S, Selwyn N, Gasevic D. Assessment in the age of artificial intelligence. Computers Education: Artif Intell. 2022;3. https://doi.org/10.1016/j.caeai.2022.100075.
36. Tan LY, Hu S, Yeo DJ, Cheong KH. A Comprehensive Review on Automated Grading Systems in STEM Using AI Techniques. Mathematics. 2025;13(17). https://doi.org/10.3390/math13172828.
37. Wang H, Dang A, Wu Z, Mac S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. Computers Education: Artif Intell, 7. https://doi.org/10.1016/j.caeai.2024.100326
38. Xiong J, Li F. A decade of reflection and thematic review on artificial intelligence's impact on educational measurement. Educational Res Rev. 2026;51. https://doi.org/10.1016/j.edurev.2026.100789.

39. Xu Q, Liu Y, Li X. Unlocking student potential: How AI-driven personalized feedback shapes goal achievement, self-efficacy, and learning engagement through a self-determination lens. Learn Motiv. 2025;91. https://doi.org/10.1016/j.lmot.2025.102138.
40. Yildirim-Erbasli SN, Dibek MI, Thomas ML, Lesoway N. (2026). When Do Undergraduate Students Prefer AI? Insights into AI Scoring and Feedback. *Behavioral sciences, 16(7).* https://doi.org/10.3390/bs16071196
41. Yousefiazar A, Ghahramanian A, Rahmani A, Feizollahzadeh H, Sadeghi MRT, Ghaffarifar S, Parizad R. Integrating theoretical and practical constructs in patient education: A comprehensive scoping review. Patient Educ Counselling. 2025;139. https://doi.org/10.1016/j.pec.2025.109226.
42. Zhan Y, Boud D, Dawson P, Yan Z. (2025). Generative artificial intelligence as an enabler of student feedback engagement: a framework. *Higher Education Research & Development, 44(5).* https://doi.org/10.1080/07294360.2025.2476513
43. Zhan Y, Yan Z. (2025). Students' engagement with ChatGPT feedback: implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 51(5).* https://doi.org/10.1080/02602938.2025.2471821
44. Zhao J, Chapman E, Sabet PGP. (2024). Generative AI and Educational Assessments: A Systematic Review. *Education Research and Perspectives, 51 https*://doi.org/10.70953/ERPv51.2412006