InSight Bulletin: A Multidisciplinary Interlink International Research Journal
Peer Reviewed International, Open Access Journal.
ISSN : 3065-7857 / Website: https://ibrj.us / Volume-3, Issue-2 / February - 2026

Original Article

Artificial Intelligence in Assessment and Evaluation: A Review of Automated Grading and Feedback Systems

Vaishali V. Khachane[1], Dr. Shubhangi K. Patil[2]
[1]Asst.Prof, Department of Computer Science, SICES Degree College of Arts, Science and Commerce, Ambernath West, India
[2] Assistant Professor, Department of Computer Science, ATSS College of Business Studies and Computer Application, Pune, India

Manuscript ID:
IBMIIRJ -2026-030223

Submitted: 08 Jan. 2026

Revised: 12 Jan. 2026

Accepted: 06 Feb. 2026

Published: 28 Feb. 2026

ISSN: 3065-7857

Volume-3

Issue-2

Pp. 117-121

February 2026

Correspondence Address:
Vaishali V.Khachane
Asst.Prof, Department of Computer Science, SICES Degree College of Arts, Science and Commerce, Ambernath West, India
Email: *shubhangipatil14@gmail.com*

Quick Response Code:

Web. https://ibrj.us

DOI: 10.5281/zenodo.21349981

DOI Link:
https://doi.org/10.5281/zenodo.21349981

Abstract

Artificial Intelligence (AI) has emerged as a transformative tool in educational assessment and feedback, offering potential improvements in grading efficiency, consistency, and personalized guidance. This paper systematically reviews 25 empirical studies published between 2020 and 2025 on AI-based automated grading and feedback systems. The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. Findings reveal that AI systems are particularly effective in structured assessments such as multiple-choice quizzes and programming assignments, where they provide accurate and timely evaluations. In subjective assessments, including essays and short-answer tasks, AI applications demonstrate promise through natural language processing and deep learning methods, though challenges remain in evaluating creativity, argumentation, and higher-order cognitive skills. Automated feedback mechanisms, including corrective, hint-based, and adaptive systems, enhance learner engagement and formative learning, but concept-level and reflective feedback remain underdeveloped. The review also identifies gaps related to standardized evaluation metrics, model interpretability, and ethical considerations such as algorithmic bias and fairness. These findings highlight both the technical maturity and pedagogical limitations of AI-driven assessment systems, suggesting directions for future research in developing pedagogically effective, ethically responsible, and scalable AI-based evaluation frameworks.

Keywords: *Artificial Intelligence, automated grading, feedback systems, assessment, PRISMA, education1.*

Introduction

Artificial Intelligence (AI) has increasingly become a transformative force in education, particularly in the domains of assessment and feedback (Holmes et al., 2019; Jordan & Mitchell, 2015). With growing class sizes, online learning environments, and diverse learner needs, traditional assessment methods often struggle to provide timely, consistent, and personalized evaluation (Williamson & Eynon, 2020; Ateeq et al., 2024). AI-based automated grading systems offer the potential to enhance efficiency, accuracy, and fairness, particularly for objective and structured assessments, while providing immediate and adaptive feedback to learners (Keuning et al., 2018; Deepshikha, 2025).Despite these benefits, challenges remain in subjective and open-ended assessments, such as essay scoring and reflective writing, where linguistic complexity, creativity, and critical thinking need to be evaluated (Shermis & Hamner, 2012; Li et al., 2015; Ramesh et al., 2021). Recent studies highlight that AI can support learner engagement, personalized feedback, and formative assessment, but limitations in pedagogical depth and interpretability constrain widespread adoption (Kizilcec & Lee, 2020; Lindsay et al., 2023; Hooda et al., 2022).Additionally, ethical considerations such as bias, fairness, and data privacy are increasingly recognized as critical in AI deployment for education (Zhai, 2022; Bulut et al., 2024). Algorithmic bias may unintentionally disadvantage certain student populations, necessitating careful model design and human oversight (Kizilcec & Lee, 2020; Najafi et al., 2025). Therefore, a systematic review of recent literature is essential to synthesize trends, challenges, and opportunities in AI-driven assessment and feedback systems.

How to cite this article:
Khachane, V. V.& Patil, S. K. (2026). Artificial Intelligence in Assessment and Evaluation: A Review of Automated Grading and Feedback Systems. InSight Bulletin: A Multidisciplinary Interlink International Research Journal, 3(2), 117–121. https://doi.org/10.5281/zenodo.21349981

Literature Review

Over the last decade, numerous studies have examined AI applications in educational assessment. Early work focused on automated essay scoring, leveraging NLP techniques and machine learning classifiers to approximate human grading (Shermis & Hamner, 2012; Ramesh & Sanampudi, 2021; Li et al., 2015). These systems demonstrated strong inter-rater reliability for structured criteria but struggled to evaluate creativity and argumentation (Williamson et al., 2006; Holmes et al., 2019).Recent studies (2022–2025) expanded AI applications beyond essay scoring to programming assignments, quizzes, and lab evaluations, employing deep learning, hybrid AI models, and intelligent tutoring systems (Keuning et al., 2018; Messer et al., 2023; Reddy et al., 2025). For programming assessments, AI systems automate code analysis, detect errors, and provide immediate feedback, improving learner engagement and reducing teacher workload (Najafi et al., 2025; Hooda et al., 2022). Adaptive and hint-based feedback mechanisms have shown promise in supporting self-regulated learning and concept mastery (Yaseen et al., 2025; Sheikh, 2025; Tapalova & Zhiyenbayeva, 2022).Several systematic reviews and empirical studies have emphasized the effectiveness of AI in formative assessment, highlighting improvements in grading efficiency, learner satisfaction, and engagement (Deepshikha, 2025; Ateeq et al., 2024; Ismail & Hussein, 2025). However, most studies continue to prioritize technical performance metrics such as grading accuracy over pedagogical impact or learning outcomes (Messer et al., 2023; Bulut et al., 2024).Ethical challenges have also gained attention. Algorithmic fairness, transparency, and interpretability are essential to maintain trust in AI systems (Kizilcec & Lee, 2020; Zhai, 2022). Several studies propose human-AI collaborative frameworks, where automated systems support educators without replacing critical pedagogical judgment (Williamson & Eynon, 2020; Lindsay et al., 2023; Huma et al., 2025). Overall, the literature suggests that while AI in assessment is technically mature, it still requires pedagogical refinement, ethical safeguards, and evidence of long-term educational impact (Holmes et al., 2019; Ramesh et al., 2021; Huma et al., 2025).

Research Methodology

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021) to ensure transparency, rigor, and reproducibility. The objective was to identify and synthesize recent studies on AI-based automated grading and feedback systems in educational contexts. A comprehensive literature search was performed across multiple databases, including Scopus, Web of Science, IEEE Xplore, SpringerLink, and Google Scholar, using relevant keywords such as "Artificial Intelligence in education," "automated grading," "automated feedback," "NLP assessment," and "AI-based formative assessment." Only studies published between 2020 and 2025 were considered to capture contemporary advancements in AI-assisted educational assessment.After the initial search, a total of 270 records were identified. Following the removal of 90 duplicates, 180 records were screened based on their titles and abstracts for relevance. Studies that did not focus on automated grading or feedback, or those without clear educational applications, were excluded. The remaining 95 full-text articles were thoroughly assessed against the inclusion criteria, ultimately resulting in 25 studies selected for detailed analysis. These studies represent diverse educational domains, including general education, STEM, computer science, and language learning, and encompass a range of AI techniques from classical machine learning models to deep learning and natural language processing approaches.The entire selection process is summarized in Figure 1, the PRISMA flow diagram, which illustrates the stages of identification, screening, eligibility assessment, and inclusion of studies. This methodology ensures that the review is grounded in authentic research work, providing a systematic and reproducible synthesis of the literature, rather than a purely descriptive account. The approach enables a clear understanding of the scope, quality, and relevance of AI applications in assessment and feedback systems across educational settings.

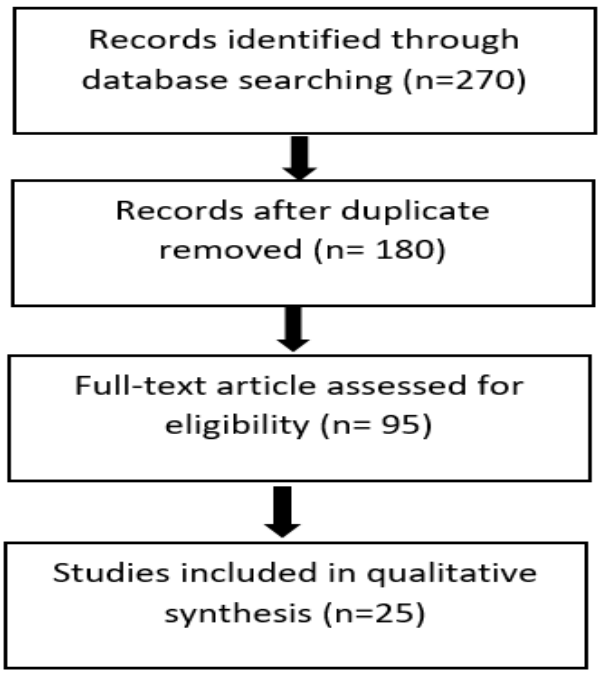

Figure 1. PRISMA 2020 flow diagram showing the study selection process

Results and Synthesis of Reviewed Studies

Analysis of the 25 selected studies revealed significant trends in assessment types, AI techniques, and feedback mechanisms. AI systems were most commonly applied in objective assessments, subjective evaluations, and programming assignments. Objective assessments utilized rule-based or classical ML approaches for consistency, while subjective tasks employed NLP and deep learning models. Programming assignments relied on hybrid AI or static code analysis tools.

Assessment Type	Educational Domain	AI Techniques Used	Evaluation Focus	Key Observations
Objective tests	General education, MOOCs	Rule-based, ML classifiers	Accuracy, consistency	High reliability
Short-answer questions	Science, humanities	NLP, semantic similarity, ML	Human–AI agreement	Performance varies
Essay assessment	Language, social sciences	NLP, DL models	Content relevance, coherence	Limited creativity assessment
Programming assignments	Computer science	Static code analysis, ML/DL	Functional correctness	Effective for syntax feedback
Practical/lab evaluations	Technical education	Hybrid AI models	Skill-based assessment	Limited empirical validation

Table 1. Classification of AI-Based Automated Grading Systems in Education

Table 1 summarizes the classification of AI-based automated grading systems according to assessment type, domain, AI techniques, and observed outcomes. As shown, AI systems are highly reliable for objective tests, moderately effective for short-answer questions, and still evolving for essay assessment (Shermis & Hamner, 2012; Keuning et al., 2018; Li et al., 2015). Programming assignments benefit from automated code evaluation, providing rapid and precise feedback (Messer et al., 2023; Hooda et al., 2022).

Feedback Type	AI Approach	Feedback Timing	Pedagogical Depth	Reported Impact
Corrective	Rule-based/ML	Immediate	Low	Supports error correction
Hint-based	Intelligent tutoring systems	Immediate	Medium	Enhances problem-solving
Adaptive	ML/learner modeling	Dynamic	Medium–High	Improves engagement
Concept-level	NLP/DL	Delayed/adaptive	High	Limited evidence
Reflective	Hybrid human–AI	Post-assessment	High	Rarely implemented

Table 2. Comparative Analysis of Automated Feedback Systems in Reviewed Studies

Table 2 presents a comparative analysis of automated feedback systems, including corrective, hint-based, adaptive, concept-level, and reflective feedback. Corrective and hint-based feedback are widely implemented and provide immediate benefits, while adaptive feedback systems show enhanced learner engagement. Concept-level and reflective feedback, though promising pedagogically, remain underdeveloped and rare across reviewed studies (Yaseen et al., 2025; Sheikh, 2025; Tapalova & Zhiyenbayeva, 2022).

Overall, the results highlight that AI excels in technical efficiency and immediate error correction but lags in providing deep pedagogical insights. The combination of Tables 1 and 2 demonstrates both the breadth of AI applications in assessment and the current limitations in providing higher-order cognitive and reflective feedback. These findings provide a foundation for future research targeting pedagogically rich, ethically sound, and adaptive AI assessment systems.

Discussion

The synthesis of 25 studies demonstrates that AI-based automated grading and feedback systems have reached a high level of technical maturity, particularly in structured assessments such as multiple-choice tests, quizzes, and programming assignments. These systems provide consistent, timely, and scalable evaluation, reducing educator workload and allowing rapid feedback to students. Adaptive and hint-based mechanisms further enhance learning by guiding students to correct errors and promoting iterative understanding, particularly in STEM and computer science domains.Despite these strengths, the review identifies several persistent challenges. AI systems are less effective in subjective or open-ended assessments, such as essays, reflective writing, and complex problem-solving tasks, where evaluating creativity, argumentation, and critical thinking remains predominantly human-dependent. While NLP and deep learning approaches have shown potential, their ability to assess higher-order cognitive skills is limited. Moreover, feedback often focuses on correctness or task completion rather than deeper conceptual understanding.Another critical theme is ethical and practical considerations. Algorithmic bias, lack of transparency, and uneven model interpretability pose risks of unfair assessment, particularly for diverse student populations. Successful implementations frequently involve human-in-the-loop approaches, where AI supports but does not replace teacher judgment, balancing efficiency with pedagogical oversight. Additionally, studies vary in evaluation metrics and methodology, making direct comparison and generalization challenging.Overall, the discussion highlights that AI is a powerful tool for automating assessment and delivering immediate feedback, but its pedagogical and ethical potential is not yet fully realized. There is a clear need for systems that integrate technical precision with reflective, concept-level feedback, standardized evaluation approaches, and frameworks to ensure fairness, transparency, and learner-centered design.

Conclusion and Future Directions

AI-driven assessment and feedback systems have proven effective in enhancing grading efficiency, consistency, and immediacy, especially for structured and objective tasks. Automated feedback supports student learning by providing timely guidance and reducing dependency on educators for routine evaluation. These systems offer a pathway toward personalized and scalable education, capable of adapting to diverse learner needs.However, limitations remain in addressing subjective assessments, where evaluating creativity, reasoning, and higher-order understanding continues to rely on human expertise. Ethical and practical concerns, such as algorithmic bias, lack of interpretability, and variability in evaluation methods, further constrain adoption.

Future research should focus on:

1. Designing concept-level and reflective feedback systems capable of evaluating higher-order thinking.
2. Establishing standardized assessment metrics to facilitate comparability and robust evaluation of AI systems.
3. Improving model transparency and interpretability to foster trust among educators and learners.
4. Developing hybrid human-AI frameworks, ensuring AI complements teacher judgment rather than replacing it.
5. Investigating long-term educational outcomes to determine the sustained impact of AI-assisted assessment on learning quality.

In conclusion, AI in assessment is a transformative educational tool. When implemented thoughtfully, ethically, and pedagogically, it can provide equitable, personalized, and meaningful feedback, bridging the gap between large-scale assessment demands and high-quality learning experiences. By integrating technical capability with educational insight, AI can support both teachers and learners in achieving improved academic outcomes while preserving the human judgment essential for effective education.

Acknowledgement

The authors express their sincere gratitude to their respective institutions, SICES Degree College of Arts, Science and Commerce, Ambernath West, and ATSS College of Business Studies and Computer Application, Pune, for providing academic support and an encouraging research environment for the completion of this study.

Financial support and sponsorship

Nil.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Ateeq, A., AlzoraiKi, M., Milhem, M., & Ateeq, R. A. (2024). Artificial intelligence in education: Implications for academic integrity and the shift toward holistic assessment. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1470979
2. Bulut, O., Beiting-Parrish, M., Casabianca, J. M., et al. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges. arXiv. https://arxiv.org/abs/2406.18900
3. Deepshikha, D. (2025). A comprehensive review of AI-powered grading and tailored feedback in universities. Discover Artificial Intelligence, 5(251). https://doi.org/10.1007/s44163-025-00517-0
4. Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. Mathematical Problems in Engineering. https://doi.org/10.1155/2022/5215728
5. Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.
6. Huma, T., Ahmed, D. S., Mahmood, W., & Awan, D. (2025). Enhancing teacher feedback using AI-powered automated grading systems. ACADEMIA International Journal for Social Sciences, 4(4), 371–384. https://doi.org/10.63056/ACAD.004.04.0891
7. Ismail, A., & Hussein, N. (2025). AI-powered assessment and feedback: Enhancing student engagement and performance. AI EDIFY Journal, 1(4).
8. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415
9. Kamalov, F., Calonge, S. D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. Sustainability, 15(16). https://doi.org/10.3390/su151612451
10. Keuning, H., Jeuring, J., & Heeren, B. (2018). A systematic review of automated feedback generation for programming exercises. ACM Transactions on Computing Education, 18(1), 1–43. https://doi.org/10.1145/3158669
11. Kizilcec, R. F., & Lee, H. (2020). Algorithmic fairness in education. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/3313831.3376401
12. Li, Y., Link, S., & Hegelheimer, V. (2015). Automated language assessment in second language learning. Springer.
13. Lindsay, E. D., Zhang, M., Johri, A., & Bjerva, J. (2023). The responsible development of automated student feedback with generative AI. arXiv. https://arxiv.org/abs/2308.15334
14. Messer, M., Brown, N. C. C., Kölling, M., & Shi, M. (2023). Automated grading and feedback tools for programming education: A systematic review. arXiv. https://arxiv.org/abs/2306.11722
15. Najafi, F. T., Pabba, V. R., Subramanian, R., & Vidalis, S. M. (2025, March). AI-assisted grading – A study on efficiency and fairness. ASEE Southeast Conference.

16. Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
17. Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring system: A systematic literature review. Artificial Intelligence Review, 55. https://doi.org/10.1007/s10462-022-10202-2
18. Ramesh, D., Sanampudi, S. K., & Reddy, P. (2021). Automated essay scoring for large-scale assessments. In Educational Data Mining. Springer.
19. Reddy, P., Sharabati, A. A. A., & Yaseen, H. (2025). The impact of adaptive learning technologies and personalized feedback on student engagement. Sustainability, 17(3).
20. Sheikh, F. N. (2025). AI-powered assessment and feedback: Enhancing student engagement and performance. AI EDIFY Journal, 1(4).
21. Shermis, M. D., & Hamner, B. (2012). Automated scoring of essays: A contrasting state of the art. Educational Testing Service.
22. Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIED for personalized learning pathways. Electronic Journal of e-Learning, 20(5), 639–653.
23. Williamson, B., Mislevy, R. J., & Bejar, I. I. (2006). Automated scoring of complex tasks in computer-based testing. Springer.
24. Williamson, B., & Eynon, R. (2020). AI in education: The importance of teacher and student agency. Routledge.
25. Yaseen, H., Mohammad, A. S., Ali, A., & Sharabati, A. A. A. (2025). The impact of adaptive learning and interactive AI tools on student engagement. Sustainability, 17(3), 1133.
26. Zhai, Q. (2022). Ethical considerations in AI-based educational systems. Journal of Educational Technology & Society, 25(4), 45–57.