



AI-Driven Assessment Systems in Higher Education: Effectiveness for Enhancing Critical Thinking and Creativity

Iván Claudio Suazo-Galdames^{1*}, Alain Manuel Chaple-Gil²

¹ Universidad Autónoma de Chile, Vicerrectoría de Investigación y Doctorados, Santiago de Chile 7500912, Chile

² Universidad Autónoma de Chile, Facultad de Ciencias de la Salud, Santiago de Chile 7500910, Chile

Corresponding Author Email: ivan.suazo@uautonoma.cl

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300624>

ABSTRACT

Received: 9 May 2025

Revised: 10 June 2025

Accepted: 14 June 2025

Available online: 30 June 2025

Keywords:

AI-based assessment, critical thinking, creativity, automated scoring, higher-order cognitive skills, machine learning evaluation, automated scoring

Artificial intelligence (AI)-based assessment systems are emerging as innovative tools to evaluate and enhance critical thinking and creativity in higher education. By leveraging deep learning algorithms, generative language models, and automated scoring techniques, these systems offer scalable, adaptive, and personalized feedback mechanisms aligned with 21st-century cognitive skill development. Despite increasing implementation, empirical evidence regarding their effectiveness remains fragmented. This systematic review synthesized the findings of original peer-reviewed studies assessing the impact of AI-driven evaluation tools on students' higher-order thinking skills. Following PRISMA 2020 guidelines, comprehensive searches were conducted in PubMed, Scopus, and Web of Science. Inclusion criteria focused on university-level interventions evaluating critical thinking and/or creativity using AI-based assessment tools. Of 234 records identified, only three studies met all eligibility criteria for final inclusion. Data were extracted using standardized forms, and risk of bias was assessed with CASP checklists. The included studies applied diverse AI systems: a BERT-based short answer grading tool, a deep-learning-powered creativity assessment platform, and a GPT-3.5-based mock interview rubric. All reported strong correlations between AI-generated scores and expert human evaluations. Outcomes indicated that AI-based assessments reliably measured cognitive indicators such as inference, originality, communication clarity, and divergent thinking. However, ethical considerations, data transparency, and researcher-participant dynamics were insufficiently addressed across studies. AI-based assessment systems consistently demonstrated effectiveness in enhancing critical thinking and creativity among university students. This systematic review identified strong correlations between AI-generated evaluations and traditional human assessments, validating their reliability across cognitive domains such as inference, originality, and clarity. Despite ethical and methodological gaps in existing studies, the evidence supported AI's potential as a valuable complement to human judgment in higher education. These findings directly address the research question and confirm that AI-based assessment tools, when implemented responsibly, can contribute meaningfully to the development of higher-order cognitive skills.

1. INTRODUCTION

The effectiveness of AI-based assessment systems in enhancing critical thinking and creativity was increasingly recognized in recent educational research. Studies consistently indicated that integrating artificial intelligence (AI) frameworks and applications into educational practices substantially improved these cognitive skills among students.

The AI-Charya framework, which integrated cognitive, affective, and psychomotor domains, successfully promoted personalized learning experiences. Students exposed to this framework demonstrated notable improvements in critical and creative thinking skills, effectively preparing them for future workforce requirements [1].

Similarly, the AI Assessment Scale (AIAS) enabled educators to design assessments featuring varying levels of AI

integration, emphasizing human input and the development of critical thinking. Pilot studies utilizing this scale revealed reductions in academic misconduct and increased student engagement, which in turn fostered innovative submissions and significantly enhanced the overall learning experience [2].

Furthermore, AI integration within Design-Based Learning (DBL) activities was found to positively influence creative self-efficacy and reflective thinking among students. Participants using AI tools such as ChatGPT and Midjourney reported improvements in their design-thinking mindset; however, it was noted that certain cognitive domains did not show statistically significant differences [3].

Despite the promising results, some research suggested caution regarding AI integration in education. While AI assessments were beneficial for enhancing critical and creative thought processes, excessive dependence on technological

tools posed risks, potentially impeding deeper cognitive engagement and independent problem-solving abilities. This duality underscored the necessity for a balanced approach when incorporating AI into educational environments, aiming to maximize cognitive benefits while mitigating potential drawbacks [4].

The rapid evolution and integration of AI into educational contexts have created new opportunities and challenges for fostering higher cognitive skills among students. Despite the growing adoption of AI-based assessment systems designed to enhance critical thinking and creativity, their actual impact on student learning outcomes and traditional pedagogical paradigms remains insufficiently understood. Questions persist regarding their capacity to foster authentic cognitive engagement, particularly in contexts where over-reliance on automated tools may compromise independent reasoning and problem-solving. In light of these uncertainties, it becomes essential to critically examine the current empirical landscape, identifying how these technologies are being implemented in higher education and to what extent they contribute to the development of advanced cognitive abilities within academic environments increasingly shaped by digital innovation.

2. METHODOLOGY

This systematic review was carried out following the guidelines established by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [5].

2.1 Eligibility criteria

Inclusion criteria:

- Original research articles evaluating AI-based assessment systems.

- Participants are university students (undergraduate or postgraduate).
- Assessments explicitly measure critical thinking and/or creativity.
- Studies published in peer-reviewed journals.
- Studies employing quantitative, qualitative, or mixed-methods research designs.

Exclusion criteria:

- Studies focusing on AI tools not explicitly intended for assessment purposes.
- Articles without clear methodological descriptions or outcomes relevant to critical thinking or creativity.
- Opinion articles, reviews, editorials, book chapters, conference abstracts without full text, conference proceedings, or other non-original empirical sources.

2.2 Strategy for identifying relevant studies

A thorough literature search was performed across several electronic databases, namely PubMed, Scopus, and Web of Science (WoS). In addition, the reference lists of the selected articles and previous systematic reviews were examined manually to identify any further relevant studies that met the inclusion criteria.

The search strategy was designed using the PICO framework, targeting studies involving university or higher education students (Population), the use of artificial intelligence-based assessment tools (Intervention), conventional assessment methods or absence of intervention (Comparison), and outcomes focusing on critical thinking, creativity, advanced cognitive abilities, or overall cognitive development. The specific search strings applied in each database are provided in Table 1.

Table 1. Search queries for each database used in the research

Database	Formulation	Filters
Pubmed	((("Artificial Intelligence" [MeSH Terms] OR "Machine Learning" [MeSH Terms] OR "Educational Measurement" [MeSH Terms] OR "Educational Technology" [MeSH Terms] OR "Computer-Assisted Instruction" [MeSH Terms] OR "Algorithms" [MeSH Terms]) AND ("Students" [MeSH Terms] OR "Education, Higher" [MeSH Terms]) AND ("Thinking" [MeSH Terms] OR "Creativity" [MeSH Terms] OR "Problem Solving" [MeSH Terms] OR "Cognition" [MeSH Terms])) OR (("AI-based evaluation" [Title/Abstract] OR "Artificial intelligence evaluation" [Title/Abstract] OR "Automated evaluation" [Title/Abstract] OR "AI-based assessment" [Title/Abstract] OR "Algorithm-based evaluation" [Title/Abstract] OR "Computer-assisted evaluation" [Title/Abstract] OR "Automated grading" [Title/Abstract] OR "Automated scoring" [Title/Abstract]) AND ("University student*" [Title/Abstract] OR "College student*" [Title/Abstract] OR Undergraduate*[Title/Abstract] OR Graduate*[Title/Abstract]) AND ("Critical thinking" [Title/Abstract] OR Creativity [Title/Abstract] OR "Higher-order thinking" [Title/Abstract] OR "Cognitive skill*" [Title/Abstract] OR "Problem-solving" [Title/Abstract] OR "Analytical thinking" [Title/Abstract]))	Filters applied: Clinical Study, Clinical Trial, Randomized Controlled Trial.
Scopus	(TITLE-ABS-KEY("AI-based evaluation" OR "Artificial intelligence evaluation" OR "Automated evaluation" OR "AI-based assessment" OR "Algorithm-based evaluation" OR "Computer-assisted evaluation" OR "Automated grading" OR "Automated scoring") AND TITLE-ABS-KEY("university student*" OR "college student*" OR "higher education student*" OR undergraduate* OR graduate*) AND TITLE-ABS-KEY("critical thinking" OR "creative thinking" OR creativity OR "higher-order thinking" OR "cognitive skill*" OR "problem-solving" OR "analytical thinking"))	AND (LIMIT-TO (DOCTYPE, "ar"))
WoS	TS=("AI-based evaluation" OR "Artificial intelligence evaluation" OR "Automated evaluation" OR "AI-based assessment" OR "Algorithm-based evaluation" OR "Computer-assisted evaluation" OR "Automated grading" OR "Automated scoring" AND TS=("university student*" OR "college student*" OR "higher education student*" OR undergraduate* OR graduate*) AND TS=("critical thinking" OR "creative thinking" OR creativity OR "higher-order thinking" OR "cognitive skill*" OR "problem-solving" OR "analytical thinking"))	

2.3 Study selection

The selection of studies involved a two-stage screening process conducted independently by two reviewers. Initially, titles and abstracts of all retrieved articles were screened against the eligibility criteria using Rayyan® website. Articles clearly irrelevant or not meeting the inclusion criteria were excluded at this stage. Remaining articles proceeded to full-text screening, independently carried out by the same two reviewers to determine final eligibility. Discrepancies between reviewers at either stage were resolved through discussion and consensus; if consensus could not be reached, a third reviewer was consulted to make the final decision.

2.4 Data extraction

Data extraction was independently conducted by two reviewers using a predefined Excel spreadsheet that included the following fields: authors and publication year, country of study, academic area or degree program, study design, sample size and characteristics, AI assessment tool description, measurement instruments for critical thinking and creativity, main findings related to effectiveness and limitations reported by author.

Disagreements in data extraction was resolved through consensus meetings or consultation with a third reviewer.

Extracted data were screened and organized in the Excel® spreadsheet and subsequently exported to RStudio® for further analysis using specific libraries suitable for qualitative and quantitative analyses. Disagreements in data extraction were resolved through consensus meetings or consultation with a third reviewer.

2.5 Method for assessing risk of bias

Two reviewers independently assessed the methodological rigor of the included studies using the Critical Appraisal Skills Programme (CASP) checklists, selected according to the study design (qualitative, quantitative, or mixed-methods). Discrepancies in the evaluations were addressed through consensus discussions, and when needed, a third reviewer was consulted to resolve remaining disagreements.

2.6 Data analysis

Results were synthesized narratively, structured by outcome domains (critical thinking and creativity). Subgroup analysis was considered based on academic disciplines, types of AI interventions, and assessment methods.

Given the systematic review nature, ethics approval was not required. However, ethical considerations reported in included studies were documented and discussed.

3. RESULTS

A total of 234 records were identified through database searches, including 220 from PubMed, 6 from Scopus, and 8 from Web of Science. Prior to the screening process, 6 duplicate records were removed, leaving 228 records to be screened by title and abstract.

During the screening phase, 197 records were excluded for the following reasons: 39 were not original research articles, 91 presented an irrelevant study design, and 67 did not meet the predefined inclusion criteria. Consequently, 31 full-text

reports were sought for retrieval, all of which were successfully obtained.

Subsequently, 12 reports were assessed for eligibility. Of these, 9 were excluded after full-text evaluation due to the following reasons: 7 were not related to research, 1 had incomplete data, and 1 did not have the full text available. Finally, 3 studies met all inclusion criteria and were included in the final synthesis (Figure 1).

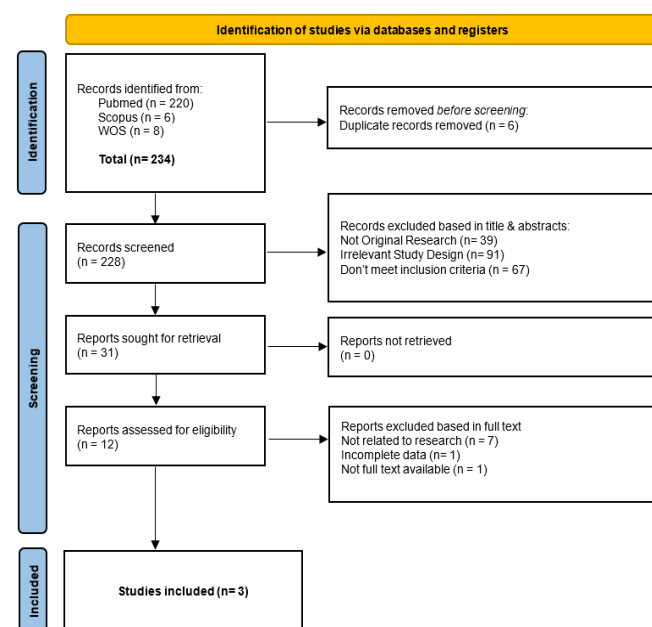


Figure 1. PRISMA flowchart

The diagram outlines the stages of exclusion due to duplication, ineligibility, and evaluation criteria, beginning with an initial pool of 234 records.

The impact of assessments based on AI tools on the development of higher cognitive skills in higher education was systematically analyzed through studies conducted between 2024 and 2025.

Mardini et al. [6] implemented an Automated Short Answer Grading (ASAG) system using Bidirectional Encoder Representations from Transformers (BERT) and Skip-Thought embeddings across multiple unspecified academic fields involving 199 participants. The AI-based assessment specifically evaluated critical cognitive indicators such as reading comprehension and inference abilities, operationalized through short-answer evaluations. When compared with traditional expert human grading, the AI-driven ASAG system produced highly correlated scores, demonstrating robust performance in assessing these higher cognitive skills.

In another study conducted by Sung et al. [7], a Computerized Creativity Assessment Tool (C-CRAT) was applied to 493 undergraduate participants from diverse academic disciplines. This tool focused on assessing divergent thinking indicators such as fluency, originality, flexibility, and elaboration. The outcomes revealed a strong correlation between AI-generated assessments and conventional paper-based divergent thinking tests, affirming the tool's validity and reliability in effectively capturing dimensions of creative thinking.

Uppalapati et al. [8] utilized a GPT-3.5-based grading system for mock interviews within the field of Engineering and Technology, involving 123 participants. This AI-driven

assessment specifically targeted cognitive indicators including professionalism, structured responses, clarity, and communication abilities. Findings indicated a strong correlation between GPT-3.5-generated rubric evaluations and human expert ratings, highlighting the potential of AI systems for objectively assessing professional competencies integral to higher cognitive skill development (Table 2).

All three studies Mardini et al. [6], Sung et al. [7], and Uppalapati et al. [8] presented a clear statement of aims and employed appropriate methodologies aligned with their research objectives. Their research designs were deemed suitable for addressing the central questions related to AI-based assessments and the development of higher cognitive skills.

The recruitment strategies were clearly described in only one of the studies [7], while the others did not provide sufficient detail to determine the adequacy of participant selection. In all cases, data collection methods were transparently reported and relevant to the outcomes assessed.

However, none of the studies explicitly discussed the relationship between researchers and participants, nor did they clearly address ethical considerations such as consent or institutional review board (IRB) approval, which introduces a potential source of bias. Despite these omissions, all studies conducted rigorous data analyses and clearly stated their findings. Moreover, each study was considered valuable for contributing to the understanding of how AI tools can support the assessment of higher-order cognitive skills in higher education (Table 3).

Table 2. Summary of studies analyzing the impact of AI-based assessments on critical/creative thinking indicators in higher education

Author Year	Title	N	Academic Area	AI assessment Tool	Critical/Creative Thinking Indicators	Comparison with Traditional Methods	Outcomes
Mardini et al. [6]	A deep-learning-based grading system (ASAG) for reading comprehension assessment by using aphorisms as open-answer-questions	199	Multiple fields (not specified)	ASAG system using BERT and Skip-Thought embeddings	Reading comprehension, inference ability (via short answer grading)	Yes (comparison with expert human grading)	BERT-based ASAG system produced scores with high correlation to human expert ratings
Sung et al. [7]	Construction and Validation of a Computerized Creativity Assessment Tool with Automated Scoring Based on Deep-Learning Techniques	493	Multiple fields (focus on undergraduate education)	Computerized Creativity Assessment Tool (C-CRAFT) using Word2Vec-based automated scoring	Divergent thinking (fluency, originality, flexibility)	Yes (compared with conventional paper-based DT test)	Automated scoring showed strong correlation with traditional DT scores and high reliability
Uppalapati et al. [8]	AI-driven mock interview assessment: leveraging generative language models for automated evaluation	123	Engineering and Technology	GPT-3.5-based grading system for mock interviews	Professionalism, structured answers, clarity, correctness, authenticity	Yes (compared with human expert ratings)	GPT-3.5-based rubric grading showed strong correlation with human ratings and effective performance summaries

Table 3. CASP risk of bias evaluation

Study	Clear Statement of Aims	Appropriate Methodology	Research Design Appropriate	Recruitment Strategy Appropriate	Data Collection Method Clear	Relationship Researcher- Participant Considered	Ethical Issues Considered	Rigorous Data Analysis	Clear Statement of Findings	Valuable Research
Mardini et al. [6]	Yes	Yes	Yes	Not clear	Yes	Not declared	Not declared	Yes	Yes	Yes
Sung et al. [7]	Yes	Yes	Yes	Yes	Yes	Not declared	Not declared	Yes	Yes	Yes
Uppalapati et al. [8]	Yes	Yes	Yes	Not clear	Yes	Not declared	Not declared	Yes	Yes	Yes

4. DISCUSSION

The analyzed studies consistently reported that AI-based assessments demonstrated strong validity and effectiveness when evaluating critical and creative cognitive skills, providing outcomes comparable to, and consistent with, traditional human grading methods.

The integration of Artificial Intelligence (AI) into educational assessment systems emerged as a transformative approach, offering innovative ways to evaluate and enhance students' critical thinking and creativity [9, 10]. Recent research consistently demonstrated that AI-based assessment systems significantly transformed traditional evaluation methods by enabling dynamic, personalized, and adaptive

learning experiences [11, 12]. These systems utilized advanced technologies such as machine learning, natural language processing, and generative AI to create interactive assessments that transcended conventional testing, focusing primarily on developing higher-order cognitive skills crucial for success in the 21st century [9].

Generative AI tools, notably platforms such as ChatGPT, were extensively employed to enhance critical analysis by requiring students to critique AI-generated outputs. For instance, postgraduate students engaging in critical evaluation of ChatGPT responses to project management-related questions demonstrated substantial improvements in their critical thinking skills [13]. However, studies also identified significant limitations, such as superficial responses and

potential biases within AI outputs, emphasizing the necessity of continued human oversight [13]. Furthermore, AI-driven adaptive learning platforms, including Intelligent Tutoring Systems (ITS), Knewton, and Carnegie Learning's MATH, provided personalized and real-time feedback, effectively promoting deeper reflection and enhanced critical thinking among learners [11, 12]. Additionally, inquiry-based learning facilitated through generative AI assessments encouraged students to explore complex, real-world problems, thereby strengthening their critical thinking skills and preparing them for authentic learning environments [11]. In STEM education contexts, AI-supported socio-scientific issue (SSI) education significantly improved critical thinking dispositions, as evidenced by a quasi-experimental study involving geology students [14].

AI-based assessment systems also effectively fostered creativity among students by providing interactive and imaginative learning environments. Generative AI tools applied in creative storytelling projects enabled prospective teachers to enhance their creative processes through self-assessment and co-evaluation activities [15]. Additionally, AI-generated simulations and prompts in language learning contexts were shown to stimulate creativity and innovative problem-solving, facilitating students' transition from critical to creative thinking [16]. Personalized learning trajectories enabled by adaptive AI systems catered to individual student needs, ensuring challenging and engaging experiences that promoted creative skill development [10]. Ethical and reflective creativity was further emphasized by studies requiring students to critically evaluate and reflect upon ethical implications of AI-generated content, thereby highlighting the importance of responsible AI usage in educational settings [17].

Despite these promising outcomes, several challenges and limitations were identified. Ethical concerns, particularly related to data privacy and algorithmic biases, posed significant issues in AI-based assessments, necessitating clear guidelines for ethical practices to safeguard student data and maintain trust in educational processes [9, 18]. Furthermore, the potential for academic integrity breaches due to excessive reliance on AI-generated content was recognized, underscoring the need for assessments resistant to AI misuse, which assess not only final outcomes but also students' interactions with AI tools throughout their learning processes [19]. Reliability and validity concerns arose due to the superficiality and inconsistency often observed in AI-generated rubrics, highlighting the essential role of human expertise in ensuring meaningful assessments [13]. Additionally, excessive reliance on AI tools risked promoting passive learning experiences, potentially undermining deeper cognitive engagement and independent critical thought [20].

To address these challenges and enhance AI-based assessments' effectiveness, recommendations included integrating AI as a supportive tool rather than a replacement for human instruction, thereby leveraging AI's strengths while retaining human judgment [13, 17]. Emphasis was placed on establishing clear ethical guidelines for AI use within educational settings, fostering AI literacy among educators and students, and ensuring responsible AI practices [18, 21]. Promoting collaboration between humans and AI was identified as critical for enhancing balanced and meaningful educational experiences [17]. Lastly, continuous monitoring and evaluation of AI-based systems were recommended to address emerging educational challenges and to optimize their

positive impacts on critical thinking and creativity [12].

The present study presented certain limitations that must be acknowledged. Firstly, the small number of studies that met the inclusion criteria restricted the generalizability and robustness of the findings. Additionally, among the included studies, details regarding recruitment strategies and ethical considerations were often insufficiently described, raising concerns about possible selection bias and ethical transparency. None of the studies clearly addressed researcher-participant relationships, nor did they explicitly report obtaining informed consent or institutional ethical approvals, which represented a significant gap that potentially affected the interpretability and validity of the outcomes. Moreover, although rigorous data analysis procedures were generally reported, the absence of explicit discussions around potential biases and limitations inherent in AI-based tools introduced further uncertainty regarding the comprehensive applicability of the findings.

Despite these limitations, the investigation significantly contributed to the understanding of AI's potential in higher education. The systematic approach provided valuable insights into how AI-based assessment systems effectively supported the development of critical thinking and creativity, demonstrating substantial comparability with traditional human-based assessments. The research highlighted specific AI frameworks and applications that successfully fostered personalized and adaptive learning experiences, aligning with contemporary educational demands. Furthermore, by clearly identifying areas for improvement, such as the need for enhanced transparency in ethical practices and recruitment methodologies, this study offered essential guidance for future research. Ultimately, despite the identified limitations, the investigation reinforced the importance of carefully integrating AI tools into educational contexts to maximize cognitive benefits while responsibly managing potential drawbacks.

Ethical Framework for AI-Based Assessments in Higher Education

Despite the promising effectiveness of AI-based assessment systems for enhancing critical thinking and creativity, the reviewed studies failed to sufficiently address essential ethical considerations such as informed consent, data protection, and researcher-participant relationships [22, 23]. These omissions signal a pressing need for a robust ethical framework to guide the responsible deployment of AI-driven educational technologies [24].

An ethical approach to AI-based assessment must prioritize three core principles: transparency, informed consent, and data protection. Transparency requires that students clearly understand how AI systems generate assessments, including the algorithms' decision-making processes and any potential biases embedded in training data [25]. Informed consent implies that participants should voluntarily agree to be evaluated using AI tools, with full awareness of the scope, limitations, and implications of the technology. Protection of personal and academic data is equally crucial, necessitating secure data handling protocols and strict compliance with institutional data governance policies [26, 27].

Institutional and international guidelines offer valuable direction for operationalizing these principles. For example, the UNESCO Recommendation on the Ethics of Artificial Intelligence [28] emphasizes the right to human oversight, algorithmic explainability, and the avoidance of algorithmic discrimination in educational contexts [23]. Similarly, the IEEE Global Initiative on Ethics of Autonomous and

Intelligent Systems provides guidance on prioritizing human well-being, ensuring accountability, and embedding ethical design from the outset of AI system development [25].

Institutions implementing AI assessments should adopt best practices that reflect these standards. These include:

- Conducting Ethical Impact Assessments before deployment [24].
- Establishing AI governance committees to oversee educational technology integration.
- Providing AI literacy training for both educators and students to foster critical engagement with these tools [23].
- Requiring clear documentation and public disclosure of AI system parameters and limitations [22].

Ultimately, embedding this ethical framework is essential to maintain trust, promote equitable educational practices, and ensure that AI-based assessments enhance rather than compromise the integrity of learning environments. As the field evolves, aligning technological innovation with ethical responsibility will be key to realizing the full potential of AI in education [25, 27].

One of the principal limitations of this review lies in the relatively low number of eligible studies, which constrains the overall robustness of the conclusions and limits the generalizability of the findings across diverse educational contexts. This scarcity of empirical evidence reduces the capacity to establish consistent patterns or draw definitive inferences regarding the effectiveness and broader applicability of AI-based assessment tools in fostering higher cognitive skills. Consequently, the findings should be interpreted with caution, acknowledging the potential for contextual variability and publication bias. This gap underscores the critical need for future research grounded in rigorous empirical methodologies, including well-defined control conditions, transparent recruitment strategies, and longitudinal follow-up. Expanding the evidence base through such studies would significantly enhance the reliability, validity, and transferability of conclusions in this emerging field.

While artificial intelligence offers significant potential to enhance assessment practices through scalability, efficiency, and consistency, its role must be understood as fundamentally complementary. AI should serve as a support tool that augments rather than replaces the nuanced judgment of educators. Particularly in qualitative assessments that involve evaluating complex competencies such as critical thinking, creativity, or ethical reasoning, human pedagogical insight remains indispensable. The interpretative richness, contextual awareness, and relational understanding that educators bring to these evaluations cannot be replicated by algorithmic systems. Ensuring that AI operates under human oversight preserves academic integrity, promotes fairness, and reinforces the central role of educators in shaping meaningful learning experiences.

5. CONCLUSIONS

This systematic review achieved its objective by synthesizing existing empirical evidence on the use of AI-based assessment systems to foster higher-order cognitive skills specifically critical thinking and creativity among university students. The included studies consistently demonstrated that AI-driven tools, such as BERT-based short-answer grading systems, deep-learning creativity assessments,

and GPT-3.5-powered rubric scoring, produced results strongly aligned with human expert evaluations. These outcomes affirm that such technologies are not only reliable in measuring critical and creative thinking, but also effective in enhancing them across diverse academic contexts.

In direct response to the research question, the findings confirmed that AI-based assessment systems hold significant potential to improve higher cognitive skills in university settings. Each of the selected studies reported measurable positive effects on indicators such as inference, originality, communication clarity, and divergent thinking. The review thus provides evidence that these tools when implemented with pedagogical rigor and ethical oversight can complement traditional assessment methods and enrich educational practices aimed at developing critical and creative competencies.

However, the limited number of included studies and the insufficient attention to ethical considerations and recruitment transparency indicate that further research is needed to generalize the findings. Future investigations should focus on expanding the empirical base, improving methodological transparency, and exploring the long-term cognitive impacts of integrating AI into assessment processes. Nevertheless, this review substantiates the value of AI assessment tools as legitimate and promising contributors to cognitive skill development in higher education.

REFERENCES

- [1] Namboothiri, S., Varghese, T., Jacob, M., Job, S., Cyriac, J. (2025). Integrating artificial intelligence with NHEQF descriptors for pedagogical excellence. *Higher Education for the Future*, 12(1): 27-50. <https://doi.org/10.1177/23476311241290894>
- [2] Furze, L., Perkins, M., Roe, J., MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4): 38-55. <https://doi.org/10.14742/ajet.9434>
- [3] Saritepeci, M., Yildiz Durak, H. (2024). Effectiveness of artificial intelligence integration in design-based learning on design thinking mindset, creative and reflective thinking skills: An experimental study. *Education and Information Technologies*, 29(18): 25175-25209. <https://doi.org/10.1007/s10639-024-12829-2>
- [4] Faisal, M. (2024). Dampak kecerdasan buatan (AI) terhadap pola pikir cerdas mahasiswa di Pontianak. *Nucleus*, 5(1): 60-66. <https://doi.org/10.37010/nuc.v5i1.1684>
- [5] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372: 71. <https://doi.org/10.1136/bmj.n71>
- [6] Mardini G, I.D., Quintero M, C.G., Vilorio N, C.A., Percybrooks B, W.S., Robles N, H.S., Villalba R, K. (2024). A deep-learning-based grading system (ASAG) for reading comprehension assessment by using aphorisms as open-answer-questions. *Education and Information Technologies*, 29(4): 4565-4590. <https://doi.org/10.1007/s10639-023-11890-7>
- [7] Sung, Y.T., Cheng, H.H., Tseng, H.C., Chang, K.E., Lin, S.Y. (2024). Construction and validation of a computerized creativity assessment tool with automated

- scoring based on deep-learning techniques. *Psychology of Aesthetics, Creativity, and the Arts*, 18(4): 493-509. <https://psycnet.apa.org/doi/10.1037/aca0000450>
- [8] Uppalapati, P.J., Dabbiru, M., Kasukurthi, V.R. (2025). AI-driven mock interview assessment: Leveraging generative language models for automated evaluation. *International Journal of Machine Learning and Cybernetics*, 1-23. <https://doi.org/10.1007/s13042-025-02529-9>
- [9] Saputra, I., Kurniawan, A., Yanita, M., Putri, E.Y., Mahniza, M. (2024). The evolution of educational assessment: How artificial intelligence is shaping the trends and future of learning evaluation. *The Indonesian Journal of Computer Science*, 13(6). <https://doi.org/10.33022/ijcs.v13i6.4465>
- [10] Onesí-Ozigagun, O., Ololade, Y.J., Eyo-Udo, N.L., Ogundipe, D.O. (2024). Revolutionizing education through AI: A comprehensive review of enhancing learning experiences. *International Journal of Applied Research in Social Sciences*, 6(4): 589-607. <https://doi.org/10.51594/ijarss.v6i4.1011>
- [11] Tariq, M.U. (2025). Advancing inquiry-based learning through generative ai-enabled assessments. In *Educational Assessments in the Age of Generative AI*. IGI Global Scientific Publishing, pp. 179-206. <https://doi.org/10.4018/979-8-3693-6351-5.ch007>
- [12] Zhao, C. (2024). AI-assisted assessment in higher education: A systematic review. *Journal of Educational Technology and Innovation*, 6(4). <https://doi.org/10.61414/jeti.v6i4.209>
- [13] Huang, C.W., Coleman, M., Gachago, D., Van Belle, J.P. (2023). Using ChatGPT to encourage critical AI literacy skills and for assessment in higher education. In *Annual Conference of The Southern African Computer Lecturers' Association*. Cham: Springer Nature Switzerland, pp. 105-118. https://doi.org/10.1007/978-3-031-48536-7_8
- [14] Liu, Q., Tu, C.C. (2024). Improving critical thinking through AI-supported socio-scientific issues instruction. *Journal of Logistics, Informatics and Service Science*, 11(3): 52-65. <https://doi.org/10.33168/JLISS.2024.0304>
- [15] Niclòs, I.P., Sanz, Y.E., Gómez, P.O., Ezpeleta, A.M. (2024). Creativity and artificial intelligence: A study with prospective teachers. *Digital Education Review*, (45): 91-97. <https://doi.org/10.1344/der.2024.45.91-97>
- [16] Roozafzai, Z.S. (2024). Artificial intelligence assistance and cognitive abilities: Harnessing AI-assisted heuristic methods for transitioning from critical to creative thinking in English language learning. *Educational Challenges*, 29(2): 339-361. <https://doi.org/10.34142/2709-7986.2024.29.2.23>
- [17] Doyle, S. (2024). Augmenting intelligence with generative AI: A guide for teaching talented students. In *Practices That Promote Innovation for Talented Students*. IGI Global Scientific Publishing, pp. 125-144. <https://doi.org/10.4018/978-1-6684-5806-8.ch006>
- [18] Abubakar, U., Falade, A.A., Ibrahim, H.A. (2024). Redefining student assessment in Nigerian tertiary institutions: The impact of AI technologies on academic performance and developing countermeasures. *Advances in Mobile Learning Educational Research*, 4(2): 1149-1159. <https://doi.org/10.25082/AMLER.2024.02.009>
- [19] Awadallah Alkhouk, W., Khlaif, Z.N. (2024). AI-resistant assessments in higher education: Practical insights from faculty training workshops. *Frontiers in Education*, 9: 1499495. <https://doi.org/10.3389/feduc.2024.1499495>
- [20] Hamdi, M. (2024). How AI is transforming and shaping the future of education. In *2024 IEEE 28th International Conference on Intelligent Engineering Systems (INES)*, Gammarth, Tunisia, pp. 000115-000116. <https://doi.org/10.1109/INES63318.2024.10629089>
- [21] Redondo-Duarte, S., Ruiz-Lázaro, J., Jiménez-García, E., Requejo, S.M. (2025). Didactic strategies for the use of AI in the classroom in Higher Education. In *Integration Strategies of Generative AI in Higher Education*. Hershey, PA, USA: IGI Global, pp. 23-50. <https://doi.org/10.4018/979-8-3693-5518-3.ch002>
- [22] Slade, S., Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10): 1510-1529. <https://doi.org/10.1177/0002764213479366>
- [23] Yan, L., Sha, L., Zhao, L., Li, Y., Martínez-Maldonado, R., Chen, G., Li, X., Jin, Y., Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1): 90-112. <https://doi.org/10.1111/bjet.13370>
- [24] Perkins, M., Furze, L., Roe, J., MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6): 49-66. <https://doi.org/10.53761/q3azde36>
- [25] Winfield, A.F., Michael, K., Pitt, J., Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]. *Proceedings of the IEEE*, 107(3): 509-517. <https://doi.org/10.1109/JPROC.2019.2900622>
- [26] Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, 16(2): 2577-2587. <https://doi.org/10.15354/sief.23.re202>
- [27] Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4): 675-690. <https://doi.org/10.1007/s10676-021-09606-x>
- [28] United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the ethics of artificial intelligence. Francia, pp. 1-21. <https://unesdoc.unesco.org/ark:/48223/pf0000380455.locale=es>