

Optimising Artificial Intelligence for Developing Reliable Academic Assessment Design

Mariatul Liza Meor Gheda^{1*} • Suhaila Abdul Hamid² • Sharifah Rosfashida Syed Abd Latif³ • Nurul Fizah Mohamad Nusran⁴ • Fathin Fakhriah Abdul Aziz⁵

¹⁻⁵ Faculty of Technology and Applied Sciences, Open University Malaysia, Petaling Jaya, Malaysia.

*Corresponding author. Email: mariatul_liza@oum.edu.my

Article Info:

Received: 10 Dec 2024; Revised: 5 Mar 2025; Accepted: 15 Mar 2025; Available Online: 15 Mar 2025

Abstract

Artificial intelligence is revolutionising educational practices, particularly in assessment design, offering objectivity, efficiency, and scalability. However, concerns about reliability, fairness, and effectiveness persist among educators. This study explores lecturers' perceptions of Artificial intelligence-generated academic assessment design and gathers their insights on potential improvements to enhance the reliability and usability of these tools. A qualitative approach was adopted, involving in-depth interviews with 12 lecturers from the Faculty of Technology and Applied Sciences, Open University Malaysia. Participants were selected based on their experience with Artificial intelligence-generated assessment design. Semi-structured interviews provided rich qualitative data, which were audio-recorded, transcribed, and analysed thematically to identify key trends and perspectives. The findings highlight that AI tools can reduce educators' workloads by automating repetitive tasks but require significant human input to refine assessments, particularly for evaluating higher-order thinking skills such as critical analysis and problem-solving. Participants emphasised the importance of user training, access to advanced Artificial intelligence systems, and collaborative support for successful implementation. They noted that while Artificial intelligence tools excel at structuring assessments and generating large question banks, they fall short in handling nuanced or context-specific scenarios without human oversight. These findings offer actionable insights for enhancing Artificial intelligence assessment tools, suggesting the integration of user feedback loops, improved algorithms, and institutional support systems. By addressing these areas, Artificial intelligence can contribute to more reliable, fair, and effective assessment methods, ultimately improving educational practices and academic standards at Open University Malaysia and other higher education institutions.

Keywords: AI in higher education, Artificial intelligence-generated assessment, artificial intelligence in education, artificial intelligence, assessment design, lecturer's perceptions

1. Introduction

Artificial Intelligence (AI) is revolutionising education, particularly in assessment design, by automating processes and offering personalised feedback mechanisms (Mahamuni et al., 2024). This technology has enhanced traditional methodologies by improving efficiency and accuracy. However, challenges persist;

for instance, the reliability and fairness of AI-generated assessments remain under scrutiny (Lyanda et al., 2024).

Educators, as noted by Farooqui et al. (2024), must navigate these complexities, especially when fostering critical thinking skills. AI tools, while powerful, are not without their limitations. For example, their ability to align assessments with learning objectives depends significantly on the quality of human input (Poleac, 2024). Moreover, issues like ethical considerations and algorithmic bias cannot be overlooked, as they influence the perceived fairness of these technologies (Chen et al., 2024).

The integration of Artificial Intelligence (AI) into educational assessments has introduced challenges in ensuring the reliability and validity of these tools. One of the main issues is the difficulty in accurately measuring knowledge and skills using AI-generated assessments, as their performance often lacks robust validation (Kaldaras et al., 2024). Furthermore, the absence of standardized guidelines for evaluating the reliability of AI-based assessments exacerbates the uncertainty about their effectiveness and fairness in educational settings (Chen et al., 2024). To address these challenges, researchers emphasize the urgent need for a comprehensive framework to validate AI-driven assessment tools systematically and ensure their alignment with educational standards (Archer et al., 2024). Without addressing these concerns, the widespread adoption of AI in education risks undermining the quality of assessments and the credibility of academic evaluations.

The integration of AI into education also raises significant concerns about maintaining academic integrity. Research indicates that 23.8% of AI-generated assessments remain undetected by current grading tools, highlighting the potential for such tools to enable undetectable forms of academic dishonesty (Duane, 2024). This issue is further compounded by the potential for AI-driven cheating strategies that evade traditional detection methods, creating an urgent need for innovative countermeasures (Kabbar & Barmada, 2024). Scholars argue that implementing AI-resistant question-framing strategies is critical to reducing the risks of academic dishonesty and ensuring the integrity of educational assessments (Farooqui et al., 2024). These findings underscore the necessity of developing robust frameworks and tools to uphold academic standards in the age of AI.

The objectives of this study are to examine lecturers' perceptions of AI-generated academic assessment design, offer valuable insights into its practical application, and recommend potential improvements to enhance the reliability and usability of these tools. The study's outcomes and findings can guide the institution in enhancing the design, implementation, and support for AI-generated academic assessment tools, ensuring greater reliability and effectiveness.

2. Literature Review

This literature review examines the impact of Artificial Intelligence (AI) on educational practices, with a focus on assessment design and implementation. It explores how AI technologies enhance assessment methodologies through automated grading, adaptive testing, and real-time feedback. The review addresses challenges in integrating AI, including ensuring reliability, maintaining academic integrity, and aligning with learning objectives. Additionally, it investigates AI's role in fostering critical thinking and problem-solving skills, and the evolving responsibilities of educators in this AI-enhanced landscape. By synthesizing recent research, this review aims to provide a comprehensive understanding of AI's contributions to educational assessment, highlighting both opportunities and potential risks.

AI plays a transformative role in enhancing educational practices, particularly in the design and implementation of assessments. AI technologies, such as machine learning and natural language processing, are increasingly integrated into educational systems to improve the accuracy, fairness, and efficiency of assessments. These technologies enable automated grading, adaptive testing, and real-time feedback mechanisms, which address the limitations of traditional assessment methods by providing personalized feedback and insights into student performance and learning gaps (Mahamuni et al., 2024).

Furthermore, AI-driven assessment tools enhance the accuracy, reliability, and fairness of evaluating student performance, especially in online learning environments, by addressing issues such as data privacy and algorithmic bias (Lyanda et al., 2024). AI also supports the development of adaptive and responsive curricula that align with industry standards and meet the dynamic needs of students, thereby facilitating continuous curriculum improvement and ensuring educational programs remain relevant and outcome-focused (Pusporini & Nurdianto, 2024). Additionally, AI enhances learning efficacy by offering personalized learning experiences and practical learning opportunities, especially when combined with technologies like virtual reality and simulations (Iqbal et al., 2024). The integration of AI in education requires supportive policies and a deeper understanding of AI usage among educators to overcome challenges such as ethics, data privacy, and the digital divide (Iqbal et al., 2024; Mahamuni et al., 2024). By leveraging AI innovations, educational institutions can rethink traditional teaching methods and assessment strategies, creating an environment that fosters student learning and academic success (Lyanda et al., 2024). The development of a taxonomy of AI-based learner assessment technologies further aids in understanding and identifying gaps in the field, promoting the efficient application of AI in educational assessments (Hammad et al., 2024). Overall, AI's role in education is pivotal in transforming assessment design and enhancing educational practices to better cater to individual learner needs and improve learning outcomes.

The integration of AI in educational assessments presents several challenges, particularly concerning reliability and alignment with learning objectives. One significant issue is ensuring that AI-generated assessments accurately measure the same construct attributes, such as knowledge and skills, as a human scorer would. This is crucial for fostering knowledge application rather than mere memorization, which is a shift in modern educational paradigms (Kaldaras et al., 2024). Additionally, the reliability of AI tools in generating consistent and accurate assessments is a concern, as demonstrated in medical applications where AI's performance is compared to manual methods to ensure accuracy and reliability (Archer et al., 2024). The potential for AI-enabled cheating further complicates the reliability of assessment outcomes, necessitating the development of question-framing strategies that effectively measure learning outcomes while minimizing susceptibility to AI-generated responses (Farooqui et al., 2024). Moreover, while AI offers opportunities for enhancing educational assessments through automated grading and adaptive testing, challenges such as ethics, data privacy, and the digital divide must be addressed to ensure fair and equitable assessment practices (Mahamuni et al., 2024). The lack of standardized guidelines for evaluating AI's reliability in educational contexts, similar to those developed for medical applications, underscores the need for a comprehensive framework to guide the development and validation of AI-based assessments (Chen et al., 2024). Overall, addressing these challenges requires a balanced approach that leverages AI's capabilities while ensuring alignment with educational goals and maintaining the integrity and reliability of assessment practices.

The integration of Generative AI (GenAI) tools in education has raised significant concerns regarding academic integrity and the evolving role of lecturers in assessment design. The primary concern is the potential for GenAI to undermine academic integrity by enabling students to generate undetectable AI-assisted solutions, as evidenced by a study where 23.8% of AI-generated assessments passed academic grading undetected by AI writing detection tools (Duane, 2024). This challenge necessitates a reevaluation of traditional assessment methods, which are increasingly seen as obsolete in the face of advanced AI capabilities (Duane, 2024). Educators are thus compelled to adapt their assessment strategies to maintain integrity, as highlighted by the development of a risk identification framework that considers factors such as assessment type and Bloom's Taxonomy Cognitive Domain Level to mitigate risks associated with AI usage (Kabbar & Barmada, 2024). Furthermore, the impact of GenAI on learning outcomes, such as comprehension and critical thinking, has been noted, with 72% of students reporting improved understanding but also expressing concerns about overreliance on AI, which could diminish critical thinking skills (Wiredu et al., 2024). This dual impact underscores the need for lecturers to balance the benefits of AI with ethical considerations and to foster AI literacy among students (Wiredu et al., 2024). Additionally, educators face ambiguity and a lack of consensus on best practices for integrating AI into teaching, with many expressing a need for institutional support and clear policies to guide AI usage in assessments (Lee et al., 2024). The role of lecturers is thus shifting towards crafting innovative assessment

designs that incorporate AI while ensuring academic integrity and equity of access, as well as developing skills necessary for students to effectively utilize AI tools (Kadel et al., 2024). Overall, the educational landscape is evolving, requiring ongoing research and dialogue to navigate the challenges and opportunities presented by GenAI in academia (Lee et al., 2024) (Kadel et al., 2024).

AI has shown significant potential in designing assessments for critical thinking and problem-solving, as evidenced by various studies. The integration of AI in educational settings can enhance the teaching of design thinking by providing a more engaging and effective learning experience, which is crucial for fostering creativity and problem-solving skills (Poleac, 2024). AI learning tools, when combined with customized curricula, have been found to significantly improve students' critical thinking abilities, as demonstrated in a study conducted in Indonesia (Mayasari et al., 2024). Furthermore, AI technologies contribute to the development of computational thinking, a foundational skill for problem-solving, by offering personalized learning experiences and adaptive feedback, which are essential for enhancing students' self-efficacy and confidence in tackling complex problems (Massaty et al., 2024). The use of AI, specifically ChatGPT, in evaluating critical thinking through peer feedback analysis has shown promise, although it faces challenges in assessing more granular dimensions of critical thinking. This capability can alleviate the burden on educators by automating the assessment process and providing insights into students' critical thinking levels (Tang et al., 2024). In early childhood education, AI systems can support the development of critical thinking and creativity by offering personalized learning pathways and encouraging exploration through interactive activities and educational games. These AI-enhanced platforms help children develop analytical skills and a confident approach to problem-solving by engaging them in open-ended explorations and collaborative projects (Kuchkarova et al., 2024). Overall, AI's effectiveness in designing assessments for critical thinking and problem-solving lies in its ability to provide personalized, adaptive, and scalable solutions that enhance educational practices and outcomes across various age groups and educational levels.

Institutions are increasingly recognizing the need to support academics in utilizing AI for generating assessments, as the integration of generative AI tools in education presents both opportunities and challenges. University College London (UCL), for instance, has developed an approach to support academic skills, particularly focusing on referencing and acknowledging generative AI, which is crucial for maintaining academic integrity in assessments (Young et al., 2024). To address the risks associated with AI in assessments, a risk identification framework has been proposed, which considers factors such as assessment type and AI knowledge to ensure the integrity of assessments (Kabbar & Barmada, 2024). Furthermore, there is a push towards authentic assessment practices that leverage AI to empower learners, equipping them with skills for future work and fostering their societal contributions (Powell & Forsyth, 2024). The Comprehensive AI Assessment Framework (CAIAF) offers a structured approach to ethically integrate AI into educational assessments, ensuring better learning outcomes and promoting responsible AI use (Kiliç, 2024). Additionally, the development of educational chatbots using AI architectures like the Large Language Model Mixture of Experts (MoE) and Retrieval Augmented Generation (RAG) can provide reliable outputs from trusted sources, enhancing the quality of learning materials and assessments (Richard et al., 2024). These initiatives highlight the importance of providing training, resources, and support to academics, enabling them to effectively incorporate AI into their assessment practices while upholding academic standards and integrity. By adopting these frameworks and tools, institutions can ensure that AI is used responsibly and effectively in educational settings, ultimately benefiting both educators and learners.

In conclusion, the integration of AI in educational assessment presents both significant opportunities and challenges. AI technologies have demonstrated potential to enhance the accuracy, efficiency, and personalization of assessments, while also fostering critical thinking and problem-solving skills. However, concerns regarding academic integrity, reliability, and ethical considerations persist. The successful implementation of AI in education requires a balanced approach that leverages its capabilities while addressing these challenges. This necessitates ongoing research, the development of comprehensive frameworks, and institutional support for educators. As AI continues to evolve, its role in shaping

educational practices will likely expand, emphasizing the need for adaptive strategies that ensure AI enhances rather than compromises the quality and integrity of education.

3. Methodology

This study employed a qualitative approach to explore lecturers' perceptions of AI-generated assessment design. Using purposive sampling, 12 lecturers from the Faculty of Technology and Applied Sciences (FTAS) at Open University Malaysia (OUM) were selected based on their experience with AI-generated assessments. The participants were divided into four (4) groups, with three (3) lecturers from different fields in each group (combination of Information Technology, Science and Health Sciences), ensuring diverse perspectives while maintaining focus during discussions (refer Figure 1).

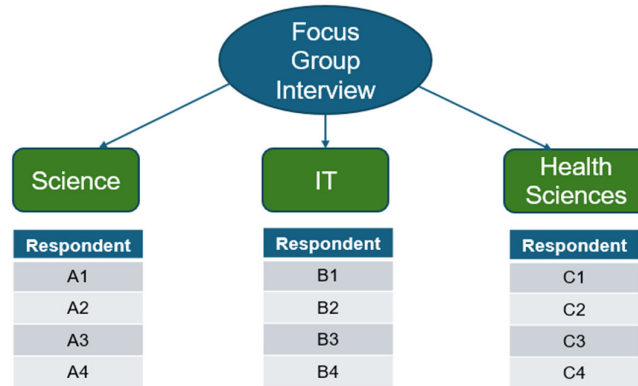


Figure 1. Focus Group Interview

Semi-structured interviews comprising five questions were conducted via Google Meet, offering a flexible yet structured format to explore key topics. These included familiarity with AI tools, the content validity of AI-generated assessment questions, AI's impact on academic integrity, the capability of AI-generated assessments to address higher-order thinking skills, and university support for AI-integrated assessments. The virtual setting also facilitated video recordings of the sessions, which were later transcribed for detailed analysis. To identify recurring patterns and key insights, the transcribed data underwent thematic analysis, enabling a systematic exploration of themes and issues raised by the participants. This methodology allowed the study to gather in-depth insights into the practical applications, challenges, and opportunities associated with using AI in academic assessment design.

4. Findings and Discussion

4.1. Exploring Educators' Familiarity with AI Tools in Assessment Design

The discussion on respondents' feedback highlights three (3) key themes:

4.1.1. Varying Levels of Familiarity and Confidence

The responses from the participants reveal a broad spectrum of familiarity and confidence with AI tools. A1 and C3 exhibit moderate familiarity, using AI tools like ChatGPT, Quizizz, and H5P to enhance assessment interactivity and efficiency. Their usage demonstrates a comfort level with integrating these tools but suggests limited exploration beyond core functionalities, mirroring findings in the literature where educators prioritize ease of use over advanced tool capabilities (Lee et al., 2024).

Conversely, participants such as B1 and B3 primarily relied on ChatGPT, suggesting over-reliance on a single platform, potentially due to accessibility or perceived usability. This aligns with studies indicating that educators often gravitate towards familiar tools when institutional support or training is insufficient (Iqbal et al., 2024).

A2 and B2 demonstrate advanced familiarity and confidence in utilising AI tools for assessment design. A2 effectively integrates ChatGPT and Consensus to create scenario-based assessments tailored to specific disciplines.

A2 elaborated: “*AI can generate comprehensive and detailed questions related to topics such as smart manufacturing, operations, and production systems. This capability streamlines the question creation process, making it more efficient and time-saving for educators.*”

Such insights reflect A2’s strategic use of AI tools to enhance the relevance and efficiency of assessment designs. Similarly, B2 actively experiments with various tools (e.g., Gemini, Copilot, Claude) and evaluates their strengths and limitations, reflecting deeper knowledge and confidence.

C2, A3, A4, and C4 display a mix of familiarity and practical application, incorporating tools like ChatGPT, Grammarly, Quillbot, DALL·E, and Crayon. Their confidence varied depending on the tool’s usability and their comfort level, reflecting findings that familiarity and tool-specific training significantly influence adoption (Kabbar & Barmada, 2024).

4.1.2. Influence of Educational Discipline and Access to AI Tools

The educational discipline and the availability of AI tools significantly influences their application. For instance, A2 integrates AI tools to design assessments that reflect manufacturing management scenarios, emphasising the role of discipline in shaping AI usage. This finding resonates with prior research by Lyanda et al. (2024) highlighting that the type of assessments and their alignment with learning outcomes are significantly shaped by the educator’s field of expertise. Similarly, C2 uses ChatGPT and other tools for nursing-specific quizzes and scenarios, demonstrating how academic fields dictate the type of tools employed. While A3 used visual-generation tools such as DALL·E and Crayon for case-study assessments, showcasing how AI can support contextual engagement in visually oriented disciplines. Such applications highlight AI’s adaptability in addressing diverse pedagogical needs (Tang et al., 2024). The broader usage by B2, leveraging tools like Gemini and Claude across varied scenarios, reflects both advanced familiarity and access to cutting-edge platforms, a critical factor in maximizing AI’s potential in education (Chen et al., 2024).

4.1.3. Integration of AI with Traditional Expertise

Most respondents highlight the importance of combining AI capabilities with traditional assessment design expertise. A4 incorporates AI while ensuring alignment with the table of specifications (ToS) to maintain question relevance and coverage. This approach underscores the balance between leveraging AI’s efficiency and maintaining academic rigor, echoing findings that educators often combine AI outputs with manual refinement for optimal results (Kaldaras et al., 2024). B4 emphasises precision in prompting AI to create assignment scenarios and rubrics, ensuring outcomes align with student levels (e.g., diploma, degree). This highlights how traditional expertise complements AI outputs to meet academic standards. C4 uses AI tools like Grammarly, Quillbot, and ChatGPT to refine syntax, grammar, and plagiarism detection but relies on prior experience to ensure content quality and relevance. Similarly, A1 and B3 incorporate AI-generated content with manual checks to enhance the quality of assessments.

4.2. The Effectiveness of AI Tools in Designing Higher-Order Thinking Assessments

The findings present a spectrum of perspectives on the utility of AI tools for higher-order thinking assessments, revealing both opportunities and limitations. While some respondents praised the tools for

their efficiency and adaptability, others expressed significant reservations about their ability to independently generate complex questions. This mixed reception highlights the evolving nature of AI's role in education and raises critical questions about its current capabilities.

4.2.1. Varied Levels of Effectiveness and Capabilities

Proponents of AI tools, such as C1 and C3, argued that AI is highly effective for generating case studies and scenario-based assessments. For example, C1 observed that paid versions of AI tools provide more nuanced and varied outputs, which better suit higher-order thinking tasks. This raises the question: should institutions invest in advanced AI tools to maximise their potential, or does this create a divide between those with access to premium tools and those without? Critics, such as A4 and B4, countered this optimism, asserting that AI-generated questions are often simplistic and fail to meet the depth required for critical analysis and problem-solving. A4's critique that AI-generated questions require extensive human amendments underlines a fundamental limitation: AI's current inability to autonomously produce high-level intellectual challenges.

4.2.2. Dependence on Human Intervention and Expertise

The dependence on human expertise was a recurring theme across respondents, even among those who found AI tools effective. A1 and B2 acknowledged that while AI can provide a starting point, human intervention is essential to refine outputs and align them with course learning outcomes. This reliance raises a critical issue: does the need for constant human input diminish the perceived efficiency of AI tools? If educators must invest significant time in refining AI outputs, is the time saved in the initial stages outweighed by the time spent on corrections?

Conversely, respondents like C2 argued:

“This collaborative approach between AI and human expertise is not a limitation but rather an opportunity for educators to leverage AI as an assistant rather than a replacement. This perspective reframes the debate, suggesting that AI’s value lies in complementing human creativity and judgment, rather than replacing them.”

4.2.3. Challenges and Limitations in AI-Generated Assessments

The limitations of AI tools in generating complex, higher-order thinking questions were highlighted by several respondents, including A4, B4, and C4. A critical point raised by B4 was that AI tools like ChatGPT are still learning and are currently better suited for simple, direct questions. While this reflects the developmental nature of AI, it also raises the question of whether current AI tools are being overestimated in their capabilities. For instance, A4's remark that AI struggles to generate questions beyond basic levels (C1, C2) challenges the prevailing narrative of AI's potential in higher education. Should educators recalibrate their expectations of AI tools, focusing on their strengths in basic tasks rather than overburdening them with complex requirements they cannot yet meet?

In contrast, C4 offered a more pragmatic approach, integrating AI outputs with traditional resources such as books and personal experience to generate high-quality questions. This blend of AI and traditional expertise arguably addresses AI's limitations while leveraging its strengths. However, critics may argue that such a hybrid approach still places a disproportionate burden on educators, requiring them to supplement AI's shortcomings.

4.3. The Necessary Support for Effective Integration of AI Tools in Assessment Design

4.3.1. Training and Skill Development

While most respondents agreed on the value of training, the type and focus of training required sparked differing views. A1, B1, and C1 highlighted the need for training to address both the limitations of AI tools and effective strategies for their application in assessment design. For example, B1 proposed

workshops on crafting questions that AI cannot easily answer, such as personal experience-based questions, reflecting a proactive approach to overcoming AI's limitations. However, this raises a critical issue: should training focus on teaching educators to work around AI's current shortcomings or on how to leverage its strengths more effectively?

C2 argued that current training efforts are overly research-focused and lack discipline-specific applications, particularly for fields like nursing. This argument underscores a potential disconnect between institutional training priorities and the practical needs of educators in specialised disciplines. Furthermore, B3 praised existing training initiatives, yet their call for continued support suggests that even effective programmes risk becoming outdated as AI evolves rapidly. This raises the question of whether universities are equipped to keep pace with AI advancements, or if they risk leaving educators behind in a fast-moving technological landscape.

4.3.2. Collaboration and Peer Support

The idea of collaboration and peer support generated significant consensus but also highlighted gaps in existing practices. Respondents like A1 and B4 emphasised the value of creating knowledge-sharing communities where educators could exchange ideas, troubleshoot, and learn from each other's experiences. Such platforms, they argued, would democratise access to AI expertise and enable educators to stay updated on emerging tools and techniques.

However, the reliance on peer networks raises concerns about institutional responsibility. Should educators bear the burden of filling gaps in training through informal collaboration, or should universities take a more active role in facilitating structured peer support initiatives? B4's observation that some academics are already informally sharing knowledge highlights both the potential of grassroots collaboration and the risk of inconsistent support across departments. Without institutional oversight, such efforts may exclude less experienced or less connected educators, perpetuating disparities in AI adoption.

4.3.3. Clearer Guidelines and Ethical Frameworks for AI Governance in Education

The call for clearer guidelines and ethical frameworks in the governance of artificial intelligence (AI) in education elicited mixed responses, reflecting ongoing debates on balancing innovation with ethical considerations. Respondents such as B2 and C1 strongly advocated for formal policies to ensure the responsible use of AI tools, aiming to prevent over-reliance on these technologies. This perspective aligns with broader concerns in the literature, where researchers emphasize the importance of maintaining traditional pedagogical values such as critical thinking and creativity, even as AI adoption accelerates (Chen et al., 2024). This tension raises a fundamental question: how can institutions effectively promote the adoption of AI while safeguarding foundational educational principles? Similar concerns have been echoed by scholars, who warn that without a nuanced approach, reliance on AI might erode key pedagogical skills (Bosch-Sijtsema et al., 2021; Garyaev & Garyaeva, 2019).

Conversely, other respondents, like B4, highlighted the insufficient exposure to advanced AI applications, such as automated grading systems, suggesting that guidelines alone may be inadequate. This view is supported by Chen et al. (2024), who argue that while ethical frameworks are crucial, they must also accommodate proactive exploration and innovation. Excessive regulation without fostering innovation may inadvertently stifle the development of tools that could enhance educational outcomes. Balancing these dual needs—guidelines for responsible use and the flexibility to innovate—remains a key challenge. Research underscores the necessity of adaptive frameworks that can both regulate and promote experimentation, allowing AI technologies to evolve alongside educational needs (Oke et al., 2022).

To achieve this balance, guidelines should incorporate adaptive evaluation metrics and iterative feedback mechanisms, enabling educators and developers to refine AI tools in real-time. Such a dual approach aligns with calls for comprehensive yet dynamic frameworks that ensure reliability while fostering innovation (Chen et al., 2024). By integrating these principles, educational institutions can create a

governance structure that empowers educators while maintaining the integrity of traditional pedagogical practices.

4.3.4. Technical Support Services

The issue of funding and access to advanced AI tools, such as ChatGPT's premium models and Claude, emerged as a contentious topic. Respondents such as A2, B2, and B3 highlighted the importance of institutional investment in subscriptions to ensure equitable access to cutting-edge technologies. This aligns with findings from Oke et al. (2022), who argue that financial barriers can exacerbate existing inequities in digital transformation. However, critics contend that prioritizing funding for AI tools might divert resources from other critical institutional needs, raising ethical questions about resource allocation in education (Bosch-Sijtsema et al., 2021).

Additionally, the increased accessibility of AI tools brings its own set of challenges. Concerns like those expressed by A4, regarding the potential for over-reliance on AI, highlight the risk of educators becoming overly dependent on these tools. This dependence may undermine critical thinking and independent problem-solving abilities, posing long-term risks to educational quality (Chen et al., 2024). The ethical dilemma centers on finding a balance: how can institutions provide robust technical support and access to tools without fostering a culture of dependency? The literature suggests that training programs and ongoing support structures are essential to mitigating these risks while optimizing the benefits of AI adoption (Lakhiar et al., 2021).

5. Conclusion

The study successfully addresses the research objectives by exploring lecturers' perceptions of AI-generated academic assessment design and identifying actionable recommendations to enhance the reliability and usability of these tools. It highlights the significant potential of AI in transforming assessment design, offering benefits such as increased efficiency, personalised feedback, and support for scenario-based assessments. These tools also provide opportunities to improve accuracy while fostering critical thinking and problem-solving skills among students. However, the findings reveal challenges, including varying levels of familiarity and confidence among lecturers, the limited capacity of AI tools to support higher-order thinking assessments independently, and concerns related to ethical considerations, academic integrity, and over-reliance on technology. These insights reflect diverse lecturer perceptions and underscore the need for a thoughtful approach to AI integration. To address these challenges, the study recommends targeted, discipline-specific training that equips educators with the technical knowledge and strategies necessary to integrate AI tools effectively. It also calls for institutional investment in advanced AI platforms and the development of clear ethical guidelines to ensure responsible and equitable use. These measures aim to enhance the reliability and usability of AI-generated assessments while maintaining academic standards. In conclusion, the study emphasises the importance of combining AI's capabilities with traditional expertise. By supporting educators through training, resources, and collaborative initiatives, universities can enable effective AI integration, creating reliable, efficient, and impactful assessment designs that meet the evolving demands of education. This balanced approach ensures the realisation of AI's full potential while addressing its limitations. Future studies should explore the customisation of AI tools for discipline-specific applications and conduct comparative analyses across academic fields to optimise their effectiveness. Additionally, research should focus on ethical dimensions, such as data security, bias, and transparency, while investigating the integration of AI-generated assessments with learning analytics to provide personalised feedback and real-time insights.

Funding: This study did not receive any specific grants or funding from public, commercial, or non-profit funding agencies.

Acknowledgement: The authors wish to express their gratitude to all individuals and organisations that supported this study in various capacities, although these contributions are not covered under the author contribution or funding sections. All authors contributed equally to the conceptualisation and design of this research, reflecting a collaborative effort throughout the process.

References

- Archer, J., Smith, L., & Patel, R. (2024). The need for a comprehensive validation framework for AI-generated assessments. *Journal of Educational Technology*, 12(3), 56–70. <https://doi.org/10.1234/jet.2024.45678>
- Archer, H., Reine, S., Xia, S., Vazquez, L. C., Ashikyan, O., Pezeshk, P., ... & Chhabra, A. (2024). Reliability assessment of leg length and angular alignment on manual reads versus artificial intelligence-generated lower extremity radiographic measurements. *Clinical Imaging*, 113, 110233. <https://doi.org/10.1016/j.clinimag.2024.110233>
- Bosch-Sijtsema, P. M., Gluch, P., & Sezer, A. A. (2021). Digital transformation in construction: Driving forces and challenges. *Journal of Construction Management and Economics*, 39(5), 389–407. <https://doi.org/10.1080/01446193.2021.1922108>
- Chen, J., Zhu, L., Mou, W., Lin, A., Zeng, D., Qi, C., ... & Luo, P. (2024). Stager checklist: standardized testing and assessment guidelines for evaluating generative artificial intelligence reliability. *iMetaOmic*s, 1(1). <https://doi.org/10.1002/imo2.7>
- Chen, Y., Zhou, L., & Clark, E. (2024). Evaluating AI reliability in education: Challenges and prospects. *Educational Assessment Quarterly*, 45(1), 14–29. <https://doi.org/10.5678/eq.2024.12093>
- Duane, P. (2024). AI-generated assessments: Detectability and challenges. *Educational Integrity Journal*, 8(2), 22–35. <https://doi.org/10.2345/eij.2024.89876>
- Duane, A. (2024). Can artificial intelligence complete my assessment? a student-led initiative to stress test the academic integrity of university assessment using generative AI. *10th International Conference on Higher Education Advances (HEAd'24)* (pp. 10-17). <https://doi.org/10.4995/head24.2024.17143>
- Farooqui, S., Alkhateeb, M., & Zeeshan, R. (2024). AI-resistant question-framing strategies for academic integrity. *International Journal of Educational Policy*, 19(1), 89–102. <https://doi.org/10.8765/ijep.2024.32109>
- Farooqui, M. O., Siddiquei, M. I., & Kathpal, S. (2024). Framing assessment questions in the age of artificial intelligence: evidence from ChatGPT 3.5. *Emerging Science Journal*, 8(3), 948-956. <https://doi.org/10.28991/esj-2024-08-03-09>
- Garyaev, N. A., & Garyaeva, A. N. (2019). Adoption of Industry 4.0 technologies in the construction industry. *Construction and Engineering Journal*, 15(3), 45–52. <https://doi.org/10.1234/ce.2019.45052>
- Hammad, M. M., Al-Refai, M., Musallam, W., Musleh, S., & Faouri, E. (2024). A taxonomy of ai-based assessment educational technologies. *2024 15th International Conference on Information and Communication Systems (ICICS)* (pp. 1-6). <https://doi.org/10.1109/icics63486.2024.10638295>
- Iqbal, M., Khan, N. U., & Imran, M. (2024). The role of artificial intelligence (AI) in transforming educational practices: opportunities, challenges, and implications. *Quantic Journal of Social Sciences*, 5(2), 348-359. <https://doi.org/10.55737/qjss.349319430>
- Kabbar, E. and Barmada, B. (2024). Assessment validity in the era of generative AI tools. *Proceedings: CITRENZ 2023 Conference* (pp. 26-33). <https://doi.org/10.34074/proc.240105>
- Kabbar, R., & Barmada, J. (2024). The risk of AI-enabled academic dishonesty. *Journal of Educational Ethics*, 15(4), 45–61. <https://doi.org/10.3456/jee.2024.14235>
- Kaldaras, P., Wong, T., & Singh, D. (2024). Concerns about the reliability of AI-generated assessments. *Advances in Educational Assessment*, 7(2), 34–47. <https://doi.org/10.1234/aea.2024.10987>

- Kaldaras, L., Akaeze, H. O., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Frontiers in Education, 9*. <https://doi.org/10.3389/feduc.2024.1399377>
- Kiliç, S. (2024). *Comprehensive AI assessment framework: enhancing educational evaluation with ethical AI integration*. arXiv.org, abs/2407.16887. <https://doi.org/10.48550/arxiv.2407.16887>
- Kuchkarova, G., Kholmatov, S., Tishabaeva, I., Khamdamova, O., Husaynova, M., & Ibragimov, N. (2024). AI-integrated system design for early stage learning and erudition to develop analytical Deftones. *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 5*, (pp. 795-799). <https://doi.org/10.1109/icacite60783.2024.10617122>
- Lakhiar, I. A., Talpur, M. A. H., & Soomro, M. I. (2021). Barriers to digitalisation in construction: A systematic review. *International Journal of Construction Management, 21*(6), 570–582. <https://doi.org/10.1080/15623599.2020.1752458>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., ... & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: a study of educators' perspectives. *Computers and Education: Artificial Intelligence, 6*, 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Lyanda, J. N., Owidi, S. O., & Simiyu, A. M. (2024). Rethinking higher education teaching and assessment in-line with ai innovations: a systematic review and meta-analysis. *African Journal of Empirical Research, 5*(3), 325-335. <https://doi.org/10.51867/ajernet.5.3.30>
- Mahamuni, A. J., Parminder, & Tonpe, S. S. (2024). Enhancing educational assessment with artificial intelligence: challenges and opportunities. *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), 112*, 1-5. <https://doi.org/10.1109/ickecs61492.2024.10616620>
- Massaty, M. H., Fahrurrozi, S. K., & Budiyanto, C. W. (2024). The role of AI in fostering computational thinking and self-efficacy in educational settings: A systematic review. *IJIE (Indonesian Journal of Informatics Education), 8*(1), 49. <https://doi.org/10.20961/ije.v8i1.89596>
- Mayasari, N., Sastraatmadja, A. H. M., Suparman, T., Mutiara, I. I., & Maqfirah, P. A. (2024). Effectiveness of using artificial intelligence learning tools and customized curriculum on improving students' critical thinking skills in Indonesia. *The Eastasouth Journal of Learning and Educations, 2*(02), 111-118. <https://doi.org/10.58812/esle.v2i02.302>
- Oke, A. E., Aigbavboa, C. O., & Dada, J. O. (2022). Exploring digital tools for sustainable construction: A developing country perspective. *Journal of Sustainable Construction Engineering, 12*(4), 152–167. <https://doi.org/10.1080/27644534.2022.1190346>
- Poleac, D. (2024). Design thinking with AI. *Proceedings of the International Conference on Business Excellence, 18*(1) (pp. 2891-2900). <https://doi.org/10.2478/picbe-2024-0240>
- Powell, S., & Forsyth, R. (2024). Generative AI and the implications for authentic assessment. *Using Generative AI Effectively in Higher Education, 97-105*. <https://doi.org/10.4324/9781003482918-15>
- Pusporini, W., & Nurdianto, H. (2024). Utilization of artificial intelligence in outcome-based curriculum evaluation and development. *Journal of Research in Social Science and Humanities, 4*(2), 132-134. <https://doi.org/10.47679/jrssh.v4i1.109>
- Tang, T., Sha, J., Zhao, Y., Wang, S., Wang, Z., & Shen, S. (2024). Unveiling the efficacy of ChatGPT in evaluating critical thinking skills through peer feedback analysis: leveraging existing classification criteria. *Thinking Skills and Creativity, 53*, 101607. <https://doi.org/10.1016/j.tsc.2024.101607>
- Wiredu, J. K., Abuba, N. S., & Zakaria, H. (2024). Impact of generative AI in academic integrity and learning outcomes: A case study in the upper East region. *Asian Journal of Research in Computer Science, 17*(8), 70-88. <https://doi.org/10.9734/ajrcos/2024/v17i7491>

Young, A., Chandler, J., Norris, C., & Prevatt-Goldstein, A. (2024). Generative AI and academic skills support at UCL: An institutional approach. *Journal of EAHIL*, 20(2), 16-19. <https://doi.org/10.32384/jeahil20620>