Education Research and Perspectives
An international journal

Volume 51, 2024, Pages 124-155

Generative AI and Educational Assessments: A Systematic Review

Jian Zhao [1†]

Elaine Chapman[1]

Peyman G. P. Sabet[1,2]

1. The University of Western Australia

2. Global Curtin, Curtin University

The launch of ChatGPT and the rapid proliferation of generative AI (GenAI) have brought transformative changes to education, particularly in the field of assessment. This has prompted a fundamental rethinking of traditional assessment practices, presenting both opportunities and challenges in evaluating student learning. While numerous studies have examined the use of GenAI in assessment, no systematic review has been conducted to synthesise the existing empirical evidence on this topic. Systematically reviewing 19 empirical studies published within 10 years, starting in 2014, this study assessed the current state of empirical evidence regarding GenAI in educational assessment practices and the future research directions required to advance this field. The findings were synthesised into four themes: (1) Educators' perceptions of GenAI in assessment practices; (2) Students' perceptions of GenAI in assessment practices; (3) Effectiveness of applying GenAI in assessment practices; and (4) Recommendations for leveraging GenAI in future assessment practices. The first three themes summarise the current empirical evidence, while the fourth theme identifies priorities for future research to guide the effective integration of GenAI into assessment practices.

[†]Address for correspondence: Graduate School of Education, The University of Western Australia, 35 Stirling Highway, Crawley, 6009, Australia. Email: jian.zhao@uwa.edu.au

Introduction

The launch of ChatGPT and the rapid proliferation of generative artificial intelligence (GenAI) has brought both transformative opportunities and unprecedented challenges to education. Research has demonstrated numerous opportunities and benefits that GenAI has brought to education, offering advantages to both educators and students. For educators, GenAI can greatly reduce their workload by supporting the effective design of lesson plans and teaching materials that well aligned with instructional outcomes (Kasneci et al., 2023; Liu, 2024; Moundridou et al., 2024). It can streamline grading processes, maintaining consistency in scoring and providing immediate and personalised feedback to students' writing (Chan & Hu, 2023).

For students, GenAI can support their learning by providing personalised and dynamic assistance, streamlining complex tasks, and fostering active engagement (Kasneci et al., 2023; Liu, 2024; Walter, 2024). For instance, GenAI can aid students' comprehension by summarising and simplifying complex materials or providing illustrative examples and foster students' problem-solving skills through step-by-step guidance for tackling intricate tasks or challenges (Kasneci et al., 2023; Liu, 2024).

Tools such as ChatGPT offer personalised learning experiences by adapting to individual students' needs, strengths, and weaknesses, enabling them to follow tailored learning paths at their own paces and preferences (Walter, 2024). Moreover, GenAI promotes engagement by creating dynamic, interactive learning environments that make learning enjoyable and impactful (Walter, 2024). Lately, language learning apps have been increasingly integrated with GenAI to simulate authentic interactions in target languages, significantly enhancing learners' language development and proficiency (Creely, 2024).

Despite the opportunities and benefits, educators and researchers have also voiced significant concerns regarding the impact of GenAI on education. These concerns primarily centre around issues such as academic misconduct (Rasul et al., 2024) and the risk of hindering students' intellectual growth and problem-solving skills (Michel-

Villarreal et al., 2023). For instance, Črček and Patekar (2023) found that of the 201 university students in Croatia, 44.7% reported using ChatGPT for university assignments. Over half of these students (55.2%) acknowledged using the tool for generating ideas, paraphrasing, summarising, and proofreading, with a concerning 18% admitting using it to write entire assignments.

Similarly, Gruenhagen et al. (2024) surveyed 337 Australian university students and revealed that over a third of them had used a chatbot for assistance with assessments, often without perceiving this as a breach of academic integrity. Adding to this, Lim et al. (2023) provided convincing evidence of GenAI in facilitating academic misconduct, reporting a substantial self-plagiarism rate of 59%. The issue is further exacerbated by the current lack of reliable AI detection tools (Elkhatat et al., 2023).

In addition to concerns about academic integrity, over-reliance on GenAI poses risks to students' intellectual growth and the development of cognitive skills (Çela et al., 2024; Zhai et al., 2024). Zhai et al. (2024) conducted a systematic review of 14 studies to examine how over-reliance on AI dialogue systems impacts educational and research contexts, with a particular focus on critical cognitive skills such as decision-making, critical thinking, and analytical reasoning. While these systems enhance academic writing and research efficiency, they frequently undermine originality, creativity, and independent critical thinking. Excessive dependence on AI tools has been linked to diminished problem-solving capabilities and increased reliance on AI-generated content.

Similarly, in an empirical study, Çela et al. (2024) surveyed 53 students from a private university in Albania to understand their experiences with AI tools and their effects on cognitive development. The study found a significant negative correlation between reliance on AI for assignments and students' problem-solving abilities, indicating that excessive dependence on AI can impede independent cognitive processes and hinder intellectual growth.

Among GenAI's most profound impacts on education is its potential to reshape assessment practices. Assessment plays a central role in

education, serving as both a measure of learning outcomes and a guide for instructional design (Fuentealba, 2011). While traditional assessment methods, characterised by standardised tests such as multiple choice questions, True or False, matching, short-answer questions and essays, have been proven effective in certain contexts (ALsabbah et al., 2022; Quansah, 2018), they often fail to address the diverse needs of learners or deliver timely and actionable feedback. The integration of GenAI in education has heralded a new era of adaptive assessments, offering a revolutionary shift from traditional assessment methods.

While concerns about the misuse of GenAI for plagiarism persist (Liu, 2024) and integrating GenAI into educational assessment raises critical challenges related to validity, fairness, ethical use, and equity (Chaudhry et al., 2023; Foung et al., 2024; Liu, 2024; Tobler, 2024), GenAI offers promising potential by automating grading (Tobler, 2024), supporting adaptive assessments (Bsharat & Khlaif, 2024), and enabling authentic, real-world tasks that align with modern educational goals (Salinas-Navarro et al., 2024).

Aim of the Study

As educational institutions and educators navigate the evolving landscape of GenAI in education, synthesising existing empirical findings is crucial to understanding its opportunities, challenges and implications for assessment practices. Recent studies have increasingly explored the perceptions of educators and students regarding GenAI tools, such as ChatGPT, and their effectiveness and efficiency in educational assessments. However, to the best of the authors' understanding, no comprehensive synthesis of these findings has been conducted.

This systematic review aims to address this gap by investigating the empirical evidence on GenAI in educational assessment practices, focusing on two key research questions: (1) What is the current state of knowledge about GenAI in educational assessment practices? (2) What future research directions are needed to advance this field? By synthesising existing research, this review aims to provide educators, policymakers, and researchers with a comprehensive understanding of the current evidence base and practical implications of integrating GenAI

into educational assessment. It seeks to inform strategies for leveraging GenAI effectively to enhance assessment practices.

Search Methods

Search Strategy

A systematic literature search was conducted in mid-August 2024, encompassing articles published up to date in three electronic databases - Google Scholar, Web of Science and Scopus. Keywords and search terms ("assessment" OR "evaluation") AND ("learning" OR "education") AND ("generative AI" OR "generative artificial intelligence") were used to search across the three databases to identify relevant studies published within the past 10 years (2014–2024). While the search spanned a decade, the majority of identified studies were published in 2023 and 2024, reflecting the rapid advancements in this research area during the last two years. A total of 2,084 articles were retrieved and Table 1 shows the number of articles retrieved from the three databases.

Table 1. Number of articles retrieved from the three databases.

Databases	Number of articles identified
Web of Science	146
Scopus	1,892
Google Scholar	46
Total	2,084

Article Selection

The process of selecting studies for this review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009), and a flow chart detailing the steps taken for article selection can be seen in Figure 1.

As a result, 19 studies were included in this review.

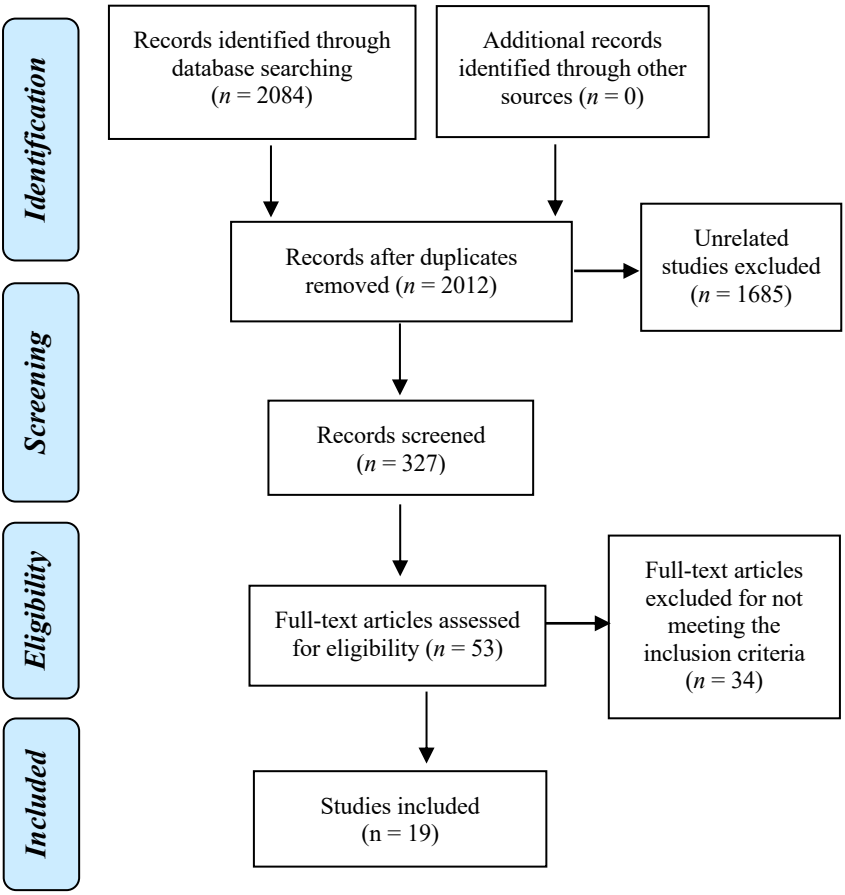

Figure 1-Flow chart of selecting papers (Moher et al., 2009, p. 267)

Articles were included in the review only if they met all of the following criteria: (1) Full-text articles available for review; (2) Written in English; (3) Published in peer-reviewed journals; (4) Focus specifically on GenAI and educational assessment; (5) Empirical studies with supporting data; and (6) Long-form articles featuring substantial analysis or discussion. Exclusion criteria were as follows: (1) Lack of full-text access; (2) Written in languages other than English; (3) Published outside of peer-

reviewed journals, such as conference proceedings and book chapters; (4) Focused on assessments unrelated to education, such as health or medical assessments; (5) Non-empirical studies, including reviews, conceptual papers, protocols, commentaries or theoretical discussions without supporting data; and (6) Short-form articles with limited depth or without substantive analysis.

Main Findings

The 19 studies included in this review have been summarised in Table 2. These empirical studies exhibit diverse characteristics in terms of regions, disciplines and research methodologies, reflecting the global and interdisciplinary nature of the exploration of GenAI in assessment practices. Over 80% of studies focus on higher education (Bernal, 2024; Bower et al., 2024; Chaudhry et al., 2023; Farazouli et al., 2024; Farooqui et al., 2024; Foung et al., 2024; Gruenhagen et al., 2024; Jukiewicz, 2024; Kizilcec et al., 2024; Liu, 2024; Mizumoto, 2023; Ogunleye et al., 2024; Panthier & Gatinel, 2023; Salinas-Navarro et al., 2024; Shahid et al., 2024; Tobler, 2024), with only three (16%) focusing on school education (Ali et al., 2023; Kerneža & Zemljak, 2023; Tang et al., 2024). Methodologically, they encompass qualitative, quantitative, and mixed methods approaches, targeting on fields such as language education (Foung et al., 2024; Liu, 2024; Mizumoto, 2023), science, technology, engineering, and mathematics (STEM) (Ali et al., 2023; Jukiewicz, 2024; Kerneža & Zemljak, 2023; Ogunleye et al., 2024; Tobler, 2024), social science and management (Farooqui et al., 2024; Shahid et al., 2024) and medical education (Panthier & Gatinel, 2023).

Table 2 – Characteristics of included articles

No	Author(s) (year)	Region	Discipline	Level of Education	Summary of Main Findings
1	Liu (2024)	China	Language	Higher Education	Teachers perceived GenAI as both a threat to assessment validity and a tool with potential benefits. Most reported inadequate institutional guidance and used personal strategies like redesigning tasks and promoting ethical AI use. Challenges included difficulty detecting plagiarism and the reliability of detection tools. Recommendations included clearer policies and professional development for leveraging AI effectively.
2	Farazouli et al. (2024)	Sweden	Multiple disciplines	Higher Education	Teachers assigned passing grades to AI-generated responses in 37.5% to 85.7% of cases. Teachers were generally more critical of student-written texts, suspecting some of being AI-generated. ChatGPT responses were noted for their high linguistic quality but lacked depth and engagement with course material. The study highlights ChatGPT's mediating role in amplifying teachers' suspicion and altering assessment practices.
3	Bower et al. (2024)	Global (Various regions)	Multiple disciplines	Higher Education	Teachers perceive GenAI as significantly impacting teaching and assessment, proposing curriculum changes (e.g., teaching AI use and critical thinking) and shifts to more personalised, ethical, and higher order thinking assessments. Motivations include performance expectancy for students and self-improvement. Awareness of AI correlates with perception of its impact.
4	Salinas-Navarro et al. (2024)	United Kingdom, Mexico	Did not specify	Higher Education	GenAI tools can transform teaching and learning by enhancing intended learning outcomes (ILOs), teaching and learning activities (TLAs), and assessment tasks (ATs) through constructive alignment. The study calls for creating AI-enhanced, human-centered learning experiences that support critical thinking, experiential learning, and authentic assessment.

5	Tobler (2024)	Switzerland	Natural science or Technology	Higher Education	The study validated the tool with a high agreement between manual and AI-based grading (Krippendorff's $\alpha = 0.818$). The tool automates grading using predefined questions, sample solutions, and evaluation instructions, providing reliable and customisable assessments. Limitations include challenges with complex questions and potential ethical concerns in AI usage.
6	Tang et al. (2024)	China	Not applicable	School Education (7th-grade level)	The study demonstrated that GPT-4 significantly outperformed GPT-3.5 and Claude 2 in scoring reliability. Well-crafted prompts enhanced model alignment with human raters, with criteria-based prompts achieving the highest agreement. Lower temperature settings produced more consistent outputs. GPT-4 achieved notable accuracy in evaluating Ideas (QWK=0.551) and Organization (QWK=0.584) but struggled with Style and Conventions.
7	Shahid et al. (2024)	Malaysia	Social science & management	Higher Education	Anxiety negatively impacts adoption readiness and attitudes toward AI-based assessment, while resistance to change has negligible effects. Adoption readiness mitigates the negative effects of anxiety on attitude but does not mediate the effects of resistance to change. Teachers with greater anxiety are less likely to adopt AI for assessment purposes, highlighting the importance of supportive measures and training to ease adoption.
8	Ogunleye et al. (2024)	United Kingdom	STEM-related disciplines	Higher Education	GenAI tools demonstrated subject knowledge, problem-solving, analytical, critical thinking, and presentation skills. ChatGPT showed strengths in critical analysis and presentation, while Bard performed better in technical implementations and referencing. Both tools struggled with complex problem-solving in construction management tasks. Findings highlight the necessity of redesigning assessments to address AI-generated solutions and incorporating ethical AI use into pedagogy.
9	Mizumoto et al. (2024)	Japan, Macau	Language	Higher Education	ChatGPT exhibited strong alignment with human raters for error detection ($\rho = 0.79$) and writing scores ($\rho = -0.63$), outperforming Grammarly in both areas. ChatGPT is promising for L2 writing assessments due to its accuracy and predictive validity but requires further validation across diverse

					datasets. Differences in error definitions across evaluators were noted as a limitation.
10	Kizilcec et al. (2024)	Australia, Cyprus, United States	Multiple disciplines	Higher Education	The study found that essay and coding assessments are perceived as most impacted by GenAI. Educators strongly preferred adapting assessments to incorporate GenAI for critical thinking, while students expressed mixed feelings, citing creativity concerns. The findings emphasised the importance of involving both groups in assessment reform, focusing on higher-order thinking and authentic learning tasks.
11	Jukiewicz (2024)	Poland	Science/ Programming	Higher Education	ChatGPT demonstrated strong alignment with teacher evaluations ($r = 0.81$) but was stricter in grading, with slightly lower scores. It excelled in repeatability (ICC ≈ 0.95) and provided meaningful, objective feedback. However, limitations include occasional hallucinations, cost, and the need for teacher intervention to address errors. The study suggests ChatGPT as a complementary tool for efficient and unbiased grading.
12	Gruenhagen et al. (2024)	Australia	Did not specify	Higher Education	More than a third of students have used AI chatbots like ChatGPT for assessments, often not perceiving this as a breach of academic integrity. Higher psychological resilience was associated with lower chatbot usage. Students predominantly used AI tools to find information or assist in analysis. Ethical concerns and mistrust of chatbot-generated information were highlighted, along with the potential for AI tools to support neurodiverse and non-native English-speaking students. Recommendations include redefining assessment practices and creating clear policies around AI use in education.
13	Foung et al. (2024)	Hong Kong, China	Language	Higher Education	Integrating traditional AI and GenAI tools (e.g., ChatGPT, Grammarly, and WeCheck!) in assessments encouraged students to critically evaluate tool affordances, promoting AI literacy and improved writing skills. Students leveraged tools differently across writing stages, such as brainstorming, grammar checking, and stylistic refinement. Equity issues regarding access to premium versions of tools were highlighted.

14	Farooqui et al. (2024)	United Arab Emirates	Management education	Higher Education	Experiential learning, recent events, and decision-making questions were found to be less susceptible to AI-generated cheating. ChatGPT performed better with subjective opinion-based questions but struggled with questions requiring physical perception or real-time data. Recommendations include designing AI-proof assessments and promoting ethical AI use among students.
15	Bernal (2024)	United States	Educational Technology, eLearning	Higher Education	GPT-4 has potential to enhance eLearning through dynamic and interactive content, including custom MCQs and instant feedback for coding exercises. It improved content relevance and adaptability, addressing individual learning needs. The Learnix platform demonstrated scalability and versatility across different educational contexts. Challenges include managing AI variability and ensuring ethical considerations.
16	Panthier & Gatinel (2023)	France	Medical Education	Higher Education	ChatGPT achieved a 91.2% success rate, demonstrating strong understanding across all question categories, with rapid response times compared to humans. Limitations include inability to interpret images, recent knowledge gaps (post-2021), and occasional errors with ambiguous or poorly worded questions. The study highlights AI's potential in medical education but emphasises it as a supplementary tool rather than a replacement for human expertise.
17	Kerneža & Zemljak (2023)	Slovenia	STEM	Primary/ Secondary	The study found varying assessment practices across subjects during remote teaching. Science teachers used oral assessments and authentic tasks less frequently than social science and vocational teachers. Teachers highlighted a lack of preparedness for AI-based assessment, emphasising the need for professional development in digital literacy and AI integration. Recommendations include fostering reading literacy skills as a foundation for AI-driven evaluations.
18	Chaudhry et al. (2023)	United Arab Emirates	Did not specify	Higher Education	The study found that ChatGPT consistently produced coherent, critical, and grammatically accurate responses, often outperforming top students' submissions. However, current plagiarism detection tools failed to identify AI-generated work as lacking integrity. The study emphasises revising

					performance evaluation models to reflect genuine student skills and integrating ethical AI usage policies.
19	Ali et al. (2023)	Singapore	Science	K-12 education	TeacherGAIA utilises GPT-4 with prompt engineering to implement four learning approaches: knowledge construction, inquiry-based learning, self-assessment, and peer teaching. The study found high fidelity to learning goals, cognitive guidance, and social-emotional support. Limitations include a lack of multimodal interactions and challenges in factual accuracy. Future development includes collaborative features and media-rich tools.

Four key themes emerged from the analysis of 19 studies on GenAI and assessment practices. These themes provide valuable insights into educators' and students' perceptions of GenAI's transformative impact on assessment methods, offer empirical evidence of its application in assessment practices, and underscore critical considerations alongside potential directions for future advancements in this field.

Theme 1: Educators' Perceptions of GenAI in Assessment Practices

Educators' perceptions of GenAI in assessment practices are pivotal in shaping its implementation in assessment and determining the extent of its integration into assessment. Educators recognised the transformative change GenAI would bring to assessment practices. Some expressed optimism, envisioning significant improvements in assessment efficiency and innovation, while others adopted a more cautious stance, raising concerns about its reliability, ethical implications, and the challenges it may pose to academic integrity.

Transformative impact and opportunities GenAI bring to assessment practices

Educators have recognised the transformative impact of GenAI tools such as ChatGPT on assessment practices. A recent survey by Bower et al. (2024) involving 318 university educators across various teaching levels, disciplines, and regions revealed that nearly two-thirds of respondents anticipated GenAI would have a significant or profound impact on assessments. This recognition is accompanied by a shared understanding of the need to adapt assessment practices in response to GenAI, motivated by a desire to enhance outcomes for both students and educators.

Educators expressed optimism about GenAI's ability to improve the efficiency of assessment (Shahid et al., 2024). They emphasised the importance of integrating GenAI into assessments to increase authenticity and engagement, aligning evaluations more closely with real-world applications (Salinas-Navarro et al., 2024). This forward-thinking approach underscores the growing recognition of

GenAI's potential to redefine traditional assessment paradigms while equipping students with the skills needed to thrive in an AI-enabled world.

In a multidisciplinary study, Kizilcec et al. (2024) found that educators strongly advocated for designing assessments that assume the use of AI. They viewed GenAI as a powerful tool for fostering higher-order thinking and critical evaluation skills, particularly in tasks such as essay writing and coding. By integrating GenAI into assessment practices, educators aimed to create more relevant and engaging evaluations that prepare students for the challenges and opportunities of an AI-enabled future.

Concerns regarding students' misuse of GenAI in assessment practices

The rise of GenAI has sparked significant concerns among educators about their potential misuse in assessments. Chaudhry et al. (2023) found that educators were particularly worried about how tools like ChatGPT might compromise academic integrity. Although AI-generated assignments can meet academic criteria, they raise ethical questions about authorship and originality. Compounding the issue is the limited reliability of current AI detection tools, which leaves educators with inadequate means to identify AI-assisted work and enforce integrity policies effectively (Chaudhry et al., 2023).

Similarly, Kizilcec et al. (2024) found that educators struggled to differentiate between student-authored and AI-generated content. This difficulty undermines their ability to accurately assess students' understanding and application of learned material, further intensifying concerns about the fairness and validity of assessment practices in an AI-driven academic landscape.

Liu (2024) echoed these concerns, noting that educators identified GenAI as a facilitator of plagiarism and a threat to the validity of assessments. Bower et al. (2024) also reported widespread apprehension among educators, who fear about GenAI's potential

to enable academic dishonesty and disrupt the integrity of traditional assessment practices.

Lack of institutional guidance, instructions and training for educators and students

Institutional guidance on the use of GenAI in assessment and relevant training remains insufficient, with discipline-specific challenges complicating these issues. Research has highlighted inadequate institutional policies in this area. Liu (2024) reported that among the ten universities represented by the 17 teachers interviewed, only one had issued guidelines on GenAI use. However, these guidelines were perceived as vague and inadequate, leaving educators uncertain about best practices for implementation.

The absence of training and professional development has further exacerbated the challenge. Shahid et al. (2024) found that university educators' anxiety about using AI systems significantly reduced their readiness to adopt AI-driven assessment tools, underscoring the critical need for structured training programs.

Similarly, Kerneža and Zemljak (2023) identified disparities in preparedness across disciplines. For example, science teachers were less equipped to implement AI-based assessments and relied more on traditional tasks and oral evaluations compared to their counterparts in the social sciences. This variation highlights the necessity of tailored training programs that address the unique needs of different disciplines while building educators' confidence in using GenAI effectively.

At the school level, Kerneža and Zemljak (2023) surveyed 1,215 primary and secondary teachers in Slovenia to explore how educators adapted assessment methods during emergency remote teaching and their preparedness for AI-driven assessments. While teachers recognised the potential of GenAI to enhance assessments, they expressed caution and a lack of readiness to fully integrate these tools into their practices.

Theme 2: Students' Perceptions of GenAI in Assessment Practices

Students' perceptions of generative GenAI in assessment are also important, as they influence how these technologies are accepted, utilised, and integrated into their learning experiences. Students share educators' goal of enhancing learning outcomes (Gruenhagen et al., 2024), but their unique perspectives also shape the role that GenAI plays in assessments.

Perceiving GenAI as a useful tool in assessments with regional disparities

Students generally view GenAI as a useful tool for improving learning and completing assessments (Alabidi et al., 2023; Gruenhagen et al., 2024). Gruenhagen et al. (2024) reported that 36.5% of students had used Chatpot for assessment-related tasks, such as information retrieval or analysing a topic or issue. Similarly, Alabidi et al. (2023) found that students valued ChatGPT's ability to provide instant, tailored feedback on their assignments, which helped them identify and address knowledge gaps. Many students appreciated how GenAI fosters deeper engagement through dynamic, context-relevant interactions that are often more stimulating than traditional paper-based exams. The gamification aspects of some GenAI platforms, such as instant scoring and visual feedback, further motivate students to participate actively in their learning.

Students from other research have also been found to be aware of GenAI and use it for coursework and personal purposes, although this varies by region. For instance, Kizilcec et al. (2024) conducted an international survey of 680 students and 87 educators across Australia, Cyprus, and the United States, revealing notable variations in awareness and usage of GenAI among students. While students in Australia and the US demonstrated high awareness and regular usage of ChatGPT for tasks such as coursework, research, professional activities, and recreation, students in Cyprus exhibited significantly lower levels of awareness and engagement with the technology. These findings illustrate the growing role of GenAI in education while highlighting regional disparities in its adoption.

Mixed perceptions of academic integrity and ethical implications
Students' views on the ethical implications of using GenAI in assessments are varied. Some students did not perceive the use of GenAI as a breach of academic integrity (Gruenhagen et al., 2024), while others raised concerns about its potential misuse (Chaudhry et al., 2023; Kizilcec et al., 2024). Gruenhagen et al. (2024) found that among students who used GenAI for assessments, some employed it to assist with parts of their assignments, including writing sections or solving multiple-choice quizzes. This presents challenges for maintaining academic integrity and mitigating plagiarism risks.

Other students expressed concerns about over-reliance on AI tools, which they believed could diminish opportunities to showcase creativity and critical thinking (Kizilcec et al., 2024). Some questioned the reliability of GenAI, expressing scepticism about its ability to accurately assess their knowledge and skills. Concerns about biases in AI algorithms and the risk of receiving generic or incorrect feedback were noted by Kerneža and Zemljak (2023), who found that students were cautious about trusting AI-generated evaluations.

In addition, some students raised issues about AI's potential to misinterpret nuanced or creative responses, leading to inaccurate evaluations. These students advocated for using GenAI as a supplement to traditional assessments rather than a full replacement (Chaudhry et al., 2023).

Equity considerations
Equity emerged as a significant concern, as students may face unequal access to premium GenAI tools (Chaudhry et al., 2023; Foung et al., 2024). Students expressed concerns about equity in the use of GenAI tools in assessments as disparities between free and premium versions of tools like ChatGPT or Grammarly create a digital divide (Foung et al., 2024). While free versions are accessible to most, premium versions offer advanced features, such as detailed writing suggestions, which are unavailable to students from low-income backgrounds. This inequity impacts students'

ability to compete fairly with peers who can afford these tools. Chaudhry et al. (2023) also noted that disparities in access to GenAI tools shape students' perceptions of fairness and influence their academic experiences. Addressing this issue requires institutions to consider policies that ensure equitable access to these technologies, thereby levelling the playing field for all students.

Theme 3: Evidence of Using GenAI in Assessment Practices

A growing body of research has investigated the effectiveness of GenAI in enhancing assessment practices through improved design and efficiency (Bernal, 2024), grading and reliability (Tang et al., 2024; Tobler, 2024), interactivity (Bernal, 2024), as well as personalisation and diversity (Anggoro & Pratiwi, 2023; Bernal, 2024).

Evidence of integrating GenAI in assessment design

Some educational researchers have begun leveraging GenAI for assessment design or taking GenAI into their assessment design. Bernal (2024) evaluated the integration of GPT-4 into the e-learning platform *Learnix*, which dynamically generates multiple-choice questions (MCQs) and provides instant feedback for coding exercises. The study demonstrated GPT-4's ability to create interactive and personalised learning experiences, including adaptive MCQs and real-time feedback tailored to individual learning needs. The platform's flexible design also supports broader applications across various disciplines, offering a scalable framework for integrating AI-driven assessments in diverse educational contexts.

Similarly, Ali et al. (2023) developed TeacherGAIA, a GenAI-powered chatbot prototype designed to promote self-directed learning and self-assessment in K-12 education. TeacherGAIA provided students with rubrics, checklists, and self-reflection prompts, along with tailored feedback to refine their understanding and skills. The tool also incorporated social-emotional support, offering encouragement and empathy to build student confidence, although challenges such as maintaining factual accuracy and mitigating over-reliance on AI-generated feedback were noted.

Other research has redesigned assessments to encourage students to use both traditional AI and GenAI tools (Foung et al., 2024). Students reported using GenAI for tasks such as planning, brainstorming, summarising, and improving writing styles while remaining aware of its limitations, including inaccuracies, biases, and problematic citations. This dual approach fostered critical awareness of AI's strengths and weaknesses in academic settings.

Evidence of GenAI's effectiveness in grading, alongside identified limitations

Research has provided evidence of using GenAI in grading across multiple disciplines. Tobler (2024) developed an automatic grading tool that uses predefined questions, sample solutions, and evaluation criteria to provide consistent and customisable assessments. The tool showed high agreement with manual evaluations, although challenges with complex questions remain.

Tang et al. (2024) evaluated large language models (LLMs) such as GPT-4 and Claude 2 for automated essay scoring (AES). Their study found GPT-4 particularly effective in evaluating ideas and organisation, though challenges persisted in assessing style and conventions. In second language (L2) writing, Mizumoto et al. (2023) compared ChatGPT with human raters and Grammarly, finding that ChatGPT aligned strongly with human evaluations for error detection (ρ = 0.79) and writing scores (ρ = -0.63), outperforming Grammarly.

Jukiewicz (2024) examined ChatGPT's capabilities in grading programming assignments in science and programming courses. The tool aligned closely with teacher evaluations (r = 0.81) and provided high-quality, objective feedback. However, its strict grading, occasional hallucinations, and reliance on teacher oversight for error correction were noted as limitations.

In medical education, Panthier and Gatinel (2023) tested GPT-4's performance on the French-language European Board of Ophthalmology (EBO) examination. GPT-4 achieved a 91.2% success rate, showcasing its ability to handle a broad range of

medical knowledge. Despite its strong performance, limitations included gaps in recent knowledge (post-2021), difficulty interpreting images, and occasional errors with ambiguous questions.

Evidence of challenges and issues in grading assessments

While substantial evidence highlights the effectiveness of GenAI in assessment practices, numerous studies have also revealed challenges in its application to grading students' work. For example, Farazouli et al. (2024) conducted a Turing Test-inspired experiment in which teachers graded texts blindly, unaware of whether they were written by students or generated by ChatGPT. Interestingly, teachers were more critical of student-written texts, often suspecting AI involvement, while ChatGPT-generated responses received higher ratings despite lacking depth. This suggests a potential bias in favour of AI-generated content due to its polished presentation.

Chaudhry et al. (2023) compared ChatGPT's responses to assessment tasks with submissions from top-performing undergraduate students across multiple courses. ChatGPT consistently produced coherent, accurate, and plagiarism-free responses, frequently surpassing students in clarity and critical thinking. However, existing plagiarism detection tools, such as Turnitin, failed to effectively identify AI-generated content, presenting significant challenges in ensuring originality in assessments.

Theme 4: Recommendations for Future Assessment Practices

Guidelines, strategies and professional development for educators to effectively leverage GenAI in assessment practices

The advancement of GenAI and its application in assessment practices present both opportunities and challenges. To maximise its potential while upholding academic standards and integrity, a number of studies have emphasised the need for relevant guidelines, strategies and professional development for educators to equip educators with the tools to address the challenges posed by GenAI

in assessment practices and to optimise its use for enhancing assessment processes.

Farooqui et al. (2024) called for educators to rethink and develop question-framing strategies to safeguard assessment integrity in the age of GenAI. They proposed creating question types resistant to AI-generated answers, such as decision-making tasks or questions involving recent events. Chaudhry et al. (2023) argued that existing performance evaluation approaches were no longer relevant to assess students' learning outcomes, urging institutions, academic regulators and teachers to revisit their strategies. They highlighted the importance of developing comprehensive policies that guide the ethical and effective use of GenAI in education.

Liu (2024) emphasised the need for clearer institutional policies alongside professional development programs to empower educators in using GenAI effectively. This sentiment was echoed by Kerneža and Zemljak (2023) and Shahid et al. (2024), who underscored the critical importance of teacher training and targeted professional development to equip educators with the skills and resources necessary for implementing AI-based assessments. Similarly, Gruenhagen et al. (2024) advocated for redefining assessment practices and establishing clear, actionable policies to guide the ethical and effective use of AI in education.

Training and support for students on how to use GenAI when completing assessment practices

As many students use AI chatbots like ChatGPT for assessments without fully understanding the implications for academic integrity (Gruenhagen et al., 2024), institutions must provide explicit, step-by-step instructions on the ethical use of GenAI. Foung et al. (2024) demonstrated the effectiveness of such training, which included in-class demonstrations to clarify appropriate uses of GenAI. These resources bridged the gap between the potential of GenAI as a learning tool and the necessity of maintaining academic standards.

Training programs should prioritise cultivating a strong understanding of academic integrity and ethical boundaries

surrounding GenAI use. These programs must also equip students with skills to critically evaluate AI-generated outputs for accuracy, relevance, and bias, ensuring responsible incorporation into academic work. Furthermore, assessments integrating GenAI should focus on developing higher-order thinking, creativity, and real-world problem-solving skills, preparing students for professional environments where AI plays an increasingly significant role.

To ensure the success of these guidelines, strategies and training, engaging both educators and students in their development is essential. This collaborative approach fosters ownership and commitment, reducing the likelihood of GenAI misuse and promoting responsible use (Gruenhagen et al., 2024). This not only includes using GenAI to help develop assessments of various types but also redesigning Assessments to Integrate AI Responsibly.

Designing and piloting new assessments to meet students' needs in an AI-driven era

The reviewed studies consistently emphasised the need to redesign assessments to integrate GenAI to measure higher-order thinking and more authentic and real-world practices. With GenAI becoming increasingly prevalent in fields such as software engineering (Ebert & Louridas, 2023), construction industry (Ghimire et al., 2024) and healthcare (Reddy, 2024), exposing students to GenAI tools and hands-on experiences will help equip them with the skills needed to become job-ready graduates (Gruenhagen et al., 2024).

Although suggestions for assessment redesign in the GenAI era are emerging (Charles Sturt University, n.d.; Chen, 2023), empirical evidence remains limited. Foung et al. (2024) conducted a study in a communication course, encouraging students to use both traditional and GenAI tools in their assessments. Their findings showed that, when guided appropriately, GenAI could enhance students' critical thinking skills. However, more research across disciplines, regions, and educational levels is needed to develop evidence-based models for redesigning assessments.

Mohammad et al. (2024) explored assessment redesign strategies to maintain academic integrity and measure learning outcomes effectively in the GenAI era. By testing ChatGPT 3.5 against various question types, they identified formats like moral reasoning tasks, subjective judgments, and prompts requiring up-to-date knowledge or personal experiences as more resistant to AI-generated responses. Nevertheless, advancements such as ChatGPT 4.0 necessitate further research to validate and refine these strategies.

Other research directions not informed by the existing literature
Given the rapid advancements in GenAI over the past two years, the existing literature has yet to explore certain critical areas. Notably, no studies have examined the long-term impact of integrating GenAI into assessment practices on educational outcomes. This presents a valuable opportunity for future research to design longitudinal studies that assess how GenAI influences student learning, critical thinking, and skill development over extended periods.

In addition, the 19 reviewed studies largely overlook the practical challenges educators face when implementing GenAI. These challenges may include various GenAI literacy levels of both educators and students, technical barriers and costs for adequate infrastructure to support GenAI tools effectively. Future research may delve into these practical considerations to provide actionable insights and recommendations for educators and institutions striving to adopt GenAI in assessment contexts.

Discussion and Conclusion

This study systematically reviewed empirical research published before mid-August 2024 on GenAI and educational assessment practices, addressing two research questions (1) What is the current state of knowledge about GenAI in educational assessment practices? (2) What future research directions are needed to advance this field? Findings related to educators' and students' perceptions

of GenAI and empirical evidence of its application addressed the first question, while recommendations for future research directions provided insights into the second.

The review revealed that while most educators and students recognise the transformative potential of GenAI in assessment practices and the effectiveness of various GenAI tools has been empirically validated, significant challenges remain. Concerns about academic integrity, ethical implications, and the limitations of these tools remain prominent. Educators face anxiety and a lack of preparedness in addressing issues related to GenAI and integrating it into assessment practices. Notably, the long-term impact of using GenAI in assessments on students' learning outcomes has yet to be explored. Practical challenges, such as varying levels of GenAI literacy among educators and students, technical barriers, and the cost of infrastructure to support effective implementation of GenAI tools, also require further investigation.

The findings highlight the critical need to develop clear guidelines, frameworks, and strategies to support the ethical and effective integration of GenAI into educational assessments. Transparency in how AI algorithms function and apply assessment criteria is vital to building trust, ensuring fairness, and safeguarding academic integrity. Comprehensive training programs are essential to equip both educators and students with the skills and confidence to navigate GenAI responsibly. Piloting innovative assessment designs that incorporate GenAI in a thoughtful and accountable manner will be crucial for enhancing evaluations of both learning processes and outcomes. Addressing the long-term impacts of GenAI on students' learning outcomes should become a priority, and technical barriers, including infrastructure and accessibility, must also be considered.

Given the ubiquity of GenAI and its proven benefits, eliminating its use in assessment practices is neither practical nor advantageous. Instead, institutions, educators, and researchers should focus on leveraging GenAI responsibly and effectively. By doing so, they can enhance students' learning experiences and outcomes,

minimise misuse, and better prepare students for careers and professional growth in an increasingly AI-driven world.

Authors

Dr. Jian Zhao is an early-career researcher at the Graduate School of Education, the University of Western Australia. Specialising in mental health measurement, assessment design, and mixed-methods research, she has contributed to projects on the mental health of Chinese international students in Australia and the prediction of self-harm, suicidal behaviours among young people in Western Australia. Her key research interests include mental health measurement, assessment design, international student mental health, coping strategies, resilience, and the prediction of suicide and self-harm.

A/Professor Elaine Chapman holds a PhD in Psychology and has over 20 years of teaching experience across three leading universities - Monash University, the University of Sydney, and the University of Western Australia. Her teaching expertise spans child and educational psychology, assessment, quantitative research design, and statistics. She serves as an editor for high-impact journals, including *Frontiers in Psychology* and the *British Journal of Educational Psychology*. She has delivered invited video presentations for SAGE on quantitative methods in education and has provided keynote addresses at conferences centred on measurement and assessment.

Dr. Peyman G.P. Sabet is a Doctor of Education candidate at the University of Western Australia with a focus on educational psychology and the internationalisation of Australian tertiary education. He also holds a PhD in language and intercultural education from Curtin University where he works as a lecturer in TESOL. Peyman has been involved in language pedagogy and linguistics for more than twenty-five years, with a wealth of teaching experience and publications in a number of peer-reviewed journals. Peyman’s research expertise lies in the areas of

Interlanguage Pragmatics, Inter/Cross-cultural Communication, Intercultural Competence, Assessment, Vague Language and second language acquisition.

References

Alabidi, S., Alarabi, K., Alsalhi, N. R., & Mansoori, M. A. (2023). The dawn of ChatGPT: Transformation in science assessment. *Eurasian Journal of Educational Research*, *2023*(106), 321–337. Scopus. https://doi.org/10.14689/ejer.2023.106.019

Ali, F., Choy, D., Divaharan, S., Tay, H. Y., & Chen, W. (2023). Supporting self-directed learning and self-assessment using TeacherGAIA, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, *9*(2), 135–147. Scopus. https://doi.org/10.1080/23735082.2023.2258886

ALsabbah, S., Almomani, J., Amani, D., & Najwan, F. (2022). Traditional versus authentic assessments in higher education. *Pegem Journal of Education and Instruction*, *12*(1), 283–291. https://doi.org/10.47750/pegegog.12.01.29

Anggoro, K., & Pratiwi, D. (2023). Fostering self-assessment in English learning with a generative AI platform: A case of Quizizz AI. *Studies in Self-Access Learning Journal*, *14*(4), 489–501.

Bernal, M. E. (2024). Revolutionizing eLearning assessments: The role of GPT in crafting dynamic content and feedback. *Journal of Artificial Intelligence & Technology*, *4*(3), 188–199. Scopus. https://doi.org/10.37965/jait.2024.0513

Bower, M., Torrington, J., Lai, J., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. *Education & Information Technologies*. https://doi.org/10.1007/s10639-023-12405-0

Bsharat, T., & Khlaif, Z. (2024). *Generative AI-powered adaptive assessment* (p. Pages: 430). https://doi.org/10.4018/979-8-3693-6397-3

Çela, E., Fonkam, M., & Potluri, R. M. (2024). Risks of AI-assisted learning on student critical thinking. *International Journal of Risk & Contingency Management*, *12*, 1–19. https://doi.org/10.4018/IJRCM.350185

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 43. https://doi.org/10.1186/s41239-023-00411-8

Charles Sturt University. (n.d.). *Rethinking assessment strategies in the age of artificial intelligence (AI)*. Retrieved November 29, 2024, from https://www.csu.edu.au/division/learning-teaching/assessments/assessment-and-artificial-intelligence/rethinking-assessments

Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT - A case study. *Cogent Education*, *10*(1). Scopus. https://doi.org/10.1080/2331186X.2023.2210461

Chen, J. (2023, July 21). *Four directions for assessment redesign in the age of generative AI*. THE Campus Learn, Share, Connect. https://www.timeshighereducation.com/campus/four-directions-assessment-redesign-age-generative-ai

Črček, N., & Patekar, J. (2023). Writing with AI: University Students' Use of ChatGPT. *Journal of Language & Education*, *9*, 128–138. https://doi.org/10.17323/jle.2023.17379

Creely, E. (2024). Exploring the role of generative AI in enhancing language learning: Opportunities and challenges. *International Journal of Changes in Education*, *1*. https://doi.org/10.47852/bonviewIJCE42022495

Ebert, C., & Louridas, P. (2023). Generative AI for software practitioners. *IEEE Software*, *40*(4), 30–38. IEEE Software. https://doi.org/10.1109/MS.2023.3265877

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, *19*(1), Article 1. https://doi.org/10.1007/s40979-023-00140-5

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, *49*(3), 363–375. https://doi.org/10.1080/02602938.2023.2241676

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. Scopus. https://doi.org/10.28991/ESJ-2024-08-03-09

Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. *Computers & Education: Artificial Intelligence*, *6*. Scopus. https://doi.org/10.1016/j.caeai.2024.100250

Fuentealba, C. (2011). The role of assessment in the student learning process. *Journal of Veterinary Medical Education*, *38*(2), 157–162. https://doi.org/10.3138/jvme.38.2.157

Ghimire, P., Kim, K., & Acharya, M. (2024). Opportunities and challenges of Generative AI in construction industry: Focusing on adoption of text-based models. *Buildings*, *14*(1). Scopus. https://doi.org/10.3390/buildings14010220

Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers & Education: Artificial*

Intelligence, *7*. Scopus. https://doi.org/10.1016/j.caeai.2024.100273

Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. *Thinking Skills & Creativity*, *52*. Scopus. https://doi.org/10.1016/j.tsc.2024.101522

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning & Individual Differences*, *103*, 102274-. https://doi.org/10.1016/j.lindif.2023.102274

Kerneža, M., & Zemljak, D. (2023). Science teachers' approach to contemporary assessment with a reading literacy emphasis. *Journal of Baltic Science Education*, *22*(5), 851–864. Scopus. https://doi.org/10.33225/jbse/23.22.851

Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers & Education: Artificial Intelligence*, *7*. Scopus. https://doi.org/10.1016/j.caeai.2024.100269

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, *21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Liu, X. (2024). Navigating uncharted waters: Teachers' perceptions of and reactions to AI-induced challenges to assessment. *Asia-Pacific Education Researcher*. https://doi.org/10.1007/s40299-024-00890-x

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*, *13*(9), Article 9. https://doi.org/10.3390/educsci13090856

Mizumoto, A. (2023). Data-driven learning meets Generative AI: Introducing the framework of metacognitive resource use. *Applied Corpus Linguistics*, *3*(3). Scopus. https://doi.org/10.1016/j.acorp.2023.100074

Mohammad-Rahimi, H., Khoury, Z. H., Alamdari, M. I., Rokhshad, R., Motie, P., Parsa, A., Tavares, T., Sciubba, J. J., Price, J. B., & Sultan, A. S. (2024). Performance of AI chatbots on controversial topics in oral medicine, pathology, and radiology. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *137*(5), 508–514. Scopus. https://doi.org/10.1016/j.oooo.2024.01.015

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moundridou, M., Matzakos, N., & Doukakis, S. (2024). Generative AI tools as educators' assistants: Designing and implementing inquiry-based lesson plans. *Computers and Education: Artificial Intelligence*, *7*, 100277. https://doi.org/10.1016/j.caeai.2024.100277

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. *Journal of Applied Learning & Teaching*, *7*(1), 46–56. Scopus. https://doi.org/10.37074/jalt.2024.7.1.28

Panthier, C., & Gatinel, D. (2023). Success of ChatGPT, an AI language model, in taking the French language version of the European Board of Ophthalmology examination: A novel approach to medical knowledge assessment. *Journal Francais d'Ophtalmologie*, *46*(7), 706–711. Scopus. https://doi.org/10.1016/j.jfo.2023.05.006

Quansah, F. (2018). *Traditional or Performance Assessment: What is the Right Way to Assessing Learners? 8*.

Rasul, T., Nair, S., Kalendra, D., Balaji, M. S., Santini, F. de O., Ladeira, W. J., Rather, R. A., Yasin, N., Rodriguez, R. V., Kokkalis, P., Murad, M. W., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI Era:A holistic framework. *The International Journal of Management Education*, *22*(3), 101041. https://doi.org/10.1016/j.ijme.2024.101041

Reddy, S. (2024). Generative AI in healthcare: An implementation science informed translational path on application, integration and governance. *Implementation Science*, *19*(1). Scopus. https://doi.org/10.1186/s13012-024-01357-9

Salinas-Navarro, D., Vilalta-Perdomo, E., Michel-Villarreal, R., & Montesinos, L. (2024). Designing experiential learning activities with generative artificial intelligence tools for authentic assessment. *Interactive Technology & Smart Education*. https://doi.org/10.1108/ITSE-12-2023-0236

Shahid, M. K., Zia, T., Bangfan, L., Iqbal, Z., & Ahmad, F. (2024). Exploring the relationship of psychological factors and adoption readiness in determining university teachers' attitude on AI-based assessment systems. *International Journal of Management Education*, *22*(2). Scopus. https://doi.org/10.1016/j.ijme.2024.100967

Tang, K.-S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning and assessment with Generative AI in secondary education. *Learning & Assessment with Generative AI in Secondary Education (February 11, 2024)*.

Tobler, S. (2024). Smart grading: A generative AI-based tool for knowledge-grounded answer evaluation in educational assessments. *MethodsX*, *12*. Scopus. https://doi.org/10.1016/j.mex.2023.102531

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education.

International Journal of Educational Technology in Higher Education, *21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7