Williams, A. (2025). Integrating artificial intelligence into higher education assessment. *Intersection: A Journal at the Intersection of Assessment and Learning, 6*(1), 128-154.

Integrating Artificial Intelligence into Higher Education Assessment

Andrew Williams, Ph.D.

Author Note

Andrew Williams, orcid.org/0000-0002-0908-0364

"I have no conflicts of interest to disclose."

Intersection: A Journal at the Intersection of Assessment and Learning
Vol. 6, Issue 1, 2025

Abstract: Generative AI has the potential to transform higher education assessment. This study examines the opportunities and challenges of integrating AI into coursework assessments, highlighting the need to rethink traditional paradigms. A case study is presented that explores AI as an auxiliary learning tool in postgraduate coursework. Students found AI valuable for text generation, proofreading, idea generation, and research but noted limitations in accuracy, detail, and specificity. AI integration offers advantages such as enhancing assessment authenticity, promoting self-regulated learning, and developing critical thinking and problem-solving skills. A holistic approach is recommended, incorporating AI into feedback, adapting assessments to leverage AI's capabilities, and promoting AI literacy among students and educators. Embracing AI while addressing its challenges can enable effective, equitable, and engaging assessment and teaching practices. Universities are encouraged to strategically integrate AI into teaching and learning, ultimately transforming the educational landscape to better prepare students for an AI-driven world.

Keywords: *artificial intelligence, higher education, assessment, student, essay*

Introduction

Artificial intelligence (AI) is a rapidly advancing field of study, with its capabilities reported to be doubling every six months (Pichai, 2023). Innovations in AI represent a disruptive technology for our society and for the higher education sector. Generative AI, large language models (LLMs), and natural language processing (NLP) technologies, therefore, have the potential to be transformative for higher education practices (Gamage et al., 2023). Unlike earlier AI systems designed to automate repetitive tasks, generative AI can create new content, provide feedback, and simulate human-like interactions. The capabilities of AI have raised concerns among educators about the impact of AI on academic integrity (Perkins, 2023), but they also offer myriad opportunities to transform teaching practices, personalize approaches to student learning and revolutionize traditional assessment modalities (Rudolph et al., 2023). This paper aims to explore the potential of generative AI to enhance and redefine the way assessments are conducted in higher education by examining how AI can be integrated into assessment design and practice. Opportunities for integrating AI into feedback and assessment practices are discussed, and a case study is presented to illustrate a practical application of this approach. Furthermore, an analysis of the student perspective of using generative AI as an auxiliary tool for their studies and assessments is presented.

AI has now become pervasive throughout modern society, with the technology utilized across multiple business sectors, entertainment, media, healthcare, and education. (Babaniyazovich, 2023; Vijayakumar, 2022). AI competency has rapidly become a desired employability skill for businesses and industries (Pauceanu et al., 2020), including healthcare and the biomedical sciences (Jain et al., 2021). Therefore, it is argued that universities have a responsibility to train and equip students with the technological and ethical competencies for using AI in the workplace (Aoun, 2024). This includes teaching students about the capabilities and limitations of generative AI and allowing students to experiment with AI in a safe and controlled environment. The inclusion of generative AI into contemporary higher education curricula and assessment may therefore be required to fulfil desired learning outcomes (Zouhaier, 2023). Given the transformative impact that AI has had on society, the integration of AI into assessments may satisfy this approach in the context of assessment authenticity and diversification (Swiecki et al., 2022).

As higher education institutions seek to improve the efficiency and effectiveness of their assessments, the application of AI technologies has gained increasing interest. Traditionally, university assessments have relied on several methods to evaluate student learning, including written exams, multiple choice questions (MCQs), short answer questions (SAQs), essays, oral presentations, and dissertations. Despite their widespread use, these traditional forms of assessment often face criticism for their inability to fully capture the breadth of student understanding and skills, and their lack of authenticity (Andrews, 2003; Walker, 2018). However, assessment remains a cornerstone of university education and is considered a key driver of student learning and performance development (Bloxham & Boyd, 2007), particularly if constructively aligned with curriculum learning outcomes (Biggs, 1996). Effective assessment design is therefore crucial to evaluate both the performance and educational development of students, while internal and external stakeholders still require academic achievement to be measured. Although traditional assessment methods have been relatively successful at fulfilling these requirements, the disruptive potential of AI poses significant challenges for the continued implementation of the standard assessment paradigm (Mislevy et al., 2012; Swiecki et al., 2022).

Authentic assessments are considered more reliable for the appraisal of student learning goals and skills development (Saher et al., 2022). In this context, authentic assessment aims to integrate realism, contextualization and problematization, thereby replicating tasks and outcomes typically undertaken in a professional environment (Villarroel et al., 2018; Wiggins, 1990). In contrast, traditional assessments tend to be limited to objective measurement at a single point in time (Koh et al., 2019) and tend to test memorization, which may not be a true reflection of a student's comprehension (Villarroel et al., 2020). Furthermore, students demonstrate a deeper understanding and more stable learning outcomes when engaged with assessments that test higher-order cognitive skills (Entwistle, 2017; Rawson et al., 2013). Well-designed authentic assessments can incorporate elements of critical thinking and problem-solving, and introduce students to real-world scenarios that prepare them for the workplace (Schultz et al., 2022). These elements should be aligned with program learning outcomes that not only include subject-specific knowledge acquisition but also opportunities to acquire transferable skills, develop critical thinking competencies, undertake self-reflection and promote an ethos of continued professional development (Meyers & Nulty, 2009; Sarkar et al., 2020; Villarroel et al., 2020). However, effectively implementing authentic assessments across a program of

study is challenging. Institutions may retain a preference for more traditional approaches to assessments that align with educators' competences and experiences, while additional guidance and instruction is often required for authentic assessments to be successfully delivered and embedded within local education cultures (Schultz et al., 2022). In addition, concerns remain regarding the impact that generative AI may have on academic integrity, particularly regarding traditional forms of assessment.

Artificial intelligence and academic integrity

Assessment in higher education is often entrenched within the standard assessment paradigm of MCQs, SAQs, and essays. These assessment formats are familiar to lecturers and students. However, they require careful design and can be time-consuming to mark and to ensure that student answers are accurately converted into a quantifiable evaluation, which often requires sophisticated marking criteria or rubrics (Kaipa, 2021; Mislevy et al., 2012; Swiecki et al., 2022). The standard assessment paradigm is also subject to uniformity, despite the varying learning needs of diverse student cohorts. This uniformity could, therefore, disadvantage some students and restrict their opportunity to demonstrate learning attainment (Gipps & Stobart, 2009). Diversifying the portfolio of assessments across a course or program of study has also been shown to increase inclusivity and provide students with multiple opportunities to demonstrate proficiency across various assessment points (Bloxham & Boyd, 2007; Hounsell et al., 2007; O'Neill & Padden, 2022).

The standard assessment paradigm may also be inauthentic, as it struggles to replicate experiences within a professional environment. This has become more acute with the rapid development of generative AI, which poses a serious challenge to the standard assessment paradigm. In particular, the traditional academic essay appears outdated and inauthentic considering the advances in computational and AI tools available to the professional writer (Swiecki et al., 2022). The standard assessment paradigm may also be particularly susceptible to AI interference, whereby students use AI tools to generate text and submit it as their own work (Yeadon et al., 2023). The higher education sector, therefore, faces significant challenges to ensure academic integrity is maintained in light of advances in AI technologies (Cotton et al., 2024; Perkins, 2023; Sullivan et al., 2023). These challenges extend to non-proctored, online exams and assessments, where cheating and AI use are difficult to detect (Geerling et al., 2023). In addition to concerns about plagiarism, there are also concerns about the over-reliance on AI technologies, the automation of knowledge acquisition and the degradation of student critical thinking and problem-solving skills (Perera & Lankathilake, 2023).

The academic performance of ChatGPT was evaluated by Yeadon et al. (2023) in response to five open-ended university physics essay questions. The ChatGPT generated answers were marked by five independent examiners and consistently received first-class grades (Yeadon et al., 2023). When ChatGPT answers were tested using Turnitin or Grammarly plagiarism detectors, the generated output was deemed to be original. These findings highlight the difficulty in discerning student and AI generated work, with significant implications for plagiarism detection and academic integrity, and suggests the need for a re-evaluation of the traditional academic essay as a reliable form of assessment. Evidence also suggests that ChatGPT can produce well-written content in response to a

wide range of prompts, as demonstrated by its performance on final exams in an MBA program and its ability to pass the Bar exam in the United States (Katz et al., 2024; Terwiesch, 2023). Therefore, the sophisticated text generation capabilities of LLMs will significantly impact the approach to written assessments in higher education (Waltzer et al., 2024).

Existing AI text detection tools face significant challenges in accurately identifying AI-generated content. Some higher education institutions (HEIs) have integrated AI detection applications, such as Turnitin, within virtual learning environments, or use standalone tools like GPTZero (https://gptzero.me/) and WinstonAI (https://gowinston.ai/) to detect AI generated work (McDonald et al., 2024). However, other institutions remain hesitant due to concerns about the accuracy and reliability of these tools, particularly the risk of false positives (Dalalah & Dalalah, 2023; Saqib & Zia, 2024; Weber-Wulff et al., 2023). Moreover, current detection tools may unfairly target students whose first language is not English, mistakenly identifying their work as AI-generated (Fröhling & Zubiaga, 2021). It is critical to ensure the fair application of AI detection software across diverse student populations, and if this cannot be reliably achieved, then detection software may be redundant. Furthermore, the rapid advancement of generative AI technologies and the release of more capable versions may easily surpass the ability of detection software, making the pursuit of effective AI detection a futile digital arms race (Elkhatat et al., 2023).

Rethinking multiple choice questions

Several studies have consistently demonstrated the ability of ChatGPT to correctly answer MCQs in subjects such as medicine, science, economics, mathematics, and law (Geerling et al., 2023; Giannos & Delardas, 2023; Newton & Xiromeriti, 2024). The performance of ChatGPT on the United States Medical Licensing Exam (USMLE) was commensurate with the level required to pass and appeared to demonstrate critical thinking and clinical insight (Kung et al., 2023). However, in a study by Gilson et al. (2022), ChatGPT 3 performance decreased as the level of question difficulty increased, suggesting that this model of ChatGPT was similar to a third-year medical student (Gilson et al., 2022). However, newer versions of LLMs are likely to deliver better performance. For example, ChatGPT 4 outperformed ChatGPT 3.5 in a 500-question neurosurgery written board examination and on an orthopedic resident performance on orthopedic assessment examination (Ali et al., 2023; Massey et al., 2023), although both versions passed the exams. In other disciplines ChatGPT 3.5 and ChatGPT 4 exceeded median performance levels when answering conceptual physics questions, with ChatGPT 4 reaching performance levels equivalent to an expert physicist (West, 2023), while on a fundamentals of engineering exam ChatGPT 4 performed with better accuracy compared to ChatGPT 3.5, with performance further increased following additional prompt refinement (Pursnani et al., 2023).

These capabilities have obvious implications for higher education, not least in ensuring the security and academic integrity of MCQ exams. MCQs that test lower-order cognition skills such as simple factual recall or knowledge may be particularly susceptible to generative AI compared to higher-order, problem-based questions (Al-Shakarchi & Haq, 2023; Wang et al., 2023). Up-scaling MCQs to test higher-order cognition may allow greater AI resistance. For example, single best answer questions provide a range of possible answers, all of which are correct, while extended matching questions pose

a series of scenarios with a range of possible options. Questions could also be designed based on the interpretation of complex data sets, image analysis or audio-video content that requires reasoning and conceptualization (Cuthbert & Simpson, 2023). Although MCQs can be carefully and thoughtfully designed to include complex or situational scenarios that evaluate higher-order cognitive understanding, (Billings et al., 2016; Kasai et al., 2023; Wood et al., 2023), it may be insufficient in the long-run to try to out-smart LLMs given the developments and improvements in their competences (Newton & Xiromeriti, 2024).

The most obvious mitigation is for non-proctored and online exams to be abandoned in favor of in-person, invigilated exams. Online examinations were already prone to cheating prior to the advent of generative AI (Farazouli et al., 2024; Newton & Essex, 2024), further supporting the benefit of in-person exams for upholding academic integrity. Another option is the replacement of closed-book exams with alternative forms of in-person assessment. For example, replacing exams with more authentic or experiential assessments, which replicate real-world experiences and professional working environments. However, these approaches ignore the potential for the integration of generative AI into both formative and summative assessment design. For example, ChatGPT can generate its own MCQ questions, which can be used to facilitate student understanding of a given subject, as a study tool or revision tool to explore concepts and to practice for summative exams (Passby et al., 2023; Surapaneni et al., 2024). This could also enable educators to design MCQs, although prompt refinement and expert oversight may be necessary to ensure quality (Indran et al., 2024; Rivera-Rosas et al., 2024). Another pragmatic alternative is to ask students to evaluate and assess the responses generated by AI. This stimulates critical thinking, engages students with the technology, and trains students on the capabilities and limitations of generative AI. Due to the current limitations of AI in contextual understanding and complex judgement, these AI applications should act to complement existing teaching modalities and human-led instruction, rather than replacing the lecturer as the primary educator. The human teacher has the ability to adapt to emotional cues, empathize, and build personal relationships with students; critical elements of teaching that cannot be replicated by current generative AI technology.

Integrating AI into coursework assessments

It is now well established that certain forms of assessment may be more impacted by generative AI than others. ChatGPT and other LLMs are capable of generating logical and reasoned text that is difficult to distinguish from human-written content (Andrews, 2003; Chaudhry et al., 2023; Fyfe, 2023; Waltzer et al., 2024). This has led many to adopt the view that the traditional academic essay is dead (Andrews, 2003; Rudolph et al., 2023; Stokel-Walker, 2022). The release of newer and improved AI models may compound this view. For example, ChatGPT 4 is thought to be capable of producing a level of linguistic diversity and sophistication that supersedes that of students, which exceeds the capabilities of previous models (Herbold et al., 2023). Liu et al. (2023) analyzed over 4,000 GPT-generated evaluative essays and demonstrated that generative AI creates sentences with more complicated grammar, although human-written essays use a wider variety of words (Liu et al., 2023). Across multiple university-level courses, ChatGPT was shown to perform better than students in response to a variety of written questions, although performance was lower for questions requiring

higher-order cognitive knowledge, such as creativity, compared to knowledge-based questions (Ibrahim et al., 2023; Williams, 2024). In response to several university-level psychology SAQ and essay questions, AI out-performed students by 5-10% on almost all submitted answers, while AI-written answers were virtually undetectable following blind marking (Scarfe et al., 2024). In this study, the only question that ChatGPT struggled with involved high-order reasoning.

Webb (2023) outlines three strategies that higher education institutions and educators can employ when dealing with the potential impact of generative AI technologies on assessments (Webb, 2023). The first is to completely avoid AI by reverting to in-person exams. The disadvantage of relying on this strategy is the lack of opportunity for students to participate in diverse forms of coursework and authentic assessments. Complete reliance on in-person assessments also poses significant logistical challenges for departments, especially where large student cohorts are involved. The second strategy is to devise assessments that current versions of generative AI tools cannot do or are not very good at doing. The danger of trying to outrun AI is that rapid developments in the technology may soon make this strategy redundant (Newton & Xiromeriti, 2024). A good example is the data and image analysis capabilities of ChatGPT 4o, which the previous 3.5 version was incapable of performing. Therefore, trying to outrun AI may be a perilous exercise. The third strategy is to adapt to and embrace AI by actively encouraging its use by students and designing assessments that incorporate the authentic use of AI (Holmes & Miao, 2023; Sabzalieva & Valentini, 2023; Sharples, 2023; Webb, 2023). This includes creating teaching and learning activities that support AI in assessment, the training and guidance of students on its responsible use, the evaluation of existing forms of assessments for ‘AI-proofing’ and ensuring academic integrity is maintained (Sabzalieva & Valentini, 2023).

Generative AI and feedback

The impact of generative AI on student assessments presents both challenges and opportunities (Adeshola & Adepoju, 2023; Sullivan et al., 2023), with the most significant of these challenges being concerns over student cheating (Chaudhry et al., 2023; Perkins, 2023; Stutz et al., 2023). Conversely, generative AI presents several opportunities to enhance student learning directly related to assessments. For example, generative AI could provide immediate and continuous feedback on student work (Guo & Wang, 2024). This feedback may take various forms, which are related to both formative and summative assignments (Table 1). AI generated feedback could also assist self-assessment, such as asking ChatGPT for recommendations on further learning, feedback for improvements on submitted work, creation of practice questions and generation of marking rubrics (Rajabi et al., 2023). However, others have cautioned that although AI may encourage students to approach assessment for learning (Carless, 2005), accuracy in generated output is inconsistent (Fuchs et al., 2024), while feedback provided by LLMs may not align with tutors’ expectations or discipline-specific course material and could be derived from inherent biases in the training dataset (Guo & Wang, 2024; Naidu & Sevnarayan, 2023). Nevertheless, some students consider feedback from ChatGPT to less biased and judgmental than human feedback (Greiner et al., 2023; Guo & Wang, 2024), while ChatGPT may be able to provide more detailed and timely feedback that encourages students to take ownership of their feedback and learning development (Chaudhry et al., 2023; Dai et al., 2023).

Table 1

Opportunities for feedback from generative AI.

LLM feedback	Benefits	Disadvantages
Idea generation	Time efficiency, large knowledge base, creativity, diverse perspectives, overcome blocks	Inhibit originality and creative thinking, relevancy, quality variation
Resource recommendation	Personalisation, tailor to individual learning needs, variety of sources, comprehensive coverage, focus on topics	Accuracy and reliability, over-dependence, inherent biases and perspectives, misalignment with curriculum
Generation of rubrics	Standardisation and consistency, time efficiency, clarify expectations	Lacks context or nuance, limits creativity and flexibility, misalignment with educators
Cross checking criteria	Time efficiency, immediate feedback, consistent application, detailed analysis of criteria, error checking	Lacks contextual understanding, over-reliance, reduces critical engagement, potential misjudgements
Provision of learning tasks	Personalisation and tailored learning, engaging and interactive, diverse task generation, immediate progress checks	Lacks context or nuance, over dependence, misalignment with curriculum, inherent biases
Progress evaluation	Immediate feedback, consistency, reduces subjective bias, provides areas for improvement	Lacks context and nuance, inconsistent feedback, misalignment with learning outcomes
Generate practice questions	Personalisation, can adjust difficulty, immediate feedback, formative assessment	Potential errors, may lack complexity, detail or nuance, over-reliance
Simulate peer review	Enhanced learning engagement, constructive feedback, preparation for real peer review	Lacks authenticity, reduces opportunity for peer interaction, trust

Innovative assessment approaches

There are two broad approaches to designing innovative assessments in the context of generative AI. The first is to design assessments that mitigate the impact of AI, either by completely eliminating AI use or by minimizing its potential application (Table 2). For example, Smolansky et al. (2023) suggest that presentations and discussions, whether pre-recorded or live are less impacted by ChatGPT (Smolansky et al., 2023). These forms of assessment require elements of design and creativity and their delivery in real time. Similarly, video-conferencing and podcasts also fulfill this brief, although generative AI could be used to generate content (Nikolic et al., 2023). Viva examinations and face-to-face presentations, on-campus exercises, practical examinations, objective structured clinical examinations (OSCEs), and clinical assessments are also considered to be less impacted by LLMs (JISC, 2024). These forms of assessment mostly avoid AI interference and, in some cases, retain authenticity where they replicate professional skills.

The second approach is to design assessments that are more progressive, by integrating AI as a fundamental element of the assessment process (Table 2). This second approach is also flexible by way of design and can include assessments that use AI as an auxiliary or assistive tool, whereby students are permitted to use AI in a defined and pre-determined way. An example would be using AI to draft an outline of an essay plan or supporting an element of the assessment, such as the generation of insights. However, this form of AI integration falls short of using AI to generate substantive student output. In addition, AI can be fully incorporated into the assignment task, whereby students would be allowed to use AI as an essential element of the assessment (Arnold, 2021; Bowditch, 2023). For example, AI could be used to create text, images, or code, as a research tool to generate ideas or solutions, or to compare content between AI and human sources. In this integral role, students would demonstrate their competencies in the responsible use of AI and simultaneously critically evaluate AI generated output to inform on the assignment task and learning outcomes.

Several of the assessments that integrate AI involve the generation of text or images, such as case studies, research insights, or assessment questions. For example, medical case studies or clinical scenarios could be generated by AI, which students then have to critically evaluate and provide a solution based on their knowledge. This type of AI-integrated coursework assesses contextual intelligence and professional competencies, as well as promoting AI literacy (Arnold, 2023; JISC, 2024). Similarly, critiquing AI output promotes AI literacy by encouraging students to evaluate the accuracy, context, and limitations of an AI generated product, while simultaneously engaging in critical thinking and independent research. Debating and refining, as part of an integrated AI assessment, allows students to present a difficult question or challenge to ChatGPT, for example, whereby students enter a Socratic debate (question and answer session) with ChatGPT and later critique and examine their own perspectives compared to those of the AI. Using generative AI as a possibility engine to generate alternative explanations or as an Exploratorium to explore and interpret data, could also be integrated into coursework assessment design (Sharples, 2023), whereby the final output may be a reflective written report based on the conversation. Debating and refining is another example that promotes critical and independent thinking, problem-solving, contextual intelligence, and AI literacy (Arnold, 2022; JISC, 2024). There are many possibilities to integrate AI in assessment, as outlined in Table 2, awaiting development, testing, and implementing.

Table 2

Assessment strategies that mitigate or integrate AI.

Assessments that mitigate AI	Assessments that integrate AI
Clinical assessments	AI case study
Data analysis	AI image generation
Dissertation	AI research, insights, solution finder
Documentary	Comparative analysis
Exhibitions	Critical analysis
Fieldwork assessment	Critique of AI output
Invigilated exams	Data visualisation
Learning journal (personal development)	Debate with AI (Socratic debate)
Live analysis	Digital field guide
Model creation	Debate and refine
Multi-media, blog, vlog	Design an assessment
Objective structured clinical examinations	Draft generator
On-campus projects	Exploratorium
Poster presentation	Patchwork assessment
Practical (laboratory) examinations	Peer (AI) assessment, review
Presentations (live online, in-person)	Possibility engine
Public conference	Prompt competition
Research project (laboratory)	Question generation
Role-play, simulations	Teaching resource creation
Viva voce	Translation
	Web design
	Writing for a style

There is a strong case that university assessments need to be transformed to promote self-regulated learning, critical thinking and problem-solving proficiencies, and encourage the responsible use of generative AI to safeguard academic integrity (Xia et al., 2024). The integration of generative AI into assessment therefore promotes a transition towards more authentic assessment (Crawford et al., 2023).

AI and assessment literacy

The disruptive nature of generative AI will have a significant impact on education and teaching, which will drive a need for better assessment and technological literacy (Firat, 2023; Popham, 2009). It has been argued that digital and AI literacy should be part of the curricula and learning outcomes of higher education programs (Chiu, 2024; Stutz et al., 2023). This encompasses multiple elements of higher education, including assessment design, teaching practices, and intelligent learning systems that support personalized learning pathways. Critical thinking, problem-solving and creativity are likely to be fundamental in establishing a successful integration of AI into education. As the TN-3 model

conceded, rather portentously, in The Complete Robot, "…things like myself can be manufactured. But nothing yet can imitate the creativity and versatility of a human brain" (Asimov, 2018).

Higher education institutions need to be aware of the training requirements of teachers in order to implement these new practices, and therefore, institutions should align their AI polices accordingly (Chan, 2023). Assessments are of foremost concern in this context, with some suggesting an interdisciplinary approach may be needed to future-proof curricula (Southworth et al., 2023) and to move away from traditional academic writing and performance-based assessments (Geerling et al., 2023; Smolansky et al., 2023). Although the theoretical framework for integrating generative AI into assessment has been discussed here and elsewhere in the literature, few examples of practical applications exist. Therefore, a case study of integrating AI into coursework assessment design is presented, together with an analysis of the student perspective of this approach.

Method

AI integration assessment design

The redesign of a coursework essay assessment occurred at a leading UK university within an established postgraduate taught MSc program. The module was taught in the first term of a one- year academic cycle, and the discipline was immunology. The previous iteration of the coursework assessment was a traditional 2000-word essay on prescribed questions provided by the course tutor. The assessment pattern for the module consisted of a 50% in-person MCQ exam (not discussed) and a 50% coursework essay.

The integration of AI into the design of the new assessment consisted of a combination of a student guidance and training workshop (in-person), peer evaluation and feedback workshop (in-person) and individual student components with and without the permitted use of generative AI (ChatGPT 3.5; (OpenAI, 2024) and according to the institution assessment policy on the use of generative AI (Figure 1). Workshop 1 involved guidance on the coursework instructions, provision of the essay question, and training on the institution's academic integrity and generative AI policies. In addition, training on the use of ChatGPT, the user interface, and an introduction to ChatGPT prompt engineering was provided. Students were then tasked with generating a draft outline of the coursework essay using ChatGPT.

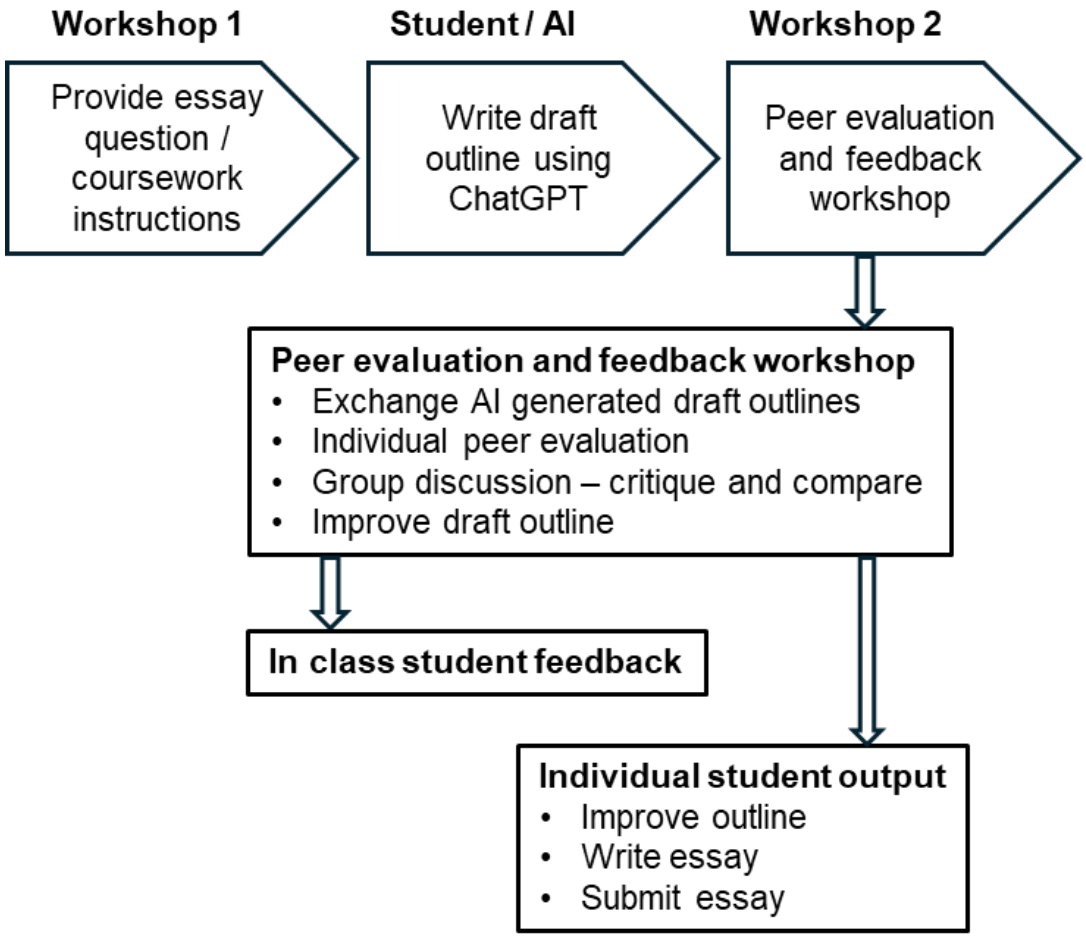
Workshop 2 consisted of students undertaking a peer evaluation of ChatGPT generated draft outlines. Firstly, students were asked to exchange generated draft outlines with peers for individual critique and evaluation. Students were then placed into small groups (3-4 people) to further discuss and compare the ChatGPT generated draft outlines, focusing on their strengths and weaknesses. Finally, students were asked to improve their ChatGPT generated draft outlines without any further assistance from ChatGPT. During workshop 2, in-course student feedback was obtained.

The remainder of the coursework assessment was undertaken individually and without further assistance from ChatGPT. Students were, therefore, asked to improve their draft outline as a basis for the final submission. Writing the essay, including text generation and editing, and the inclusion of

supporting figures or diagrams, was undertaken without further assistance from ChatGPT. Final versions were then submitted for summative assessment.

Figure 1

Schematic of AI-integrated coursework design



In-course evaluation

During workshop 2, students were asked their opinions and perspectives on the use of generative AI in their assessments and university studies. This was performed in class as part of the interactive workshop using the online polling software Mentimeter (https://www.mentimeter.com/). A total of 20 students participated in the Mentimeter polling. Students were asked 8 questions, including aspects of generative AI use, such as its capabilities and limitations (Table 3), which facilitated their training on the use of ChatGPT for this assessment component. In addition, an open discussion allowed students to provide further opinions on their experiences of using ChatGPT for the generation of their coursework essay draft outline, focusing on the advantages and disadvantages of using ChatGPT as an integral part of their assessment.

Table 3

In course survey questions

Number	Question
1	Have you used ChatGPT or other generative AI tools?
2	What generative AI tools have you used?
3	How confident are you using generative AI tools, for example ChatGPT?
4	Would you consider using generative AI in your studies?
5	What aspects of your learning could generative AI be useful for?
6	What are the capabilities of generative AI?
7	What are the limitations of generative AI?
8	Do you agree, using generative AI for my assessments is an act of plagiarism?

Intended course learning outcomes

The intended course learning outcomes relating to the coursework assessment were: 1) discuss the central themes of the immune system and explain how the immune system can be divided into the innate and adaptive immune systems, and 2) appreciate and apply your knowledge of the limitations and capabilities of generative AI technologies for the generation of human-like text. These intended learning outcomes were evaluated in relation to the survey and group discussion responses from students.

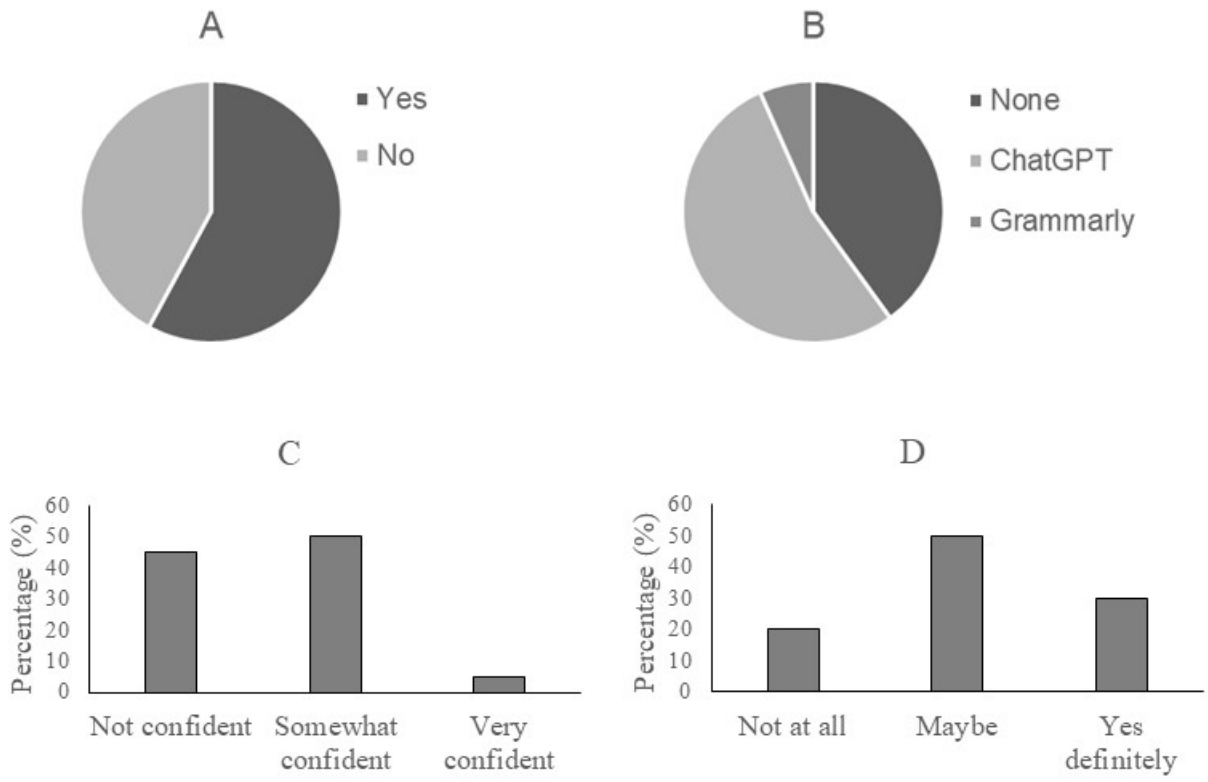
Results

Student use of generative AI

A series of questions (Table 3) were asked to gain an understanding of how students engage with and perceive the use of generative AI. Of the students surveyed, 57.9% had previously used generative AI technologies (Figure 2A), with most students (53%) using ChatGPT (88% of those students who were currently using generative AI) and a minority (6.6%) using the AI writing assistant Grammarly (Figure 2B). When asked about their confidence in using generative AI tools, such as ChatGPT, only 5% of students revealed they were very confident using this technology, and a further 50% revealed they were somewhat confident, with the remainder (45%) not being confident (Figure 2C). When asked whether they would consider using generative AI in their studies, 30% revealed they would, 50% answered maybe, while the others (20%) stated they would not use generative AI in their studies at all (Figure 2D).

Figure 2

Survey of students on using generative AI.



The next set of questions was aimed at understanding what aspects of their learning students would consider using generative AI for and what they considered to be the capabilities and limitations of generative AI technologies. Students revealed that they would primarily use AI for planning and structuring assignments, such as essays, for drafting coursework assignments and for checking spelling or summarizing text. Other uses included the generation of ideas, to undertake initial research on a subject, as a revision tool, and to use generative AI as a proofreading tool or to correct writing (Table 4). When asked what the capabilities of generative AI are, students considered academic support and summarizing information as one of the most useful applications, in addition to several other capabilities such as the generation of text, images or mock exam questions. Students considered accuracy and out-of-date information to be the primary limitations of generative AI, as well as lack of critical thinking, detailed responses, and the generation of superficial answers.

Table 4

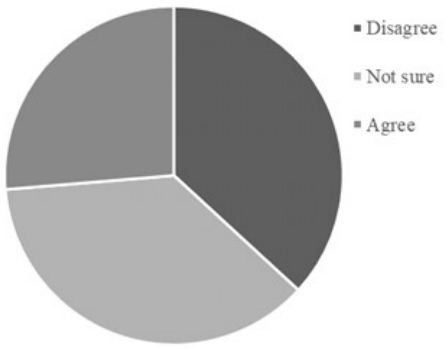
Student perspective of the capabilities and limitations of generative AI.

What aspects of your learning could generative AI be useful for?	What are the capabilities of generative AI?	What are the limitations of generative AI?
Planning/structure	Academic support	Accuracy
Drafting	Summarising information	Out of date
Grammar	Generating exam questions	Lack of critical thinking
Summarising	Generating text	Unable to provide references
Ideas/prompting	Grammar correction	Unable to check source of information
Research	Generating images	
Revision	Answer questions	Unhelpful/irrelevant information
Proofreading/correcting	Provide feedback	
	Creativity	Superficial responses
	Saves time	Lacks detail
	Planning	
	Information sources	

The final question of the in-class survey concerned the potential use of generative AI as an act of plagiarism. When asked whether they thought using generative AI in their assessments was an act of plagiarism, only 26% of respondents agreed, with 37% not being sure (Figure 3). Of the remaining respondents, 37% disagreed that using generative AI in their assessments was an act of plagiarism.

Figure 3

Perspective on plagiarism and AI use in assessments.



The second workshop provided a further opportunity to discuss the merits of integrating AI into the existing coursework assessment and to evaluate whether the coursework re-design met the intended

learning outcomes. Following an open discussion, students provided their opinions on the advantages and disadvantages of integrating AI into the new assessment design (Table 5). Providing a basic outline of the essay was considered a useful starting point for the design of the draft essay outline. It was agreed that integrating AI into the coursework provided a valuable opportunity to learn about the capabilities and limitations of the technology, indicating this intended learning outcome was met. Students were also enthusiastic about learning effective prompt engineering to generate more nuanced and useful outputs. However, it was thought that the draft outlines generated by ChatGPT were uniform and generic when compared across the cohort, and the outputs generated were too superficial to meet the subject-specific learning outcome, further enabling students to consider the limitations of AI in this context. Students were, therefore, critical of its limitations, with some questioning the accuracy of generated content.

Table 5

Advantages and disadvantages of AI in the coursework assessment.

Advantages	Disadvantages
• Provided a basic outline – useful starting point • Advantageous in designing essay outline • Engages students with AI capabilities/limitations • Students were interested in better prompt design	• ChatGPT output was similar between students • Output was quite generic / superficial • Did not provide the detail some students were expecting • Students were critical of its limitations • Some students were less confident in its accuracy

Discussion

Generative AI encompasses advanced technologies such as natural language processing (NLP), machine learning, and neural networks, which are able to generate convincing and sophisticated human-like text. These capabilities open new avenues for developing more authentic assessments that directly integrate AI (Lawrie, 2023; Saher et al., 2022; Villarroel et al., 2020). Generative AI has found utility in several aspects of higher education, including as a personal learning tool, digital tutor, as an automated and predictive grading system, for immediate and elaborate feedback on student assignments, and for the generation of formative exercises and assessment questions (Owan et al., 2023). This paper focused on the incorporation and integration of generative AI directly into coursework assessment strategies and presented a case study of AI integration into a postgraduate taught module coursework assessment at a leading UK university.

The case study outlines the design and implementation of a coursework assessment that integrates generative AI as an auxiliary agent. The approach fulfilled the intended learning outcomes associated with the assessment, including training students on the capabilities and limitations of generative AI. This represents an innovative re-design of a familiar coursework format that embraces the

transformative potential of AI (Naidu & Sevnarayan, 2023; Nikolic et al., 2023). The caveat is that students could, in theory, still use AI to generate substantive text for their summative essay submission and, therefore, engage in plagiarism. However, the associated training for students on the limitations of generative AI, their personal experience of using ChatGPT, engaging students in peer evaluation of ChatGPT output, and empowering students to take responsibility for their own feedback and learning, should help mitigate instances of academic misconduct. Training students to acknowledge the use of AI through conventional citations and transparent referencing could further avoid unnecessary plagiarism (Perkins, 2023). Incorporating an additional element to the assessment, such as an in-person vivo voce or verbal presentation, would further allow students to substantiate submitted work as their own.

When challenging existing education frameworks, it is important to collaborate with students to understand existing gaps in knowledge and skill sets. It is also important to establish the reasons for change and the benefits that transformative projects have for student learning (Billings et al., 2023). While multiple opportunities exist to foster student-staff partnerships in this area, more research is needed on the student perspective. One aspect of the current study was to understand the extent to which students are using AI technologies and to identify levels of AI literacy. Although only a relatively small cohort of students was surveyed, it was interesting to note that only 57% of students had used generative AI tools, with only 5% of students declaring they were confident using these technologies. This highlights the necessity to gauge student technological literacy and provide training and guidance on the proper use of generative AI. This is particularly important where AI is being integrated into coursework assessment design and practice (Owan et al., 2023; Owen & Wasiuk, 2021).

The responses from students indicated they used AI for planning, drafting, and structuring written work. Students were also using ChatGPT for the generation of ideas, research and revision. However, students were critical of the accuracy of generated responses, out-of-date answers, and unverifiable sources of information. These are familiar negative attributes of existing generative AI technologies, which are likely to necessitate the continued supervision and validation of AI generated output (Rudolph et al., 2023; Sallam, 2023). Despite ChatGPT providing a useful starting point and draft outline for an essay, students were critical of the superficial and generic nature of the generated output and were cautious about the accuracy of the information provided. This may reflect the limited utility of AI for assisting students in achieving subject-specific learning outcomes. However, they were positive about the opportunity of using ChatGPT for their assessment, as this engaged them with its capabilities and limitations, thereby fulfilling the second intended learning outcome for the assessment.

Promoting student AI literacy is crucial for student educational development in an increasingly AI-driven society, and embedding AI literacy into curricula is considered advantageous for future employability and to supplement traditional teaching and learning practices (Southworth et al., 2023; Walter, 2024). Students had contradictory perceptions of what constitutes plagiarism when using AI for assignments (Firat, 2023; Kumar & Raman, 2022), with the survey indicating that 36% of students did not consider using AI for assessments an act of plagiarism. This further highlights the necessity for embedding AI literacy into curricula. However, there are concerns that students will become over-reliant on generative AI, which could diminish the same creativity and critical thinking skills that are important for employability and professional practice (Rajabi et al., 2023). This likely necessitates

proper training and guidance for students, not least on the capabilities and limitations of generative AI but also on its ethical application (Mitra & Chitra, 2024; Wood & Moss, 2024). Aligning this guidance with curriculum learning outcomes and recognizing the student perspective (Johnston et al., 2024), are important considerations for its success.

As AI becomes embedded into higher education, enhancing educator AI literacy is also essential (Xia et al., 2024), especially with the rapid advancements in the capabilities of this technology. However, integrating AI into teaching and learning environments should not replace the teacher as the primary expert for knowledge acquisition, the main arbiter of learning (Dillenbourg, 2016; Sharples, 2023), or as the primary provider of academic judgements on students' work (Adıgüzel et al., 2023). For example, plagiarism detection software is currently not adequate to reliably detect generative AI output or distinguish between AI and human generated text (Ardito, 2024; Chaka, 2024). AI detection software is likely to become further redundant due to advancements in generative AI models. Rather, a robust and reliable assessment framework that integrates AI while retaining academic integrity is recommended. Furthermore, the application of AI detection software is incompatible with teaching students the ethical and judicious use of generative AI and may foster a culture of accusation, rather than engagement (Ardito, 2024). Therefore, institutions should be prepared to promote professional and technological development by training teachers on the applications of AI in education, the potential for integrating AI into curricula and assessment strategies, and the ethical considerations of AI (Sperling et al., 2024; Wilton et al., 2022). University teachers enter an AI-centric environment from diverse backgrounds and varying levels of technological expertise. Professional development concerning AI literacy should, therefore, be equitable and accessible. Although there is a lack of research on teacher's AI literacy, it has been argued that understanding how AI functions, its impact on student employability, and its ethical implications is necessary for the effective inclusion of AI into higher education environments (Sperling et al., 2024).

Generative AI's ability to deliver instant and personalized feedback may significantly impact student learning, and educators need to be aware of these capabilities. Immediate feedback enables students to promptly understand their mistakes and misconceptions, fostering a more effective learning process (Nikolopoulou, 2024). Personalized feedback, tailored to each student's needs, could make the assessment process more meaningful, impactful, and timely. Sharples (2023) further advocates for generative AI to contribute to a process of social learning, whereby learning goals, activities, and ideas are shared conversationally to reach agreements (Sharples, 2023). However, there are ethical concerns surrounding the analysis of sensitive student data by AI, while ensuring data security and maintaining student confidentiality are critical in education settings (Elliott & Soifer, 2022). Nevertheless, many opportunities exist to customize generative AI tools to provide feedback and act as a student learning resource, such as acting as an idea generator, task generator or to create bespoke assessment questions (Guo & Wang, 2024; Rajabi et al., 2023). Coupled with the iterative characteristics of generative AI models and the ability to receive immediate and recurring feedback, AI offers students an accessible way to complement and augment traditional teaching and learning methods.

As AI assumes more assessment-related tasks, educators must adapt to new responsibilities focused on facilitating learning and providing nuanced, human-oriented support. This provides an opportunity for

the creation of more holistic assessment models that consider a broader range of student skills and competencies. It also provides an opportunity to re-focus assessment on higher-order cognitive tasks, such as critical thinking, problem-solving, complex evaluation, and creativity (Adeshola & Adepoju, 2023; Gamage et al., 2023). These approaches could be more authentic by aligning assessments with the societal and occupational applications of AI. Assessing higher-order cognitive proficiencies may also mitigate some of the challenges surrounding academic integrity and plagiarism. The disruptive impact of AI on assessments may also devalue traditional assessment models, such as MCQ exams and standard written essays. Rather, AI offers more diverse and authentic assessment models, a reconfiguration of learning outcomes that embed AI and assessments that promote skills and competencies that are valued by employers (Chiu, 2024; Gorichanaz, 2023; Xia et al., 2024).

Conclusion

Integrating generative AI into coursework assessment holds significant promise for transforming higher education. Universities should ensure that adequate training and guidance are provided to educators and students on the use of AI technologies, as AI literacy will be essential for the success of this approach. Integrating AI into university assessments should be a holistic exercise, rather than a complete replacement of traditional assessment paradigms. AI offers many opportunities for educators to innovate their teaching practices and for students to supplement and enrich their learning experiences. AI is likely to transform the way feedback is delivered and how students receive and internalize their feedback. Integrating AI will allow a more authentic approach to assessment models and promote assessment diversity that caters for diverse learning requirements. Although academic integrity and plagiarism are real concerns for higher education institutions, AI may provide a means to refocus assessment that evaluates higher-order cognitive proficiencies. By embracing AI's potential and addressing the associated challenges, higher education institutions can pave the way for more effective, equitable, and engaging assessment practices that better prepare students for the complexities of the modern world (Aoun, 2024). Higher education institutions should review and rethink their assessment policies to adapt to the challenges posed by generative AI. Additionally, institutions are recommended to offer more interdisciplinary programs and teaching approaches to prepare students for the evolving assessment landscape (Southworth et al., 2023). By taking an innovative and progressive approach to assessments integrating AI, universities will be perceived as forward-thinking and fit to train students for an AI-driven society.

References

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments,* 32(10), 6159-6172. https://doi.org/10.1080/10494820.2023.2253858

Adıgüzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*. *15*(3), https://doi.org/10.30935/cedtech/13152

Al-Shakarchi, N. J., & Haq, I. U. (2023). ChatGPT performance in the UK medical licensing assessment: how to train the next generation? *Mayo Clinic Proceedings: Digital Health, 1*(3), 309-310. https://doi.org/doi:10.1016/j.mcpdig.2023.06.004

Ali, R., Tang, O. Y., Connolly, I. D., Zadnik Sullivan, P. L., Shin, J. H., Fridley, J. S., Asaad, W. F., Cielo, D., Oyelese, A. A., & Doberstein, C. E. (2023). Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. *Neurosurgery, 93*(6), 1353-1365. https://doi.org/10.1101/2023.03.25.23287743

Andrews, R. (2003). The end of the essay? *Teaching in Higher Education, 8*(1), 117-128. https://doi.org/10.1080/1356251032000052366

Aoun, J. (2024, July 01). How high ed can adapt to the challenges of AI. https://www.chronicle.com/article/how-higher-ed-can-adapt-to-the-challenges-of-ai?sra=true

Ardito, C. G. (2024). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*. 1-18. https://doi.org/10.1002/tl.20624

Arnold, L. (2021, 7th August 2024). Authentic Assessment Top 'Reflective' Trumps! https://lydiaarnold.net/2021/04/08/authentic-assessment-top-reflective-trumps-cards/

Arnold, L. (2022, November 14). Expanded Assessment Top Trumps. https://lydiaarnold.net/2022/11/14/expanded-assessment-top-trumps/

Arnold, L. (2023, June 12). Striking a Balance: Integrating AI into Assessment Practices. https://lydiaarnold.net/2023/06/12/striking-a-balance-integrating-ai-into-assessment-practices/

Asimov, I. (2018). The Complete Robot (1st Edition ed., Vol. Satisfaction Guaranteed). HarperVoyager.

Babaniyazovich, A. A. (2023). The impact of artificial intelligence on human life in the future. *International Multidisciplinary Journal for Research & Development, 10*(10), 444-448.

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher education, 32(3), 347-364. https://doi.org/10.1007/BF00138871

Billings, M., DeRuchie, K., Haist, S., Hussie, K., Merrell, J., Paniagua, M., Swygert, K., & Tyson, J. (2016). Constructing written test questions for the basic and clinical sciences. Philadelphia: National Board of Medical Examiners (NBME).

Bloxham, S., & Boyd, P. (2007). Developing effective assessment in higher education: a practical guide: a practical guide. McGraw-Hill Education (UK).

Bowditch, I. (2023, September 13). Assessment Menu: Designing assessment in an AI enabled world. https://nationalcentreforai.jiscinvolve.org/wp/2023/09/12/designing-assessment-in-an-ai-enabled-world/

Cabral, A., Fuller, S., De Wilde, J., Khama, K., & Melsen, M. (2023). Curriculum enhancement through co-creation: fostering student-educator partnerships in higher education. *International Journal for Students as Partners, 7*(2), 111-123. https://doi.org/10.15173/ijsap.v7i2.5280

Carless, D. (2005). Prospects for the implementation of assessment for learning. Assessment in Education: Principles, *Policy & Practice, 12*(1), 39-54. https://doi.org/10.1080/0969594042000333904

Chaka, C. (2024). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning and Teaching, 7*(1), 1-12. https://doi.org/10.37074/jalt.2024.7.1.14

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, 38. https://doi.org/10.1186/s41239-023-00408-3

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), 2210461. https://doi.org/10.1080/2331186X.2023.2210461

Chiu, T. K. (2024). Future research recommendations for transforming higher education with generative AI. Computers and Education: Artificial Intelligence, 6, 1-9. https://doi.org/10.1016/j.caeai.2023.100197

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of university teaching & learning practice, 20*(3), 1-19. https://doi.org/10.53761/1.20.3.02

Cuthbert, R., & Simpson, A. I. (2023). Artificial intelligence in orthopaedics: can chat generative pre-trained transformer (ChatGPT) pass section 1 of the fellowship of the royal college of surgeons (trauma & orthopaedics) examination? *Postgraduate Medical Journal, 99*(1176), 1110-1114. https://doi.org/doi:10.1093/postmj/qgad053

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gašević, D., & Chen, G. (2023, July 10-13). Can large language models provide feedback to students? A case study on ChatGPT 2023 IEEE International Conference on Advanced Learning Technologies (ICALT), Orem, UT, USA.

Dalalah, D., & Dalalah, O. M. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education, 21*(2), 100822. https://doi.org/10.1016/j.ijme.2023.100822

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education, 26*, 544-560. https://doi.org/10.1007/s40593-016-0106-z

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), 17. https://doi.org/10.1007/s40979-023-00140-5

Elliott, D., & Soifer, E. (2022). AI technologies, privacy, and security. Frontiers in Artificial Intelligence, 5. https://doi.org/10.3389/frai.2022.826737

Entwistle, N. (2017). Teaching for understanding at university: Deep approaches and distinctive ways of thinking. Bloomsbury Publishing.

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

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Fröhling, L., & Zubiaga, A. (2021). Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover. *PeerJ Computer Science, 7*, e443. https://doi.org/10.7717/peerj-cs.443

Fuchs, A., Trachsel, T., Weiger, R., & Eggmann, F. (2024). ChatGPT's performance in dentistry and allergyimmunology assessments: a comparative study. *SWISS DENTAL JOURNAL SSO–Science and Clinical Topics, 134*(2), 1-17. https://doi.org/10.61872/sdj-2024-06-01

Fyfe, P. (2023). How to cheat on your final paper: Assigning AI for student writing. *AI & SOCIETY, 38*(4), 1395-1405. https://doi.org/10.1007/s00146-022-01397-z

Gamage, K. A., Dehideniya, S. C., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: more opportunities than concerns? *Journal of Applied Learning and Teaching, 6*(2), 358-369. https://doi.org/doi:10.37074/jalt.2023.6.2.32

Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has aced the test of understanding in college economics: Now what? *The American Economist, 68*(2), 233-245. https://doi.org/10.1177/05694345231169654

Giannos, P., & Delardas, O. (2023). Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. *JMIR Medical Education, 9*(1), e47737. https://doi.org/doi:10.2196/47737

Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). How does ChatGPT perform on the medical licensing exams? The implications of large language models for medical education and knowledge assessment. MedRxiv, 2022.2012. 2023.22283901. https://doi.org/10.1101/2022.12.23.22283901

Gipps, C., & Stobart, G. (2009). Fairness in Assessment. In C. Wyatt-Smith & J. J. Cumming (Eds.), Educational Assessment in the 21st Century: Connecting Theory and Practice (pp. 105-118). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9964-9_6

Gorichanaz, T. (2023). Accused: How students respond to allegations of using ChatGPT on assessments. Learning: *Research and Practice, 9*(2), 183-196. https://doi.org/10.1080/23735082.2023.2254787

Greiner, C., Peisl, T. C., Höpfl, F., & Beese, O. (2023). Acceptance of AI in semi-structured decision-making situations applying the four-sides model of communication—An empirical analysis focused on higher education. *Education Sciences, 13*(9), 865-875. https://doi.org/10.3390/educsci13090865

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. Education and Information Technologies, 29(7), 8435-8463. https://doi.org/10.1007/s10639-023-12146-0

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific reports, 13*(1), 18617. https://doi.org/10.1038/s41598-023-45644-9

Holmes, W., & Miao, F. (2023). Guidance for generative AI in education and research. UNESCO Publishing.

Hounsell, D., Falchikov, N., Hounsell, J., Klampfleitner, M., Huxham, M., Thomson, K., & Blair, S. (2007). Innovative assessment across the disciplines. An analytical review of the literature. Final Report. HEA.

Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafni, B., Adnan, W., Alhanai, T., AlShebli, B., & Baghdadi, R. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific reports, 13*(1), 12187. https://doi.org/10.1038/s41598-023-38964-3

Indran, I. R., Paranthaman, P., Gupta, N., & Mustafa, N. (2024). Twelve tips to leverage AI for efficient and effective medical question generation: a guide for educators using Chat GPT. *Medical Teacher*, 1-6. https://doi.org/10.1080/0142159X.2023.2294703

Jain, M., Goel, A., Sinha, S., & Dhir, S. (2021). Employability implications of artificial intelligence in healthcare ecosystem: responding with readiness. *foresight, 23*(1), 73-94. https://doi.org/10.1108/FS-04-2020-0038

JISC. (2024). Artificial Intelligence - Embrace artificial intelligence (AI) with confidence. JISC. Retrieved 7th August 2024 from https://www.jisc.ac.uk/innovation/artificial-intelligence

Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity, 20*(2). https://doi.org/10.1007/s40979-024-00149-4

Kaipa, R. M. (2021). Multiple choice questions and essay questions in curriculum. Journal of Applied *Research in Higher Education, 13*(1), 16-32. https://doi.org/10.1108/JARHE-01-2020-0011

Kasai, J., Kasai, Y., Sakaguchi, K., Yamada, Y., & Radev, D. (2023). Evaluating gpt-4 and chatgpt on japanese medical licensing examinations. arXiv preprint arXiv:2303.18027. https://doi.org/10.48550/arXiv.2303.18027

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2024). Gpt-4 passes the bar exam. *Philosophical Transactions of the Royal Society A, 382*(2270), 20230254. https://doi.org/http://dx.doi.org/10.2139/ssrn.4389233

Koh, K., Delanoy, N., Thomas, C., Bene, R., Chapman, O., Turner, J., Danysk, G., & Hone, G. (2019). The role of authentic assessment tasks in problem-based learning. Papers on Postsecondary Learning and Teaching, 3, 17-24. https://doi.org/10.11575/pplt.v3i1.53144

Kumar, V. R., & Raman, R. (2022, March 26). Student Perceptions on Artificial Intelligence (AI) in higher education. 2022 IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, USA.

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., & Maningo, J. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS digital health, 2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lawrie, G. (2023). Establishing a delicate balance in the relationship between artificial intelligence and authentic assessment in student learning. *Chemistry Education Research and Practice, 24*(2), 392-393. https://doi.org/10.1039/D3RP90003G

Liu, Y., Zhang, Z., Zhang, W., Yue, S., Zhao, X., Cheng, X., Zhang, Y., & Hu, H. (2023). ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models. arXiv preprint arXiv:2304.07666. https://doi.org/10.48550/arXiv.2304.07666

Massey, P. A., Montgomery, C., & Zhang, A. S. (2023). Comparison of ChatGPT–3.5, ChatGPT-4, and orthopaedic resident performance on orthopaedic assessment examinations. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 31*(23), 1173-1179. https://doi.org/10.5435/JAAOS-D-23-00396

Meyers, N. M., & Nulty, D. D. (2009). How to use (five) curriculum design principles to align authentic learning environments, assessment, students' approaches to thinking and learning outcomes. *Assessment & Evaluation in Higher Education, 34*(5), 565-577. https://doi.org/10.1080/02602930802226502

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of educational data mining, 4*(1), 11-48. https://doi.org/10.5281/zenodo.3554641

Mitra, N. K., & Chitra, E. (2024). Glimpses of the Use of Generative AI and ChatGPT in Medical Education. *Education in Medicine Journal, 16*(2), 155-164. https://doi.org/10.21315/eimj2024.16.2.11

Naidu, K., & Sevnarayan, K. (2023). ChatGPT: An ever-increasing encroachment of artificial intelligence in online assessment in distance education. *Online Journal of Communication and Media Technologies, 13*(3), e202336. https://doi.org/10.30935/ojcmt/13291

Newton, P. M., & Essex, K. (2024). How common is cheating in online exams and did it increase during the COVID-19 pandemic? A systematic review. *Journal of Academic Ethics, 22*(2), 323-343. https://doi.org/10.1007/s10805-023-09485-5

Newton, P. M., & Xiromeriti, M. (2024). ChatGPT performance on MCQ exams in higher education. A pragmatic scoping review. *Assessment & Evaluation in Higher Education, 49* (6), 781-798. https://doi.org/10.1080/02602938.2023.2299059

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education, 48*(4), 559-614. https://doi.org/10.1080/03043797.2023.2213169

Nikolopoulou, K. (2024). Generative Artificial Intelligence in Higher Education: Exploring ways of harnessing pedagogical Practices with the assistance of ChatGPT. *International Journal of Changes in Education, 1*(2), 103-111. https://doi.org/10.47852/bonviewIJCE42022489

O'Neill, G., & Padden, L. (2022). Diversifying assessment methods: Barriers, benefits and enablers. *Innovations in education and teaching international, 59*(4), 398-409. https://doi.org/10.1080/14703297.2021.1880462

OpenAI. (2024). openai.com/chatgpt. https://openai.com/chatgpt

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(8), em2307. https://doi.org/10.29333/ejmste/13428

Owen, J., & Wasiuk, C. (2021). An agile approach to co-creation of the curriculum. *International Journal for Students as Partners, 5*(2), 89-97. https://doi.org/10.15173/ijsap.v5i2.4475

Passby, L., Jenko, N., & Wernham, A. (2023). Performance of ChatGPT on Specialty Certificate Examination in Dermatology multiple-choice questions. Clinical and experimental dermatology, 722–727. https://doi.org/doi:10.1093/ced/llad197

Pauceanu, A. M., Rabie, N., & Moustafa, A. (2020). Employability under the fourth industrial revolution. *Economics & Sociology, 13*(3), 269-283. https://doi.org/10.14254/2071-789x.2020/13-3/17

Perera, P., & Lankathilake, M. (2023). AI in higher education: A literature review of chatgpt and guidelines for responsible implementation. https://doi.org/10.47772/IJRISS.2023.7623

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of university teaching & learning practice, 20*(2), 07. https://doi.org/10.53761/1.20.02.07

Pichai, S. (2023, February 6). An important next step on our AI journey. Google. https://blog.google/technology/ai/bard-google-ai-search-updates/

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into practice, 48*(1), 4-11. https://doi.org/10.1080/00405840802577536

Pursnani, V., Sermet, Y., Kurt, M., & Demir, I. (2023). Performance of ChatGPT on the US fundamentals of engineering exam: Comprehensive assessment of proficiency and potential implications for professional environmental engineering practice. *Computers and Education: Artificial Intelligence, 5*(100183), 1-12. https://doi.org/10.1016/j.caeai.2023.100183

Rajabi, P., Taghipour, P., Cukierman, D., & Doleck, T. (2023, May 4-5). Exploring ChatGPT's impact on post-secondary education: A qualitative study. Proceedings of the 25th Western Canadian Conference on Computing Education, Vancouver, BC, Canada.

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review, 25*(4), 523-548. https://doi.org/10.1007/s10648-013-9240-4

Rivera-Rosas, C. N., Calleja-López, J. T., Ruibal-Tavares, E., Villanueva-Neri, A., Flores-Felix, C. M., & Trujillo-López, S. (2024). Exploring the potential of ChatGPT to create multiple-choice question exams. *Educación Médica, 25*(100930), 1-8. https://doi.org/10.1016/j.edumed.2024.100930

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Sabzalieva, E., & Valentini, A. (2023). ChatGPT and artificial intelligence in higher education: Quick start guide. UNESCO. https://eduq.info/xmlui/handle/11515/38828

Saher, A.-S., Ali, A. M. 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

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(6), 887-907. https://doi.org/10.3390/healthcare11060887

Saqib, M. B., & Zia, S. (2024). Evaluation of AI content generation tools for verification of academic integrity in higher education. *Journal of Applied Research in Higher Education.* https://doi.org/10.1108/JARHE-10-2023-0470

Sarkar, M., Overton, T., Thompson, C. D., & Rayner, G. (2020). Academics' perspectives of the teaching and development of generic employability skills in science curricula. *Higher Education Research & Development, 39*(2), 346-361. https://doi.org/10.1080/07294360.2019.1664998

Scarfe, P., Watcham, K., Clarke, A., & Roesch, E. (2024). A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. *PloS one, 19*(6), e0305354. https://doi.org/10.1371/journal.pone.0305354

Schultz, M., Young, K., K. Gunning, T., & Harvey, M. L. (2022). Defining and measuring authentic assessment: a case study in the context of tertiary science. *Assessment & Evaluation in Higher Education, 47*(1), 77-94. https://doi.org/10.1080/02602938.2021.1887811

Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *Learning: Research and Practice, 9*(2), 159-167. https://doi.org/10.1080/23735082.2023.2261131

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023, July 20-22). Educator and student perspectives on the impact of generative AI on assessments in higher education Proceedings of the tenth ACM conference on Learning @ Scale, Copenhagen Denmark.

Southworth, J., Migliaccio, K., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: *Artificial Intelligence, 4*, 100127. https://doi.org/10.1016/j.caeai.2023.100127

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in Teacher Education: A scoping review. Computers and Education Open, 6(100169), 1-13. https://doi.org/10.1016/j.caeo.2024.100169

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays-should academics worry? *Nature, 613*, 620-621. https://doi.org/10.1038/d41586-022-04397-7

Stutz, P., Elixhauser, M., Grubinger-Preiner, J., Linner, V., Reibersdorfer-Adelsberger, E., Traun, C., Wallentin, G., Wöhs, K., & Zuberbühler, T. (2023). Ch (e) atGPT? An anecdotal approach addressing the impact of ChatGPT on teaching and learning Giscience. *GI Forum, 1,* 140-147. https://doi.org/10.1553/giscience2023_01_s140

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17

Surapaneni, K. M., Rajajagadeesan, A., Goudhaman, L., Lakshmanan, S., Sundaramoorthi, S., Ravi, D., Rajendiran, K., & Swaminathan, P. (2024). Evaluating ChatGPT as a self-learning tool in medical biochemistry: A performance assessment in undergraduate medical university examination. *Biochemistry and Molecular Biology Education, 52*(2), 237-248. https://doi.org/10.1002/bmb.21808

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Terwiesch, C. (2023). Would Chat GPT3 get a Wharton MBA? A prediction based on its performance in the operations management course. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania, 45.

Vijayakumar, A. (2022). Potential impact of artificial intelligence on the emerging world order. F1000Research, 11. https://doi.org/10.12688/f1000research.124906.2

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840-854. https://doi.org/10.1080/02602938.2017.1412396

Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2020). Using principles of authentic assessment to redesign written examinations and tests. Innovations in education and teaching international, 57(1), 38-49. https://doi.org/10.1080/14703297.2018.1564882

Walker, J. (2018). Time to ditch the traditional essay! *Journal of Writing in Creative Practice, 11*(2), 267-273. https://doi.org/10.1386/jwcp.11.2.267_1

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(15), 1-29. https://doi.org/10.1186/s41239-024-00448-3

Waltzer, T., Pilegard, C., & Heyman, G. D. (2024). Can you spot the bot? Identifying AI-generated writing in college essays. *International Journal for Educational Integrity, 20*(11), 1-18. https://doi.org/10.1007/s40979-024-00158-3

Wang, Y.-M., Shen, H.-W., & Chen, T.-J. (2023). Performance of ChatGPT on the pharmacist licensing examination in Taiwan. *Journal of the Chinese Medical Association, 86*(7), 653-658. https://doi.org/doi:10.1097/JCMA.0000000000000942

Webb, M. (2023, July 12). Generative AI - a primer. An introduction to generative artificial intelligence technology and its implications on education. JISC Reports. Retrieved 7th August 2024 from https://www.jisc.ac.uk/reports/generative-ai-a-primer

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19(*1), 1-39. https://doi.org/10.1007/s40979-023-00146-z

West, C. G. (2023). AI and the FCI: Can ChatGPT project an understanding of introductory physics? arXiv preprint arXiv:2303.01067, 1-13. https://doi.org/10.48550/arXiv.2303.01067

Wiggins, G. (1990). The case for authentic assessment. *Practical assessment, research, and evaluation, 2*(1). https://doi.org/10.7275/ffb1-mm19

Williams, A. (2024). Comparison of generative AI performance on undergraduate and postgraduate written assessments in the biomedical sciences. *International Journal of Educational Technology in Higher Education, 21*(52), 1-22. https://doi.org/10.1186/s41239-024-00485-y

Wilton, L., Ip, S., Sharma, M., & Fan, F. (2022, July 27-31). Where is the AI? AI literacy for educators International Conference on Artificial Intelligence in Education, Durham, UK.

Wood, D., & Moss, S. H. (2024). Evaluating the impact of students' generative AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning, 17*(2), 152-167. https://doi.org/10.1108/JRIT-06-2024-0151

Wood, D. A., Achhpilia, M. P., Adams, M. T., Aghazadeh, S., Akinyele, K., Akpan, M., Allee, K. D., Allen, A. M., Almer, E. D., & Ames, D. (2023). The ChatGPT artificial intelligence chatbot: How well does it answer accounting assessment questions? *Issues in Accounting Education, 38*(4), 81-108. https://doi.org/10.1016/j.xops.2023.100324

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education, 21*(40), 1-22. https://doi.org/10.1186/s41239-024-00468-z

Yeadon, W., Inyang, O.-O., Mizouri, A., Peach, A., & Testrow, C. P. (2023). The death of the short-form physics essay in the coming AI revolution. *Physics Education, 58*(3), 1-13. https://doi.org/10.1088/1361-6552/acc5cf

Zouhaier, S. (2023). The Impact of Artificial Intelligence on Higher Education: An Empirical Study. *European Journal of Educational Sciences, 10*(1), 17-33. https://doi.org/10.19044/ejes.v10no1a17

About the Authors

Andrew Williams, Professor, Division of Medicine, University College London,
andrew.williams@ucl.ac.uk