RaY

Research at the University of York St John

For more information please contact RaY at
ray@yorksj.ac.uk

Embracing a new world: authentic assessment designs in an age of generative artificial intelligence (GenAI)

Feng Su, Michael O'Dea, Margaret Wood & Eva Seidl

Published online: 29 Jan 2026.

Submit your article to this journal ⏎

Article views: 68

View related articles ⏎

View Crossmark data ⏎

Embracing a new world: authentic assessment designs in an age of generative artificial intelligence (GenAI)

Feng Su [a], Michael O'Dea [b], Margaret Wood [c] and Eva Seidl [d]

aSchool of Education and Social Sciences, Liverpool Hope University, Liverpool, UK; bDepartment of Computer Science, University of York, York, UK; cSchool of Education, Language and Psychology, York St John University, York, UK; dDepartment of Translation Studies, University of Graz, Graz, Austria

ABSTRACT

Higher education academic practices are increasingly shaped by technologies. The recent advancement of generative artificial intelligence (GenAI) has posed both challenges and opportunities. In this paper, drawing on relevant literature through a rapid review, the authors argue that rethinking assessment design in an era of artificial intelligence is essential for preparing students for a rapidly evolving world. GenAI can help reconceptualise authenticity through students' transparent, purposeful interactions with this technology. In this shifting landscape, higher education must embrace authenticity as hybrid and relational, fostering assessment practices that value insight, context, and co-constructed meaning in technologically mediated learning environments. To illustrate how authentic assessment designs can be realised in the age of GenAI, four scenarios are presented: (1) the integration of GenAI within a social science assessment task; (2) GenAI-assisted language education; (3) the use of GenAI in a computer database assignment in computer science; and (4) the adoption of GenAI-enhanced assessment design in medical education. These scenarios demonstrate possibilities for aligning assessment designs in higher education with present technological realities. The paper also highlights the importance of developing GenAI literacy among staff and students, positioning such literacy as a foundational capability for effective and ethical engagement with AI-enabled assessment designs.

AI won't replace humans but humans with AI will replace humans without AI. (Lakhani 2023)

The recent advancement of generative artificial intelligence (GenAI) has posed both opportunities and challenges in higher education. AI has brought assessment and academic integrity in higher education to the fore. According to a Higher Education Policy Institute report (HEPI 2025) of their annual survey of UK students on GenAI use, this has dramatically increased over the past year, with some 88% of students reporting that they have used GenAI for assessments in the 2025 survey compared to 54% in the 2024 survey. The key implication of this is that GenAI cannot be ignored and that it is already part of the student learning process. This highlights the imperative for universities to directly address the use of GenAI in all of its applications, both positive and negative. Higher education institutions and academics need to find ways to embrace and leverage the affordances of AI on assessment, and this study suggests that authentic assessment designs may offer some potential solutions.

Farrokhnia et al. (2024, 460) explain that 'Artificial Intelligence (AI) refers to the science and engineering of creating systems that are capable to perform tasks commonly associated with intelligent beings such as learning, judgement, and decision-making'. For Swaffield (2011, 434) '"Authentic assessment" refers to the assessment of learning that is conducted through 'real world' tasks requiring students to demonstrate their knowledge and skills in meaningful contexts'. Ashford-Rowe, Herrington, and Brown (2013) suggest

that there are eight critical elements of authentic assessment based on their literature review – an authentic assessment should be challenging; the outcome of an authentic assessment should be in the form of a performance or product (outcome); authentic assessment design should ensure transfer of knowledge; meta-cognition as a component of authentic assessment; the assessment requires a product or performance that could be recognised as authentic by relevant industry stakeholders; the role of the assessment environment and the tools used to deliver the assessment task; the importance of formally designing is an opportunity to discuss and provide feedback; and the value of collaboration. McArthur (2023) argues that we should move from thinking in terms of either the so-called real world, or the world of work, to focus our justification for authentic assessment on its social value. McArthur (2023) has offered three important points in our conceptual rethinking and practical application of authentic assessment. Firstly, from the real world/world of work to society; secondly, from task performance to why we value the task; and thirdly, from the status-quo of the real-world/world of work to transforming society.

Authenticity in higher education emphasises genuine student engagement, original thought, and real-world relevance in learning and assessment. GenAI-enabled assessment challenges traditional notions of authenticity, since students may use AI to help create their work. But instead of making learning less authentic, GenAI can help us see authenticity in new ways through students' transparent, purposeful interactions with AI. In this evolving landscape, higher education must embrace authenticity as hybrid and relational, fostering assessment practices that value insight, context, and co-constructed meaning in technologically mediated learning environments.

Surveying the landscape: a rapid review

To inform our thinking and writing of this paper, we adopted a rapid review approach to explore the following question – what might authentic assessment designs in higher education look like in the context of GenAI? Hamel et al. (2020) define a rapid review as a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting specific methods to produce evidence for stakeholders in a resource-efficient manner, while also maintaining quality and credibility. Our choice of the rapid review approach was informed by the benefits of resource efficiency combined with rigour to ensure confidence in the credibility of the findings. The following two selected higher education journals were used for the rapid review – *Innovations in Education and Teaching International* and *Assessment & Evaluation in Higher Education*. These were selected because they are two established higher education journals with a focus on academic practice, assessment and evaluation in higher education.

The following keywords were used for the literature search – 'authentic', 'assessment*', 'artificial intelligence*' OR 'ChatGPT' AND 'higher education*' OR 'university*'. The selection of these keywords was determined by the overarching research question. These keywords were therefore selected as likely to focus the search on literature most relevant to our purposes. In addition, we searched only for articles where the above search terms appeared in the article's title, abstract or keywords. Another key inclusion criterion was peer reviewed research articles published in these two journals between 2022 and 2024. The rationale for defining this time period was that one of the main artificial intelligence products 'ChatGPT' developed by OpenAI was launched in 2022. The literature search produced an initial result of 55 peer reviewed articles from these two journals that met the above criteria. After further examination of each of these articles by reading the full text, the following 12 articles were identified as being appropriate for the aim and purpose of this rapid review as shown in Table 1.

In addition, the following blogs were also considered as part of the rapid review process – WonkHE and Times Higher Education: Campus. We included blogs as part of our review because they are popular platforms for scholars to quickly state emerging positions, explore ideas in a less formalised format, and engage in open-ended discussions in a more flexible, conversational style, allowing for engagement with current debates and theories. The search criteria used for the selection of the following blog pieces were 'AI' OR 'GenAI' in the title and a reference to assessment or pedagogy in the initial paragraphs of the article. The research terms for blog pieces are more general and wider because blogs are more discursive and they do not always use keywords. The most recent blog pieces published in 2024 and 2023 were included in the review as shown in Table 2.

Table 1. Journal articles included in the rapid review by title, author, year of publication, journal and context of study.

	Article title	Author(s)	Year	Journal	Context of study
1	Developing evaluative judgement for a time of generative artificial intelligence	Margaret Bearman, Joanna Tai, Phillip Dawson, David Boud & Rola Ajjawi	2024	Assessment & Evaluation in Higher Education	Conceptual discussions
2	ChatGPT performance on multiple choice question examinations in higher education: A pragmatic scoping review	Philip Newton & Maira Xiromeriti	2024	Assessment & Evaluation in Higher Education	Scoping review
3	A SWOT analysis of ChatGPT: Implications for educational practice and research	Mohammadreza Farrokhnia, Seyyed Kazem Banihashem, Omid Noroozi & Arjen Wals	2024	Innovations in Education and Training International	SWOT analysis
4	A critical review of GenAI policies in higher education assessment: a call to reconsider the 'originality' of students' work	Jiahui Luo	2024	Assessment & Evaluation in Higher Education	University policy discourse analysis
5	From authentic assessment to authenticity in assessment: broadening perspectives	Rola Ajjawi, Joanna Tai, Mollie Dollinger, Phillip Dawson, David Boud & Margaret Bearman	2024	Assessment & Evaluation in Higher Education	Conceptual discussions
6	Artificial Intelligence (AI) literacy – an argument for AI literacy in education	Siu-Cheung Kong, Satu-Maarit Korte, Steve Burton, Pigga Keskitalo, Tuija Turunen, David Smith, Lixun Wang, John Chi-Kin Lee & Mhairi C. Beaton	2024	Innovations in Education and Training International	Comment
7	Chatting and cheating: Ensuring academic integrity in the era of ChatGPT	Debby R. E. Cotton, Peter A. Cotton & J. Reuben Shipway	2024	Innovations in Education and Training International	Reflections
8	Challenging cheating in higher education: a review of research and practice	Malcolm Tight	2024	Assessment & Evaluation in Higher Education	Literature review
9	Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices	Alexandra Farazouli, Teresa Cerratto-Pargman, Klara Bolander-Laksov & Cormac McGrath	2024	Assessment & Evaluation in Higher Education	Experimental study
10	Interactive oral assessment case studies: An innovative, academically rigorous, authentic assessment approach	Monica Ward, Fiona O'Riordan, Danielle Logan-Fleming, Dervila Cooke, Tara Concannon-Gibney, Marina Efthymiou & Niamh Watkins	2024	Innovations in Education and Training International	Case studies
11	Designing the digital in authentic assessment: is it fit for purpose?	Juuso Henrik Nieminen, Margaret Bearman & Rola Ajjawi	2023	Assessment & Evaluation in Higher Education	Scoping review
12	Designing assessment in a digital world: an organising framework	Margaret Bearman, Juuso Henrik Nieminen & Rola Ajjawi	2023	Assessment & Evaluation in Higher Education	Conceptual framework

Table 2. Blog pieces included in the rapid review by title, author, publication date and blog platform.

	Blog article title	Author(s)	Publication date	Blog platform
1	Access to premium AI services is a significant concern for students	Naima Rahman and Gunter Saunders	11/9/2024	WonkHE
2	Why are London-based students using AI more than the rest of UK	Sunday Blake	9/9/2024	WonkHE
3	Higher education's engagement with AI will have to keep a keen eye on the world of work	Eve Alcock, Ailsa Crum and Nick Watmough	20/2/2024	WonkHE
4	'Your work must be your own'. What does that mean?	Jim Dickinson	12/2/2024	WonkHE
5	There's no point comforting ourselves over AI and cheating when we don't know what cheating is.	Jim Dickinson	2/2/2024	WonkHE
6	Of course you can't detect students' use of AI. So what next?	Jim Dickinson	3/8/2023	WonkHE
7	The real risk of GenAI is a crisis of knowledge	Joshua Thorpe	27/6/2023	WonkHE
8	A four-step process to embedding AI literacy in business courses	John Murphy	10/10/2024	Times Higher Education: Campus
9	AI potential 'squandered by universities' risk focused approach'	John Ross	12/8/2024	Times Higher Education: Campus
10	10 tips when instigating an assessment transformation programme	Steve Briggs	17/9/2024	Times Higher Education: Campus
11	Why I ban AI use for writing assignments	James Stacy Taylor	12/7/2024	Times Higher Education: Campus
12	Assessment curriculum design can't ignore students' use of AI	Dorottya Sallai	23/7/2024	Times Higher Education: Campus
13	Students fear over reliance on AI devalues HE	Juliette Rowsell	7/8/2024	Times Higher Education: Campus
14	AI exam answers undetected and outperform students at UK campus	Juliette Rowsell	27/6/2024	Times Higher Education: Campus

Potential benefits and risks of using AI tools in higher education

The rapid review shows that there are potential benefits and potential risks of using AI tools such as ChatGPT in higher education (Cotton, Cotton, and Shipway 2024). Some of the potential risks pointed to include the risk of plagiarism and how universities might deal with this challenge to the security of assessment practices. An interesting study by Farazouli et al. (2024) which focused on the use of AI chat-bots and university teachers' assessment practices found that 'ChatGPT may have an impact on university teachers' assessment practices in amplifying their criticality and suspicion and challenging their trust in students' texts' (372). Farrokhnia et al. (2024) suggested that 'Students may utilise ChatGPT due to its promising capabilities without realising that it may lead to plagiarism' (467). One of the benefits high-lighted in the literature is the opportunity offered by tools such as ChatGPT to generate personalised assessments. As such, educators could base 'customised exams or quizzes for each student […] on their individual needs and abilities' (Cotton, Cotton, and Shipway 2024, 229). Another potential opportunity of GenAI lies in the fact that it might enable enhanced dialogue, joint reflection and partnering with students for a better understanding of the implications of the use of AI in assessment (Bearman et al. 2024; Luo 2024; Tight 2024).

GenAI tools can improve assessment for students with disabilities by enhancing accessibility through text-to-speech, summarising complex texts, generating personalised study materials, and translating or rephras-ing instructions. GenAI tools offer features like customisable font settings and help to overcome language barriers. By integrating with assistive technologies, GenAI tools provide alternative ways to access and process information, creating more equitable opportunities for students with diverse needs during assess-ments. For instance, Zhao, Cox and Chen's (2025) research shows that GenAI can support students with dis-abilities, particularly those with dyslexia, in comprehending learning materials by improving their reading efficiency and effectiveness.

Some authors point to the importance of a mixed methods approach to assessment in order to mitigate some concerns of using GenAI tools. For example, some have advocated the inclusion of a viva or oral element in the assessment design. Ward et al. (2024) report on the use of interactive oral assessment 'as an online, innovative authentic assessment approach that prepares students for professional life, combats plagiarism and promotes academic integrity' (930). With the relationship of assessment and the digital in focus, Bearman, Nieminen, and Ajjawi (2023) present an 'organising framework, which articulates this nuanced relationship between the digital and assessment design' (291). The framework has three purposes: 'as a tool for improvement; as a means to develop and credential digital literacies; and as a means to develop and credential uniquely human capabilities' (291). A critical scoping review of the higher education literature by Nieminen, Bearman, and Ajjawi (2023, 529) noted that 'in order to fit its purpose of preparing students for the digital world, the digital needs to be designed into authentic assessment in meaningful ways' (529). Context, meaning and purpose have also been defined as critically relevant, in that assessment design choices 'need to be made about what dimensions or aspects of authenticity matter in a particular assess-ment' (Ajjawi et al. 2024, 506). Farazouli et al. (2024) remind us of the importance of considering different assessment cultures in different disciplines.

Some academics have argued that the use of GenAI is a double-edged sword. Greater incorporation of GenAI into learning and teaching runs the risk of over reliance and uncritical use of it by students. Concern has been raised that overuse will result in an underdeveloped ability in students to engage with arguments due to uncritical use by students (Sallai 2024). For instance, AI tools offering quick sol-utions and ready-made information can discourage students from engaging in the cognitive processes essential for critical thinking. Risko and Gilbert (2016) describe this as cognitive offloading which involves using external tools to reduce the cognitive load on an individual's working memory. While this can free up cognitive resources, it may also lead to a decline in cognitive engagement and skill development. Following this line of argument, it might be suggested that students have been trained to think extrin-sically and instrumentally about the outcomes of their work. Arguably, it could make it difficult for them to see the intrinsic value and self-transformative value that comes from undertaking challenging tasks. GenAI exacerbates this as it removes problems (Thorpe 2023). Some academics report that prohibiting the use of GenAI for certain assessments and/or at lower levels of study may be acceptable and reason-able (Sallai 2024; Taylor 2024).

Research shows that academics cannot detect all AI outputs at present and it is likely that it will become increasingly difficult to distinguish from human outputs. Figures show a high proportion of students using GenAI (Sallai 2024). The Digital Education Council Global AI Student Survey, 2024, reports 86% of students regularly use GenAI in their studies (Rowsell 2024). This is likely to be an underestimate, as students are likely to be underreporting in surveys. Many students are scared to declare their use and many are unclear of what is acceptable use (Dickinson 2024; Rowsell 2024).

Students are progressing faster in their use of GenAI than academics (Ross 2024). Although the majority are aware of university guidelines, they are still unclear as to what is acceptable use in their studies and regard many university guidelines as unhelpful in this respect (Ross 2024). It appears that students are probably not using GenAI in inappropriate and dishonest ways, but are using it to manage workload (Blake 2024) and figures suggest only 5% of students are using GenAI to 'cheat' (Dickinson 2024). This suggests that a shift of focus from detecting AI content and emphasising academic misconduct to embracing and integrating GenAI into the teaching and assessment process would be more appropriate. There has been some development towards this shift. For instance, the AI Assessment Scale proposed by Perkins, Roe, and Furze (2025) is now well adopted by practitioners and is directly addressing this issue. However, in our view, there is still some way to go to embrace and integrate GenAI into the assessment process.

A recent study by Luo (2024) on generative AI (GenAI) in higher education assessment found that 'higher education policies tend to frame the problem of work originality from the angle of academic misconduct … a more inclusive approach that extends beyond viewing originality through the narrow lens of surveillance should be reflected in future higher education policies' (661). In this regard, Tight (2024) asserts that students are frequently over-assessed which is why he makes a call for doing less better. By reducing the coverage of programmes, courses, and assessments, students and teachers would profit from 'more time to build relationships and develop academic integrity' (920). Since academic integrity and authenticity in authorship, i.e. originality, are closely linked, 'having fewer but more meaningful assessments' (Tight 2024, 917) is reported as beneficial for students and teachers alike.

At the same time, students need to be aware of concerns regarding the quality of GenAI outputs. Bearman et al. (2024) note the increasing importance of 'evaluative judgement'. They argue that

> [t]he relationship between evaluative judgement and generative AI is more than just the application of human judgement to machine outputs. We have a collective responsibility, as educators and learners, to ensure that humans do not relinquish their roles as arbiters of quality. (893)

This shared responsibility is also key in Tight's (2024) call for stronger student-teacher relationships which might strengthen academic integrity and reduce the propensity for academic misconduct. It is highlighted as a key responsibility of educators to design meaningful, creative, enjoyable assessments which are relevant for students' future careers in a digital world (Ward et al. 2024). Considering that a sense of self is 'a human characteristic, not possessed by machines in the foreseeable future', Bearman et al. suggest that 'educators should consider how assessment can orient students towards a future self in a digital world' (2023, 299).

Our analysis of the literature shows that universities' responses to GenAI are slow and often they do not address key issues. A risk-based approach to the use of AI by universities in learning and teaching means teaching is focused upon teaching students about GenAI and academics are not using GenAI to teach (Ross 2024). Academics have been discouraged from using GenAI to teach, by this approach. Studies show only 37% of academics use AI as a teaching tool and it is mainly to inform students about academic integrity (Ross 2024; Murphy 2024). GenAI is now integrated into everyday life, and therefore universities need to accept it and to work with employers to develop AI literate graduates (Sallai 2024). Wong et al. (2022, 1341) have noted how 'the emergence of graduate attributes has been attributed to the marketisation of higher education' and they are marketed by universities as advantageous for employability. Nartey (2024, 10) has noted 'As technology continues to evolve, the relevance of AI in the job market becomes increasingly apparent, necessitating a thorough examination of how GenAI can contribute to the development of competencies essential for career readiness'. There is an argument therefore for alignment of GenAI with the embedding of employability skills in the curriculum, although this is not unproblematic for, as Wong et al. (2022, 1341) have noted 'Although it is generally agreed that students' development of graduate attributes should be embedded into the curriculum, staff have reportedly struggled with the lack of time, resource and confidence to embrace this approach'.

However, there is concern about equitable access to GenAI tools for all (Rahman and Saunders 2024) which raises questions about equity, justice concerns and inclusivity in the assessment of students. Rahman and Saunders (2024) for example discussed the issue of access to premium AI services, where 'academic success could be influenced by a student's ability to afford advanced AI tools'. They noted that 'the advanced tools and resources available to those who can afford these services can potentially impact academic performance'. Such questions about variance in access to tools is an important one to consider in terms of inclusivity and equity in opportunities to achieve success. To address this issue, some universities in the UK have purchased enterprise licences of GenAI applications for student and staff use. Overall, the literature points to a need for universities to support the AI literacy of staff and students and to consider issues of inclusive practice in terms of access to GenAI tools. In addition, there was a recognition in the literature that there are environmental ethical dimensions in the use of AI (Baum and Owe 2023)

If GenAI can critique itself, how will assessors know that AI tools have not been used in student assessment? To address this, it appears that academic tutors may need to rely on a combination of process, content analysis, and direct assessment methods, rather than purely on AI detection tools, to determine if AI tools are used in student work. Other proposals to address this important issue have encompassed both practical and theoretical solutions. At a practical level, AI assessment frameworks aimed at both allowing tutors to clarify what types of AI use are acceptable when undertaking assessments, and encouraging student transparency of use have been developed and adopted. On a more theoretical level, concepts such as the postplagiarism world (Eaton 2023; Kenny 2025) and generativism (Pratschke 2024) are pointing towards the need for paradigm changes in the way assessments are, not just developed, but conceptualised, within a wider reconsideration of learning and teaching processes and even educational aims and outcomes. Eaton (2023) defined a 'postplagiarism world' as an era in human society and culture in which advanced technologies such as artificial intelligence and neurotechnology become a normal part of life, including how we teach, learn, communicate, and interact on a daily basis. In Pratschke's view (2023), generativism describes the symbiotic approach to teaching and learning with GenAI. Generativism is characterised by dynamic, personalised and adaptive learning experiences that involve co-creation, co-facilitation and co-assessment using GenAI tools. Generativism embraces the agility enabled by AI, and it encourages educators to rethink how they design learning experiences that are 'social, collaborative, community-oriented and human-centred' (Pratschke, 2023, 8). In our view, these are the beginning of a move towards AI enlightened pedagogies, which involve not just the adaptation, integration and incorporation of AI in the learning and teaching practice, but the transformation of it.

In the next section, we offer some examples of GenAI assessment designs at a practical level.

GenAI in practice: scenarios of authentic assessment designs in the age of generative artificial intelligence

In this section, four scenarios drawn from practice in higher education settings are recounted. These scenarios illustrate aspects of authentic assessment and our thinking about these was informed by reading about authentic assessment design, for example in the work of Ashford-Rowe, Herrington, and Brown (2013). In the first scenario, AI is designed into an assessment task with real world relevance. Scaffolded by a series of critical questions, students are required to critically evaluate the veracity and quality of an AI generated response to a specific topic. In the second scenario, the assessment task requires a clear outcome in the form of an audio or video of an informal conversational role play activity for which the use of GenAI is permitted in line with specific protocols in place. The third scenario concerns an assessment task for students in a computer database class on a Computer Science course. This task challenged students to produce work of better quality than an attempt at the same task produced by ChatGPT output. The tutor engaged students in discussion of the assessment criteria and shared the rationale for updating the marking criteria to stress higher order thinking and metacognition through the requirement to reflect on their learning process and work. The fourth scenario illustrates how AI is adopted in assessment design in the area of medical education. Whilst the scenarios are drawn from different and varied contexts, subjects, and student groups and serve to illustrate varied approaches to assessment design, they each include elements of 'authentic' assessment as discussed above, and each makes use of GenAI from the scale of 'AI-assisted editing' to 'AI task completion, human evaluation' (Perkins et al. 2024).

Scenario 1: the integration of GenAI within a social science assessment task

This is a final year undergraduate assignment in social sciences at a university in England. Students are asked to explore the future trends of education in 50 years in a particular context in light of ongoing technological, cultural and socio-political changes in a 2,500-word essay. As part of this assignment, students are invited to use generative AI technology such as ChatGPT to produce a 500-word response on the topic. Students are allowed to use any text generative AI programmes of their choice for this assignment. Students are then required to critically evaluate responses from generative AI technology with their own takes and reflections on the topic (2,000 words). It is made clear that students are not allowed to use AI for the second part of the assignment for their own reflections.

When students evaluate responses from AI, they need to consider the following questions – Is the AI response supported by evidence? Is the response fully truthful? How do they assess the quality of the supporting evidence cited in the AI response? Are there any inaccuracies? Is there any bias in the answer? What are the strengths and weaknesses of the AI's response? Do they agree or disagree with what the AI suggested? If they do not agree with the AI response, what are their own takes on the question? When students submit their work, they are required to specify which generative AI programme they used, and the specific AI prompt they used. In order to complete this assignment, students would need to have some basic knowledge and experience of using generative AI programmes. Also, students needed to have some basic training on writing an AI prompt. For example, students were recommended to ask the AI a specific question that requires it to reflect, discuss, explain or synthesise relevant information within the required word count.

Students' feedback on this assessment design showed that the majority of the cohort appreciated the clear instructions on how GenAI could be used in this assignment. Such assessment design enabled students to learn about GenAI and its potential applications. Students found the initial training on using some relevant GenAI tools, e.g. ChatGPT, Gemini, Copilot, Deepseek, were the key to develop their confidence and understanding of such technology. At the same time, a few students were concerned about potential unequal access to some advanced GenAI functions through paid subscription services. In addition, some students were critical of the accuracy of generated responses, out-of-date answers, and unverifiable sources of information.

By attending to the digital literacy needs of students, this assessment attempted to equip them with the necessary skills to understand and use AI tools to enhance inclusivity. Indeed, as Kong et al. (2025, 2) have argued 'As AI-based technologies used in education increase, the development of AI literacy becomes essential'. However, the issue of unequal access discussed by Rahman and Saunders (2024) for example, is illustrated here in terms of disparities in affordability of some advanced GenAI capabilities creating 'an uneven playing field', which runs counter to considerations of inclusion and equal opportunities for all students to succeed.

Scenario 2: GenAI assisted language education

The background to this scenario is translation and interpreting students' language education at an Austrian university for a German language course entitled 'Intralingual textual competence'. As a scholarly practice and a field, translation and interpreting oriented language learning and teaching (TILLT) promotes heightened awareness of the critical importance of students' linguistic skills, their cross-cultural competence and mediation skills for their vocational future in a wide range of fields of professional activity. As such, assessment tasks in the TILLT classroom should be designed to support students with processes of enculturation in a field of study and professional activity with great social relevance, the latter being stressed by McArthur (2023). In light of these considerations, this assessment scenario focuses on students' ability development in critical reading and critical thinking.

The formative assessment task consists of two reading assignments and the production of an audio or video file in groups of three. The required reading was one text about legal literacy (10 pages) by a didactics expert in legal education and another one (12 pages) about the challenges of teaching legal translations by a translation studies scholar. After the first phase of intensive, critical reading and information acquisition students were asked to imagine the following scenario – imagine that a law student and the authors of text 1 and text 2 meet each other at a panel discussion on legal literacy. Later, at the cafe, they engage in a lively conversation and exchange of ideas. Students then have to produce an audio or video file (5–7 min long) in

which they take on the roles of the three panel discussion participants. They should not simulate the formal public discussion but instead their concern is with the informal conversation at the buffet. In this assessment, students are allowed to use a generative AI programme in their preparation but they are required to specify how they have used AI for this purpose.

When reflecting on this assessment task, students reported that it was much easier for them to identify as the persons whose roles they should take for the production of a recorded informal discussion when they spent time reading the required texts on their own, without the help of AI. They were able to see things through the authors' eyes and to engage deeply with the arguments put forward in the reading assignments. One group of students chose an AI tool for the audio production because they did not like to hear their own voices. This seems difficult for two main reasons. First, this AI generated recording sounded very impersonal and without any warmth in tone, as compared to the lively voices of the other student groups. Second, as prospective translators or interpreters, students should get used to recording and evaluating their own voices, since their voice is one of their main tools when they earn their living as interpreters for clients who, in situations where the stakes are high, could not understand each other without the help of professional linguistic and cross-cultural language service providers.

This scenario has been included as a further illustration of assessment with a focus on student critical engagement with AI within creative and engaging learning strategies designed to stimulate informed informal conversations and critical thinking. The development of critical thinking is arguably an essential part of constructing an educated citizenry (Nixon 2008).

Scenario 3: pre-release review of assessments through GenAI

The assessment task is for a database systems class, on a second year Computer Science undergraduate module. The submission element of the assessment is a report of the database design process for their chosen database. In this scenario the assessment is written and then submitted to GenAI (ChatGPT) prior to release to the students. The outputs of the GenAI are then reviewed and the assessment marking criteria adjusted to stress higher order thinking skills. Greater weighting is placed on solution match to case study details, description of process and innovation in solution proposed and implemented.

Students are given free choice on what the system, subject and content were, but clear guidelines of the outputs are specified. The assessment is in 2 parts, the first is the design and specification of the system and the second is the implementation. Part 1 the design and specification is pre-submitted to ChatGPT by the lecturer. An example similar to a case study used within the taught classes is used. The ChatGPT output is then presented to the class and run through. Students are informed that their submissions need to be superior to the ChatGPT example to pass the module. Where their work needs to be superior is noted in class and annotated on the example output and a Q&A session follows. The updated marking criteria stresses clear exposition of process and justification of decisions made. Evaluation and analysis of choices when decisions had to be made is given extra weighting. Marks are allocated for students' own reflection of what they have attempted to include in their submissions which are over and above the basic requirements for the system and how that improves the end result of the system, in what way and for what reason.

The aim of the adjustment to the assessment process by pre-submission to ChatGPT prior to release to the students was primarily to ensure clarity of permissible GenAI use and also transparency of the assessment marking process. At the point of release the university where the assessment was being delivered had not adopted a traffic light system or an assessment scale for use of GenAI in assessments, so this was an attempt at providing guidelines, such as by stressing in the demonstration and following Q&A session the weighting of marks, i.e. that it rewarded higher order thinking skills. That the assessment was designed to be an authentic assessment meant that, upon analysis, the assessment was suited towards such a treatment with a relatively straightforward adjustment of the marking criteria required. It exhibited many of the characteristics of Ashford-Rowe et al.'s (2013) eight critical elements: it was challenging in that it was a complex multi-part assessment based upon a scenario drawn from a real-life situation, the outcome was a product, and the process was part of the assessment and marks were allocated to this, and there was an element of metacognition in the assessment with evaluation, analysis and justification required. Overall, students responded positively to the assessment and the quality of work submitted was good, with the marker not having to second guess the issue of plagiarism and being able to focus on the creative and innovative aspects of the student work.

This scenario illustrates a response to current debates about uses of GenAI and challenges of this new technology for student learning. This is especially in relation to the need for students to be discerning when using AI tools to access information and to be able to critically evaluate its veracity, particularly as AI can be prone to fabrication. In this scenario higher order thinking skills were foregrounded, and students were actively challenged to develop these skills to develop the quality of their work. This assessment task promoted students' use of evaluative judgement, something which Bearman et al. (2024) have suggested is urgently needed.

Scenario 4: integrating an AI assessment into medical education

This is a second-year undergraduate assignment in applied medical sciences at a London university (Williams 2023). In the previous assessment students were tasked with writing a grant proposal on a chosen human disease. However, a new approach is being adopted using AI-generated grant proposals created by Chat GPT. Students are required to evaluate these AI-generated proposals using specific criteria, including scientific accuracy, background coverage, rationale, and experimental feasibility. This shift in assessment design aims to foster critical thinking and evaluation skills among the students. AI has an integral role in the assessment and students are encouraged to leverage AI extensively.

When students evaluate AI-generated proposals, they need to use the following criteria – (1) scientific accuracy: is the science described an accurate representation of the field? Provide your own critical evaluation; (2) scientific background: does the grant application cover the background science in sufficient detail? Make any additions you may deem appropriate; (3) rationale for the study: does the grant describe the rationale for the study clearly? Is this rationale provided in enough depth and is it convincing? Provide your own critical evaluation; (4) experimental design: do you think the experimental design described is appropriate and feasible? Justify your reasons; (5) outcome of proposed research: what will be the ultimate benefit to our understanding of the disease or benefit to patients? Provide your own summary of the implications of this proposed research study.

Williams (2023) reflected that this shift in the above assessment design aims to foster students' critical thinking and evaluation skills. The decision to embrace AI in assessments was influenced by existing case studies in the sector and close collaboration with module leads and students. There were a number of challenges in implementing this new assessment design, for example, some students' unfamiliarity with the GenAI tool ChatGPT. To address this, the module team used workshops and formative exercises to train students in using ChatGPT and to consider ethical considerations. Despite some of the challenges, Williams (2023) noted that many students welcomed the opportunity to utilise the GenAI tool as part of the assessment which allowed them to learn about ChatGPT's capabilities and limitations from first-hand experience.

Williams (2023) offers some helpful pointers regarding incorporating AI in future assessments. One of these is to 'Think about building generative AI skills in both your staff and students', which is an important point to consider. Arguably the development of AI literacy for students is often discussed in the literature and alongside this there is a need to also enable staff to gain the necessary skills. Another of the five suggestions from Williams is to 'use it as an opportunity to move away from that traditional standard assessment paradigm of MCQs, SEQs and essays', a point which is illustrated in the varied assessment strategies exemplified across these scenarios.

Concluding remarks: implications for higher education academics and managers

GenAI has the potential to not just enable authentic assessment, but by its very nature if incorporated in ways such as those demonstrated by the four scenarios, the assessment becomes authentic almost by doing so. At the same time, it needs to be recognised that, in order to adopt AI effectively for learning and teaching in higher education, priority needs to be given to the development of academics' and students' generative AI literacy which enables them to become critically aware of the affordances and limitations of using AI in a particular learning context. AI literacy (JISC 2024) refers to the ability to understand and use generative AI responsibly and ethically across three fundamental areas: terms, tools, and tasks. O'Dea and O'Dea (2023) suggest that 'to not keep on top of AI literacy runs the risk of disadvantaging both our graduates and our society', and they further argue that.

> Our graduates cannot fear AI but should understand it, so that we can, as a society, contain the worst excesses of the technology and harness its abilities for the benefit of all. Universities must provide AI literacy training to their students.

Therefore, it is important for higher education institutions to develop academics' and students' AI literacy as part of their learning and teaching strategies. GenAI enabled authentic assessment designs need to reflect the real world, and world of work. When this happens, authentic assessment enabled by GenAI has the potential to transform students' learning.

As the literature has shown, the most pressing issue is the need to implement and standardise clear guidance on when, for what purposes, and in what ways students may use GenAI in assessments. AI Assessment frameworks, such as those exemplified by Perkins et al. (2024) are now sufficiently mature for these to be adopted on an institutional level. Not only would these ensure clarity of use for students (and staff), but they would also provide a tool to enable greater transparency of use than is currently available. However, whilst such measures are a step forward, to address the pressing current needs, it appears that we are now at a point where educational leaders and management need to consider some of the fundamentals of learning and teaching. Approaches to education such as those set out in the work on postplagiarism (Eaton 2023) and generativism (Pratschke 2024), incorporate and address new and disruptive technologies, such as GenAI, in such a way as to integrate educational theory and practice with the realities of the ever-changing technologies we are seeing now and the rapidity of their development and adoption, not just within the education realm, but in the industry.

There is no 'silver bullet' to addressing the disruption to the education process precipitated by GenAI, but there perhaps is a set of 'silver buckshot' measures. These would operate at a number of levels. At the operational level for academics and students, authentic assessment can be a way of educators addressing and adapting to the changes required in assessments. It could, and should, go hand in hand with the implementation of AI literacy programmes for students and staff, and the adoption of AI assessment frameworks, which would function at the operational level to address the adaptation of authentic assessment designs to the changed landscape. This process could include the integration of GenAI in curriculum design and collaboration of different stakeholder groups, such as academics, students, external professional bodies and industries. In addition, there needs to be clear guidance for both academics and students on when and how GenAI tools can be used in student assessment. At a strategic level, for institutional leaders and policy makers, an examination of their approaches to the fundamentals of higher education practice needs to take place. In this debate the ideas and concepts proposed in the work on postplagiarism and generativism need to be considered. If higher education is to prepare students for the future, a debate about embracing GenAI as part of their current education needs to be taking place now.

Acknowledgements

Disclosure statement

Notes on contributors

Feng Su is Associate Professor in Education and Head of the School of Education and Social Sciences at Liverpool Hope University, UK. The prime focus of his current research and writing is academic practice and professional learning in higher education.

Michael O'Dea is Lecturer in Computer Science at University of York, York, UK. His research interests are centred around Computer Science education and AI Literacy.

Margaret Wood is Associate Professor in Education at York St John University, York, UK. Social justice and inclusive approaches to the development of academic practice are connective themes across her body of work.

Eva Seidl is Lecturer in German Language and Culture at University of Graz, Graz, Austria. Her research focuses on scholarship of teaching and learning (SoTL) and translation – and interpreting-oriented language learning and teaching (TILLT).

ORCID

Feng Su 🆔 http://orcid.org/0000-0003-0317-4033
Michael O'Dea 🆔 http://orcid.org/0000-0003-2112-0194
Margaret Wood 🆔 http://orcid.org/0000-0001-5067-1978
Eva Seidl 🆔 http://orcid.org/0000-0002-4094-6381

References

Ajjawi, R., J. Tai, M. Dollinger, P. Dawson, D. Boud, and M. Bearman. 2024. "From Authentic Assessment to Authenticity in Assessment: Broadening Perspectives." *Assessment & Evaluation in Higher Education* 49 (4): 499–510. https://doi.org/10.1080/02602938.2023.2271193.

Alcock, E., A. Crum, and N. Watmough. 2024, February 20. "Higher Education's Engagement with AI Will Have to Keep a Keen eye on the World of Work. *WonkHE*. Accessed December 6, 2024. https://wonkhe.com/blogs/higher-educations-engagement-with-ai-Will-Have-to-keep-a-keen-eye-on-the-world-of-work/.

Ashford-Rowe, K., J. Herrington, and C. Brown. 2013. "Establishing the Critical Elements That Determine Authentic Assessment." *Assessment & Evaluation in Higher Education* 39 (2): 205–222. https://doi.org/10.1080/02602938.2013.819566

Baum, S. D., and A. Owe. 2023. "Artificial Intelligence Needs Environmental Ethics." *Ethics, Policy & Environment* 26 (1): 139–143. https://doi.org/10.1080/21550085.2022.2076538.

Bearman, M., J. H. Nieminen, and R. Ajjawi. 2023. "Designing Assessment in a Digital World: An Organising Framework." *Assessment & Evaluation in Higher Education* 48 (3): 291–304. https://doi.org/10.1080/02602938.2022.2069674

Bearman, M., J. Tai, P. Dawson, D. Boud, and R. Ajjawi. 2024. "Developing Evaluative Judgement for a Time of Generative Artificial Intelligence." *Assessment & Evaluation in Higher Education* 49 (6): 893–905. https://doi.org/10.1080/02602938.2024.2335321.

Blake, S. 2024, September 9. "Why Are London-based Students Using AI More than the Rest of the UK." *Wonkhe*. https://wonkhe.com/wonk-corner/why-Are-london-based-students-using-ai-more-than-the-rest-of-the-uk/.

Briggs, S. 2024, September 1. "10 Tips When Instigating an Assessment Transformation." *Times Higher Education Campus*. Accessed December 6, 2024. https://www.timeshighereducation.com/campus/ten-tips-When-instigating-assessment-transformation-programme.

Cotton, D. R. E., P. A. Cotton, and J. R. Shipway. 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.

Dickinson, J. 2023, August 3. "Of course You Can't detect students' use of AI. So What Next? *WonkHE*. Accessed December 6, 2024. https://wonkhe.com/blogs/of-course-You-cant-detect-students-use-of-ai-So-What-Next/.

Dickinson, J. 2024, February 12. "There's No point comforting Ourselves over AI and Cheating When We don't Know What Cheating Is." *Wonkhe*. https://wonkhe.com/wonk-corner/theres-No-point-comforting-Ourselves-over-ai-and-cheating-When-We-dont-know-What-cheating-Is/.

Eaton, S. E. 2023. "Postplagiarism: Transdisciplinary Ethics and Integrity in the age of Artitficial Intelligence and Neurotechnology." *International Journal for Educational Integrity* 19 (23): 1–10. https://doi.org/10.1007/s40979-023-00144-1.

Farazouli, A., T. Cerratto-Pargman, K. Bolander-Laksov, and C. McGrath. 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.

Farrokhnia, M., S. K. Banihashem, O. Noroozi, and A. Wals. 2024. "A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research." *Innovations in Education and Teaching International* 61 (3): 460–474. https://doi.org/10.1080/14703297.2023.2195846.

Hamel, C., A. Michaud, M. Thaku, B. Skidmore, A. Stevens, and B. Nussbaumer-Streit. 2020. "Defining Rapid Reviews: A Systematic Scoping Review and Thematic Analysis of Definitions and Defining Characteristics of Rapid Reviews." *Journal of Clinical Epidemiology* 129:74–85. https://doi.org/10.1016/j.jclinepi.2020.09.041

Higher Education Policy Institute (HEPI). 2025. "Student Generative AI Survey 2025." HEPI Policy Note 61. https://www.hepi.ac.uk/2025/02/26/student-generative-ai-survey-2025/.

Kenny, K. 2025, January 15. "We Have to Rethink Academic Integrity in a 'Post-plagiarism Era." *Times Higher Education Campus*. https://www.timeshighereducation.com/campus/We-Have-rethink-academic-integrity-postplagiarism-era.

Kong, S.-C., S.-M. Korte, S. Burton, P. Keskitalo, T. Turunen, D. Smith, L. Wang, J. C.-K. Lee, and M. C. Beaton. 2025. "Artificial Intelligence (AI) Literacy – An Argument for AI Literacy in Education." *Innovations in Education and Teaching International* 62 (2): 477–483. https://doi.org/10.1080/14703297.2024.2332744

Lakhani, K. R. 2023. "AI Won't Replace Humans – but Humans with AI Will Replace Humans without AI." *Harvard Business Review*. https://hbr.org/2023/08/ai-wont-replace-humans-but-humans-with-ai-Will-replace-humans-without-ai.

Luo, (Jess), J. 2024. "A Critical Review of GenAI Policies in Higher Education Assessment: A Call to Reconsider the 'Originality' of Students' Work." *Assessment & Evaluation in Higher Education* 49 (5): 651–664. https://doi.org/10.1080/02602938.2024.2309963

McArthur, J. 2023. "Rethinking Authentic Assessment: Work, Well-Being, and Society." *Higher Education* 85:85–101. https://doi.org/10.1007/s10734-022-00822-y.

Murphy, J. 2024, October 10. "A Four-Step Process to Embedding AI Literacy in Business Courses." *Times Higher Education Campus*. https://www.timeshighereducation.com/campus/fourstep-process-embedding-ai-literacy-business-courses.

Nartey, E. K. 2024. "Guiding Principles of Generative AI for Employability and Learning in UK Universities." *Cogent Education* 11 (1): 1–27. https://doi.org/10.1080/2331186X.2024.2357898.

Newton, P., and M. Xiromeriti. 2024. "ChatGPT Performance on Multiple Choice Question Examinations in Higher Education: A Pragmatic Scoping Review." *Assessment & Evaluation in Higher Education* 49 (6): 781–798. https://doi.org/10.1080/02602938.2023.2299059

Nieminen, J. H., M. Bearman, and R. Ajjawi. 2023. "Designing the Digital in Authentic Assessment: Is It fit for Purpose?" *Assessment & Evaluation in Higher Education* 48 (4): 529–543. https://doi.org/10.1080/02602938.2022.2089627

Nixon, J. 2008. *Towards the Virtuous University: The Moral Bases of Academic Practice.* New York, NY and Abingdon, Oxon: Routledge.

O'Dea, X., and M. O'Dea. 2023, June 9. "Teaching AI Literacy: How to Begin." *Times Higher Education Campus*. https://www.timeshighereducation.com/campus/teaching-ai-literacy-How-begin.

Perkins, M., L. Furze, J. Roe, and J. MacVaugh. 2024. "The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment." *Journal of University Teaching and Learning Practice* 21 (6): 1–18. https://doi.org/10.53761/q3azde36.

Perkins, M., J. Roe, and L. Furze. 2025. "Reimagining the Artificial Intelligence Assessment Scale (AIAS): A Refined Framework for Educational Assessment." *Journal of University Teaching and Learning Practice* 22 (7): 1–26. https://doi.org/10.53761/rrm4y757.

Pratschke, M. (2023). 'Generativism: the New Hybrid.' Unpublished. https://arxiv.org/pdf/2309.12468

Pratschke, M. 2024. *Generative AI and Education: Digital Pedagogies, Teaching Innovation and Learning Design.* Cham, Switzerland: (Springer briefs in Education) Springer.

Rahman, N., and G. Saunders. 2024, September, 11. "Access to Premium AI Services Is a Significant Concern for Students." *Wonkhe*. https://wonkhe.com/blogs/access-to-premium-ai-services-ls-a-significant-concern-for-students/.

Risko, E. F., and S. J. Gilbert. 2016. "Cognitive Offloading." *Trends in Cognitive Sciences* 20 (9): 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Ross, J. 2024, August, 12. "AI Potential 'Squandered by Universities' Risk Focused Approach." *Times Higher Education Campus*. https://www.timeshighereducation.com/news/ai-potential-squandered-universities-risk-focused-approach.

Rowsell, J. 2024, August 7. "Students Fear over Reliance on AI Devalues HE." *Times Higher Education Campus*. Accessed December 6, 2024. https://www.timeshighereducation.com/news/students-fear-over-reliance-ai-devalues-higher-education.

Rowsell, J. 2024, June 27. "AI Exam Answers Undetected and Outperform Students at UK Campus." *Times Higher Education Campus*. Accessed December 6, 2024. https://www.timeshighereducation.com/news/ai-exam-answers-undetected-and-outperform-students-uk-campus.

Sallai, D. 2024, July 23. "Assessment Curriculum Design Can't Ignore Students' Use of AI." *Times Higher Education Campus*. Accessed December 6, 2024. https://www.timeshighereducation.com/blog/assessment-and-curriculum-design-cant-ignore-How-students-use-ai.

Swaffield, S. 2011. "Getting to the Heart of Authentic Assessment for Learning." *Assessment in Education: Principles, Policy & Practice* 18 (4): 433–449. https://doi.org/10.1080/0969594X.2011.582838

Taylor, J. S. 2024, July 12. "Why I Ban AI Use for Writing Assignments." *Times Higher Education Campus*. https://www.timeshighereducation.com/campus/why-l-ban-ai-use-writing-assignments.

The Joint Information Systems Committee (JISC). 2024. "HE Generative AI Literacy Definition." https://nationalcentreforai.jiscinvolve.org/wp/2024/07/23/He-generative-ai-literacy-definition/.

Thorpe, J. 2023, June 27. "The Real Risk of GenAI Is a Crisis of Knowledge." *Wonkhe*. https://wonkhe.com/blogs/the-real-risk-of-generative-ai-ls-a-crisis-of-knowledge/.

Tight, M. 2024. "Challenging Cheating in Higher Education: A Review of Research and Practice." *Assessment & Evaluation in Higher Education* 49 (7): 911–923. https://doi.org/10.1080/02602938.2023.2300104

Ward, M., F. O'Riordan, D. Logan-Fleming, D. Cooke, T. Concannon-Gibney, M. Efthymiou, and N. Watkins. 2024. "Interactive Oral Assessment Case Studies: An Innovative, Academically Rigorous, Authentic Assessment Approach." *Innovations in Education and Teaching International* 61 (5): 930–947. https://doi.org/10.1080/14703297.2023.2251967.

Williams, A. 2023. "Developing an AI Assessment in the Division of Medicine." Digital Assessment at UCL. https://reflect.ucl.ac.uk/digital-assessment/2023/07/30/developing-an-ai-assessment-in-the-division-of-medicine/.

Wong, B., Y-L. T. Chiu, M. Copsey-Blake, and M. Nikolopoulu. 2022. "A Mapping of Graduate Attributes: What Can We Expect from UK University Students?" *Higher Education Research and Development* 41 (4):1340–1355. https://doi.org/10.1080/07294360.2021.1882405

Zhao, X., A. Cox, and X. Chen. 2025. "The Use of Generative AI by Students with Disabilities in Higher Education." *The Internet and Higher Education* 66 (101014): 1–11. https://doi.org/10.1016/j.iheduc.2025.101014.