

RESEARCH ARTICLE:

Early Adaption of Assessments Using Generative Artificial Intelligence and the Impact on Student Learning: A Case Study

Ling Wang¹, Shengchen Li² and Yue Chen³

Received: 14 June 2024 | Revised: 30 October 2024 | Published: 06 December 2024

Reviewing Editors: Dr. Charlie Reis, Xi'an Jiaotong-Liverpool University

Abstract

The emergence of generative artificial intelligence (GenAI) has brought new challenges for learners and educators, along with new possibilities. Assessment especially has provided opportunity for innovation using GenAI, but the fast pace of change as technology develops has meant a paucity in the research on the impact of these changes and adaptations. This paper intends to cover this gap by reporting on a case study of a prompt adaptation of GenAI in an assessment in a year-two module in artificial intelligence which is delivered to students in three-degree programmes in a Sino-British university, at the end of 2022. Taking a mixed method qualitative methodology approach, the study collected data from both staff and students, lecturing and studying on the module which was then analysed using NVIVO software. The study highlights that the adapted assessment had a positive impact on the student learning and development experience whilst simultaneously developing within the students an awareness of ethical issues associated with GenAI use. At the same time, it has raised many questions for academic staff, emphasizing the wider impact on learning and teaching that GenAI brings.

Keywords: GenAI, critical thinking; evaluative judgement; research skills; student learning

Introduction

The advent of Generative AI (GenAI) has been a significant landmark in artificial intelligence and machine learning, impacting multiple sectors across the world, such as education (Chan and Colloton, 2024). GenAI has the potential to revolutionize education by providing unique opportunities to improve teaching and learning methods (Jacobsen, 2023), providing customized learning experiences (Chiu, 2023). With this technology, educators can create personalized educational resources that cater to the specific needs, learning preferences, and speeds of each learner. Assessment in higher education is also impacted by GenAI. One implication is its tendency to invent plausible, confident responses that seem believable at first glance but do not stand up to detailed scrutiny, as well as the invention of references (Alkaissi and McFarlane, 2023). This triggered assessment changes in higher education in the GenAI era. Initially universities were caught off guard by the sudden widespread student use in assessment with GenAI, resulting in an increase anxiety by those designing assessments (Baidoo-Anu and Ansah, 2023). A demand for increased assessment security measures with a return to in-person invigilated exams became increasingly common. Whilst some may have taken this path as a resistance to change, for others some higher education institutions perceive it as an opportunity to transform assessment (Ajjawi *et al.*, 2023: 232).

AI can be described as the simulation of human intelligence in machines that are designed to think and act like humans (Chan and Colloton, 2024), and has been around since the 1950/60s, though general applications were limited. It is the advent and expansion of Big Data and the Internet of Things (or IoT) that has created a platform to use AI in a number of areas, which not only includes industry but also within education (Chan and Colloton, 2024). Put simply, Big Data refers to vast amounts of data that are too complex and large for traditional data processing systems to handle (Favaretto *et al.*, 2020), whilst IoT is the network of physical objects and devices

¹Xi'an Jiaotong-Liverpool University, lynlingwang821@gmail.com | <https://orcid.org/0000-0003-0272-2974>

²Xi'an Jiaotong-Liverpool University, shengchen.li@xjtlu.edu.cn | <https://orcid.org/0000-0002-2488-298X>

³Queen Mary University of London, yue.chen@qmul.ac.uk | <https://orcid.org/0000-0002-3136-008X>

which connect with each other through systems such as Wi-Fi and Bluetooth, as well as generate, exchange, send, and receive data (Li *et al.*, 2015). Perhaps the best example of IoT in education is the use and development of Smart Classrooms, whilst Big Data has been used now in most institutions to influence decision making in key areas. Whilst traditional AI has focused on tasks such as decision-making and pattern-recognition (Ruiz and Fusco, 2023), the recent emergence of more advanced GenAI technologies has allowed the generation of text, image and music from verbal prompts (Mishra *et al.*, 2024). ChatGPT, along with other tools such as Copilot, DALL-E, Perplexity and Elixir, are specific examples of GenAI tools generating human-like text based on conversational inputs. ChatGPT is arguably the most well-known GenAI tool, and its ability to engage in meaningful interactions and produce coherent responses demonstrates the practical applications within the broader AI landscape.

The recent emergence of GenAI is the most recent challenge facing higher education. Its integration into higher education offers transformative opportunities, yet it also presents distinct challenges, especially in the domain of assessment (Chan and Colloton, 2024). GenAI is changing the nature of how assessment takes place, what is assessed, and how students may use the technology in their submissions (Epstein *et al.*, 2023, Lodge *et al.*, 2023). The fast evolution of GenAI and resultant rapid pace of change has meant a paucity in research on its use in assessments, and impact of changes and adaptations to current assessment (Ajjawi *et al.*, 2023; Moorhouse *et al.*, 2023). An exploratory study of the literature between November 2022 and March 2023 by Baidoo-Anu and Ansah (2023) highlighted the potential benefits and drawbacks of ChatGPT in promoting teaching and learning. Benefits identified included the potential for the promotion of personalized and interactive learning, the potential for adaptive learning, and for changes to feedback and marking with the potential for continuous feedback to inform learning and teaching. Challenges included the lack of human interaction, information generated often being incorrect, lack of contextual understanding, limitations with complex subjects, and privacy, ethical and academic integrity issues.

This paper focuses on a case study of a year-two artificial intelligence module in a Sino-British university, the module leader of which undertook a very swift adaption of previous assessment as a result of the sudden 'disruption' of GenAI in learning, teaching and assessment in higher education. The aim of this study is to examine how the quickly-adapted assessment has impacted student learning, and how academic integrity was managed. The study has also informed an AI-enhancement project that the university has subsequently undertaken, with the aim of making all modules and programmes AI-enabled by 2026. This study therefore outlined three research questions:

- i. To what extent the changes to the module assessment enhances student learning and development through using GenAI?
- ii. How were academic integrity issues addressed?
- iii. Do adapted assessment effectively evaluate the intended module learning outcomes?

This paper adds to the available research on the impact and potential of GenAI on assessment design through the discussion of a case study of a year-two undergraduate module in artificial intelligent delivered to computing and technology students in a Sino-British university in the beginning of 2023, which quickly adapted its assessment in response to the technology.

Context of the Study

The study took place at a Sino-foreign English Medium of Instruction University in China, established in 2006 through a joint venture between an elite Chinese university and a Russell Group UK university. The institution is an independent entity accredited to issue both a UK degree and a Chinese degree at undergraduate level (a double degree), though at postgraduate level, only the UK degree is issued. The university needs to meet the requirements of both the UK and Chinese quality assurance and regulatory systems, and has an annual monitoring visit, with a reaccreditation visit every four years. The university is comprehensive in that it delivers a wide range of degrees, but almost half are in the STEM subjects. The undergraduate degrees are four years, with the first year mainly delivering modules required by the Chinese government. The second year therefore can be thought of as similar to the first year of a UK degree, and it is not unusual for modules to be shared across degree programmes. The module within this case study is a year-two Artificial Intelligence module with over 250 students delivered by the Department of Intelligent Science within the School of Advanced Technology. It is shared by three-degree programmes; BEng Digital Media Technology, BSc Information and Computing Science, and BSc Information Management and Information Systems. The module is 5 credits (150 hours of study, including self-study) with three assessments. The adapted assessment is the first assessment, a coursework assignment, worth 15% of the marks.

The second assessment is also a coursework, worth 15% of the marks. The third assessment is a final exam worth 70%. The first coursework within the study assesses three of the six module learning outcomes, namely:

- i. Demonstrate knowledge of the basic principles of artificial intelligence
- ii. Gain experience in the related AI area such as experiment design and result analysis
- iii. Acquire the fundamental knowledge of modern AI concepts and technologies including computer vision, natural language processing and machine learning.

The adapted assessment is the coursework 1 in 2022-2023. Some minor amendments on the weighting of coursework 1 from 2021-2022 to 2022-2023 include higher weight in critical thinking (from 10% to 15%) and organisation (10%-15%), less weight in language use. This adapted assessment was a written coursework that asked students to use GenAI whilst writing up mini research that they had undertaken using GenAI. Students were tasked with finding ways to mislead ChatGPT, leading to off-topic or nonsensical essays. This required them to engage deeply with the GenAI's functioning and limitations, encouraging a critical evaluation of its outputs and potential weaknesses. Students then needed to identify ways to fix the issues identified, backed up by the literature. Students were given some instruction in prompt engineering prior to undertaking the assignment, so that they had what the module leader described as 'the tools to cope'. By analysing how GenAI can be misled, students develop a nuanced understanding of GenAI capabilities and limitations, fostering critical analysis skills.

This particular case study is part of a wider school level project to look at to what extent AI can be incorporated into the curriculum. In addition, in spring 2024 the institution launched an AI-enhanced learning and teaching initiative highlighting existing good practice for sharing and identifying modules that need enhancement as a pilot study, with lessons learned becoming case studies to assist with the initiative to make all degree programmes AI-enhanced by 2026. The module in the case study has since been included in the AI-enhanced initiative.

Methodology

To explore the research questions, a case study method (Denzin and Lincoln, 2011) was employed, which allows for a comprehensive examination of the results and outcomes of an adapted assignment which adopted using GenAI quickly after ChatGPT was launched. The case study approach enables the collection of contextual data and facilitates the exploration of the nuanced impacts of GenAI on assessment and student learning. Although quantitative methods are often associated with larger sample sizes and statistical generalization, such as the class size in the module within the case study, qualitative methods were used as they offered unique strengths in exploring complex, context-dependent phenomena such as the impact of GenAI on assessment practices and student learning. A qualitative approach was chosen for this study because it attempts to make sense of students and staff experience, perception, beliefs, attitudes and behaviour in a given cultural context (Pacho, 2015), in this case using GenAI for student learning. It was important for the study to focus on students who had little or prior knowledge of GenAI, and to capture representations from across the three-degree programmes. The depth of insight by the qualitative research is critical for developing a comprehensive, nuanced understanding of the educational implications of the adapted assessment with GenAI on the learners' development which might be overlooked by purely quantitative measures. The three research questions indicate the exploratory nature of the study which the qualitative methods fit appropriately with attempts to explore the contextual elements and generate new insights and understandings with detailed picture of how and why such outcomes occur.

The study uses purposive sampling to select a sample of 10 student participants from the student body studying in the module (Palinkas *et al.*, 2015: 534; Patton, 2002: 264). Table 1 provides information for each participant in the study. Purposive sampling is a non-probability technique used to select informants with specific knowledge or characteristics relevant to a study (Tongco, 2007). The students were self-identified as having limited or no experience of GenAI and covered the three-degree programmes within the larger cohort, to ensure a comprehensive understanding of the impact across different student groups (Uprichard, 2013). All participants had individually completed the adapted coursework and were asked to provide diverse perspectives on their experiences. Only one staff member was included out of a teaching team of two, who was the module leader and designed the assessment. All members of the study gave their informed consent to take part, and all participants were anonymized to respect confidentiality.

Table 1: Participants' Demographic Information

Participant ID	Gender	Programme	Study Level/Role	Previous Experience	GenAI
A	M	BSc Information and Computing Science	Undergraduate	Limited experience	
B	M	BSc Information and Computing Science	Undergraduate	No experience	
C	M	BSc Information and Computing Science	Undergraduate	Limited experience	
D	F	BEng Digital Media Technology	Undergraduate	Limited experience	
E	M	BSc Information Management and Information Systems	Undergraduate	No experience	
F	M	BSc Information and Computing Science	Undergraduate	Limited experience	
G	M	BSc Information and Computing Science	Undergraduate	No experience	
H	M	BSc Information Management and Information Systems	Undergraduate	No experience	
I	M	BSc Information and Computing Science	Undergraduate	Limited experience	
J	M	BEng Digital Media Technology	Undergraduate	No experience	
K	M	Dept of Intelligent Science	Module Leader	Some experience	

The study employs multiple methods for data collection from both staff and students learning and teaching on the module, to ensure a robust and triangulated understanding of the research questions (Denzin, 2012: 82). Observable data was collected through a document analysis of the students' coursework (Bowen, 2009: 36). Bowen (2009) indicated that document analysis is useful as it is both time and cost-effective, uses the widespread availability of documents, and is unobtrusive. However, Bowen (2009: 28) also suggests that it should be used in conjunction with other data collection methods. This study also included unobservable data through individual stimulated recall interviews (SRIs) as described by Gass and Mackey (2016: 53). These involved both the staff who lectured and adapted the coursework assessment of the module, as well as the ten identified students who studied and completed the re-designed coursework. SRI allows for a variety of interview question types, and this study used a semi-structured structure to elicit the participants' perceptions and self-reflections to the extent to which the promptly adapted assessment has accounted for GenAI, promoted learning and development of learners, and safeguarded academic integrity (Gass and Mackey, 2016: 67). The data collected from documents and SRIs was analysed using thematic analysis (Braun and Clarke, 2012). The SRIs were conducted in Chinese as the student's native language, enabling easy understanding of the interview questions and smooth, accurate expression of students' perceptions and practices. Data was transcribed into English using the 'speech-to-text conversion' by the Tencent meeting TranSmart function. Translated transcriptions of interviews were then shared with the participants to check for accuracy. Initial codes of accuracy checked data were generated and relevant data to each code is identified and gathered. A review of all draft codes was undertaken to discern vagueness, overlap, and inconsistency. With this review, final codes were defined with clear definitions and names.

Findings

The data collected was analysed using NVivo, which enabled codes (themes) and where appropriate sub-codes (sub-themes) to be identified (Mortelmans, 2019), highlighted in Table 2. These themes and sub-themes are discussed in the following sections.

Table 2: Themes and sub-themes

Themes	Sub-themes
Critical Thinking	Problem identification
	Evaluative judgement
Research Skills	Source evaluation
	Prompting
Academic Integrity	Ethics
	Plagiarism

Critical thinking

The theme of critical thinking was identified, with the two sub-themes of problem identification and evaluative judgement. The assignment asked students to 'mislead' ChatGPT, and to identify where it provides 'incorrect' information. It then asked students to 'fix the issues' identified, grounding those arguments or 'fixes' within the literature.

Problem identification: In the field of computing and engineering, the ability to accurately identify problems is a fundamental skill that underpins the entire problem-solving process (Harlim and Belski, 2013). By identifying problem, learners develop a solid understanding of the issues at hand which could empower them to formulate targeted solutions. Identifying problem also cultivates critical thinking and analytical skills (Loh, 2020), as students are required to analyse and evaluate the complex issues so that they could target the issues accurately before solving. This skill is particularly crucial in the field of computing and engineering, where students are often tasked with addressing intricate technical challenges and designing innovative solutions. The adapted assessment asked students to find ways to mislead ChatGPT. By doing so, students need to discern the weaknesses of ChatGPT first. High achieving students tended to learn about the weaknesses of ChatGPT through various sources, including relevant academic papers from credible academic sources, news and articles from frequently visited websites worldwide, and highly-played videos from popular international platforms.

Student A in the SRI stated:

"I focused on investigating ChatGPT weaknesses as I believe I can check out its benefits through ChatGPT fairly easily. I can't simply learn about its disadvantages simply via any GenAI or other shortcuts, but searching papers online instead... This might be one key focus of this coursework in my view."

Student H also commented on how he knew about the weaknesses of ChatGPT:

"I watched a video about fundamental principles of ChatGPT by its development and design team including supervising learning. Followed by undertaking research with my teacher for a while, I started to understand the shortcomings of ChatGPT which provided me ideas of writing this report".

Evaluative judgement: The adapted assessment in the case study expected students to undertake analysis of data and facts from various sources for both summarizing drawbacks of GenAI and recommending sensible solutions. Evaluative judgement, the capability to weigh up the quality of work of self and others, is an important skill for university graduates. Bearman *et al.* (2024) urge that there is an urgent need for university graduates to leverage the disciplinary knowledge with evaluative judgement to differentiate trustworthiness from the 'hallucinatory'. Such competency has become increasingly important with the widespread use of GenAI, where the technology gives confident responses which later may turn out to be incorrect or misleading (Alkaissi and McFarlane, 2023). Assessments should be designed in the direction of promoting learners' evaluative judgement of the GenAI work (Bearman and Luckin, 2020).

Students' reports and the SRIs demonstrated how learners have established and built up their evaluative judgement competency in the assessment preparation and completion process. Some students evaluated the reliability of ChatGPT output by discussing its flaws in accurate processing and interpretation of information. Student J commented that:

"A significant reason why ChatGPT is off-topic is that it is an artificial intelligence that does not have the ability of humans to think and judge right from wrong. Therefore, some data flaws are one of the reasons why ChatGPT writes off-topic."

Student G had already doubted ChatGPT, saying:

"I do not fully trust or rely on ChatGPT. Before setting prompts in it, I had already learnt about the question (drawbacks of ChatGPT) via other channels and acquired some knowledge... I can evaluate the correctness, accuracy and/or quality of ChatGPT outputs. Therefore, I'm not concerned much that it will mislead me".

The benefits that the students saw in the assignment are summarized by Student D, stating that:

"I developed my evaluative judgement skill in the process of outlining and writing this report with the help of ChatGPT".

This was further supported by Student I who thought that:

"GenAI sets higher threshold of users who are expected to master solid base knowledge and critical evaluation to discern correct and incorrect information, reason why incorrect".

Finally, the value of academic staff was recognized, with Student B highlighting that:

"I learnt from this assessment that knowledge shared by teachers is more credible than that from GenAI as teachers deliver the most fundamental and most essential key concepts of the subject".

Research skills

The students would not have been able to complete the assignment without demonstrating a reasonable level of research skills. Although the students received some instruction in prompt engineering, they still needed to decide how to mislead ChatGPT effectively in the assignment. Students exercised their skill in crafting effective and precise queries to interact with and mislead ChatGPT. Whilst evaluating whether responses from ChatGPT were authentic or not, students then needed to source and identify relevant academic sources to help them with the 'fixes'. By preparing and completing this assessment, students developed their source evaluation skills by assessing the credibility, relevance, and reliability of information sources. Such skill is crucial for students as they need to ensure the information used for research is accurate and trustworthy.

Source evaluation: This module was one of the very first few year-two modules in the BEng Digital Media Technology, BSc Information and Computing Science, and BSc Information Management and Information Systems degrees with written coursework assessment, with traditional mid-semester and end of semester assessments dominating. Before that, students only practiced academic writing in their English for Academic Purposes (EAP) modules, which all students take in the first two years, meaning they have very limited experience of academic essay writing. Some student participants shared how their capability of information and literature searching and source evaluation has been promoted during this coursework preparation and completion. Student A stated:

"I had a quick check of the academic papers recommended by ChatGPT, finding that one paper does not exist at all, and mismatch between paper titles and authors in another paper".

Student J commented on the importance of preparation for the assignment:

"Prior to drafting the essay, I learnt some theories and principles of GenAI which significantly help develop some ideas of this coursework."

Whilst Student A shared:

"Based on my impression of AI hallucination, I did not ask ChatGPT of references, but I searched myself at google scholar instead which is credible and trustworthy in my view."

Student G commented that:

"I used ChatGPT in summarizing key ideas of each paper I searched so that I can decide whether I will read that paper in details and cite it in my essay later". "It saves time in research comparing with my previous approach of literature review without ChatGPT".

The module leader suggested that the adapted assessment enabled learner's opportunities to undertake part of the research with the assistance of ChatGPT. He thought that learners have improved their literature review capacity through efficient and targeted paper searching and effective source evaluation, so as to:

'understand the paper with in-depth thoughts at higher and a later stage'.

Prompting: Developing suitable and relevant prompts is one common way of engaging with ChatGPT. Each student participant shared his individual progress of prompting while preparing and finalizing the coursework. The approach to prompting is highlighted with the following four students. Student G shared:

"most of my prompts were asking ChatGPT to give me some themes and ideas based on which I can set up a framework to structure my essay. This assessment underpinned my understanding of GenAI."

Student C recalled that:

"I remember I kept adapting the prompts until I got the information I asked for... I view it as a way of improving the academic writing proficiency because I need to adjusting the subject, predicate and object continuously to make my prompts understood by the ChatGPT sufficiently easy to generate the responses".

Taking a more pragmatic approach, Student E highlighted that:

"My prompting approach was to screen out all the irrelevant information first, feed the ChatGPT relevant, easily understood and targeted information only to help it generate useful response".

The importance of asking the right prompts was also highlighted with students recognizing that prompt quality impacted ChatGPT answers. The difficulty of doing this effectively was summarized by Student H who reflected that not knowing how to prompt properly meant:

"It is a challenge for me to interact with ChatGPT effectively".

Academic integrity

This theme revolved around issues to do with academic integrity, which have often dominated the early discussions of GenAI, and especially ChatGPT. This has meant that often GenAI is seen as a cheating tool, rather than a learning tool, with potential impact on the overall impression of the value of higher education by the public (Sullivan, Kelly and McLaughlan 2023). Whilst introducing students to the potential and limitations of ChatGPT as a learning tool, it also explored students' views on the impact of ChatGPT on academic integrity.

Ethics: The module leader explained that in the previous year's assignment, a short essay discussing the future development of automatic music generation with literature review, there was an assumption that students had probably used GenAI in many of the answers submitted, though it was difficult to prove. As a result, the assessment was re-designed so that it incorporated the technology and taught students how to use it ethically, through summarizing issues of ChatGPT with possible solutions. The adapted assessment showed the module leader's adaptability and creativity in embracing GenAI while safeguarding academic integrity. The module leader suggested that:

"GenAI definitely would be a type of tool to help student learning, just like calculator in the last century. The adapted assessment has successfully prevented students from academic offences by setting the task of deceiving ChatGPT. It is not a task that students can simply take information from ChatGPT for direct use in their essay or use ChatGPT to write an essay for him or her, as this is a complicated question expecting students to interrogate ChatGPT continuously with a series of prompts, evaluate the appropriacy and relevancy of prompts and add his or her thoughts."

Student participants acknowledged that they were fully aware of the ethical considerations and responsible use of ChatGPT for this assignment. They used different approaches in preparing this report in alignment with academic integrity. Student I said:

"I was somewhat concerned about academic integrity, but not that much, because the quality of ChatGPT output is rather limited and the ideas from ChatGPT for different users are similar.... I am more concerned about the accuracy and quality of ChatGPT output, or the implication of ChatGPT on learners' learning motivation. I know some of my classmates are heavily dependent on ChatGPT for the assignments in the absence of learning subject knowledge".

Plagiarism: Students recognized the importance and need of appropriate referencing in using ChatGPT, but felt it was difficult to enact. This is best summed up by Student J who said it was:

"essential to cite the source of codes shared in public to acknowledge the work of the author. However, it was rather difficult to reference it if I get the coding from ChatGPT with no clear authorship".

By having identified issues with sources using ChatGPT, and by using ChatGPT as an integral part of the assignment, students seemed to have increased their awareness of defending academic integrity using ChatGPT as an assistive tool in preparing and writing up this assignment. Student F highlighted that he had:

“learnt more about it (plagiarism) by the assignment than in other lessons he studied in”

This lesson then can be seen as an added, if unintended, value of this adapted assessment.

Discussion and Implication for Practice

The paper starts by highlighting three research questions that the team wanted to address as a result of the adapted assessment. GenAI clearly enhanced the student learning experience in that it introduced them to both the limitations and potential of the technology in learning and teaching. The advantages of using GenAI to develop and enhance written assessments demonstrated that there is the potential GenAI can suggest improvements and generate relevant content that aligns with students' specific learning levels through self-learning. This individualized assistance might help learners better understand complex abstract concepts and could potentially improve their writing skills (Cope *et al.*, 2021: 1241). This paper's results are in line with the current work of Mai and Hanh (2024), which highlights how GenAI tools, such as ChatGPT, could markedly foster in-depth subject exploration. Additionally, Tiwari *et al.* (2024) also highlight the advantages of incorporating AI into educational settings, noting its effectiveness in improving the quality of learning and making academic activities more efficient. The integration of GenAI into higher education, and especially into assessment, therefore, offers transformative opportunities as well as challenges with regard to assessments (Chan and Colloton, 2024), recognising that students have access to sophisticated AI tools, as in the case of the adapted assessment, there is a need to redefine what assessments should look like, and how to evaluate it. The adapted assessment in the case study has demonstrated to be a successful, innovative assessment in the early introduction of ChatGPT. It empowers learners to interrogate GenAI, discuss with peers actively to generate and synthesize their own work through utilizing and evaluating sources (Burr, 2015). Therefore, the integration of ChatGPT in this assessment has promoted an environment where students as active learners engage with the assessment to co-construct knowledge through interacting with ChatGPT.

Lodge *et al.* (2023) are quite clear in suggesting that AI represents an urgent need for change, as it influences not only how student learning can be assessed, but what is valuable to assess. They continue that this change (in assessments) needs to also recognize what and how students learn. This is echoed by Chan and Colloton (2024) who recognize the need to redesign assessments that do not focus on the final product such as grades, but the process of learning. To do so requires a clear understanding of what GenAI is capable of, and its limitations. It needs a shift in assessment design which enables assessment to engage students in learning and embrace GenAI for 'rich portrayals' of student learning (Lodge *et al.*, 2023: 1). The adapted assessment did just this as it required students to interact actively with ChatGPT, leading to the development of critical thinking skills. As students attempted to find ways to uncover the unreliability of ChatGPT, many of them essentially attempted to make clear its fundamental working principles through interrogating ChatGPT and/or learning from other sources, with the goal of identifying its limitations. Students engaged in an adapted assessment that required them to critically interact with ChatGPT and develop critical thinking skills. Attempting to deceive ChatGPT and discern its drawbacks with potential solutions has enhanced evaluative judgement, as students critically evaluated ChatGPT's outputs and discerned accurate information from flawed information, which is an essential skill in the era of pervasive AI tools (Bearman *et al.*, 2024).

The common approach for learners to develop evaluative judgement through assessment in the AI-ubiquitous world is assessing GenAI outputs (Ziebell and Skeat, 2023). In this study, by preparing and finalizing the assignment report, student participants worked with ChatGPT and asked it to self-identify existing issues. Students attempted to discern the quality and appropriacy of its output by inquiring of themselves whether the output from ChatGPT was credible and convincing enough to put in the assignment, and the extent to which they trusted the output. They also asked themselves what new ideas they could further develop based on the GenAI output, and in doing so were honing their evaluative skills (Bearman *et al.*, 2024). Lodge *et al.* (2023) argue that AI requires multiple, inclusive and contextualized approaches to assessment, and that one single assessment can account for all the possible uses of AI by students. Again, Chan and Colloton (2024) seem to agree as they argue for multiple assessment sources. Bearman *et al.* (2024) also argues for multiple assessment strategies, and in doing so learners also develop the evaluative judgement competency about the processes of GenAI. In this regard the adaptive assessment in the study was just one of three assessments, with the other two being more traditional in nature. It is therefore arguable that an adaption of all assignments in the module would have been a more rewarding experience for students (and staff), but perhaps would not have had the safety need of ensuring students did not fail a module due to radical change, when it has historically had a good pass rate. The study also found that

students have developed their research skills through the assignment preparation and completion which tasked them in gathering information from various sources including GenAI, critically evaluating the credibility of the sources and synthesizing the information into sophisticated arguments. This mirrors the study by Baidoo-Anu and Ansah (2023) who highlighted the significance of integrating GenAI into facilitating personalized and interactive learning in research and general learning processes.

It is evident that the adapted assessment addressed the first two research questions in that students had an enhanced learning and teaching experience, and that academic integrity issues were identified and built into the assessment. It is, however, more difficult to determine whether the adapted assessment fully evaluated the module learning outcomes it addressed. Students passed the assessment, and indeed the overall module, demonstrated the knowledge of AI and gained experience in its use, but this can be seen as a narrow interpretation of AI as the assessment was limited to testing GenAI rather than other facets. The module leader (participant k) indicated that when the learning outcomes were originally developed, the intention was for students to be able to demonstrate knowledge and have experience of more than GenAI. If the assessment was used in future years, thought would need to be given to amending the learning outcomes to represent actual practice. Alternatively, the experience of this study may suggest that the assessment needs additional components if the learning outcomes are not modified and stay the same.

Implications for Practice and Future Research

The study's findings suggest some practical recommendations for educators and higher education institutions. The module leader was one of the very few academics in the university who responded swiftly, encountering the challenges by the GenAI to assessment, teaching and learning in higher education. Most academics in the university either resisted change or waited for instruction on what they should do in this new world. Indeed, many universities have struggled to keep up with the changes that GenAI brings (Ajjawi *et al.*, 2023). The risk and adaptability in being innovative in changing the assessment proved successful and effective in judging the extent to which learners have achieved intended learning outcomes, whilst safeguarding academic integrity. Higher education educators therefore need to keep a positive open-minded attitude towards the introduction of GenAI in education by embracing it and using it ethically as an effective tool in the learning, teaching and assessment processes. The findings from this study have implications for learning, teaching and assessment in higher education practices, especially but not exclusively in the realm of assessment design and assessment implementation. The adapted assessment has revealed a tangible potential in developing and promoting critical thinking and research skills of learners. Thus, it is suggested that educators are encouraged to facilitate their teaching and assessment design in a manner that encourages learners to explore GenAI tools, as there are potential benefits in not only furthering students' knowledge, but also in the skills and experience they gain in the entire process. This aligns to the stance by Bearman *et al.* (2023: 298) that such skills are reshaped in the 'ways of knowing, doing and being within a digitally mediated society'.

However, it is also important to acknowledge the potential challenges and issues associated with the use of GenAI tools. One significant challenge is the reliability of GenAI-generated content, as the tools can sometimes produce 'hallucinations' (Alkaissi and McFarlane, 2023). It therefore highlights the necessity for students to develop AI literacy, including the competence of critically validating and evaluating GenAI generated information (Zhai *et al.*, 2024). By understanding the limitations, students can better navigate and utilize GenAI technologies in their future professional practices. Thus, assessment can be redesigned and transformed in promoting desirable learning behaviours, developing subject-specific and professional competencies, as well as disciplinary graduate attributes. By doing so, learners develop and underpin their analysis, evaluation and problem-solving skills in the process of interrogating GenAI content, which may be important to future professional practices in the digital world. Moreover, the study also highlights the significance of academic integrity in the era of GenAI. Attempts to prevent learners from using GenAI tools to discourage plagiarism or cheating by reversing back to the controlled exam format are highly controversial as they are limited and fail to account for a world where students might be using GenAI as a professional tool. Instead, a GenAI re-designed assessment that incorporated and critically analysed GenAI content was promoted. A joint effort of promoting a culture of integrity and transparency by educators, higher education institutions and learners is desired, and providing clear and comprehensive guidelines for using GenAI ethically and responsibly is important. It is believed that what's more important is for educators to offer opportunities for learners to implement those principles and guidelines in their learning and assessments.

Despite the results of this paper, there are a number of limitations of this study. Firstly, the small sample size, although rich in context, may not be generalizable to larger populations (Morse,1994). As this study was undertaken in a specific unique context of a Sino-British joint venture, the findings may not be representative or applicable in a different setting. This is especially as students were studying in a second language, English, and so issues relating to language as well as use of GenAI need to be considered. This limitation therefore could hinder the ability to apply the findings to different educational settings or diverse student groups. This is also the case with the second identifiable limitation of the study, that the results may be discipline specific. The students in this study may not have been familiar with GenAI, nor the assessment type, but they are familiar with using technology as part of their degrees. Results may be different with students in other disciplines, such as social sciences, who may be more used to the assessment type, but not using technology. A third limitation is that these students were year-two undergraduate students, and results may be different with year three or four students, or with postgraduate students who would have been studying longer and therefore possibly more familiar with assessment techniques. A final limitation is that the use of purposive sampling and small sample size may increase the risk of overlooking important variables or characteristics that could influence the study's outcomes. A different, though arguably less detailed methodology may have resulted in different themes and sub-themes being identified.

Future research in this area could explore the long-term impact of GenAI multiple assessments on learners learning and development in the specific, or a different educational setting. By developing longitudinal studies, valuable insights could be provided into how continuous interaction with AI tools influences students' cognitive and metacognitive skills, including critical thinking, problem-solving, and learner autonomy. Mixed-method research can also be employed to investigate the effectiveness of GenAI in achieving learning outcomes and enriching learning experience in both depth and breadth. Examining learners' and educators' perceptions of GenAI over a certain period of time chronologically so as to explore how their perceptions have implicated their enactments in learning, teaching and assessment in such a GenAI-mediated world, could be another area for future research. Additionally, there is a need for research into the pedagogical implications of AI in diverse educational contexts. Studies could examine how AI tools can be tailored to meet the unique needs of different student populations in various disciplines, including those using English as medium of instruction (EMI).

Conclusion

The integration of GenAI into assessments marks a shift in the way we approach teaching, learning and assessment, especially in the context of higher education. This case study, which focused on the early adaptation of assessments using GenAI in a Sino-British university, reveals insights into how GenAI can enhance student learning outcomes and transform assessment practices. Lecturers need to develop a clear understanding of what the technology offers their learners and how it might be embraced to enhance learning. The investigation of early adaptation of assessments using ChatGPT in this study has shown promising impacts on student learning and development, particularly in enhancing critical thinking, evaluative judgement, research skills and prompting. These findings highlight the potential of GenAI to impact educational practices, provided that ethical considerations and academic integrity are carefully managed. Educators, learners and higher education institutions must work collaboratively to integrate GenAI in a manner that supports and enriches student learning, preparing them for a future where AI tools are ubiquitous. Further research is needed to expand on these findings and explore the broader implications of GenAI in higher education.

References

- Ajjawi, R., Tai, J., Boud, D. and Jorre de St Jorre, T. 2023. *Assessment for Inclusion in Higher Education: Promoting Equity and Social Justice in Assessment*. London: Routledge.
- Alkaissi, H. and McFarlane, S. I. 2023. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, 15(2): 1-4.
- Baidoo-Anu, D. and Ansah, L. O. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, 7(1): 52-62.
- Bearman, M. and R. Luckin. 2020. Preparing University Assessment for a World with AI: Tasks for Human Intelligence. In: Bearman, M., Dawson, P., Ajjawi, R., Tai, J. and Boud, D. eds. *Re-Imagining University Assessment in a Digital World*. Switzerland: Springer Nature, 49-63.

Bearman, M., Nieminen, J. H. and Ajjawi, R. 2023. Designing Assessment in a Digital World: An Organising Framework. *Assessment and Evaluation in Higher Education*, 48(3): 291-304.

Bearman, M., Tai, J., Dawson, P., Boud, D. and Ajjawi, R. 2024. Developing Evaluative Judgement for a Time of Generative Artificial Intelligence. *Assessment and Evaluation in Higher Education*, 49(6): 1-13.

Bowen, G. A. 2009. Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2): 27-40.

Braun, V. and Clarke, V. 2012. Thematic Analysis. In: Cooper, H., Camic, P. M., Long, D. L., Panter, A. T., Rindskopf, D. and Sher, K. J. eds. *APA Handbook of Research Methods in Psychology: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological*. Washington: American Psychological Association, 57-71.

Burr, V. 2015. *Social Constructionism*. London: Routledge.

Chan, C. and Colloton, T. 2024. *Generative AI in Higher Education: The ChatGPT Effect*. London: Routledge.

Chiu, T. K. 2023. The Impact of Generative AI (GenAI) on Practices, Policies and Research Direction in Education: A Case of ChatGPT and Mi-Journey. *Interactive Learning Environments*, 1-17. Available: <https://doi.org/10.1080/10494820.2023.2253861> (Accessed 10 September 2024).

Cope, B., Kalantzis, M. and Sears, D. 2021. Artificial Intelligence for Education: Knowledge and Its Assessment in AI-Enabled Learning Ecologies. *Educational Philosophy and Theory*, 53(12): 1229-1245.

Denzin N. K. and Lincoln Y. S. 2011. *The SAGE Handbook of Qualitative Research. 4th ed.* Thousand Oaks, CA: Sage.

Denzin, N. K. 2012. Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2): 80-88.

Epstein, Z., Hertzmann, A., Investigators of Human Creativity, Akten, M., Farid, H., Fjeld, J., Frank, M. R., Groh, M., Herman, L., Leach, N. and Mahari, R. 2023. Art and the Science of Generative AI. *Science*, 380(6650): 1110-1111.

Favaretto, M., De Clercq, E., Schneble, C. O. and Elger, B. S. 2020. What is Your Definition of Big Data? Researchers' Understanding of the Phenomenon of the Decade. *PloS One*, 15(2): 1-20.

Fischer, J., Bearman, M., Boud, D. and Tai, J. 2024. How Does Assessment Drive Learning? A Focus on Students' Development of Evaluative Judgement. *Assessment and Evaluation in Higher Education*, 49(2): 233-245.

Gass, S. M. and Mackey, A. 2016. Characterisation of Stimulated Recall Interview. In: Gass, S. M. and Mackey, A. ed. *Stimulated Recall Methodology in Applied Linguistics and L2 Research*. New York: Routledge, 42-76.

Harlim, J. and Belski, I. 2013. Long-Term Innovative Problem-Solving Skills: Redefining Problem Solving. *International Journal of Engineering Education*, 29(2): 280-290.

Jacobsen, P. 2023. Why ChatGPT Will Change Higher Education for the Better. Available: <https://fee.org/articles/chatgpt-will-change-higher-ed-for-the-better/> (Accessed 09 September 2024).

Li, S., Xu, L. D. and Zhao, S. 2015. The Internet of Things: A Survey. *Information Systems Frontiers*, 17: 243-259.

Lodge, J., Howard, S., Bearman, M., Dawson, P. 2023. Assessment Reform for the Age of Artificial Intelligence. <https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf> (Accessed 10 September 2024).

Loh, W. L. 2020. Sharpening Critical Thinking in Problem Identification in Design and Technology Education. Available: <https://dl.designresearchsociety.org/drs-conference-papers/drs2020/researchpapers/123/> (Accessed 18 June 2024).

Mai, D. H. and Hanh, N. V. 2024. The Use of ChatGPT in Teaching and Learning: A Systematic Review through SWOT Analysis Approach. *Front. Educ.* 9: 1-17. Available: <https://doi.org/10.3389/educ.2024.1328769> (Accessed 22 June 2024).

Mishra, P., Oster, N. and Hensriksen, D. 2024. Generative AI, Teacher Knowledge and Educational Research: Bridging Short and Long-Term Perspectives. *TechTrends*, 68: 205-210.

Moorhouse, B. L., Yeo, M. A. and Wan, Y. 2023. Generative AI Tools and Assessment: Guidelines of the World's Top-Ranking Universities. *Computers and Education Open*, 5: 1-10.

Morse, J. M. 1994. *Critical Issues in Qualitative Research Methods*. London: Sage Publications.

Mortelmans, D. 2019. Analyzing Qualitative Data Using NVivo. In: Van den Bulck, H., Puppis, M., Donders, K. and Van Audenhove, L. eds. *The Palgrave Handbook of Methods for Media Policy Research*. London: Palgrave Macmillan, 435-450.

Pacho, T. 2015. Exploring Participants' Experiences Using Case Study. *International Journal of Humanities and Social Science*, 5(4): 44-53.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N. and Hoagwood, K. 2015. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5): 533-544.

Patton, M. Q. 2002. *Qualitative Research and Evaluation Methods*. California: Sage.

Ruiz, P. and Fusco, J. 2023. Glossary of Artificial Intelligence Terms for Educators. Available: <https://circls.org/educatorcircls/ai-glossary> (Accessed 22 May 2024).

Sullivan, M., Kelly, A. and McLaughlan, P. 2023. ChatGPT in Higher Education: Considerations for Academic Integrity and Student Learning. *Journal of Applied Learning and Teaching*, (6:1): 1-10.

Tai, J., Ajjawi, R., Boud, D., Dawson, P. and Panadero, E. 2018. Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work. *Higher Education*, 76(3): 467-481.

Tiwari, C. K., Bhat, M. A., Khan, S. T., Subramaniam, R. and Khan, M. A. I., 2024. What Drives Students toward ChatGPT? An Investigation of the Factors Influencing Adoption and Usage of ChatGPT. *Interactive Technology and Smart Education*, 21(3): 333-355.

Tongco, M. C. 2007. Purposive Sampling as a Tool for Informant Selection. *Ethnobotany Research and Applications*, 5: 147-158.

Uprichard, E. 2013. Sampling: Bridging Probability and Non-Probability Designs. *International Journal of Social Research Methodology*, 16(1): 1-11.

Zhai, C., Wibowo, S. and Li, L. D. 2024. The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review. *Smart Learning Environments*, 11(1): 3-17.

Ziebell, N. and Skeat, J. 2023. How is Generative AI being Used by University Students and Academics? Available: https://education.unimelb.edu.au/_data/assets/pdf_file/0010/4677040/Generative-AI-research-report-Ziebell-Skeat.pdf (Accessed 11 September 2024).