ASCILITE 2024

Navigating the Terrain:
Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies

Evaluating an online assessment framework through the lens of Generative AI

Amanda White
University of Technology Sydney

Elaine Huber
University of Sydney

Since the general public's ability to access generative AI in November 2022, higher education faces significant challenges and opportunities. This paper examines how generative AI affects a framework for quality online assessment, which includes academic integrity, authenticity, equity of access, information integrity, quality feedback, and student experience. Drawing on feedback from survey respondents and focus group/interview participants from a range of stakeholder groups, the study provides qualitative insights on the impact of generative AI on assessment. Findings highlight both potential benefits, such as improved performance and efficiency, and challenges, including threats to academic integrity and equity of access. The paper explores the implications of generative AI and discusses the trade-offs educators and assessment designers face when designing assessments that must balance innovation and preparing students for the generative AI-enhanced workplace, with the need for integrity and fairness.

Keywords: quality assessment, generative AI, online assessment, evaluation

Introduction

Generative AI (GenAI) constructed on Large Language Models (LLMs) have taken the academic landscape by storm since becoming more publicly accessible in November 2022. We recognise that GPT-1 by OpenAI was launched in June 2018, but until November 2022, GPTs were not easily used by the lay person. The technology is experiencing exponential improvement – from being reported as poor at passing accounting exams in the very early days (Wood et al., 2023) to passing professional certification exams after previously failing (Steinhardt, 2023). Researchers found that in one global consulting firm's implementation of ChatGPT, the impact was staff working 25% faster and with a 40% improvement in performance (Dell'Acqua et al., 2023). Higher education institutions are grappling with the implications of GenAI and the definition of learning outcomes, integration into curriculum, student expectations and assessment security and integrity (Lodge et al. 2023).

A framework for quality online assessment
A comprehensive framework for online assessment published by Huber et al. (2024) identified 10 components that should be considered when designing assessment that is completed, delivered, and/or submitted online). The framework suggests six design considerations, namely academic integrity, authenticity, equity of access, information integrity, quality feedback and student experience. The framework also recognises that the design and implementation of online assessment is affected by constraints, including accreditation requirements, institutional assessment policies, the scale or size of the cohort and the resources available.
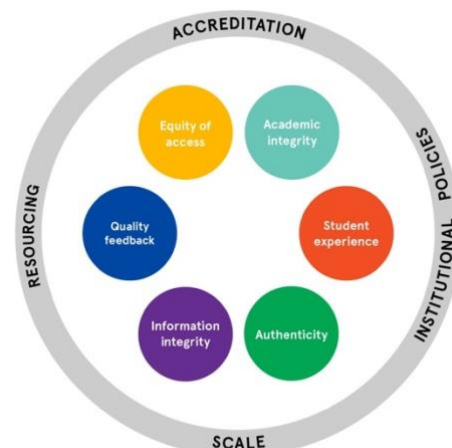

Figure 1 - Framework for online assessment

It is important to note that there is no "assessment unicorn" that rates highly across every single component, within the constraints. Instead, the framework provides educators and assessment designers with the opportunity to consider the trade-offs that are made when designing assessment within constraints. This paper considers how the components of this framework are affected by GenAI and what impact this might have on the decisions made by educators and assessment designers in designing assessment tasks.

Research method

We conducted an online survey (n= 213) and focus groups/interviews (n = 46) with participants representing five stakeholder groups, students, educators, industry experts, institutional decision makers and accreditation body representatives. In this paper, we present insights from qualitative analysis of the focus groups and interviews underpinned by the literature on GenAI. They refer to GenAI impact on assessment in general, and our framework components in particular.

The impact of Generative AI on a framework for quality assessment design

Academic integrity

It may be broadly recognised that the most common response in the media and higher education newsletters and blogs to the rise of GenAI is the concern related to academic integrity – did the students who turned in the work complete it themselves? (Sullivan et al., 2023). The ability of GenAI to create high quality and well thought out work threatens the assessment security and validity of marks received in assessment tasks where student identity is not assured. The Australia Higher Education Standards Framework (Threshold Standards) (Cth) under Section 1.4 requires accredited bodies to be confident that "students have demonstrated the learning outcomes for the course of study". However, it is impossible to detect the use of GenAI with any level of accuracy (Farrelly & Baker, 2023), and along with contract cheating and notes-sharing websites, it poses a large-scale threat to an institution's assurances of student learning. A student in a focus group noted the need for complexity, by stating that assessments may have to be more complex to 'filter out those who are otherwise using ChatGPT'. Students commented that they felt the use of ChatGPT was just another example of 'outsourcing' and therefore was inauthentic and unfair. Institutional policy makers liked the idea of a "two-lane approach" by categorising assessment into two types – one where assessments are secured and one where AI is allowed. Educators tended to agree particularly with a return to traditional invigilation of summative assessment approaches given the inability to accurately detect GenAI use. Accreditation body representatives suggested using industry-specific case studies as a means of overcoming impersonation and ChatGPT concerns.

Authenticity

Authenticity refers to the ability of an assessment task to reflect activities that students may be conducting in the workplace. Across all industries, from consulting to human resources to healthcare, organisations are considering how to integrate GenAI into their work (Ooi et al., 2023). Therefore, educators and assessment designers must explore whether GenAI is being used within that industry and whether staff are able to upskill or build authentic work-integrated learning. A challenge may be that external organisations may have reservations in providing students with access to proprietary GenAI tools and organisational data, or they may be expensive. During the focus groups and interviews, employers and accrediting bodies did not mention whether GenAI would be important in ensuring assessment tasks were authentic. Students acknowledged an awareness that AI will likely be a staple of future work, and they expect assessments that engage with AI to reflect these future workplace capabilities. Policy makers also acknowledged the need for enhancing AI competency to produce workplace ready graduates. Both students and educators suggested that designing more complex assessments can overcome AI influences.

Equity of access

Although focus group participants did not raise specific equity issues related to GenAI, it has been raised in the media as an issue (Jisc, 2024). Institutional policy makers discussed more broadly the role of infrastructure, including technology, available to students. The cost of accessing GenAI tools is multi-faceted. Firstly, there are differences in access costs and quality of GenAi tools. ChatGPT has two free versions – 3.5 and 4.o, while 4.0 with its larger data set and tools associated with ChatGPT Plus have a subscription cost. Claude.ai is another GenAI tool with a free option, as well as a subscription-based Pro model. These costs place the "higher quality" Gen AI tools out of reach of many students, thus creating competitive advantages for students with greater financial resources. In addition, these tools require internet access, which not all students may have at home, requiring travel to campus for access. The equity solution is to have higher education institutions provide access to these tools using enterprise purchasing power for cheaper costs. However, providing access though institution license might be controversial for the education institutions' missions and the cost may still be prohibitive.

Information integrity

This component focuses on the security of student and assessment information – a risk that remains active in an online assessment environment. When students use GenAI services that are not provided through their institution, there is a greater risk of their personal and assessment data being vulnerable to a security breach. Users may breach the law by uploading or inputting copyrighted information into GenAI, while the output is owned by the GenAI provider. Accrediting body representatives noted this as a major concern with input data entering the public domain. Institutional policy makers argued there will be a need for policies that address the unique challenges posed by GenAI, such as ensuring that AI-generated content is used ethically and does not compromise the integrity of assessments. They further suggested the need of building equity of access in the assessment design process instead of fully relying on infrastructure.

Quality feedback

GenAI has been found to be a useful tool for students in reviewing their work prior to submission of the assessment task in the areas of computer programming (Oliveria et al., 2023) and marketing analytics (Jürgensmeier and Skiera, 2024). It is the ease with which a non-expert user can use GenAI / LLMs that makes it an attractive tool for students to use on-demand. A student in a focus group shared 'Since you're alone, like you can't ask anyone for help. So that's when ChatGPT Comes into play'. However, they do recognise that GenAI may not always be correct 'I get wrong answers from ChatGPT'. An educator participant reported that with hundreds of students (large scale), providing very detailed feedback is difficult because of marking time constraints and that 'leveraging AI to give that feedback would be really important'. Institutional policy makers told us that they would like to see more use of data analytics enabled by AI to guide our assessment feedback practices.

Student experience

GenAI might allow students to create a structure for an assessment task, provide more detail on what is expected. GenAi is useful if the assessment details contain assumed knowledge that non-English speakers may not have, or, linking to the other components of the framework. They could be used to provide a more authentic task to draw greater student interest and provide on-demand feedback to improve student learning. Such benefits have been described in the literature (see Farrelly & Baker, 2023) and conversely, voiced as concerns by students unless they have an experience that is safe, responsible, accessible, and adaptive, consistently integrated throughout their education by knowledgeable, AI-proficient educators to help them reach their future goals (Jisc, 2024). Whilst not specifically referring to GenAI, institutional policy makers and accreditation body participants mentioned cognitive load as contributing to a (negative) student experience of assessment. Introducing multi-layered use of GenAI could contribute to this higher cognitive load and thereby needs attention.

Accreditation requirements

Many degrees/courses/programs have requirements from external accrediting bodies such as law, engineering and accounting. These requirements may include specifications related to assessment such as a minimum

proportion that is identity verified assessment. Such requirementsmay become more specific with the use of GenAI in education and workplaces. This could be in terms of the types or security around assessment, or potentially a requirement to include the application of GenAI within the curriculum. Institutional policy makers commented that accreditation requirements sometimes reflected outdated views of what is required in industry which inhibits innovation in assessment. Conversely, accreditation body participants noted that it is institutional policies and industry-based assessors, who may struggle to accept new approaches, that inhibit adoption of innovative education and assessment formats. Neither group commented on how GenAI could be used to overcome these challenges.

Institutional assessment policies

The Tertiary Education Quality and Standards Agency (TEQSA) is Australia's higher education regulator and they have oversight over every higher education provider – both universities and independent higher education institutions. TEQSA is requiring all providers to respond to a Request For Information (TEQSA, 2024), within which institutions are to provide details on their position on generative AI and how it affects teaching, learning and assessment. There is ongoing work related to how GenAI will be integrated into curriculum (Readman, 2024) and policies on assessment design (Liu & Bridgeman, 2023). Academic integrity and misconduct policies may also require adjustment to cover proper and improper use of GenAI.

Scale associated with the assessment task

Assessment tasks can become challenging with scale – with large cohort sizes assessment can become more complex in terms of implementation, support, moderation of marking. GenAI could provide benefits by allowing greater support for students through easier access to feedback. Educators mentioned this in our focus groups but were unclear about the ethical implications and were looking to institutional policies for guidelines. Lodge et al.'s (2023) recommendations to TEQSA of implementing programmatic assessment could also contribute to dealing with assessment at scale.

Resourcing

The cost of providing equitable access to GenAI, however, may be prohibitive. For example, pricing to the most accurate and up to date version of ChatGPT is approximately A$35/month. Even if institutions received a 50% discount for bulk licence access, an institution with 40,000 students would have an additional $8.4 million in costs. In addition to licencing costs, resources must also be provided to ensure that: 1) staff (both academic and professional) have the skills and training necessary to understand and use GenAI, 2) the time to carefully integrate GenAI into curriculum and assessment is considered, 3) there are mechanisms to enhance the ability to detect improper usage and the workload capacity to gather evidence and report such improper usage as part of academic integrity and misconduct processes. If the digital capability and AI literacy within an institution is raised, possible advantages as observed in industry could be achieved (Dell'Acqua et al., 2023).

Conclusion and future implications

In conclusion, GenAI is a technological advancement that appears to be an unstoppable and (currently) uncontrolled force impacting the world. It has created an unavoidable need for change – not just in a particular field of study, type of higher education institution or geographic region – but across tertiary, secondary and workplace education environments. As an educator focus group participant stated, 'it's kind of been the kick up the bum that we needed to do a lot of the things that we were planning to do in assessment'. This paper explores the multifaceted impact of GenAI using a framework for quality online assessment and demonstrates that designing and adjusting assessments is a challenging and complex process. Educators and assessment designers must navigate these complexities to leverage the potential of GenAI while mitigating the risks and trading-off the framework elements durign the assessment design. Building capacity across the academic institutions for such a rapidly improving technology has significant challenges, especially in the face of the financial deficits faced by many higher education institutions. As institutions and educators grapple with these issues, we believe future research and dissemination activities need to focus on several key areas. First, develop alternatives to high stakes closed book exams as a solution to integrity concerns and consider programmatic assessment (Lodge et al., 2023) with secured assessment points. Next, provide effective

resourcing solutions to widescale changes to assessment and integration of education on ethical use of GenAI into the curriculum. Lastly, develop GenAI-based tools to provide student feedback and conduct evaluation of newly established institutional policies designed to govern GenAI use across the academic sector.

References

Huber, E., Harris, L., Wright, S., White, A., Raduescu, C., Zeivots, S., Cram, A. & Brodzeli, A. (2023). Towards a framework for designing and evaluating online assessments in business education. *Assessment & Evaluation in Higher Education*, *49*(1), 102–116. https://doi.org/10.1080/02602938.2023.2183487

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4573321

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, *13*(11), 1109. https://doi.org/10.3390/educsci13111109

Higher Education Standards Framework (Threshold Standards) (Cth) (2021), https://www.legislation.gov.au/F2021L00488/latest/text

Jisc. (2024, May 14). *Students want to see generative AI integrated throughout curriculum, despite increased concerns around ethics and equity*. https://www.jisc.ac.uk/news/all/students-want-to-see-generative-ai-integrated-throughout-curriculum-despite-increased-concerns-around-ethics-and-equity

Jürgensmeier, L., & Skiera, B. (2024). Generative AI for scalable feedback to multimodal exercises. *International Journal of Research in Marketing*. https://doi.org/10.1016/j.ijresmar.2024.05.005

Liu & Bridgeman (2023, December 08). *Embracing the future of assessment at the University of Sydney*. Teaching@Sydney. https://educational-innovation.sydney.edu.au/teaching@sydney/embracing-the-future-of-assessment-at-the-university-of-sydney/

Lodge, J. M., Howard, S., Bearman, M., Dawson, P. & Associates. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality Standards Agency. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

Oliveria, E., Rios, S. & Jiang, Z. (2023) AI-powered peer review process. In T. Cochrane, V. Narayan, C. Brown, K. MacCallum, E. Bone, C. Deneen, R. Vanderburg, & B. Hurren (Eds.), *People, partnerships and pedagogies*. Proceedings ASCILITE 2023. Christchurch. https://doi.org/10.14742/apubs.2023.482

Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T. L., Kar, A. K., Lee, V. H., Loh, X. M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., … Wong, L. W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*. 1-32. https://doi.org/10.1080/08874417.2023.2261010

Readman, K. (2023, February 07). *GenerativeAI@UTS: effective, ethical engagement*. University of Technology Sydney. https://lx.uts.edu.au/blog/2023/02/07/generative-ai-uts-engagement/

Steinhardt, S. J. (2023). *Latest Version of ChatGPT Passed a Practice CPA Exam*. NYS Society of CPAs. https://www.nysscpa.org/article-content/latest-version-of-chatgpt-passed-a-practice-cpa-exam-052323#sthash.Z2xBxFpV.dpbs.

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

Tertiary Education Quality and Standards Agency. (n.d.). *Request for information: Addressing the risk of artificial intelligence*. Retrieved July 19, 2024, from https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence/request-information-addressing-risk-artificial-intelligence

Wood, D. A., Achhpilia, M. P., Adams, M. T., Aghazadeh, S., Akinyele, K., Akpan, M., Allee, K. D., Allen, A. M., Almer, E. D., Ames, D., Arity, V., Barr-Pulliam, D., Basoglu, K. A., Belnap, A., Bentley, J. W., Berg, T., Berglund, N. R., Berry, E., Bhandari, A., … Zoet, E. (2023). The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions? *Issues in Accounting Education*, *38*(4). https://doi.org/10.2308/ISSUES-2023-013

ASCILITE 2024

Navigating the Terrain:

Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies