Rethinking assessment in response to generative artificial intelligence

Jacob Pearce[1], Neville Chiavaroli[1]

[1] Tertiary Education, Australian Council for Educational Research, Camberwell, Australia.

Corresponding author:

Jacob Pearce, PhD
Principal Research Fellow
Specialist & Professional Assessment, Tertiary Education
Australian Council *for* Educational Research
p: +03 92775508
m: +614 01853880
a: 19 Prospect Hill Road, Camberwell, VIC 3124, Australia
w: www.acer.org e: jacob.pearce@acer.org

Word count of text: 878

- We are responsible for the content of this submission.
- Both authors contributed equally to the conceptualisation, drafting, editing and final version of this submission. Both authors are responsible for its contents and approved the final version.
- We have not excluded any qualified authors in the production of the manuscript or its revisions.
- This manuscript has not been published and is not under consideration in the same or similar form in any other journal.
- We declare no potential conflicts of interest that may affect the information presented in the manuscript.

Rethinking assessment in response to generative artificial intelligence

5 Pull-out Points

- Generative AI has significant implications for the way we assess our students and trainees.

- We can no longer rely on non-invigilated assessments and submitted 'artifacts' to demonstrate student learning.

- We see an important distinction between 'assisted' assessments and 'unassisted' assessments.

- The time has come for the 'rehabilitation' and re-acceptance of the oral format.

- Generative AI … can add fidelity and nuance to assisted assessment, while facilitating a greater focus and purposefulness to unassisted assessment.

Rethinking assessment in response to generative artificial intelligence

The use of decision-making support tools during assessments, such as electronic differential diagnosis in examinations,[1] is just the tip of the iceberg when it comes to how technology is currently changing assessment practice. We say this, because it appears we've reached a transformative stage in the development of artificial intelligence (AI). So-called generative AI is here. OpenAI's ChatGPT has burst onto the scene, and other tools are coming. ChatGPT-3 is a language model designed to understand and generate natural language responses to a wide range of prompts and questions. It was trained on a massive amount data, corresponding to approximately 45 terabytes of text, or hundreds of billions of words. At least this is what ChatGPT told us when we asked. And just as we are all trying to understand the ramifications of all this for education and assessment, the multimodal ChatGPT-4 has been released that can 'see' and understand images, highlighting just how fast this technology is advancing.

Generative AI has potentially radical and significant implications for the way we assess our students and trainees. One AI tool has already shown to be capable of passing the US Medical Licensing Exam.[2,3] Another AI tool has performed reasonably well in a Royal College of Radiologists examination, involving the interpretation of radiographs.[4] Even more significantly, at its launch ChatGPT-4 was shown to have performed in the top 10% on a range of well-known high-stakes examinations. The ability of these latest generations of AI to respond convincingly to assessment tasks, and draft questions and rubrics on specialist topics, is certainly impressive. We are at the point where we can no longer rely on non-invigilated assessments and submitted 'artifacts' to demonstrate student learning and competence.

This is bringing many long-term demands on educators, course coordinators, and curriculum designers, forcing us to rethink assessment approaches. It certainly calls into question the role and place of written assessment tasks, and makes us reconsider what value they offer or will continue to offer. We see an important distinction between 'assisted' assessments and 'unassisted' assessments. The former is assessment that allows the candidate to draw on tools and resources, including textbooks, the internet, decision-making support tools,[1] and now, generative AI models; in many ways, representative of real-life clinical practice. Unassisted assessment, on the other hand, refers to times when we may wish to assess our students' and trainees' knowledge and understanding without access to such resources, such as for certification or summative contexts when the independent (and verifiable) demonstration of clinical knowledge and reasoning is a necessary element of the relevant competencies. We would argue that oral assessment methods such as the viva play an important role for such unassisted assessment.

Oral assessment methods have been undervalued in many assessment circles for some time, in part due to their perceived poor reliability, lack of standardisation and potential for assessor bias.[5] But as many assessment scholars have pointed out, the value or 'utility' of an assessment format depends on many different and competing factors.[6] In certain circumstances, the intrinsic characteristics of oral assessment – in particular its mode of direct communication, interactivity and flexibility – come to the fore and make it a particularly apt choice for unassisted assessment. For those occasions when we need direct and verifiable evidence of our learners understanding and competence, the oral assessment format is a relatively straightforward and highly useful method.

Beyond security and assurance of learners' independent knowledge and understanding, the oral format has another specific advantage. Since candidate responses can be clarified in real time, deep probing of knowledge, understanding and higher-order thinking is possible. Such examiner interaction can be standardised effectively with thoughtful design and appropriate training.[7,8] The recent emergence of clearer guidelines for their use provides support for more valid application of the oral format in medical education assessment.[9,10]

With the recent increase and facilitation of virtual assessment through convenient platforms, and the new challenge to un-invigilated assessment formats posed by AI, we think the time has come for the 'rehabilitation' and re-acceptance of the oral format as a highly valuable and unique form of assessment in medical education. As with any assessment format, this requires that assessment designers and practitioners be thoughtful in planning and designing oral assessments, including clarifying the different types of prompting practices, and educating all stakeholders on how to best operationalise this assessment format.

Generative AI need not threaten the validity or trustworthiness of our assessments in either formative or summative contexts. Rather, it can add fidelity and nuance to assisted assessment, while facilitating a greater focus and purposefulness to unassisted assessment. The role and value of generative AI within assisted assessments should be further explored. The usefulness of these tools is quickly becoming apparent, with many rightly considering how to use AI to scaffold learning. Building the use of AI into our *assisted* assessment tasks may advance learning, help identify misconceptions and oversight, and suggest avenues for solutions to problems. For example, AI can be used as a confirmatory resource or text generation tool, akin to the way doctors routinely use their phones to consult clinical practice guidelines, or standardise reports through the use of document templates. The other side of such strategic inclusion of AI in assisted assessment is the opportunity to place greater emphasis on the assessment quintessential human skills, such as judgment, evaluation, and critical reasoning through *unassisted* assessments.[11]

Although generative AI does not warrant a reactive dissolution of established assessment practices, it does call for a rethinking of assessment as we know it, with a sharpening focus on its purpose in context. Initially, we propose that the oral assessment format should be 'reclaimed' as it can play an important role in meeting the certification function of assessment in unassisted contexts. Further, access to AI in assessment contexts where assisted tasks make sense, will also need to be embraced as part of a comprehensive assessment system.

References

1. Sibbald M, Pugh D, Sherbino J, Morin M, Norman G, Monteiro S. Does Allowing Access to Electronic Differential Diagnosis Support Threaten the Reliability of a Licensing Exam? Med Ed. 2023 in press/DOI

2. Tseng V. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. PLOS Digit Health. 2023 Feb; 2(2): e0000198.

3. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large Language Models Encode Clinical Knowledge [Internet]. arXiv; 2022 [cited 2023 Mar 9]. Available from: http://arxiv.org/abs/2212.13138

4. Shelmerdine SC, Martin H, Shirodkar K, Shamshuddin S, Weir-McCall JR. Can artificial intelligence pass the Fellowship of the Royal College of Radiologists examination? Multi-reader diagnostic accuracy study. BMJ. 2022 Dec 21;e072826.

5. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. Med Teach. 2005;27: 294–297. https://doi.org/10.1080/01421590500126437.

6. Van der Vleuten C. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ. 1996;1: 41–67.

7. Tekian A, Yudkowsky R. Oral examinations. In: Downing SM, Yudkowsky R (eds) Assessment in Health Professions Education. New York: Routledge; 2009.

8. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. Adv Health Sci Educ. 2010;15: 277-289.

9. Pearce J, Chiavaroli N. Prompting Candidates in Oral Assessment Contexts: A Taxonomy and Guiding Principles. Journal of Medical Education and Curricular Development. 2020 Jan;7:238212052094888.

10. Pylman S, Ward A. 12 tips for effective questioning in medical education [published online April 16, 2020]. Med Teach. 2020. doi:10.1080/0142159X.2020. 1749583

11. Bearman M, Luckin, R. Preparing university assessment for a world with AI: Tasks for human intelligence. In: Re-imagining university assessment in a digital world. 2020. p. 49–63.