Medical Teacher

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/imte20

Artificial Intelligence in Health Professions Education assessment: AMEE Guide No. 178

Ken Masters, Heather MacNeil, Jennifer Benjamin, Tamara Carver, Kataryna Nemethy, Sofia Valanci-Aroesty, David C. M. Taylor, Brent Thoma & Thomas Thesen

To cite this article: Ken Masters, Heather MacNeil, Jennifer Benjamin, Tamara Carver, Kataryna Nemethy, Sofia Valanci-Aroesty, David C. M. Taylor, Brent Thoma & Thomas Thesen (09 Jan 2025): Artificial Intelligence in Health Professions Education assessment: AMEE Guide No. 178, Medical Teacher, DOI: 10.1080/0142159X.2024.2445037

To link to this article: https://doi.org/10.1080/0142159X.2024.2445037

MEDICAL TEACHER
https://doi.org/10.1080/0142159X.2024.2445037

MEDICAL TEACHER
Taylor & Francis
Taylor & Francis Group

AMEE GUIDE

Artificial Intelligence in Health Professions Education assessment: AMEE Guide No. 178

Ken Masters[a], Heather MacNeil[b], Jennifer Benjamin[c], Tamara Carver[d], Kataryna Nemethy[e], Sofia Valanci-Aroesty[f], David C. M. Taylor[g], Brent Thoma[h] and Thomas Thesen[i]

[a]Medical Education and Informatics Department, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman; [b]Department of Medicine, University of Toronto, Toronto, Canada; [c]Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA; [d]Institute of Health Sciences Education, Faculty of Medicine and Health Sciences, McGill University, Montreal, Canada; [e]Baycrest Academy for Research and Education, Baycrest Academy for Research and Education, Toronto, Canada; [f]Royal College of Physicians and Surgeons of Canada, Ottawa, Canada; [g]College of Medicine, Gulf Medical University, Ajman, United Arab Emirates; [h]School of Medicine, Toronto Metropolitan University, Toronto, Canada; [i]Department of Medical Education, Dartmouth College Geisel School of Medicine at Dartmouth, Hanover, NH, USA

ABSTRACT

Health Professions Education (HPE) assessment is being increasingly impacted by Artificial Intelligence (AI), and institutions, educators, and learners are grappling with AI's ever-evolving complexities, dangers, and potential. This AMEE Guide aims to assist all HPE stakeholders by helping them navigate the assessment uncertainty before them. Although the impetus is AI, the Guide grounds its path in pedagogical theory, considers the range of human responses, and then deals with assessment types, challenges, AI roles as tutor and learner, and required competencies. It then discusses the difficult and ethical issues, before ending with considerations for faculty development and the technicalities of AI acknowledgment in assessment. Through this Guide, we aim to allay fears in the face of change and demonstrate possibilities that will allow educators and learners to harness the full potential of AI in HPE assessment.

ARTICLE HISTORY
Received 9 December 2024
Accepted 17 December 2024

KEYWORDS
Artificial intelligence; assessment; medical education; health professions education

1. Introduction

1.1. Background

Educators are familiar with the adage that 'assessment drives learning' [1], and, in an ideal world, assessment results are measures of learning. While the Internet impacted learning and assessment in Health Professions Education (HPE), Artificial Intelligence (AI) introduces a new perspective on learning and assessment. Our focus in this Guide is on AI in HPE assessment.

Although the idea of thinking machines and AI has existed for decades [2,3], the public release of the Generative AI (GenAI) tool, ChatGPT 3.5 on 30 November 2022 caught the world's attention and fired both learners' and teachers' imaginations. (It is also for that reason that, although this Guide speaks of 'AI', most examples will be from the sub-set of GenAI.)

Since then, the publication surge covering AI in HPE followed patterns found in other education literature, as plagiarism, impact on written work, cheating, and similar concerns were discussed [4]. As the AI models advanced, concerns increased, especially as the systems scored increasingly higher marks in local, national, and international academic examinations; simultaneously, however, the AI tools' ability to assist teachers and learners in formative and summative assessment was recognised [4].

As HPE institutions have grappled with AI's impact on assessment, and have developed strategies for using and coping with it, some trends have emerged. The first is the rather deterministic realisation that AI is here to stay, and its power is advancing. The second, more closely related to assessment, is that we cannot simply plug AI tools into our current assessment methods and carry on as before: our assessment methods need to be adjusted, some scrapped, and new methods introduced. Because AI is ever-advancing, this will not be a once-off exercise, and should be revisited repeatedly.

In this light, there is a need to assist HPE educators in using AI in assessment [5]. The required re-orientation and embarking on the path to effective AI use in HPE assessment will not be without difficulty; this AMEE Guide attempts to lead readers along that path.

Practice points

- As AI usage in HPE increases, so will its impact on AI assessment.
- Through pedagogical theory and best practices, educators need to recognise the potential, the possible problems, and know how correctly use AI in HPE assessment.
- This AMEE Guide helps them begin that journey.

CONTACT Ken Masters itmeded@gmail.com Medical Education and Informatics Department, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman

Supplemental data for this article can be accessed online at https://doi.org/10.1080/0142159X.2024.2445037.

1.2. This guide's structure

The Guide covers several topics, and we have attempted to arrange them in a logical order. We begin with the pedagogical theories and human responses to AI, and then discuss the conceptual view of AI's impact on personal and flexible assessment, standardisation, and assessment types. We then move to more directly applicable topics of AI as tutor and learner, and competency-based assessment. Lastly, we deal with difficult and ethical issues, leading into faculty development and ending with a discussion of acknowledging AI usage in HPE assessment.

Each topic follows a similar (but not identical) layout: introducing the issue, referring to the relevant theory, discussing the potential, providing examples and best practices, and indicating limitations. Some topics are supported by material in the Appendices.

Because this Guide focuses on AI's use in assessment, it will not describe general or digital assessment in any depth, as this has already been covered in other publications [5,6]. Simultaneously, this Guide will not see AI as a solution to address all educational challenges. Rather, this Guide will focus on identifying areas in HPE Assessment that can be enhanced by AI usage, while simultaneously acknowledging possible problems caused by AI usage in HPE assessment.

As our focus is on education, rather than technology, we begin with the educational theoretical frameworks in the next topic.

2. Theories and frameworks

The explosion of AI tools for assessment compels educators to consider how best to harness their potential in pedagogically-informed assessment. Institutions should provide explicit guidance on the acceptable AI use in assessments, and foster an environment of critical inquiry and transparency through modelling best practices when using GenAI, including biases and limitations. To enable this, we will visit the foundational theory-based approaches for technology integration in formative assessment ('assessment *for* learning') and summative assessment ('assessment *of* learning') [6,7], and highlight the need for a pragmatic, theory-based approach with their use.

2.1. Two pedagogical models

We begin by describing two commonly-used models: Technology, Pedagogy, Content, Knowledge Framework (TPACK) [8] and the Passive, Interactive, Creative, Replacement, Amplification, Transformation Framework (PICRAT) [9,10], both previously successfully studied in relation to online education in HPE [11,12].

TPACK considers three types of knowledge: Technological (TK), Pedagogical (PK) and Content (CK), and describes how educators should examine *how* (i.e. pedagogy) and *what* (i.e. content) is taught, as well as *why* technology integration might be used to enhance the learning experience. TPACK provides a helpful framework to conceptualise how we can train educators to use AI with a pedagogically-informed approach.

While TPACK provides an overview of integrating technology, the PICRAT framework [10] provides guidance on how to integrate AI in assessment through the learner's (PIC) and educators' perspective (RAT):

- *Passive:* learners are impacted passively by assessment strategies (e.g. time spent on module, number of completed assignments or quiz attempts, without direct interaction or personalized feedback),
- *Interactive:* learners interact and learn with AI (e.g. Socratic type assessment to gain deep understanding), and
- *Creative:* learners create content using AI (e.g. generating feedback to written assignments).
- *Replacement:* educators replace assessments using AI tools (e.g. MCQ generation, virtual patients),
- *Amplification:* improve assessment using AI (e.g. personalized learner feedback based previous assessment outcomes), and
- *Transformation:* assessment that cannot be achieved without the use of AI (e.g. large data analysis to personalize learning plans and feedback).

As shown in the rest of this Guide, understanding these models allows educators to situate their AI usage in assessment more comfortably within evidence-based teaching frameworks.

2.2. Two frameworks

From considering how technology aids in achieving learning outcomes, we now consider how to assess learner competency by considering two frameworks: Programmatic Assessment and the Competency-Based Medical Education (CBME) Core Components Framework.

Programmatic assessment is a method that collects information in a meaningful way to gain a holistic picture of the learner's competence and development [6,7]. It emphasises continuous, integrated assessments that capture many competencies over time. By aligning formative assessments with this model, educators can ensure that learners receive ongoing, diverse feedback across various skills, supporting gradual competency development without high-stakes pressure.

The Core Components Framework describes five components of CBME [13]. They include the existence of outcome competencies that are clearly articulated, competencies that are sequenced progressively to facilitate a developmental approach, incorporating tailored learning experiences and competency-focused instruction that facilitate competency development, and using Programmatic Assessment to document and perform a holistic review of learners' progress.

While CBME has been broadly embraced within HPE, it does not include time variable education, a key component of competency-based education in the broader literature. AI, with its ability to synthesise vast amounts of learner data, can help track learners' progress and provide recommendations for personalised learning and development through intelligent tutoring systems (ITS), simulations, and peer assessments to support ongoing learning [6].

In addition to these, Bloom's Taxonomy [14] provides a framework underpinning the tenets of cognitive learning

and can be used for applying AI in formative and summative assessments [15].

Within these theories and frameworks, educators need to model best practices for AI use in assessment by ensuring learner anonymity, reducing the impact of algorithm bias through transparency, being explicit about learner use of AI, supporting learner autonomy and modelling critical evaluation of AI outputs [16].

Just as our starting point with the Guide was not technology, but pedagogy, so the next topic considers humans (learners and faculty), and their overall responses to AI in assessment.

3. Human responses

This Guide largely views AI in assessment positively, reflecting the authors' biases and AI use in their teaching and learning, following and demonstrating best practices, and reporting on shared successes.

While this topic does not argue *against* using AI, readers should recognise that not all learners and teachers enthusiastically embrace AI [17,18], and there are valid reasons for caution. This section aims to balance enthusiasm with caution, while addressing common concerns.

From the previous topic, it is evident that sound pedagogical models and frameworks already exist for AI's incorporation; teachers and learners can ground their usage within these frameworks, allowing them to overcome misconceptions and fears. In addition, educators can apply some basic principles of the Unified Theory of Acceptance and Use of Technology (UTAUT) [19] to assess human receptivity to technology in general, and AI in particular. In the specific context of HPE, the Medical Artificial Intelligence Readiness Scale for Medical Students (MAIRS-MS) [20] can gauge learner readiness. Similarly, PICRAT (outlined above) is a useful starting point for evaluating and integrating AI into learning and assessment [21].

There are, however, human responses to consider and, perhaps, discuss in the classroom. We will briefly cover two: anthropomorphising AI, and balancing enthusiasm with fear.

3.1. Anthropomorphising AI

A challenge to introducing AI to HPE assessment is the tendency to anthropomorphise AI. Although discussions around AI sentience and rights are ongoing [22], AI anthropomorphising often manifests on social media reactions, such as those perceiving the voice of ChatGPT4o as 'flirty' [23] or with a voice perceived to be 'female'; from there, other system characteristics system are viewed through that perception. This misinterpretation overlooks the design intention of creating a user-friendly computer Audio User Interface, exhibiting the user-friendliness that we require from computer Text User Interfaces and Graphical User Interfaces.

Similarly, although gender and other biases exist in AI tools, allowing interruptions of ChatGPT4o's 'female' voice has been criticised as gender-biased, ignoring the fact that interrupting a machine output has been a feature of personal computing for more than 50 years and that all these characteristics exist also for the 'male' voices.

These anthropomorphic criticisms are repeated in various ways (e.g. criticisms of NotebookLM's podcast feature) and hold little value, akin to assigning gender or racial characteristics to a computer mouse based on design.

It is important for learners to remember, and educators to stress, that AI systems, regardless of their performance, are not human, and do not possess gender, race or other human demographic traits. Projecting one's own demographic perceptions onto AI systems reflects personal biases and detracts from its utility. Referencing the established theories outlined above (especially UTAUT) can help reinforce the purely technological nature of AI, no matter how human-like it is. This is especially important when students receive feedback from or interact with AI assessment systems: they should always bear in mind that they are not interacting with a human, but a machine.

3.2. Balancing extremes: Enthusiasm vs. fear

Educators: Reactions to AI range from excessive enthusiasm to deep fear, ironically, leading to similar results: misuse of AI tools. For example, 'AI-detectors' are employed by both those fearing learner misuse and by those who overly-embrace AI. (These tools are discussed in more detail later in this Guide.)

There are, however, legitimate concerns of learners' losing skills, inappropriately using AI for assessments, and fears of faculty's replacement by AI or using AI to review and grade learners' work. Some educators face challenges navigating an overwhelming array of tools, evolving institutional policies, and complexities of data-sharing, surveillance and security. The future prospect of Artificial General Intelligence (AGI) adds to the uncertainty [22].

Learners: Learners are affected similarly, and their fears are compounded by fears of AI-detectors' misclassifying their original work as AI-generated [24].

If educators can begin by recognising these issues, have open discussions with colleagues, learners, institutions and organisations, and apply their pedagogical frameworks diligently, then much of the fear can be allayed, and the focus can be on education.

With this brief overview as context, we move into the next topic, in which we examine ways in which AI systems can be personalised to better support individual learners' assessment.

4. Personalised and flexible summative assessment

4.1. Introduction

In HPE, there is a growing shift toward flexible curricula that accommodate diverse learning paces, preferences, and individual skills. Traditional, rigid structures no longer meet learners' needs who benefit from more adaptive, accessible and learner-centred approaches.

4.2. The value

Personalised assessments are particularly valuable in HPE, where phase completion (such as moving from preclinical to clinical training, specialised continuing professional

development) requires evidence of specific competencies. By providing real-time feedback and competency evaluations, on-demand assessments support personalised learning paths, enabling learners to progress through training phases at their own pace. These assessments enable learners to demonstrate readiness for each phase when they feel prepared, aligning with their individual learning paths and career trajectories, rather than a set timeline [4]. This approach supports progression based on demonstrated competency rather than time, confirming a learner's readiness to advance [25].

As AI-driven assessments continue to evolve, they offer a promising pathway for preparing learners for phase completion and certification, ultimately ensuring readiness for the demands of professional practice.

Certification in HPE requires meeting standardised competency levels, often through practical evaluations and exams. On-demand assessments make this process adaptive, offering frequent, competency-based evaluations aligned with real-world clinical skills. As learners progress, they receive timely feedback on their readiness for professional practice, helping them meet certification standards at their own pace [7], while also benefiting from objective, AI-driven feedback that minimises potential educator bias.

There are some Best Practices for on-demand assessments; these and a few case studies are covered in Appendix 2.

4.3. Limitations

While this competency-based approach offers flexibility, it is not without drawbacks. Solely focusing on rigidly-defined competencies may overlook the nuances of learning, including integrating overlapping skills and the holistic development needed for real-world practice [26].

In addition, there are ethical concerns of data collection and protection, algorithm bias in personalised learning and resource allocation, as discussed later in this Guide. Finally, it is important to emphasise that the practice must be grounded in theory, not driven by the desire to use technology, but rather, by the desire to effectively implement sound assessment principles that are difficult to implement without AI.

While promoting personalised assessment, it is also crucial to maintain standards, and AI can be used to ensure standardisation. How this is done will be covered in the next topic.

5. Assessment standardisation

While personalised assessment is an effective tool for learning, standardised assessments are foundational in HPE because they ensure that learner evaluations are fair, consistent, and comparable across institutions. They are essential for certifying that graduates meet the requisite competencies for patient safety, contributing to public trust in healthcare professionals. Examinations such as the United States Medical Licensing Examination (USMLE) set uniform benchmarks to assess foundational knowledge and clinical skills.

However, advancements in precision education, driven in part by GenAI, are introducing new possibilities for tailoring individualised assessments pertinent to career trajectories [27]. This shift challenges educators to balance the uniformity of standardised assessments with the benefits of personalised assessments discussed above. Striking this balance is critical to maintaining fairness, and avoiding inconsistencies in rigour or content that may lead to perceived or actual discrepancies in competence evaluations.

Individualised assessment aligns with the concept of precision medical education, in which the learning experience is tailored to each learner's current level and specific needs to enhance learning outcomes [27]. For HPE, precision learning offers a paradigm shift that brings both opportunities for tailored assessment and challenges in maintaining assessment rigour and standardisation.

5.1. Integrating GenAI into assessment

Standardised tests like the USMLE rely on a rigorous validation process, rooted in classical test theory (CTT) and item response theory (IRT), which includes extensive pre-testing and equating to ensure comparability across test versions. Questions generated by GenAI, particularly if created in real-time or without established psychometric validation, present challenges for reaching equivalent rigour in high-stakes settings. Furthermore, the 'black box' nature of many GenAI algorithms, resulting in educators' not really understanding how they generate output, adds additional complexity, as it may be difficult to ensure that AI-generated items consistently adhere to tested educational standards and avoid unintended biases.

Despite the fact that LLMs are not yet sufficiently accurate to consistently generate reliable, high-stakes assessment items without human oversight [28], there are free tools that can assist in generating high-quality MCQs [29,30]. Overall, LLMs show promise in handling complex medical knowledge, and with technological advancement and rigorous evaluation, they could soon achieve the standards required for creating standardised assessments [31]. As such, HPE researchers should assess the accuracy and reliability of different LLM models and configurations in generating assessment items. This will likely necessitate new paradigms for research and evaluation [32].

For clinical assessments, GenAI can improve the standardisation of clinical exams like the OSCE by utilising speech-to-text transcription to capture learner-patient interactions and analyse these based on established rubrics [33]. This approach eliminates variability inherent between human raters, a known limitation of current OSCE scoring. Fine-tuning LLMs, where models are trained on specific datasets reflecting desired assessment rubrics, has been shown to improve scoring reliability [34]. If aligned with standardised evaluation criteria, the use of fine-tuned LLMs that have been proven to score reliability across different sites can potentially enhance consistency and fairness across institutions [35].

Until LLMs consistently demonstrate reliability much better than human raters, however, a 'human-in-the-loop' approach remains essential. This requires human oversight in the generation and scoring of assessment items, ensuring that the quality and accuracy of AI-supported assessments meet the rigorous standards required in HPE.

5.2. Conclusion

AI-generated items show great promise and can be used, but it is essential that they meet both the standards of conventional psychometric validation and the public's trust, and so high-stakes assessments should ensure that such items have been reviewed by humans, or continue to rely on well-established test items that have undergone extensive validation. This ensures that the standards of fairness, reliability, and equity in HPE assessment remain uncompromised as new AI-driven tools are integrated into HPE. More rigorous and systematic research is needed before GenAI can be reliably used in high-stakes standardised assessments, but its potential to enhance consistency, fairness, and scalability is promising.

Currently, AI's most appropriate role at present may lie in low-stakes, formative assessments and adaptive learning environments, where personalisation can effectively support learning without the implications of high-stakes evaluation. We now turn to this topic.

6. Personalised and flexible formative assessment

6.1. Introduction

As HPE shifts toward personalised and flexible learning, formative assessment has become a crucial element in guiding and supporting individual learning trajectories. Unlike summative assessments, used to determine phase completion or certification, formative assessments offer continuous, low-stakes evaluations that enable learners to track their progress, identify gaps in knowledge, and receive tailored feedback. With advancements in AI, formative assessments can now be highly adaptive and responsive, offering real-time feedback that aligns with each learner's unique needs and pace of learning [4].

Balasooriya [36] imagines a future where learners are paired with an AI tutor, one who knows their learning preferences, strengths, weaknesses and career trajectory, one that follows them throughout their career, from undergraduate to postgraduate to continuing professional education as 'Personalized Digital Learning Companions … as a partner rather than merely another technological tool' [36].

While a little futuristic, this vision aligns with personalised and flexible programmatic formative assessment, encouraging supporting feedback throughout the learner's journey. Several educational frameworks inform the implementation of formative assessments in personalised learning environments particularly when integrating AI tools. Self-Regulated Learning (SRL) principles, for instance, provide valuable insights into designing assessments that encourage learners to monitor and adjust their strategies through feedback [37]. These frameworks underscore the importance of designing formative assessments that not only track progress but also actively foster autonomy and personalised growth.

To make this practical in HPE, AI can play a significant role in implementing adaptive, personalised formative assessments. For example, custom AI models like those described by Kıyak et al. [29,30] can generate multiple-choice questions tailored to individual learning needs. By leveraging tools that create case-based scenarios or adapt content complexity based on learner performance, educators can ensure that assessments remain relevant to the learner's current skill level. These AI-driven tools, detailed further in Appendix 3, highlight the potential for using GenAI in both question creation and targeted feedback.

6.2. Best Practices for AI-driven formative assessment

To implement personalised formative assessments effectively in HPE, the following best practices leverage the AI tools and case studies detailed in Appendix 3:

1. Align Assessments with Competencies for Certification: Clearly define learning objectives that map directly to phase and certification competencies. Tools like Gradescope can automate the alignment process by using AI to ensure consistency across formative and summative assessments, streamlining evaluation and feedback.
2. Leverage AI for adaptive formative assessments: Adaptive tools such as Labster and Quizlet can tailor assessments to individual learning progress. For example, virtual patient simulations (Case 1 in Appendix 3) dynamically adjust complexity based on learner performance, ensuring that students are both challenged and supported in real-time.
3. Provide timely, actionable feedback: AI-powered platforms like Socratic by Google or AI-driven peer review systems (Case 3 in Appendix 3) provide immediate, targeted feedback. This enables learners to identify areas for improvement and make course corrections promptly, enhancing self-regulation and engagement.
4. Integrate reflective practice with AI support: Encourage learners to reflect on their assessments using AI-guided prompts. For example, the feedback loops in tools like PeerWise or Anatomy Identification Quizzes (Case 2 in Appendix 3) help students consolidate knowledge and strategize for improvement.
5. Use AI to facilitate collaborative learning: Peer-based tools like PeerWise promote collaborative learning by enabling students to create and critique each other's work with AI-enhanced guidance. This dual-feedback approach fosters deeper understanding and critical thinking.

6.3. Limitations

Despite the potential benefits of personalised formative assessments, some challenges must be addressed to ensure effective implementation:

- Maintaining Consistency and Fairness in Feedback: With personalised learning, there is a risk of feedback inconsistency, which can impact fairness. To mitigate this, formative assessment criteria should be clearly defined, and feedback mechanisms standardised to ensure that all learners receive equitable guidance, regardless of their individual pathways.
- Balancing Personalised Feedback with Workload: For educators, providing individualised feedback can increase workload significantly, especially in large cohorts. AI tools that automate basic feedback processes can alleviate this

burden, but they may not replace the need for nuanced, human feedback. Continuous training and support for educators are essential to balance the efficiency of AI with the depth of instructor-led feedback [16].

As personalised formative assessments evolve with AI-driven tools, they are reshaping how we evaluate learning, fostering autonomy and adaptability. However, traditional assessment formats like the take-home essay must now navigate this evolving landscape, finding ways to incorporate AI while maintaining their value in gauging complex thought processes and academic rigour. With this understanding, we now turn to examine the role of the take-home essay.

7. The take-home essay

The take-home essay, or written assignment, has been crucial to HPE, laying the foundations for academic work. Assignments might begin with initial institution application essays, and follow with various research projects. Irrespective of learner level, well-documented processes exist [38,39], beginning with a question, assimilating the known, citing and referencing, adding new information (e.g. empirical data), then a discussion and conclusions.

The intellectual processes of writing papers requires high-level thinking in Bloom's Taxonomy [14] and are similar to those described by Constructivism, in which learners (initially with close, but ever-reducing, guidance) construct their argument, drawing together material, data, and ideas, and then construct answers [40,41]. For that reason, with the advent of ChatGPT 3.5, through which text 'prompts' could be used to generate original text responses [42], perhaps the greatest impact it had on assessment in education was on the take-home written assignment.

Despite the 'hallucinations' problem with AI, learners and educators alike realised that LLMs had forever changed the take-home assignment. Institution and individual reactions were frequently negative, including banning LLMs [43], insisting on rough drafts of essays written by hand or in class with no access to the Internet [44], resignations [45], or widespread use of 'AI-detection' software, followed by the outright failing of assignments suspectedly written by or with the aid of LLMs [46].

In these reactions, one common goal was the retention of the take-home assignment. In doing so, we appear to have forgotten the point of the take-home assignment or essay: the aim is not to write the essay, and it never has been.

So, what is the point of the essay? A clue to answering this question lies in a statement made by Professor Mumm, who awarded all his students an 'X' (incomplete) grade: 'I have to gauge what you are learning[,] not a computer' [46].

All assessments aim to gauge learning, and, until now, largely because of its Constructivist nature, the take-home essay has been a reasonably good method to gauge learning and understanding of complex ideas and methods. It does, however, also have a flaw: it permits mostly undetectable cheating, the amount of which is mitigated only by time and money. LLMs exposed this weakness and created a two-fold problem:

- Firstly, the professor and all others who ban AI usage are correct in thinking that AI renders the standard take-home essay mostly irrelevant in gauging learning.
- Secondly, although there may be an argument for specifying *minimal* usage of AI in take-home essays, when we do this, we are no longer preparing learners for the real world, in which they will be expected to use AI tools competently.

7.1. Solutions

Possible technological solutions (in line with current AI-detection methods) do exist in the form of closed writing tools that give access to AI features, but track learner usage and provide reports. These are still in their infancy, are sometimes expensive, are relatively easy to trick, and enter a murky area of ethical learner tracking and over-surveillance, including their brainstorming and rough drafts. Although they could be used to guide student AI prompt formulation, they might also encourage educators to rely too heavily on machine-generated reports and statistics (similar to the way in which some currently confuse 'similarity-checkers' with 'plagiarism-checkers' already).

A pedagogical solution is not to block or minimise AI and retain the essay, but rather to *embrace* AI, and find other ways to gauge learning, especially those that use AI. We can begin by re-thinking the essay: instead of seeing it as a single, stand-alone and complete product, see it as *one* aspect of an assessment process; instead of seeing the learner as an *author*, see the learner as a *director*, using multi-modal AI research and input systems to create a range of multi-modal AI outputs (e.g. text, video, audio) to demonstrate that the learner has addressed the problem or answered the question in such a way that indicates an understanding and control of the subject matter. The essay becomes a component of a greater whole.

This will, however, require changes to this assessment type, and, given the essay's relationship to published literature, will probably require journals to change their perceptions of the research paper. But that is a topic for another Guide.

For this Guide, the next steps involve designing these new assessments including AI usage. These will be discussed in more detail in the next topic.

8. Other assessment types

AI's role in rendering the standard essay irrelevant when assessing learning, reinforces the need to rethink traditional assessment methods. Traditional approaches, such as MCQs and short-answer tests, struggle to effectively measure higher-order thinking (analysis, application, evaluation, creation). These traditional approaches typically provide a momentary snapshot of performance, overlooking learning processes [47]. They also tend to be uniform, ignoring learners' diverse backgrounds, and fail to reflect real-life authentic scenarios and tasks [47].

Assuming learners will use GenAI, assessments should be designed to minimise opportunities for its unethical use and leverage its affordances. Alternative assessments should require learners to determine the necessary information and skills to solve problems, mimicking real-world situations and constraints [48]. They should help learners develop practical skills, understand processes, promote

discipline-specific behaviours, and connect to existing knowledge [48].

Constructivism supports authentic assessment by positioning learners as active constructors of knowledge engaged in real-world tasks that replicate professional challenges, requiring deep understanding, and complex problem-solving [49,50]. Universal Design for Learning (UDL) also supports alternative assessments as it encourages optimising relevance, value, and authenticity of activities.

8.1. Examples

Learners can demonstrate their learning by analysing, applying, evaluating, or creating content in a manner that uses GenAI to incorporate its affordances or minimise opportunities for it unethical use. For example:

- *Annotated portfolio*: GenAI can summarise content, provide feedback, and proofread content in students' annotated portfolios.
- *Virtual patients*: GenAI can create virtual patients, including diagnostic results, scenarios, and simulated conversations that provide individualised formative feedback [51]. See also Dartmouth's Patient Actor App [52].
- *Infographic or Poster*: GenAI can generate an outline, design the layout, generate images, and provide feedback.

For more examples of formative alternative assessments in HPE that utilise GenAI, please refer to Appendix 4.

8.2. Best practices

Alternative assessments that incorporate GenAI should adhere to best practices, consider:

- Designing assessments that mirror real-world scenarios, foster practical skills and discipline-specific behaviours, and emphasise higher-order thinking such as analysis, evaluation, and creation—that GenAI alone cannot easily replicate.
- When designing assessments that incorporate GenAI's affordances, clearly articulate how and when GenAI can be used considering any guidelines set by your institution, program, or department. Consider how learners may incorporate GenAI in their professional contexts, for example, as a brainstorming tool, as an editor, as a tutor.
- Using rubrics and exemplars that reflect GenAI-specific criteria. For example, emphasise transparency by requiring documentation of AI usage. Promote critical evaluation, and verification of GenAI output. How to Use AI Responsibly *EVERY* time, provides a memorable mnemonic and framework when adopting GenAI, acknowledging our role as users of the technology (https://www.aiforeducation.io/ai-resources/how-to-use-ai-responsibly-every-time)

8.3. Limitations

While alternative assessments provide numerous benefits, they are time-intensive to design and implement, and may face resistance from learners due to their non-traditional nature. Additionally, using GenAI requires a critical approach to address biases, equity, and access—especially as paid versions can create disparities. Educators may also need to explain GenAI's value and limitations to institutions, ensuring its responsible use to uphold educational integrity.

9. AI as tutor

Producing fair, standardised and personalised formative assessment is resource-intensive. Through intelligent tutoring systems (ITS), AI provides a path to personalised learning and adaptability and the ability to cater assessments to learners' needs.

9.1. Advantages

AI using ITS has significant implications for formative assessments through personalised learning activities, adaptability, automation, and feasibility to promote self-regulated learning and performance prediction of the learner [53,54], providing frequent feedback to learners in real-time to develop specific skills, leading to enhanced learning outcomes while providing significant benefits for learning [55,56]. With their ability to analyse the learner's performance, ITS can recognise a learner's strengths and weaknesses and adapt to different preferences and learning paces, catering to struggling learners' needs, or challenging more advanced learners.

In addition, they reduce extraneous load by providing solutions to complex and unfamiliar concepts tailored to the learner's knowledge base [57]. Formative assessments using ITS can guide the learner to improve their learning outcome through data-based decision-making by allowing for change in mental representations, learning new skills, understanding the need for supervision versus being autonomous, and their ability to discover, understand, reflect, and review the instructional content [47].

To successfully use ITS, the HPE needs to have mastery of the Pedagogical Content Knowledge, as addressed earlier. AI can be leveraged to provide peer feedback in large classrooms as a solution to provide higher-level learning with adaptable platforms [47]. The assessments and analytics performed by AI enable feedback and contribute to continual learning and provide the ability to track the learners' progress and challenges over time.

Virtual patients (VP) can be programmed to simulate authentic patient-clinician interactions to guide the learner through interactive questioning to develop skills such as clinical and management reasoning, which require the learner to integrate knowledge with decision-making. This creates a tailored feedback loop that helps learners build self-awareness in their clinical reasoning skills, encouraging them to practice, reflect, and refine their approach with each new case.

9.2. Examples

Examples include the Virtual operative assistant [58] and the AI Patient Actor [59] (see Appendix 1 for more examples). These simulated patients provide learners with realistic environments to practice core skills of history-taking, clinical reasoning, and diagnostic decision-making training

in communication skills, and receive immediate feedback (by LLM analysis of the interaction transcript against a rubric) necessary for formative assessment.

If educators do not have access to advanced programming skills, or wish to have more control over the simulated environment, custom GPTs can be used successfully to create simple and effective ITS, with the ability to engage in interactive dialogue to promote real-time engagement, sustain focus through their user-friendly interface, and clarify complex concepts [51]. Their customisability allows the educator to personalise the responses, reduce bias and hallucinations, and offer an open-access solution through scalability and the ability to be translated into multiple languages.

9.3. Limitations

Like all simulations, AI tutors and VP do not currently fully replicate human-patient interactions, nuanced emotional responses, complex social cues, and the unpredictability of real-life patient behaviour, so should be used for building foundational skills, rather than substitute hands-on experience with patients.

In addition, learner-analytics raises concerns about learner safety and privacy, and these systems' emotional adaptability to learners' needs. Peer assessment using AI is still in its infancy, and will need advanced consensus approaches by assessors and weighted aggregates.

Although the AI tutor is crucial, the most important person is the learner. With that in mind, we turn next to the AI learner.

10. AI as learner

AI has been found to be particularly useful for learners as a more knowledgeable other and as a feedback source.

As originally described by Lev Vygotsky [40], there is considerable value in conversations with a more knowledgeable other. A conversation among learners about a common topic extends the knowledge and understanding of all contributors, assuming there is some common ground of understanding or agreement. AI is well-placed to frame and extend debate within a group in this way. Additionally, learners learn best when they are prompted to recognise a difference between their existing knowledge and new ideas (*dissonance*), and are empowered to *elaborate* their explanations for that difference in order to *organise* it into a story that makes sense [60]. AI, in the guise of either a tutor or fellow learner, can do this.

10.1. Examples

A chatbot can rapidly and easily produce a study guide covering the main concepts in material that is being studied. Relatedly, documents can be pasted into a chatbot to be analysed and summarised. The data can then be used by the chatbot to provide the information for a concept map. This could be as plain text (as provided by ChatGPT), or as a concept map, either from nodes provided by a chatbot and pasted into mind-mapping software, or from a dedicated AI concept map generator (for example Coggle or Text2MindMap). The concept map is then used for formative assessment or revision purposes.

AI can enhance collaboration between learners, either on streaming platforms like Zoom, or on whiteboard platforms like Stormboard, Conceptboard, or Coggle. This allows the collaborative engagement of AI, allowing learners to learn from the tool and each other. Although the software is designed to be used by individuals working remotely, in practice, learners sit around a table with their laptops, editing the whiteboard screen which is displayed to the whole room.

A powerful aspect of AI is its ability to ask and respond to questions about a piece of text. It takes some time to develop prompt-writing skills [61], but it is an excellent method of giving learners feedback on their knowledge, or even their communication skills. Using ChatGPT, the bot can be asked to give a history and findings for a patient with (for example) hypertension, and then it can be asked to 'be' that patient in a consultation with the learner. It can conclude a session by providing critical and constructive feedback on communication skills.

Another possibility is to use a variation of the AI LLM debate [62] using AI systems to debate a clinical case, and have learners analyse the debate, commenting on the strengths and weaknesses of each position; this can be used for learning and assessment.

10.2. Best practices

Most publications in this area focus on using AI as a supportive tool in preparing for formal examinations. A good example is Kung et al. [63], who showed that their ChatGPT model could ask and answer questions set at the level of USMLE Steps 1, 2 and 3, and crucially, give reasonable explanations for the answers it chose. The key to all usage of AI, though, is in training learners (and faculty) in its sensible use, and being appropriately critical of the information it provides [64].

10.3. Limitations

The main limitation is the need for a parallel world in which the learner acts with human teachers, learners and patients, not least because at the front line, healthcare is practised on and with, people.

There is the secondary issue of developing bad habits in consultations, or of failing to develop incisive thinking: the chatbot may point out errors and misconceptions, but is easier to ignore than a professor or physician at the patient's bedside.

11. Competency-based assessment

Competency-based education (CBE), sometimes called outcomes-based education (OBE), is derived from the belief that the assessment of learners should determine whether or not they have the knowledge, skills and attitudes to meet graduates' expectations in their field [25, 65]. It requires assessments to be 'authentic' to learners' expectations in their future working context [66]. It is, however, exceptionally challenging, and frequently leads to imperfect implementation. The implementation and assessment

of competency-based HPE is particularly difficult, due to the complex, unpredictable, and the clinical environment' demanding nature [67].

To overcome some of these issues, a Core Components framework has been developed [13]. This framework requires a consensus of expected outcome competencies which should be delivered sequentially and progressively through tailored learning experiences. The instruction should be focussed on the acquisition of specific competencies and the assessment should be programmatic: small, frequent, relatively low-stakes [6,7]. The instruction and assessment of competencies has attracted particular interest [68,69].

11.1. Examples

Three of the core components seem to be particularly amenable to AI usage.

11.1.1. Tailored learning experiences:

The value of formative assessment lies in tailored responses, and, similar to the way in which social media suggest videos and posts based on user responses, so AI can suggest learning materials, tutorials or other interventions based upon demonstrated learner competency and level [70,71].

11.1.2. Competency-focused instruction

Similarly, formative assessment can lead to adjustment in the formal instruction, summarising and organising feedback conversations into assessments [72,73]. It also provides faculty with insight into trainees' learning needs and how to improve feedback [74]. Additionally, it is increasingly being leveraged as an *alternative* to human observation – multiple studies have been described utilising simulators to monitor performance and provide real-time learner feedback based on auditory, visual, or kinaesthetic inputs [4].

11.1.3. Programmatic assessment:

As Programmatic Assessment requires the integration of many different assessment types [75,76], AI can help in assessing the development of clinical reasoning, analysing data, saving time in review, and identifying learners ready for promotion or at-risk.

11.2. Best practices

There is limited AI usage within HPE assessment to date [5], and effective tool development will require the secondary use of educational data to facilitate the training of effective models. Best practices in this area include ensuring the participation of learners in their development, being clear regarding the intended actions of AI tools, evaluating their quality, and avoiding focus on factors that are unactionable or nonmodifiable by learners [77].

11.3. Limitations

The first limitation is the need for the simple expedient of a human individual to ensure that the AI-generated conclusions for high-stakes events are correct and appropriate.

The second limitation focuses on accessibility and ethical use of sensitive data once it appears online [78]. The use of LLMs, and gifting them with sensitive data potentially multiplies the risk [77].

With these in mind, we move into some of the difficult areas raised by using AI in HPE assessment.

12. Difficult areas and ethical issues

12.1. Introduction

AI will create challenges for assessors. This is particularly true as we begin to navigate the use of AI by learners, whose reaction to AI in education was widespread and almost immediate [17,18].

University-specific policies on GenAI do exist [79–81], but may not always be specific to HPE or current with the latest advances in the field. As a result, it is crucial for HPE educators to be aware of, and address, these issues as they arise.

The list of ethical concerns regarding AI in HPE assessment continues to grow as AI is applied in new ways. Some issues have already been raised, and, in this section, we outline some commonly identified challenges and suggest some best practices.

12.2. The source information

A controversial issue of GenAI is sourcing original training data, usually by large-scale harvesting of publicly accessible material, and usually without acknowledgment or compensation. Since this material might be useful for all assessments, there may be pressure on institutions to compensate the sources. With many models (e.g. LLMs), however, tracing the source is not possible, leaving institutions reliant on systems that may be deemed to be inherently unethical. (Traditional AI (e.g. Clinical Decision Support Systems) usually indicate their information sources, so institutions can follow through on ethical questions with the respective companies.)

12.3. Bias

AI algorithms are derived from existing datasets that may be discriminatory or biased [16]. The causes of bias are many, including a legitimate decision by population groups to not share their information for training data. Consequently, healthcare professionals trained and assessed on these systems may lack awareness and competence in subtleties related to that population group. If the AI is used to generate examination questions, it is possible that biases can make their way into these examination questions [82], resulting in examination questions that are either unrepresentative and/or play into inaccurate stereotypes of structurally marginalised communities.

Clarity regarding when and how learner data can be used to develop or expand AI models will be important to maintain trust and ensure the ethical use of data [77].

12.4. Faculty use of AI

When using an AI system for formative or summative grading or feedback, several issues arise:

- Usage transparency: Just as institutions require learner transparency of GenAI usage, faculty usage of GenAI requires transparency. Institutions may also wish to have an 'opt-out' possibility for learners and faculty.
- Learner Privacy: Learner data is an attractive source for HPE institutions [77]. Natural Language Processing (NLP) models have been developed to organise narrative assessment data by sub-competencies [83,84] and to predict performance [85]. The validity of the results is not yet conclusive, but an equally pressing concern is the unconsented use of learner data used as training data for the models. Depending upon the model, this problem can be mitigated by using institutional licenses, Application Program Interfaces (APIs), or disabling such training through settings. If this is not possible, then fully informed consent is required.
- Response validation: GenAI responses should be verified for accuracy, appropriateness, and adherence to established guidelines and best practices, even more crucial when dealing with clinical work.
- Significant work should be done exploring the validity of the use of AI tools for higher-stakes educational and clinical assessments.
- Appeals: the appeals process must be adapted to address issues specifically related to inappropriate AI usage.

12.5. GenAI detection

As described earlier in this Guide, learners have been quick to use AI in their work, and AI can draft compelling, high-quality reflections with minimal work by the learners, undermining the use of active teaching methodologies [86]. AI detectors exist, but should be avoided, for several reasons:

- They 'are neither accurate nor reliable' [87,88], particularly when learners use obfuscation techniques like careful prompting, or 'humanising' tools like Twixify, and they have a bias towards classifying outputs as human-written [88].
- They tend to disadvantage people writing in their second language by flagging their writing as AI-generated with greater frequency [89].
- AI prohibition would discourage the use of an important new tool that could have major benefits [90].

12.6. Learner use

If institutions require, or even encourage, learners to use AI tools in assessments, some issues need to be addressed:

- Policies regarding learners' appropriate AI usage within their educational programs are required.
- Educators should create a shared understanding by defining their capability, and modelling accountability with AI use.
- Guarding against 'gaming' the system: while trying to guess the graders' priorities and idiosyncrasies is not new, using publicly available systems may encourage learners to study AI idiosyncrasies more closely and aim their assessment responses at those, rather than providing their own solutions.
- Access equity: equitable access must be provided; this includes training, licenses, and hardware [16, 91].
- Dependency and over-reliance: while the future will require healthcare professionals who use the tools with ease, there is a danger of their developing an over-reliance, so assessments should balance allowing for circumstances in which AI systems may not be available.

12.7. Limitations

Given the speed at which the field of AI is moving, it will be difficult for institutions to keep pace with policies on appropriate and effective AI usage in assessment. As AI tools become ubiquitous and training data becomes increasingly important, it will be necessary to adhere to best practices related to the sharing of educational assessment data. Educators and learners need to adopt a pragmatic view of AI, where they embrace uncertainty and imperfection with openness and honesty in the world of distributed knowledge of GenAI [92].

12.8. The future

While educators appear to have softened from their November 2022 initial negative reaction to a more embracing and educating approach to using AI in assessment, the sense is that the big change is over, and now we can begin adapting to the new and stable scenario.

While our focus is on AI in HPE assessment, there is a broader ethical concern: addressing the rapid AI changes to come. We are beginning a steep curve of AI development that will affect all future HPE assessments. We can argue about the form of the future, whether or not it will be AGI [22], but this much is clear: the rate of change will increase beyond our current response, and there is a strong ethical imperative to live with the change, direct the change, and even control the change, so that we ensure HPE assessment remains relevant and effective.

On that basis, we turn to the next topic of Faculty Development.

13. Faculty and staff development

As we prepare for the anticipated changes discussed in the previous topic, Faculty Development in AI-supported assessment is essential for change management and technology acceptance by our faculty, staff, and learners, and this will require dedicated approaches from institutions. Unfortunately, a lack of faculty development across all technology-enhanced learning is common [12]. This may be due to decreased learning and teaching opportunities, mentorship, and/or institutional support. This leads to increased teaching resistance, educators avoiding technology usage in their teaching, and missed opportunities to use, learn, model, and promote scholarship in the area.

Our need for faculty development is especially true of the rapidly evolving field of AI-enabled assessment. As noted by Sam Altman: 'These are the stupidest the models will ever be' [93], so, our challenge is to provide pedagogically informed training and support for this rapidly evolving and radically different way of providing assessment.

We can begin by building upon the Technology, Pedagogy, Content, Knowledge, Framework TPACK framework [8,9], as discussed earlier, in designing and ensuring faculty development competency development as follows:

- TK: knowing how to use LLMs such as ChatGPT for assessment, including how to access them, their benefits and imitations, and prompt engineering techniques.
- TCK: knowing how to use them in a specific clinical context (e.g. neurology).
- TPK: using LLMs to enhance our pedagogical approach to assessment (e.g. enabling multimodal and authentic and precision assessment).
- TPCK: putting this all together (e.g. using LLMs in a neurology context to enhance our pedagogically informed assessment strategies).

By explicitly modelling these frameworks in providing faculty development around GenAI assessment, we can leverage the excitement and decrease the fear of technology acceptance, while implicitly teaching about pedagogical design.

Many faculty may feel confused about how and when to use GenAI in assessment while maintaining security, privacy, and ethical use, and the earlier discussion and other literature may help [16]. Some examples of institutional policies can be found at sites maintained by Lance Eaton and Tracy Mendolia [79, 81].

Whatever the policies require, institutions should ensure access to GenAI tools for assessment, both for faculty and learners. Policies without access will be meaningless. Secondly, being explicit about GenAI use is very important. By being explicit about faculty use and encouraging open discussion, especially in high-stakes assessments, we can model ethical and practical usage to our learners, noting expectations for both learner and faculty as we learn together.

When institutions have established these, they will need to consider how best to provide training on AI for faculty, staff, and learners. While single-session approaches may provide a foundational understanding of how GenAI may work in advancing assessment strategies, it will not suffice as GenAI rapidly changes. Institutions should also look to microlearning updates, communities of practice (CoPs) [94], and modelling use of GenAI assessments within other faculty development offerings. Sharing successes and challenges is equally important to demystifying GenAI usage. Some institutions may be able to offer, or sponsor their faculty to complete, longer courses and certification.

Overall, institutions should understand that their faculty will have different affinities and risk-acceptance to using non-traditional assessment methods, and that, unless there are conversations around everyone's assumptions about AI use [95], it will be difficult to advocate for institutional support of newer methods and tools. Institutions must consider different levels of education and advocacy at the micro (faculty and staff), meso (institution), and macro (regulatory and professional body) levels when providing faculty development.

When opening conversations about AI use in assessment, it is important that we acknowledge and cite our use of GenAI tools and output, as discussed in the next topic.

14. Acknowledgment and citing

14.1. Background

As seen throughout the Guide, AI is used in faculty construction and learner completion of assessments, and, as learners increasingly use AI in their assessments, they will need to acknowledge and cite these systems, but exactly how this is to be done is not completely resolved. As we saw earlier in the Guide, the future status of the essay is in flux, but there will still be similar requirement for work submission, so these issues still must be addressed.

Currently, there are two opposing perspectives on citing AI usage:

1. AI-created content is not original; it is taken from elsewhere and thus should be credited and cited, just as one would cite any author, to avoid plagiarism [90].
2. AI is not human; it cannot write or take responsibility for its writing. Therefore, AI cannot be an author, so we cannot cite its output. Proponents of this point of view argue that there is no record of the interaction with the machine unless you save your chat, and because the output cannot be perfectly re-created [96]. Another problem with citing AI's output is that reference lists usually give the readers a list of publications they can access, but GenAI outputs are not currently easily publicly available.

Within this background, many (but not all) institutions have policies on usage [79, 81]. Because the focus is currently on written assignments, and if one's institution does not have a policy, or it is not comprehensive enough, then a useful starting point is to refer to current practice and policies from academic publishers, as these can help when formulating guidelines for formative and summative assessment.

14.2. Acknowledgment

Most publishers have guidelines for acknowledging AI usage academic work. For example, Elsevier states that '… authors should disclose in their manuscript the use of AI and AI-assisted technologies …' [97], arguing that this maintains trust and transparency between authors and readers. This is useful, and appears aimed at LLMs, as there is no need to disclose grammar checkers or reference annotators.

Although more specific examples are given below, in general, acknowledgment of AI usage should be stated as one would a disclosure: in a separate paragraph, stating as a minimum, that AI was used in the preparation of the work, the date, to what extent, for what task, and whether authors reviewed the final output. In all of these, it is

essential to use the correct words in the acknowledgment: created, enhanced, edited, reviewed, assisted, etc.

14.3. Citations

A common style used as a guide is the APA, in which non-retrievable communications are cited as personal communications, but, as an AI tool is not a person, it has a slight modification:

> *Author*: creator of LLM. (*year of the version you used*). *title*: name of the LLM
>
> (*date used* version) [large language model] *URL of LLM*
>
> Example (from McAdoo 2024 [98]):
>
> OpenAI. (2024). *ChatGPT* (July 6 version) [Large language model]. https://chat.openai.com/chat
>
> For other examples of other styles, please see Appendix 5.

14.4. Limitations

As has been observed, using GenAI tools in academic work brings significant considerations regarding acknowledgment and citations. These matters are still evolving, with policies and best practices adapting as the technology advances. Institutions and publishers alike are grappling with the complexities of integrating AI into scholarly work, and it is crucial to stay informed about the latest standards to ensure correct citation and acknowledgment in your work.

15. Conclusions

This AMEE Guide on the use of AI in HPE assessment has covered a broad range of issues, beginning with pedagogical theory, human responses, through a range of assessment types and situations, and ending with the acknowledgment of AI. Although the authors' personal biases towards AI have influenced their work and approaches, we have attempted to maintain a balance, so that educators can approach the challenges, problems and solutions with full and realistic awareness of the possible positive and negative impact of AI. In this way, we have aimed to provide HPE educators with a Guide to assist them in their journey of successfully using AI for HPE assessment.

Acknowledgements

The authors gratefully acknowledge comments made by reviewers on a previous version of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Notes on contributors

Dr. Ken Masters, PhD HDE FDE, is Associate Professor of Medical Informatics, Medical Education and Informatics Department, Sultan Qaboos University, Sultanate of Oman. He teaches Artifical Intelligence and medical informatics ethics and has published several papers and AMEE Guides related to these topics. He is a member of AMEE's TEL Committee. ORCID: 0000-0003-3425-5020.

Heather MacNeil, MD, BSc(PT), MScCH (HPTE), FRCPC, is Associate Professor and Co-instructor Educational Technology for Health Practitioner Education, Faculty of Medicine, University of Toronto, Canada, and Director of Stroke Rehabilitation, Division of PM&R, Sinai Health System, Toronto, Canada, and Interim Assistant Dean, Clinical Faculty Development, Toronto Metropolitan University School of Medicine, Canada. ORCID: 0000-0001-9842-3578.

Jennifer Benjamin, MD, MS, is Associate Professor Complex Care, Huffington Department of Education Innovation and Technology, Director of Technology Center for Research Innovation and Scholarship in Health Professions Education, Co-Director Faculty College, Texas Children's Hospital, Baylor College of Medicine. ORCID: 0000-0001-6085-5973.

Tamara Carver, PhD, is Associate Professor at McGill University's Institute of Health Sciences Education and the Director of the Office of Ed-TECH (Education Technology and E-Learning Collaboration for Health), leading innovative digital learning initiatives and research in health professions education. ORCID: 0000-0003-3410-0897.

Kataryna Nemethy, Msc BMC, is Manager eLearning and Ed Tech at Baycrest Academy for Research and Education, Baycrest Academy for Research and Education, and Co-instructor Educational Technology for Health Practitioner Education, Faculty of Medicine, University of Toronto, Canada. ORCID: 0000-0002-8502-1593.

Sofia Valanci-Aroesty, MD, PhD, is Program advisor for learning strategy at the Royal College of Physicians and Surgeons of Canada, Canada. ORCID: 0000-0002-0593-5353.

David CM Taylor, BSc MEd MA PhD EdD, is Director of the Center for Innovation and Leadership in Health Professions Education, and Professor of Medical Education and Physiology, Gulf Medical University, UAE. ORCID: 0000-0002-3296-2963.

Brent Thoma, MD, PhD, is Clinical Professor in the School of Medicine at Toronto Metropolitan University; Professor of Emergency Medicine at the University of Saskatchewan. ORCID: 0000-0003-1124-5786.

Thomas Thesen, PhD, is Associate Professor of Medical Education & Computer Science, Director of Medical Learning Sciences, Director of Geisel Academy of Educators & Scholars, Department of Medical Education & Computer Science, Geisel School of Medicine at Dartmouth, USA. ORCID: 0000-0002-3793-438X.

ORCID

Ken Masters http://orcid.org/0000-0003-3425-5020
Heather MacNeil http://orcid.org/0000-0001-9842-3578
Jennifer Benjamin http://orcid.org/0000-0001-6085-5973
David C. M. Taylor http://orcid.org/0000-0002-3296-2963
Brent Thoma http://orcid.org/0000-0003-1124-5786

References

1. Hedberg J, Corrent-Agostinho S. Creating a postgraduate virtual community: assessment drives learning. Educ Media Int. 2000; 37(2):83–90. doi:10.1080/095239800410360
2. McCarthy J. A proposal for the Dartmouth summer research project on artificial intelligence. Hanover (NH); 1955. Available from: http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.
3. Turing A. Computing machinery and intelligence. Mind. 1950; LIX(236):433–460. doi:10.1093/mind/LIX.236.433
4. Gordon M, Daniel M, Ajiboye A, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach. 2024;46(4):446–470. doi:10.1080/0142159X.2024.2314198

5. Ang C-S, Ito S, Cleland J. Navigating digital assessments in medical education: findings from a scoping review. Med Teach. 2024;:1–16. doi:10.1080/0142159X.2024.2425033
6. Torre D, Schuwirth L. Programmatic assessment for learning: A programmatically designed assessment for the purpose of learning: AMEE Guide No. 174. Med Teach. 2024; Oct 5;:1–16. doi:10.1080/0142159X.2024.2409936
7. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478–485. doi:10.3109/0142159X.2011.565828
8. Mishra P, Koehler MJ. Technological pedagogical content knowledge: a framework for teacher knowledge. Teachers Coll Record. 2006;108(6):1017–1054. doi:10.1111/j.1467-9620.2006.00684.x
9. Youm J, Corral J. Technological pedagogical content knowledge among medical educators: what is our readiness to teach with technology? Acad Med. 2019;94(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 58th Annual Research in Medical Education Sessions):S69–S72. doi:10.1097/ACM.0000000000002912
10. Kimmons R, Graham CR, West RE. The PICRAT model for technology integration in teacher preparation. Contemp Issues Technol Teach Educ. 2020;20(1):176–198.
11. Masters K, Correia R, Nemethy K, et al. Online learning in health professions education. Part 2: tools and practical application: AMEE Guide No. 163. Med Teach. 2024;46(1):18–33. doi:10.1080/0142159X.2023.2259069
12. MacNeill H, Masters K, Nemethy K, et al. Online learning in health professions education. Part 1: teaching and learning in online environments: AMEE Guide No. 161. Med Teach. 2024;46(1):4–17. doi:10.1080/0142159X.2023.2197135
13. Van Melle E, Frank JR, Holmboe ES, et al. A core components framework for evaluating implementation of competency-based medical education programs. Acad Med. 2019;94(7):1002–1009. Jul. doi:10.1097/ACM.0000000000002743
14. Bloom BS, Engelhart MD, Furst EJ, et al. Taxonomy of educational objectives: the classification of educational goals. Handbook I: cognitive domain. London: Longman Green & Co; 1956.
15. Page E, Meyers G, Billings E. Theory to practice: an assessment framework for generative AI. Intersection. 2024;5(4):114–126. Oct 2 [cited 2024 Oct 28]; Available from: https://aalhe.scholasticahq.com/article/124250-theory-to-practice-an-assessment-framework-for-generative-ai
16. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No.158. Med Teach. 2023;45(6):574–584. doi:10.1080/0142159X.2023.2186203
17. Lo CK, Hew KF, Jong MS. The influence of ChatGPT on student engagement: a systematic review and future research agenda. Comput Educ. 2024;219:105100. doi:10.1016/j.compedu.2024.105100
18. Fütterer T, Fischer C, Alekseeva A, et al. ChatGPT in education: global reactions to AI innovations. Sci Rep. 2023;13(1):15310. doi:10.1038/s41598-023-42227-6
19. Venkatesh VM, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q. 2003;27(3):425. doi:10.2307/30036540
20. Karaca O, Çalışkan SA, Demir K. Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – development, validity and reliability study. BMC Med Educ. 2021;21(1):112. doi:10.1186/s12909-021-02546-6
21. Mehta N, Agrawal A, Benjamin J, et al. Pedagogy and generative artificial intelligence: applying the PICRAT model to Google NotebookLM. Med Teach. 2024;:1–3. doi:10.1080/0142159X.2024.2418937
22. Masters K, Herrmann-Werner A, Festl-Wietek T, et al. Preparing for Artificial General Intelligence (AGI) in Health Professions Education: AMEE Guide No. 172. Med Teach. 2024;46(10):1258–1271. doi:10.1080/0142159X.2024.2387802
23. Zeff M. people are going to fall in love with the flirty GPT-4 omni. Gizmodo. 2024; [cited 2024 Nov 11]. Available from: https://gizmodo.com/people-are-going-to-fall-in-love-with-the-flirty-gpt-4-1851473938
24. Sallam M, Al-Mahzoum K, Almutairi Y, et al. Anxiety among medical students regarding generative artificial intelligence models: a pilot descriptive study. 2024; [cited 2024 Aug 19]. doi:10.20944/preprints202408.1215.v1
25. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638–645. doi:10.3109/0142159X.2010.501190
26. Hodges BD. A tea-steeping or i-Doc model for medical education? Acad Med. 2010;85(9 Suppl):S34–S44. doi:10.1097/ACM.0b013e3181f12f32
27. Triola MM, Burk-Rafel J. Precision medical education. Acad Med. 2023;98(7):775–781. doi:10.1097/ACM.0000000000005227
28. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. jMIR Med Educ. 2023;9: E45312. doi:10.2196/45312
29. Kıyak YS, Kononowicz AA. Case-based MCQ generator: a custom ChatGPT based on published prompts in the literature for automatic item generation. Med Teach. 2024;46(8):1018–1020. doi:10.1080/0142159X.2024.2314723
30. Kıyak YS, Coşkun Ö, Budakoğlu Iİ, et al. ChatGPT for generating multiple-choice questions: evidence on the use of artificial intelligence in automatic item generation for a rational pharmacotherapy exam. Eur J Clin Pharmacol. 2024;80(5):729–735. doi:10.1007/s00228-024-03649-x
31. OpenAI. Introducing OpenAI o1 [Internet]. 2024 [cited 2024 Sep 15]. Available from: https://openai.com/o1/.
32. Rüdian S. Exploratory and confirmatory prompt engineering. In zenodo; 2024 [cited 2024 Sep 16]. doi:10.5281/ZENODO.12549309
33. Shakur AH, Holcomb MJ, Hein D, et al. Large language models for medical OSCE assessment: a novel approach to transcript analysis [Internet]. arXiv; 2024 [cited 2024 Nov 19]. Available from: http://arxiv.org/abs/2410.12858.
34. Xiao C, Ma W, Song Q, et al. Human-AI collaborative essay scoring: a dual-process framework with LLMs [Internet]. arXiv; 2024 [cited 2024 Nov 19]. Available from: http://arxiv.org/abs/2401.06431.
35. Holderried F, Stegemann-Philipps C, Herrmann-Werner A, et al. A language model–powered simulated patient with automated feedback for history taking: prospective study. JMIR Med Educ. 2024;10:E59213. doi:10.2196/59213
36. Balasooriya C. AI in education: a futuristic vision. Med Teach. 2024;46(7):986–988. doi:10.1080/0142159X.2024.2352160
37. Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory Practice. 2002;41(2):64–70. doi:10.1207/s15430421tip4102_2
38. Lingard L. Writing an effective literature review: part I: mapping the gap. Perspect Med Educ. 2018;7(1):47–49. doi:10.1007/S40037-017-0401-X
39. Coverdale JH, Roberts LW, Balon R, et al. Writing for academia: Getting your research into print: AMEE Guide No. 74. Med Teach. 2013;35(2):E926–e934. doi:10.3109/0142159X.2012.742494
40. Vygotsky LS. Mind in society: the development of higher psychological processes. Cole M ,John-Steiner V ,Scribner S ,Sourberman E, editors. Cambridge (MA): Harvard UP; 1978.
41. Von Glaserfeld E. Cognition, construction of knowledge, and teaching. Washington: National Science Foundation; 1988.
42. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv; 2023 [cited 2023 Mar 30]; doi:10.48550/ARXIV.2303.12712
43. Yang M. New York City schools ban AI chatbot that writes essays and answers prompts. New York: The Guardian [Internet].; 2023 Jan 6 [cited 2024 Jul 8]; Available from: https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt
44. Huang K. Alarmed by A.I. Chatbots, universities start revamping how they teach. The New York Times [Internet]. New York; 2023 Jan 16 [cited 2024 Jul 8]; Available from: https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html.
45. Livingstone VI. Quit teaching because of ChatGPT [Internet]. TIME. 2024 [cited 2024 Oct 5]. Available from:https://time.com/7026050/chatgpt-quit-teaching-ai-essay/.

46. Verma P. A professor accused his class of using ChatGPT, putting diplomas in jeopardy. Washington Post [Internet]. 2023 May 19 [cited 2023 Aug 13]; Available from: https://www.washingtonpost.com/technology/2023/05/18/texas-professor-threatened-fail-class-chatgpt-cheating/.
47. Swiecki Z, Khosravi H, Chen G, et al. Assessment in the age of artificial intelligence. Comput Educ Artif Intel. 2022;3:100075. doi:10.1016/j.caeai.2022.100075
48. TMU. Designing assessments [Internet]. Toronto Metropolitan University (TMU); 2021. [cited 2024 Aug 4]. Toronto, Canada. Available from: https://www.torontomu.ca/learning-teaching/teaching-resources/assessment/
49. Koh KH. Authentic assessment. In: Oxford research encyclopedia of education. 2017. [cited 2024 Aug 4]. doi:10.1093/acrefore/9780190264093.013.22
50. Janisch C, Liu X, Akrofi A. Implementing alternative assessment: opportunities and obstacles. Educ Forum. 2007;71(3):221–230. doi:10.1080/00131720709335007
51. Masters K, Benjamin J, Agrawal A, et al. Twelve tips on creating and using custom GPTs to enhance health professions education. Med Teach. 2024;46(6):752–756. doi:10.1080/0142159X.2024.2305365
52. Dartmouth Libraries. AI @ Dartmouth [Internet]. 2024 [cited 2024 Nov 18]. Available from: https://ai.dartmouth.edu/patient-actor.
53. Lin C-C, Huang AYQ, Lu OHT. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart Learn Environ. 2023;10(1):1–22. doi:10.1186/s40561-023-00260-y
54. González-Calatayud V, Prendes-Espinosa P, Roig-Vila R. Artificial intelligence for student assessment: a systematic review. Appl Sci. 2021;11(12):5467. doi:10.3390/app11125467
55. Schildkamp K, van der Kleij FM, Heitink MC, et al. Formative assessment: a systematic review of critical teacher prerequisites for classroom practice. Int J Educ Res. 2020;103:101602. doi:10.1016/j.ijer.2020.101602
56. Evans DJR, Zeun P, Stanier RA. Motivating student learning using a formative assessment journey. J Anat. 2014;224(3):296–303. doi:10.1111/joa.12117
57. Shalong W, Yi Z, Bin Z, et al. Enhancing self-directed learning with custom GPT AI facilitation among medical students: a randomized controlled trial. Med Teach. 2024:1–8. doi:10.1080/0142159X.2024.2413023
58. Fazlollahi AM, Bakhaidar M, Alsayegh A, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. JAMA Netw Open. 2022;5(2):e2149008. doi:10.1001/jamanetworkopen.2021.49008
59. Thesen T, Alilonu N, Stone S. AI patient actor: an open-access generativeai app for communication training in health professions. Medical Sciences Educator. 2024.
60. Taylor DCM, Hamdy H. Adult learning theories: implications for learning and teaching in medical education: AMEE Guide No. 83. Med Teach. 2013;35(11):e1561–e1572. doi:10.3109/0142159X.2013.828153
61. Heston TF, Khun C. Prompt engineering in medical education. IME. 2023;2(3):198–205. doi:10.3390/ime2030019
62. Khan A, Hughes J, Valentine D, et al. Debating with more persuasive LLMs leads to more truthful answers [Internet]. arXiv; 2024 [cited 2024 Nov 12]. Available from: http://arxiv.org/abs/2402.06782.
63. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. Dagan A, editor. PLOS Digit Health. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198
64. Alkhaaldi SMI, Kassab CH, Dimassi Z, et al. Medical student experiences and perceptions of ChatGPT and artificial intelligence: cross-sectional study. JMIR Med Educ. 2023;9:e51302. doi:10.2196/51302
65. Harden RM. AMEE Guide No. 14: outcome-based education: part 1-an introduction to outcome-based education. Med Teach. 1999;21(1):7–14. doi:10.1080/01421599979969
66. Villarroel V, Bloxham S, Bruna D, et al. Authentic assessment: creating a blueprint for course design. Assess Eval Higher Educ. 2018;43(5):840–854. doi:10.1080/02602938.2017.1412396
67. Van Melle E, Hall AK, Schumacher DJ, et al. Capturing outcomes of competency-based medical education: the call and the challenge. Med Teach. 2021;43(7):794–800. doi:10.1080/0142159X.2021.1925640
68. Kinnear B, Santen SA, Kelleher M, et al. How does TIMELESS training impact resident motivation for learning, assessment, and feedback? Evaluating a competency-based time-variable training pilot. Acad Med. 2023;98(7):828–835. doi:10.1097/ACM.0000000000005147
69. ten Cate O, Gruppen LD, Kogan JR, et al. Time-variable training in medicine: theoretical considerations. Academic Medicine. 2018;93(3S):S6–S11. doi:10.1097/ACM.0000000000002065
70. Grandinetti J. Examining embedded apparatuses of AI in Facebook and TikTok. AI Soc. 2021;38(4):1–14. doi:10.1007/s00146-021-01270-5
71. Alawneh YJ, Al-Momani T, Salman FN, et al. A Detailed Study Analysis of Artificial Intelligence Implementation in Social Media Applications. In: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) [Internet]. [cited 2024 Oct 14]. p. 1191–1194. doi:10.1109/ICACITE57410.2023.10182840
72. Nazaretsky T, Mejia-Domenzain P, Swamy V, et al. AI or human? Evaluating student feedback perceptions in higher education. In: Ferreira Mello R, Rummel N, Jivet I, Pishtari G, Ruipérez Valiente JA, editors. Technology enhanced learning for inclusive and equitable quality education. Cham: Springer Nature Switzerland; 2024. 284–298. doi:10.1007/978-3-031-72315-5_20
73. Hooda M, Rana C, Dahiya O, et al. Artificial intelligence for assessment and feedback to enhance student success in higher education. Math Prob Eng. 2022;2022(1):1–19. doi:10.1155/2022/5215722
74. Spadafore M, Yilmaz Y, Rally V, et al. Using natural language processing to evaluate the quality of supervisor narrative comments in competency-based medical education. Acad Med. 2024;99(5):534–540. doi:10.1097/ACM.0000000000005634
75. Torre D, Daniel M, Ratcliffe T, et al. Programmatic assessment of clinical reasoning: new opportunities to meet an ongoing challenge. Teach Learn Med. 2024;0(0):1–9. doi:10.1080/10401334.2024.2333921
76. Baartman LKJ, Quinlan KM. Assessment and feedback in higher education reimagined: Using programmatic assessment to transform higher education. Perspect Policy Pract Higher Educ. SRHE Website. 2024;28(2):57–67. doi:10.1080/13603108.2023.2283118
77. Thoma B, Spadafore M, Sebok-Syer SS, et al. Considering the secondary use of clinical and educational data to facilitate the development of artificial intelligence models. Acad Med. 2024;99(4S Suppl 1):S77–S83. doi:10.1097/ACM.0000000000005605
78. Masters K, Taylor D, Loda T, et al. AMEE Guide to ethical teaching in online medical education: AMEE Guide No. 146. Med Teach. 2022;44(11):1194–1208. doi:10.1080/0142159X.2022.2057286
79. Mendolia T. University policies on generative AI [Internet]. Padlet. 2024 [cited 2024 Nov 12]. Available from: https://padlet.com/cetl6/university-policies-on-generative-ai-m9n7wf05r7rdc6pe.
80. Caulfield J. University policies on AI writing tools | overview & list [Internet]. Scribbr. 2023 [cited 2024 Mar 14]. Available fromhttps://www.scribbr.com/ai-tools/chatgpt-university-policies/.
81. Eaton L. Syllabi polices for generative AI [Internet]. Google Docs. 2024; [cited 2024 Nov 20]. Available from: https://docs.google.com/spreadsheets/d/1lM6g4yveQMyWeUbEwBM6FZVxEWCLfvWDh1aWUErWWbQ/edit?usp=sharing&usp=embed_facebook&urp=gmail_link&usp=embed_facebook
82. Artsi Y, Sorin V, Konen E, et al. Large language models for generating medical examinations: systematic review. BMC Med Educ. 2024;24(1):354. doi:10.1186/s12909-024-05239-y
83. Booth GJ, Hauert T, Mynes M, et al. Fine-tuning large language models to enhance programmatic assessment in graduate medical education. J Educ Perioper Med. 2024;26(3):E729. doi:10.46374/VolXXVI_Issue3_Moore
84. Booth GJ, Ross B, Cronin WA, et al. Competency-based assessments: leveraging artificial intelligence to predict subcompetency content. Acad Med. 2023;98(4):497–504. doi:10.1097/ACM.0000000000005115
85. Maimone C, Dolan BM, Green MM, et al. Utilizing natural language processing of narrative feedback to develop a predictive

model of pre-clerkship performance: lessons learned | perspectives on medical education. Perspect Med Educ. 2023;12(1):141–148. doi:10.5334/pme.40
86. Li Y, Sha L, Yan L, et al. Can large language models write reflectively. Comput Education: Artif Intel. 2023;4:100140. doi:10.1016/j.caeai.2023.100140
87. Chaka C. Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: a literature and integrative hybrid review. J Appl Learn Teach. 2024;7(1):115–126. doi:10.37074/jalt.2024.7.1.14
88. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, et al. Testing of detection tools for AI-generated text. Int J Educ Integr. 2023;19(1):26. doi:10.1007/s40979-023-00146-z
89. Liang W, Yuksekgonul M, Mao Y, et al. GPT detectors are biased against non-native English writers. Patterns (N Y). 2023;4(7):100779. doi:10.1016/j.patter.2023.100779
90. Jarrah AM, Wardat Y, Fidalgo P. Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say? Online J Commun Media Technol. 2023;13(4):e202346. doi:10.30935/ojcmt/13572
91. Cambridge D, Wenger-Trayner E, Hammer P, et al. Theoretical and practical principles for generative AI in communities of practice and social learning. In: Buch A, Lindberg Y, Cerratto Pargman T, editors. Framing futures in postdigital education: critical concepts for data-driven practices [Internet]. Cham: Springer Nature Switzerland; 2024. p. 229–239. [cited 2024 Oct 14]. doi:10.1007/978-3-031-58622-4_13
92. Fawns T. An entangled pedagogy: looking beyond the pedagogy—technology dichotomy. Postdigit Sci Educ. 2022;4(3):711–728. doi:10.1007/s42438-022-00302-7
93. Unconfuse me. Episode 6: OpenAI CEO Sam Altman on the future of AI [Internet]. 2024 [cited 2024 Nov 11]. Available from: https://www.youtube.com/watch?v=PkXELH6Y2IM.
94. De Carvalho-Filho MA, Tio RA, Steinert Y. Twelve tips for implementing a community of practice for faculty development. Med Teach. 2020;42(2):143–149. doi:10.1080/0142159X.2018.1552782
95. Bernson-Leung ME, MacNeill H. Big assumptions in online and blended continuing professional development: finding our way forward together. J Contin Educ Health Prof. 2024;44(3):211–216. doi:10.1097/CEH.0000000000000528
96. Basbøll T. Why you shouldn't cite, acknowledge, or credit an AI with authorship [Internet]. Inframethodology. 2023; [cited 2024 Aug 4]. Available from: https://inframethodology.cbs.dk/?p=6171
97. Elsevier. The use of AI and AI-assisted technologies in writing for Elsevier [Internet]. n.d. [cited 2024 Nov 19]. Available from: https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier.
98. McAdoo T. How to cite ChatGPT [Internet]. 2024 [cited 2024 Aug 4]. Available from: https://apastyle.apa.org/blog/how-to-cite-chatgpt.