

Generative AI in higher education

Paradigm shifts in assessment

Michael Rowe
School of Health and Social Care
University of Lincoln

Some operational assumptions

- Assessment is an attempt to *infer* learning from observable behaviours.
- All assessment is prone to error because our inferences can only ever *approximate* what a student knows.
- All assessment suffers from a threat to *validity* because of the errors inherent in the process.
- Learning is defined by *change*, and so assessment processes must be cognisant of change over time.

What is different?

- Language as a fundamental technology that affects all other technologies.
- Scale and speed of content production.
- Language as a user interface.
- Cross-disciplinary integration.
- Society-wide access.

Standard assessment paradigm (SAP)

“A predefined set of items (e.g., problems or questions) is used to infer claims about students’ proficiency in one or more traits. The data used for these inferences are typically sparse, and student learning may not be the focus of the assessment.”

Assessments in the SAP are:

- *Onerous* to design and implement.
- *Discrete* snapshots rather than change over time.
- *Uniform* and unaware of individual competence.
- *Inauthentic*, adhering to the culture of schooling.
- *Antiquated*, assessing skills that (becoming) obsolete.

AI-supported assessment

- *Onerous Sustainable*. AI can generate assessment tasks, connect peers (MKOs), score (some types) of tasks, improve writing.
- *Discrete Continuous*. AI is always on, constantly aware of students' observable behaviours, making more accurate inferences over time.
- *Uniform Adaptive*. AI can modify assessments based on individual student ability.
- *Inauthentic Authentic*. AI can be used to simulate tasks typically performed by members of a community of practice.
- *Antiquated Modern*. AI is being integrated into all aspects of society, so assessments need to take that into account.

Changes to LLMs

- Multimodality: Translation between media formats.
- Fine-tuning: Narrowly focused models (e.g. Medicine, Law).
- Embeddings: Personalised models customised with private data (e.g. emails, notes, presentations).
- Plugins: Additional skills that we can make use of through natural language (e.g. computation, internet access).
- System message: Software developers creating personas (e.g. personal tutor, coach, study guide, assessor).

Universal 'anything' machines

- These features enable the creation of highly customisable, contextually rich, personally-aware characters which are high-level experts in a wide range of disciplines, that we interact with through natural language.
- A network of generative AI systems interacting with each other as increasingly autonomous agents embedded across society.
- How will society change when every person has access to a personal physician, teacher, lawyer, accountant, etc? When we have expertise on demand.
- How will our assessment need to change so that it's fit for purpose?

Principles of assessment design

- Don't use assessment design to solve the 'ChatGPT problem'. Use ChatGPT to solve the 'assessment design' problem.
 - Assessment typically suffers from issues with validity and reliability.
 - LLMs can help address issues with assessment, that are far wider than addressing issues with LLMs in assessment.
- Design tasks that require students to engage ChatGPT in:
 - Debate / critical thinking.
 - Fact-checking.
 - Problem-solving.

Example of assessment task

Physiotherapy training in some countries typically requires that students spend 1000 hours on placement. However, there are concerns that this requirement has no basis in the evidence base. I would like to engage in a debate with you about this training requirement. You will take the position that 1000 hours of clinical practice is necessary, and I will argue that it is not necessary. Each of us will give one reason for our position, and then debate the point. We will continue for 5 rounds. You must not converge towards my position. Does this make sense?

- **Debate ChatGPT:** You must write at least 5 responses to the arguments presented by ChatGPT.
- **Copy the transcript** of the debate and paste it into a Word document.
- **Annotate the transcript**, highlighting the following:
 - Point out the strong and weak claims that both you and ChatGPT have made.
 - Explore the truthfulness of claims that you and ChatGPT have made. This means you will need to provide source material for both your, and ChatGPT's claims. Do not rely on ChatGPT to provide these sources.
 - Note any errors in argument or logic on both sides.
 - Identify factual inaccuracies and provide corrections in the footnotes.

Faculty development framework

- Agency / Accountability
 - Users are responsible for their use of AI.
 - Provide guidelines for decision-making.
- Authenticity
 - Personalisation: Link the work to personal experience.
 - Transparency: Explain the use of AI.
- Awareness
 - AI will become more invisible over time.
 - Increased capacity to use AI unknowingly.
 - Responsibility to understand.

Implications of AI-supported assessment

- Potential movement of accountability from teachers to software developers.
- Changes to the pedagogic relationship, where activities around assessment that influence learning, are removed.
- Increasing surveillance of student behaviour, and the implications for control and compliance.
- The introduction of AI-supported assessment practices will introduce new challenges.
- However, we may also find that inference of learning by AI may be more accurate and defensible.

AI and learning

- The conversation in higher education has so far tended to focus on the challenges around assessment.
- Little has been said about using AI to support learning.
 - Learning by students.
 - But also, learning by lecturers.
- We have an opportunity to enhance our own practice as teachers, using AI to be more efficient and effective.
- (Teacher + AI) + (Student + AI) = Better outcomes.

Thank you.

<https://www.mrowe.co.za/blog/>