Additional resources

'Thinking through genAI' – Frameworks for thinking and acting: evaluating and enacting

Dr Mary Beth Benbenek and Dr Cassie Lowe

This information pack highlights some recent literature, resources and tools related to genAI in higher education. It is shared to inform and encourage reflection on current thinking, rather than to promote any particular educational practices, assessment methods, policies or viewpoints.

Contents

UNIVERSITY OF CAMBRIDGE
Cambridge Centre for Teaching and Learning

1. The ‘Two Lane’ approach

“we do not foresee a viable middle ground between the two lanes. It needs to be assumed that any assessment outside lane 1 (i.e. that is un-secured) may (and likely will) involve the use of AI” (Liu, 2023)

Lane 1: examples of assured ‘assessment of learning’

- In-class contemporaneous assessment e.g. skills-based assessments run during tutorials or workshops
- Viva voces or other interactive oral assessments
- Live simulation-based assessments
- Supervised on-campus exams and tests, used sparingly, designed to be authentic, and for assuring program rather than unit-level outcomes

Lane 2: examples of human-AI collaboration in ‘assessment as learning’

- Students use AI to suggest ideas, summarise resources, and generate outlines/structures for assessments. They provide the AI completions as an appendix to their submission.
- Students use AI-generated responses as part of their research and discovery process. They critically analyse the AI response against their other research. The AI completion and critique provided as part of the submission.
- Students initiate the process of writing and use AI to help them iterate ideas, expression, opinions, analysis, etc. They document the process and reasoning behind their decisions. Students design prompts to have AI draft an authentic artefact and improve upon it. They document the process and reasoning: initial prompt, improvements, sources, critiques. The documented process demonstrates learning, is graded, and is more heavily weighted than the artefact.

Liu, D. & Bidgeman, A. (2023) ‘What to do about assessments if we can’t out-design or out-run AI?’ The University of Sydney.

2. The Artificial Intelligence Assessment Scale (AIAS)

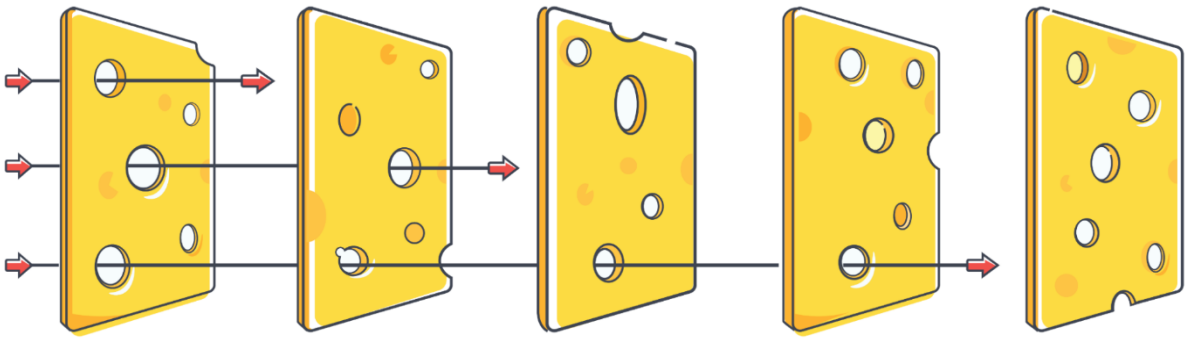
Scale Levels and Descriptions

1	**No AI**	The assessment is completed entirely without AI assistance. This level ensures that students rely solely on their knowledge, understanding and skills. **AI must not be used at any point during the assessment.**
2	**AI-assisted idea generation and structuring**	AI can be used in the assessment for brainstorming, creating structures, and generating ideas for improving work. **No AI content is allowed in the final submission.**
3	**AI-assisted editing**	AI can be used to make improvements to the clarity or quality of student-created work to improve the final output, but no new content can be created using AI. **AI can be used, but your original work with no AI content must be provided in an appendix.**
4	**AI task completion, human evaluation**	AI is used to complete certain elements of the task, with students providing discussion or commentary on the AI-generated content. This level requires critical engagement with AI-generated content and evaluating its output. **You will use AI to complete specified tasks in your assessment. Any AI-created content must be cited.**
5	**Full AI**	AI should be used as a 'co-pilot' in order to meet the requirements of the assessment, allowing for a collaborative approach with AI and enhancing creativity. **You may use AI throughout your assessment to support your own work and do not have to specify which content is AI-generated.**

Perkins *et al*. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(6), 1-18.

3. Assessment validity and the Swiss Cheese Model

'...validity can also be building of chains of evidence in support of judgements about student capability (St-Onge *et al*. 2017). This means that no single act of assessment can address all potential validity problems, however multiple types of assessment, at multiple time points, with multiple assessors can better address the challenges found in a single assessment (Van der Vleuten *et al*. 2012). This is compatible both with the notion of programmatic assessment, and with the Swiss Cheese Model (Reason, 2000) that has been adapted to the context of integrity and cheating by Rundle, Curtis, and Clare (2020). They argue that any one approach to address cheating will have 'holes' but that many layers of different types of approaches will better address the deficiencies of any one layer.'



Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, *49*(7), 1005–1016.

4. Assessment Framework for Generative AI (Blooms Taxonomy)

Abridged representation of the assessment framework for generative AI

Bloom's level	Knowledge dimension	Goal of assessment	Formative assessment with the inclusion of AI	Summative measure example (with AI)
Remember [Recall information]	**Procedural** [Recall]	Remember	Generate a list of vocabulary terms and definitions on a topic and ask AI to turn them into flashcards for studying	Test: learner is provided the definitions and must fill in the term
Understand [Demonstrate understanding of facts]	**Metacognitive** [Predict]	Infer personal reaction of information / situation	Ask AI to describe the way in which a particular organisation might respond to a new HR policy	Draft a memo to those with this role explaining the policy and how it applies to them / why it matters
Apply [Use information or solve problems in new situation]	**Factual** [Respond]	Answer procedural questions	Ask AI to detail the stages of grief (Elizabeth Kubler Ross)	Respond to a scenario where an individual is grieving: what stage are they in and what do you do
Analyse [Break information into parts]	**Conceptual** [Differentiate]	Compare and contrast, separate, classify, characterize, distinguish, examine	Ask AI to outline similarities and differences between federalists and anti-federalists	Video presentation: 'why I am a federalist' and then flip to 'why I am an anti-federalist'
Evaluate [Make judgements and defend opinions]	**Procedural** [Judge]	Assess, synthesize, interpret, conclude, predict, justify, critique	Ask AI to identify strengths and weaknesses of a speech / presentation. Ask AI to evaluate 6th-grade student essays and assign writing skill levels.	Write an argumentative essay and have AI provide a critique. Revise based on AI feedback.
Create [Develop new idea or solution]	**Metacognitive** [Create]	Personal transformation, mapping self-improvement	Ask AI to provide 5 sources on becoming a better teacher, nurse, counsellor, etc. Reflection flip: reflect upon a recent difficulty of failure. Ask AI to offer solutions.	Use AI research to build a career improvement plan. Rank solutions and discuss why one makes the most sense or offer an additional solution.

Page, E., Meyers, G., & Krahe Billings, E. (2024). Theory to practice: An assessment framework for generative AI. Intersection: A journal at the intersection of assessment and learning, 5(4), 114-126.

5. QAA genAI guidance on strengths and weaknesses of assessment types

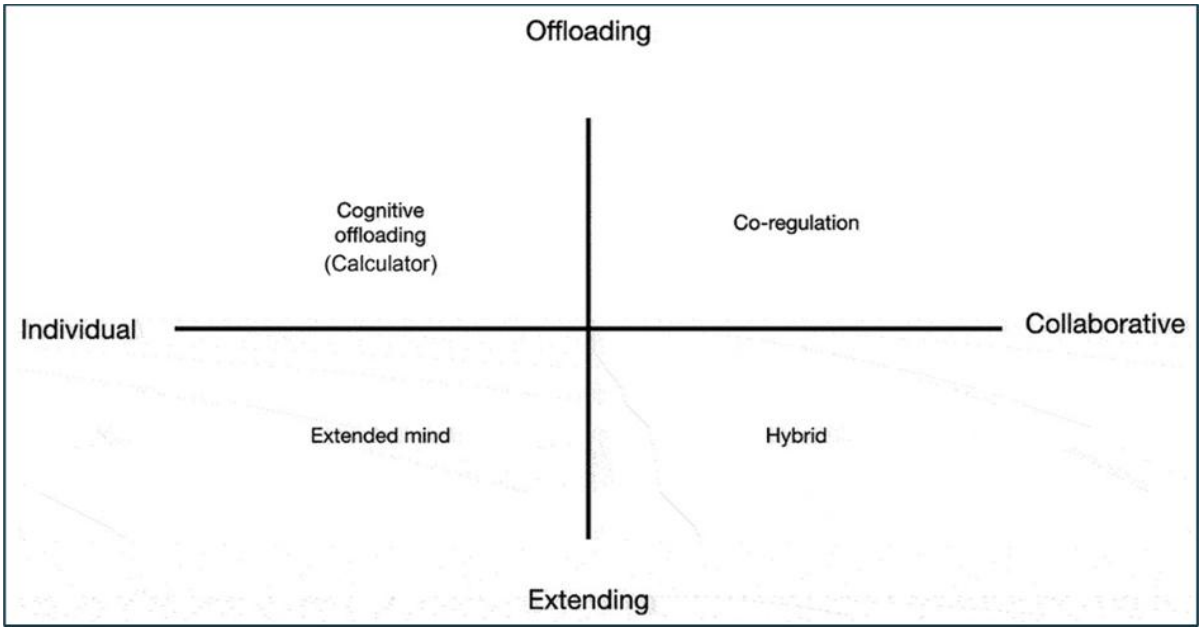
At a glance: strengths and weaknesses of different assessment types

Assessment type	Strengths	Weaknesses	Academic integrity	Sustainability
Invigilated unseen examinations (handwritten)	**Security** – no (or only controlled) access to external physical or digital sources **Synoptic** – tests different learning outcomes via structure of paper and candidates' choice of questions **Volume** – can assess large numbers of students in parallel	**Accessibility** – challenging for students with certain characteristics to access, plus rapidly decreasing number of students adequately prepared to handwrite large amounts of text **Authentic** – may only test a narrow range of knowledge / competencies **Resources** – places significant demands on a provider's estate	High	Low
Invigilated unseen examinations (digital)	**Distribution** – exams can be delivered offsite with appropriate digital security **Security** – with correct software, access to external physical or digital sources can be limited by the provider **Synoptic** – tests different learning outcomes via structure of paper and candidates' choice of questions **Volume** – can assess large numbers of students in parallel	**Accessibility** – there are challenges, though these may be easier to mitigate through technology **Security** – even with digital proctoring, remote candidates can access digital assessment via other devices **Authentic** – depending on assessment design, it may only test a narrow range of knowledge / competencies **Resources** – need to continual investment in digital security software	Medium	Medium
Observed examinations	**Authentic** – opportunity to apply competencies / knowledge to a range of realistic scenarios **Synoptic** – tests a wide range of competencies and understanding from different parts of the programme **Volume** – can assess substantial numbers of students in parallel	**Security** – scheduling demands may mean that students assessed early in cycle can transmit information to those taking the assessment later	High	High

Oral examinations	**Authentic** – tests competencies that can be used in interviews, presentations and meetings **Synoptic** – tests a wide range of competencies and understanding from different parts of the programme	**Accessibility** – stressful for some students and challenging for those with certain characteristics (e.g. speech or hearing disabilities) **Resources** – consumes considerable amount of staff time **Volume** – suitable only for individuals or small groups at a time	High	Low
Coursework that integrates generative AI by design – use of these tools is part of the assessment brief and outputs are critiqued or reflected upon	**Authentic** – learning by doing, including using generative AI tools, can be built into assessment design **Detection** – detection is not necessary if using the AI tools is part of the assessment **Resources** – can repurpose existing assessments **Synoptic** – can be designed to test a range of knowledge / competencies and different elements of a module / programme	**Accessibility** – need to ensure fair access to AI tools for all students **Resources** – initial investment of time and ongoing review to take account of developments in AI tools	High	High
Hybrid submissions – in which the use of generative AI tools are not part of the assessment brief (e.g. essays, dissertations)	**Authentic** – principally for those students who wish to continue to postgraduate education and beyond but also other careers **Synoptic** – test the ability to synthesise knowledge and evidence from across a programme	**Accessibility** – need to ensure fair access to AI tools for all students **Authentic** – less obvious relevant for those exiting academic but still develops evaluation of evidence and synthesis **Detection** – heavily dependent on student declaration even at current state of evolution of tools and will be even more difficult when integrated within licensed software **Resources** – need to invest more resource in developing foundational academic skills on which to scaffold ethical use of generative AI	Low	Medium

QAA (2023, July). Reconsidering assessment for the ChatGPT era: QAA advice on developing sustainable assessment strategies.

6. Typology model for generative AI in higher education



- Cognitive offloading refers to the use of genAI to 'think' on behalf of the individual, i.e., AI is not integrated into thinking, only used to lighten workload.
- Extended mind describes the AI functioning as a cognitive prosthesis for the individual person that enlarges or enhances what the learner can do (e.g., smartphones extending the memory of humans).
- Employing genAI as a co-regulator of learning positions it as a 'coach' – one that provides explanation, feedback and clarifications when specifically prompted.
- Hybrid learning positions genAI as an active collaborator with shared agency, engaging in iterative, co-constructive, dialogic support for both cognitive and metacognitive processes in ways that can transform how the learner thinks beyond the immediate task.

It is important that if AI is going to be integrated into the learning process, it is done so thoughtfully to ensure the student's own development of the required knowledge and skills.

The authors suggest that in this regard AI can function as a scaffold - withdrawn as learners gain competence - or as a reverse scaffold, introduced once core outcomes have been mastered.

Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? Learning: Research and Practice, 9(2), 117–124.

7. Discursive changes vs structural changes

“Discursive changes” lead to “enforcement illusions”.

Structural changes “reshape the underlying framework of the task, constraining or opening the student’s approach in ways that are built into the assessment itself” (Corbin *et al*. 2025d: 1093).

Examples of structural changes include iterative asynchronous tasks, live in-person assessments, and synchronous supervised activities.

What is important is that the assessment structures align with the intention of what needs to be measured (i.e., assessment validity). For example, “if we want to develop a student’s ability to think deeply and develop complex arguments over time, an asynchronous format may be appropriate, but we would need to build in structural assessment elements that capture the development process rather than just the final product.” (Corbin *et al*. 2025d: 1093).

Corbin, T., Dawson, P., & Liu, D. (2025d). Talk is cheap: why structural assessment changes are needed for a time of genAI. *Assessment & Evaluation in Higher Education*, *50*(7), 1087–1097.