

A Review of the Implications for Assessment in the Generative Artificial Intelligence Era

Author:
Peter Gillis – Learnovate Centre

Version 1.1
May 2024

Table of Contents

Note on the use-case of higher education.....	4
Executive Summary	5
University positions on Generative Artificial Intelligence (GAI) and assessment.....	7
Introduction	7
A review of Times Higher Education (THE) top ranking universities' GAI guidelines	8
University College London (UCL)	10
Monash University, Melbourne	12
Quality and Qualifications Ireland (QQI)	15
The Hong Kong Polytechnic University	16
Georgia Institute of Technology	17
Discussion	18
AI tools, cannot be used (<i>UCL Strand A</i>).....	19
AI tools can be used in an assistive role (<i>UCL Strand B</i>).....	19
AI has an integral role (<i>UCL Strand C</i>).....	21
Conclusion.....	21
Research looking at Generative AI (GAI) in academic assessments.....	22
Introduction	22
The problem	22
Researching assessment approaches	23
Conclusion.....	25
Generative AI (GAI) tools to support assessment	26
A meta study	26
A GAI tool for knowledge-grounded answer evaluation	27
A commercial GAI assessment & feedback platform	28
Conclusion	29
Academic dishonesty and plagiarism.....	30

Academic Dishonesty pre-GAI.	30
Academic Dishonesty in the GAI era	32
Conclusion	33
Closing Remarks	34
Chat GPT says:.....	34
References	37

Note on the use-case of higher education.

In this report we focus on the higher education space. At Learnovate we take a use-case approach to research to keep us focussed on our activities. This never precludes other use-cases, we merely focus on one and then look to see how much change might be required when applying the concept to a different domain. In this case it might well be asked why we did not focus on Corporate Learning or on Schools (K-12).

We may produce iterations in these areas in the future, but we focused the initial research on third level because of the following features: Third level students do tend to have more high-stake, summative, and formative assessments than corporate learning and, to a lesser extent, schools. These assessments are often over a period of time, for example weekly assignments. Also, as adults, third level students were seen as having more access and ability to use, or misuse, Generative Artificial Intelligence (GAI) in assessments.

In relation to assessment in Corporate Learning, “Level 2: Learning”, from a Kirkpatrick viewpoint would be the only category that would fall under assessment as it is looked at in this context. GAI stands to become an invaluable work tool to assist employees in achieving impact at Level 3: Behaviour and Level 4: Results, potentially requiring L&D to rethink more than just learning assessment.

We did take a brief look at research related to assessment and GAI, as it pertains to Schools (K12). The research that was found cited research that was carried out at higher education level and inferred that similar would apply in Schools (K12), (Mintz, Holmes, Liu, & Perez-Ortiz, 2023; Sanabria-Z, Alfaro-Ponce, Argüelles-Cruz, & Ramírez-Montoya, 2023).

However, we did find an interesting call for further general GAI research in the Schools (K12) space. Koh & Doroudi, (2023) in the introduction to a GAI-dedicated issue of ‘Learning: Research and Practice’ identified five areas which require further investigation:

1. More empirical studies, too much opinion at this early stage.

2. The ethical aspects of GAI. For example, what is at stake if students and teachers over-rely on tools that are developed by large companies with commercial or political motives?
3. The issue of free and easy access to GAI technologies for all learners.
4. There is much more to GAI than Large Language Models (LLM). Future work should look at how the range of these tools will affect teaching and learning.
5. GAI has widespread implications for 'what' we learn. More work is needed.

Executive Summary

We are entering an era of opportunity and threat when it comes to assessment in higher education. But irrespective of which way you might feel about the potential impact of Generative Artificial Intelligence (GAI) on assessment, one thing is for sure, GAI is here to stay.

At the start of the Covid pandemic the world's ability to adapt and move education online in a very short time, due to necessity, was perceived as an epic achievement. Now most are reflecting on lessons learned and looking to innovate education going forward. However, there are still a lot of educators who reverted to the more comfortable and familiar pre-pandemic approaches. The introduction of GAI is very similar, a groundswell of immediate activity has been required to deal with its rapid and pervasive introduction, and to some extent to try and stop it. However, this rapid rise of GAI can be an opportunity to reimagine higher education assessment and to align it better to meet the needs of the 21st century.

In this report we look at the guidelines currently in place in higher education institutions. The business of educating students cannot stop, so we do report on a lot of academic integrity, protection type approaches which are necessary as we work out the GAI era. Some institutions are already looking further forward to understand how we assess in a world that will require students to collaborate with GAI in the workplace, and to focus on the higher order and transversal skills required.

In the next section we identify that it is too early to have a wealth of academic research related to GAI and assessment. However, we do report on a couple of academic research studies undertaken in the realm of GAI and assessment. Both studies look at approaches to assessment design. The first proposes a practical framework that is based on Bloom's taxonomy to evaluate the ease with which assessments can be completed successfully by GAI, those that cannot tend to be higher up the taxonomy, involving aspects like creativity. The second study experimented with the introduction of oral Vivas in a chemistry subject. While the students found it challenging, they also found it rewarding and looked to have its use continued beyond the experiment. Our call and the call of others is that more experimental research is urgently required in this space.

We report on research, and a product available, in the space of GAI tools to support educators. The tools aim to bring utility, but in looking at the tools there is little said about the issue of cheating.

Finally, we conclude with a look at the core reasons for academic dishonesty. Cheating is a motivation issue. Higher education cheating has been around a long time, with rates of around 70% being reported in the early '90s. There is research looking at the base reasons students cheat, such as to achieve an award or to look good in the eyes of their peers. But there are downsides such as poor self-image or a propensity to cheat in later life. Throughout this report we refer to the identification of tools that can help us move forward and we see motivation as one of these.

As mentioned, we refer to different tools. Experimental research into these tools should be undertaken as a matter of urgency to catch the GAI wave. A proposed Intelligence Augmentation (IA) approach would seem appropriate as part of the process of reinventing assessment for the GAI era.

University positions on Generative Artificial Intelligence (GAI) and assessment.

Introduction

In this section we look at higher education institutions' stances on assessment in the emerging era of GAI. We also look at the guidelines universities currently provide to support their staff in the design of relevant assessment tasks. As will be shown the stances and guidelines vary, from embracing GAI to preventing its use.

The University of Birmingham provide a promising, overarching view of GAI, as it applies to higher education, when they state “ *To suggest that AI will have no impact on the evolution of higher education is not only denial but existentially dangerous. Despite all the challenges of AI, we must acknowledge that higher education will thrive alongside any AI evolution as long as we learn to adapt our pedagogy and assessment strategies*”. (Benuyenah, 2023). The article continues by giving three reasons why this will be so, (1) Universities will always set new policies to counter cheating; (2) students want to learn, and (3) there are tools available to deal with Intentional Academic Dishonesty (IAD). However, these three seem to conflict with the positive nature of the original statement. They attempt to prevent and identify cheating rather than embrace AI as a tool to be leveraged in student development. This is a view that will be reported again and again in this section. Because AI is so recent and evolving at a rapid pace, a lot of institutions may feel there is a need to take the more protective stance, but in this section we also see evidence of others seeking to embrace and work with GAI.

Jason Lodge, (2023), Associate Professor of Educational Psychology at University of Queensland also provides a broad category review of the options available when dealing with GAI in assessment. He refers to six options as follows:

1. Ignore GAI and hope it will go away. Far-fetched as it may seem the worlds of technology and education have had many other false dawns, like Massive Open Online Courses (MOOCS) and others.

2. Ban GAI. This option is more than likely unfeasible given the ability of students to understand new technology affordances including the ability to work around AI detection tools.
3. Invigilate assessments. Beyond proctoring exams or reverting to hand-written timed exams, Lodge proposes there are more creative ways to ensure assessment outputs reflect genuine evidence of learning outcomes, for example oral examinations.
4. Embrace GAI. Design the use of GAI into assessments. While this sounds like a workable option there is a way to go when aspects such as copyright, equity and level of GAI skills are considered.
5. Design Around. This aspect involves understanding the weaknesses of GAI and designing assessments that minimise the effectiveness of GAI in assisting students. However, with the pace of evolution in GAI this may be a short-lived solution at any stage.
6. Rethink. Is the arrival of GAI an opportunity to rethink assessment? An opportunity to investigate the validity of assessing learning at a particular moment in time and making judgement calls based on the outcome. This is not easy, but GAI may be the impetus to start this journey for the longer term.

Benuyenah, (2023), and Lodge, (2023) provide differing high-level views of GAI's role in the realm of academic assessment. In the following sections we review current activities, from across a selection of universities and higher education institutes, as they tackle the immediate approaches to GAI and assessment.

A review of Times Higher Education (THE) top ranking universities' GAI guidelines

Moorhouse, Yeo, & Wan, (2023) carried out a review of the top 50 THE ranked universities' GAI guidelines for instructors. On review, only 30 of the universities were found to have GAI guidelines. Further applying the criterion of 'dedicated to instructors' in the format of guidelines, and 'issued at university level', reduced this

number to 23. This highlights how new this area is, where over half the top 50 ranked universities have not yet published GAI guidelines.

In looking to the 23 universities included, and specifically focussing on assessment, the paper reports three main themes emerged under Academic Integrity: Plagiarism, Acknowledgement of GAI and Detection of GAI. On the last point there was a split, with less institutions encouraging the use of detection tools, and more discouraging the use based on privacy and GDPR aspects.

The paper also looked at what trends emerged related to “Advice on Assessment Design”. Five themes emerged as follows:

Test Assessments using GAI tools: Instructors should get familiar with the GAI tools, how they might deal with their assessment task. Instructors need to understand the GAI tools and how they improve with the refinement of prompts and are evolving with each iteration. The advice was against putting student papers through GAI tools, from a privacy perspective.

Re-design Assessment Tasks: Across the institutions that provided advice on re-design (N=17) four categories emerged. Include elements of critical thinking, incorporating contextual elements such as connecting responses to course content and lived experience, authentic assessments where, for example, they apply the concepts in question to real world applications, and providing alternative ways for students to represent their knowledge such as presentations, infographics, and others.

Focus on Process and Staged assessment design: Students submit formative pieces of progression that would be evaluated across the duration of the module and build toward a summative submission. The submissions could be notes, drafts, etc., and could be peer-assessed. Time to complete the tasks is identified as a challenge with this approach but it is also noted this approach could reduce final exam anxiety.

Incorporate GAI tools in assessment process: 10 of the universities propose that another alternative is to incorporate GAI into the assessment. Recognising that GAI is a tool that exists and incorporating it can build digital literacy skills for students. The approaches suggested using GAI as a formative assessment aid, or for the students to

generate GAI responses to a question and then critique it. However, it was noted the use of GAI should be acknowledged, and to be aware that requiring students to use GAI may have ethical and privacy concerns.

Use in-class assessments: Four universities refer to leveraging in-class, synchronous assessments, oral or written, maybe as part of assessment, reviewing the weighting these assessments have toward final grade. However, there are warning notes provided too. There may be a limit to what students can demonstrate in these approaches and there also may be diversity, equity, and inclusion issues around such approaches. “These changes could raise more difficulties than they solve”, (Moorhouse et al., 2023).

The third element of the research looked at channels to use to communicate with students regarding GAI and what to communicate. In relation to channels, a statement in syllabi or course outlines, engaging with students in open discussions, and through the librarians. In terms of what should be communicated to students related to GAI use in assessments the following five themes emerge: 1) setting clear expectations, 2) partnering with students to develop class policy, 3) the ethics and limitations of GAI 4) the importance of originality, and 5) the importance of college learning, intellectual struggle, and process (Moorhouse et al., 2023).

University College London (UCL)

Staff at University College London (UCL) worked together to produce broad advisory guidelines for assessment. The guidelines are cognisant of the need to prepare students for future work in an AI-supported world and at the same time seeking to maintain academic integrity. They acknowledge the fast pace of change in this area understanding that constant revision will be required as AI advances. Their advice manifests in a broad three strand fashion, akin to a maturity model. The three are *a) AI tools, cannot be used, b) AI tools can be used in an assistive role, and c) AI has an integral role*. The advice is to adopt one of these approaches in consultation with staff and students. Students are reported as being confused as what constitutes academic

integrity in a world that is rapidly becoming AI enabled. Therefore staff should clearly explain to students which format is involved in their module from the outset, and again closer to exam time. (UCL, 2023).

In category *a) AI tools, cannot be used*, UCL refer to mainly assessments that take place in either a classroom situation or a lab. They are seen as more useful for the more basic skill levels of knowledge retention, understanding, and applying knowledge. The recommended activities included:

- In-person unseen examinations
- Class tests
- Vivas
- Some laboratories and practicals
- Discussion-based assessments.

Category *b) AI tools can be used in an assistive role* looks to allow the use of AI tools for specific aspects of an assessment, for example data analysis. However, it is important to point out that the AI tool use is one part of the assessment to be built on by the student and not the assessment itself. The part assessment activities in this category are where AI is used specifically for:

- drafting and structuring content;
- supporting the writing process in a limited manner;
- as a support tutor;
- supporting a particular process such as testing code or translating content;
- giving feedback on content, or proofreading content.

The final category is *c) AI has an integral role*. This level allows for the use of AI tools to undertake assessments where students tackle complex problems and potentially generate creative solutions. In these forms of assessment the ability to leverage AI tools well is an embedded aspect of the assessment.

Some examples include:

- comparing content (AI generated and human generated);
- analysing and/or reframing content;

- researching and seeking answers;
- creating artwork (images, audio and videos);
- playing a Socratic role and engaging in a conversational discussion

(UCL, 2023).

UCL is currently taking this broad approach, leaving it to individual lecturers to decide which approach best aligns to their module and the students' capabilities. The three broad bands do allow for flexibility and avoid being overly prescriptive to lecturers. Of course, as they say themselves, this is for now and who knows how quickly it will need to be reviewed, but UCL is committed to preparing students for an AI world of work. UCL also state that will not be deploying any AI detection software and advise their lecturers to avoid doing the same, with reference to students' privacy and GDPR rights.

Monash University, Melbourne

Monash University's AI in Education Learning Circle, a collegiate group of educators from across the university, also refer to a three strand advisory on assessment approaches. The focus is on ensuring assessment is of human abilities and/or knowledge. The three strands are: a) Integrating AI collaboration into the assessment, b) Designing assessment tasks that make the use of AI less relevant, c) Designing assessment tasks that need to demonstrate individual human knowledge and bracketing out AI. In effect, while the order is different, these three strands are reasonably aligned with the UCL view. Monash take a different approach to UCL in expanding on this advice. Instead of examples they provide a table of seven assessment activities, providing a view on how easy or complex it is for GAI to be used to assist in such tasks and recommendations around how to redesign such activities accordingly (see table 1 below).

Difficulty for AI technologies	Task	Suggestion
Easy	Closed answers (e.g., MCQ or short answer exam questions that ask students to define, list or reproduce)	ChatGPT is good at synthesising information to produce convincing answers to closed or abstract questions. Where feasible, consider alternative formats that align with unit learning outcomes.
Easy	Essays without personalised application	ChatGPT can produce convincing text in many different genres and styles as requested (although it may find accurate referencing difficult). Modify questions to require personalised application. Combine with modifications such as assessing process or personal reflection.
Medium	Essays with personalised application	ChatGPT is poor at some kinds of personalised application or contextualisation of ideas. However, students may also find this type of writing difficult and scaffolding this kind of writing is important. Combine with modifications such as assessing process or personal reflection and scaffold students' abilities to write in this way.
Complex	Showing individual working process	ChatGPT does not reveal how it produces its outputs, so it cannot produce an account of its own process. Replicating that would be significant work for students. Consider introducing staged submissions where students explicitly respond to feedback on drafts, or where they are asked to submit a portfolio of the work they have done on the way to their final submitted work.
Complex	Showing collaborative working process	Producing a misleading account of how students worked together to produce an output would involve significant work. Consider group tasks (with or without a final, individual submission) where students are asked to reflect on the process of completing the work and what they have learned from it.
Complex	Multimodal artefacts (e.g. a document that	ChatGPT currently only produces text. This text could be combined with other AI tools (e.g. DALL-E) but in many cases this requires thoughtful engagement with the material to compose something that is more than the sum of its parts. Consider asking students to produce images (e.g. an

	contains text + images) as creative response.	infographic) or video recordings as part or all of a response to a question. These should not just be reproducing text in a different format (e.g. a recording of reading text aloud) but should make use of the format as an alternative way of expressing ideas. Note: Multimodal work is often more ambiguous than text and may require more marker training and time. It may also require more technical support.
Complex	Interview / viva voce	Currently not possible for AI technologies to replicate. Consider asking students to participate in a live (synchronous) conversation (on campus or online) about what they have learned or about their submitted work. Note: This may require scheduling for each student, and may involve a greater time commitment for marking.

Table 1, Assessment tasks and capability for GAI to assist in generating responses. (Monash University, 2023).

For Semester 1, 2024, Monash advise the responsibility for the assessment type rests with the Chief Examiners responsible for the module. The Chief Examiner is responsible for deciding which of the three approaches outlined above is appropriate for the assessment giving the “why,” “what” and “how” of each assessment to guide students. If the Chief Examiner indicates an assessment, in line with c) above, where no use of GAI should be adopted, they must provide an educational rationale as to why it is not appropriate. For their part, in assessments where students are allowed leverage GAI, they must acknowledge what aspects of the assessment were supported and using what tool.

To support their staff in adapting to the assessment landscape in the GAI era, Monash produced four short guideline documents. The first document is “Does my assessment need to change?” including, among other advice, guidance on how to test your current assessment through asking GAI to take the test and then grading it. The second and third are how to design GAI “in” and “out” of assessments respectively, and the fourth provides links to further resources (Monash University, 2024).

Quality and Qualifications Ireland (QQI)

QQI is the state agency responsible for promoting the quality, integrity, and reputation of Ireland's further and higher education system. The National Academic Integrity Network (NAIN) working group, was established by QQI as a response to the evolving and developing field of Gen AI. In July 2023 NAIN released a report to “support educators in their understanding of the potential uses of Generative AI, both in supporting learning for their students, and most critically, in providing a potential ‘short cut’ to students in the fulfilment of required tasks for assessment” (MacLaren et al., 2023).

The guidance provided in the report, related to assessment in the GAI environment, is at a high level but at the same time pragmatic in providing direction. The approach looks at four audiences, everybody, lecturers, students, and programme managers and institutional leaders.

Five aspects are covered in the **Everybody** category; 1) people need to be aware the GAI outputs can be impressive and not just limited to text, 2) GAI is now being embedded in everyday tools such as Microsoft Office, 3) understanding that GAI is not a subject matter expert, but based on mimicry and reproduction, and answers may not be factual, 4) there are concerns around how the GAI models are trained, including copyright and labelling processes, 5) it is important to stay up to date with institutional policy as it is likely to be revised frequently.

For **Lecturers**, NAIN provide six guidelines as follows; 1) be familiar with GAI tools, what they can and cannot do. 2) Review your current assessments and redesign or modify accordingly. If GAI use is to be allowed, ensure students are aware of what is and is not allowed. 3) Work with your colleagues to ensure assessments target appropriate learning outcomes and skills. In addition, that the assessment is appropriate and potentially develops higher order thinking skills. 4) Take in the programme level view to make sure the volume and range of assessment types is appropriate. 5) Provide consistency across the assessment process including academic integrity policies, pointing out the risks around academic misconduct to students. 6)

Build rapport with your students around the positive aspects of learning, integrity, trust, knowledge discovery, etc.

Students should be aware of four aspects. 1) While GAI outputs can be attractive and easy to use, they have limitations including providing convincing responses that can turn out to inaccurate or false. 2) As above, be as aware of what GAI cannot do as what it can, and potential ethical issues. 3) Learning is based on student development of knowledge and skills in line with the learning outcomes; students should only submit work they have done and acknowledge any sources leveraged in the process. 4) Students need to avoid any misconduct related to assessment, if they are struggling they should not look to GAI for answers they should discuss this with their lecturer or student counsellor.

Programme Managers and Institutional Leaders are provided with four guidelines. 1) Keep training related to GAI and integrity up to date and provide appropriate resources. 2) Review assessment strategies, with colleagues, on a regular basis. 3) Refresh policies regularly. 4) Develop an ethos and culture that values integrity, honesty and trust, (MacLaren et al., 2023).

The Hong Kong Polytechnic University

The Educational Development Centre (EDC) at the Hong Kong Polytechnic University designed an approach that they felt could be applied across disciplines and schools. The guidance is provided in the form of four Directions for assessment.

Direction 1 looks to have students evidence their learning by means of applying the learning to a specific instance, possibly a context, or a location that is local. GAI tends not to be good at this approach. In addition the student is asked to go beyond written text and possibly to show portfolio-type evidence to support their submission via photographic evidence or illustrations.

Direction 2 is designed for courses where the first direction might prove challenging. This approach requires students to include references back to specific lectures or discussions that were held in class.

Direction 3 proposes an oral 'viva' type approach with minimum focus on the presentation aspect leaving more for students to defend their work. This differs from a presentation which might still be susceptible to having GAI prepare both professional slides and an associated script. This approach also opens up the opportunity to assess critical thinking, potentially by playing devil's advocate in the questions asked.

Direction 4 refers to academic staff partnering with senior students, who may well be more familiar with the affordances of GAI, to design new approaches to assessment that GAI will struggle to complete.

As Julia Chen, (2023), director of the EDC points out, these approaches are in a way ambivalent to the use of GAI and are more concerned with designing assessment approaches that can meaningfully provide students with the opportunity to "demonstrate understanding via application of knowledge in situated contexts and to employ critical thinking and communication skills when completing and defending their assessment".

Georgia Institute of Technology

Georgia Tech have adopted a slightly different approach when it comes to thinking about GAI and how it relates to academic assessment. The approach is a more philosophical approach to understanding what GAI is, and what is meaningful assessment of students is and could be.

In relation to GAI and its affordances, Lee & Soylu (2023) remind us the information returned may not be factual, and when queried, Chat GPT points to its own limitations "*limited understanding of context, can potentially generate offensive or harmful content, has limited ability to fact-check or verify information, lacks empathy, and has limited knowledge on certain topics (especially if the subject matter is complex)*". In addition, understanding that while on the surface GAI outputs are impressive it does not mean GAI understands what it says.

Building on the previous paragraph the attention switches to the students and looks more at the opportunities than the threats. As mentioned above the outputs from GAI can be questionable, Lee & Soylu (2023) see the potential to progress assessment to

encompass the aspects of critical thinking, integration and synthesising of information into assessments. In addition, there has been a long running concern around the development of students' transversal and 21st century skills, an aspect Learnovate are well familiar with and have researched in projects like Assess21C, SkillTrack and others. Does the introduction of GAI provide space to assess meaningfully these skills in a more structured way? To achieve this students and tutors will need training in how to work with GAI and specifically in designing quality prompts.

Lee & Soylu, (2023) finish out their contribution referring Dede, Etemadi, & Forshaw's, (2021) work on **intelligence augmentation (IA)**. The concept is built on the complementary approach of GAI and humans working together each doing what they are good at to achieve a better outcome. In this approach, GAI is good at formulaic decision making and humans are uniquely good at judgement and other human-based abilities. This approach could potentially free up time to develop and meaningfully nurture those aspects that GAI, as far as we can tell, will not be able to assist with.

Discussion

This research report is being written at an early stage in the GAI/education assessment journey. As such, we must accept that some of the guidelines are of an 'emergency' nature, dealing with a paradigm that has landed very quickly and spread very fast. Indeed, parallels could be drawn with the reactions that were required to shift to digital learning and assessment during the pandemic. This manifests in the number of guidelines taking a defensive stance to GAI and educational assessment, when GAI is now a fact of life and evolving rapidly.

The calculator analogy gets used quite a bit in the literature, and while not a perfect fit it is a technology that was introduced in the education space and initially was defended against. Such is the pace of evolution it would be expected that the guidelines reported here will change quite rapidly and may have done so already. However, the three-strand approach designed by UCL, reported on earlier, is akin to a maturity model of approaches to GAI adoption. As such it provides a good lens to compare the different guidelines covered. Table 2 below, shows that when analysed in this fashion there are some broad similarities across the guidelines provided by institutions. In the following sections we review this alignment and discover some discrepancies among these strands.

AI tools, cannot be used (*UCL Strand A*)

The two most popular approaches reported here are in-class examinations and viva type oral defences or discussions of work. In-class examinations are an approach that would be familiar to most and does feel retrograde, there have even been some calls to have these done as hand-written exams. The Viva oral approach is one that is a discrepancy as this approach is put forward by some as an approach to block cheating through GAI and assess students' knowledge by defending their work while others propose a similar approach under 'Strand C - AI has an integral role'. The slight difference appears to be that the orals in this strand appear to be around defending a piece of submitted work which it is assumed is done without GAI assistance.

AI tools can be used in an assistive role (*UCL Strand B*)

The main reported aspect in this middle territory is to lean into GAI as means of formative assessment but of course linked to this must be the guardrails to prevent GAI doing too much. Therefore, the other component seen here is the need for students to in some way document their journey in a fashion that would not suit GAI. The approaches include notes, drafts and processes followed, all of which could be assessed. As a middle approach this is what it feels like and ultimately sounds like a heavy effort on all concerned to ensure fairness and equity.

Institution(s)	Allow GAI as part of assessment?		
	No	Yes	
	UCL Strand A - AI tools, cannot be used,	UCL Strand B - AI tools can be used in an assistive role	UCL Strand C - AI has an integral role.
UCL	<ul style="list-style-type: none"> - In-person unseen examinations - Class tests - Vivas - Some laboratories and practicals - Discussion-based assessments. 	<ul style="list-style-type: none"> - drafting and structuring content; - supporting the writing process in a limited manner; as a support tutor; - supporting a particular process such as testing code or translating content; - giving feedback on content, or proofreading content. 	<ul style="list-style-type: none"> - comparing content (AI and human generated); - analysing and/or reframing content; - researching and seeking answers; - creating artwork (images, audio and videos); - playing a Socratic role and engaging in a conversational discussion
THE Top 50 Universities	Use in-class assessments: Four universities refer to leveraging in-class, synchronous assessments, oral or written, maybe as part assessment.	Focus on Process and Staged assessment design: formative pieces of progression that would be evaluated across the duration of the module. could be notes, drafts etc and could be peer assessed. Using GAI as a formative assessment aid, or for the students to generate GAI responses to a question and then critique it.	Include elements of critical thinking, incorporating contextual elements, authentic assessments where, alternative ways for students to represent their knowledge such as presentations, infographics, and others. Incorporate GAI tools in assessment process: it can build digital literacy skills for students.
Monash University	Interview / viva voce	Showing individual working process Showing collaborative working process	Essays with personalised application Multimodal artefacts (e.g. a document that contains text + images) as creative response.
The Hong Kong Polytechnic University	Direction 3 proposes an oral 'viva' type approach with minimum focus on the presentation aspect leaving more for students to defend their work. Direction 4 refers to academic staff partnering with senior students, who may well be more familiar with the affordances of GAI, to design new approaches to assessment that GAI will struggle to complete.	The student is asked to go beyond written text and possibly to show portfolio type evidence to support their submission via photographic evidence or illustrations.	Direction 1 looks to have students evidence their learning by means of applying the learning to a specific instance, possibly a context, or a location, that is local. Direction 2 This approach requires students to include references back to specific lectures or discussions that were held in class.
Georgia Tech			progress assessment to encompass the aspects of critical thinking, integration and synthesising of information into assessments. Students and tutors will need training in how to work with GAI and specifically in designing quality prompts. Intelligence Augmentation (IA). GAI and humans working together.

Table 2. Synopsis of institution guidelines provided, aligned to the UCL three strand framework.

AI has an integral role (UCL Strand C)

While the approaches to permitting GAI tools are similar across institutions, the underlying motivation does appear to vary. Some propose approaches in this category because GAI will struggle to provide help. Others propose similar approaches but appear to be driven more by the opportunity to reshape the entire assessment process for the better.

The most common approach under this category encompasses the inclusion of critical thinking and application aspects to assessments, in effect what is termed 'authentic assessments'.

The other popular approach in this category is the multimodal artefacts approach where beyond text assessments could include images, infographics, etc.

It is worth noting that Georgia Tech do not have guidelines that fit in either strands A or B, instead their approach is one of accepting GAI is here and how can its arrival be harnessed to improve learning.

Conclusion

As mentioned, a lot of the guidelines here will be short lived but may be necessary as we adjust. In the report we have also covered useful guidelines in terms of practical advice such as those provided by the NAIN group.

The introduction of GAI is a once in a lifetime opportunity to re-imagine assessment in education for the better. However, this is not to play down the importance of 'Knowledge' and we must consider how we reimagine learning so students gain knowledge, skills, and attitudes through meaningful engagement with the topic at hand. From all the work reviewed I believe the Intelligence Augmentation (IA) (Dede, C. Etemadi, A., & Forshaw, 2021) approach, as cited by Georgia Tech, provides the direction for further research including the need for students and tutors alike to learn more about how to work with GAI.

Research looking at Generative AI (GAI) in academic assessments

Introduction

In the previous section we looked at the guidelines released by academic institutions related to GAI and academic assessment. Here, we look at the research in the area, as mentioned it is early in the GAI and Assessment journey and as a result there is a limit to the amount of research that has been completed at this point. There is little doubt that there is plenty of research ongoing and this picture will change. In this section we look at research available in the space, some is generic, and some is topic-specific. Similar to the last section there is a divide between protective, and embracing, approaches.

The problem

One of the main reasons for concern related to the misuse of GAI in academic assessments is that if students cheat, they are harming themselves. They are likely to struggle in the world of work where employers believe they possess certain knowledge, skills and attitudes based on their qualification. On the flip side, “In the future, success in many careers will require working with LLMs or GAI programs effectively” (Ahmad, Murugesan, & Kshetri, 2023).

Memarian & Doleck, (2023) undertook a meta study of research related to GAI in the realm of education, of the 63 papers reviewed only four fell under the thematic area of assessment. In looking at the original four documents a common concept is that GAI is not going away and therefore the development of appropriate policies and guidelines is what is required; academic institutions need to adapt to GAI rather than try to ban or prevent it (Ahmad et al., 2023; Cotton, Cotton, & Shipway, 2024; Perkins, 2023; Tlili et al., 2023). Such policies need to look at ethical and social implications and that any tools leveraged are designed to be accessible, inclusive, and effective for all learners (Ahmad et al., 2023).

There are warnings too, since Covid there has been an increase in dishonesty and plagiarism in academic assessments, and even at the current capabilities of GAI it is all but impossible to distinguish the output from human to GAI-generated scripts (Cotton et al., 2024; Perkins, 2023). To make their point, Cotton et al., (2024) inform the reader at the end of their document that the paper was written by GAI with prompts from the research team!

One of the main functions of higher and further education is to equip their students for the role of employment. GAI is already being adopted at pace in the world of work. All but one of the papers reference the need for academia to embrace this new paradigm, and train their students for such roles, including from an ethical perspective and understanding the limitations of GAI (Ahmad et al., 2023; Cotton et al., 2024; Tlili et al., 2023).

In conclusion to the main meta review document Memarian & Doleck, (2023) do point to the need for future research to “take a more empirical rather than opinionative stance and study ChatGPT in education”, again a reminder of how recent this phenomenon is.

Researching assessment approaches

There are some pockets of empirical research to be found and in this section we will look at two.

Nguyen Thanh et al., (2023) have designed, developed, and tested **a framework to better understand the ability of GAI to complete authentic assessments**. The use case for the research was an introductory-level economics course, centred on gross domestic product (GDP) and economic growth, however the researchers believe the approach is applicable across many domains of education. Authentic assessment requires students to apply their learning to situations and/or real-world scenarios. It is widely accepted as an approach to fostering the development of higher order thinking along with a deep understanding of the topic in hand. In a separate project, Crawford, Cowling, & Allen, (2023) also promote a similar authentic assessment approach, citing an example where GAI provided a credible answer to the question “Does authentic

leadership apply to Apple?”, but did not perform so well when asked the more authentic assessment style question, “Determine and defend the ‘best’ leadership style for Apple if it were to post losses”.

The framework leveraged Bloom’s taxonomy to define question categories, with the six cognitive levels being: “remember”, “understand”, “apply”, “analyse”, “evaluate” and “create” (Practice & Bloom, 2008). The framework is seen as having multiple functions including assessing the ability of GAI tools to answer specific question types, and to leverage GAI in the development of assessment questions and associated rubrics.

While there is not the capacity within this document to review the full methodology, it should be noted that the research developed both numerical and text questions for each of the six categories of Bloom’s and had them cross marked by four experts. The purpose of this research was to compare the GAI tools; ChatGPT-4, ChatGPT-3.5, Microsoft Bing and Google Bard. The findings showed GAI tools “excel at lower levels of Bloom's taxonomy but show weaker performance at higher levels, particularly in ‘create’ tasks”. GAI also performed better with ‘numeric’ better than ‘text-based’ questions. However, there is a salutary point when the authors refer to the most recent GAI in the study, “ChatGPT-4 performs at a high level across nearly all educational goals, even achieving a minimum of the CR mark band for the ‘create’ level of learning goals. This impressive capability presents a critical challenge for educators and universities, especially in fields like statistics or econometrics that heavily rely on data analysis” (Nguyen Thanh et al., 2023).

In the second example Gardner & Giordano, (2023) experimented with the use of **oral exams in Physical Chemistry**. Ironically it was the Covid pandemic rather than the introduction of GAI that inspired this research. However, we have already cited several times in this report, oral exams, or vivas, are seen as one tool of assessment in a GAI environment. In fact Gardner & Giordano, (2023) refer to oral exams as ‘a tool’ among others to choose from, possibly the way we need to embrace in the future.

The research was carried out in two institutions, the students did find the experience challenging and required them to adapt a new strategy of study involving group study where they were able to practice oral presentation among their peers. In addition, the

students reported on the anxiety of having to present their data and the subsequent depth of understanding of the subject matter that was required. However, overwhelmingly the students reported the experience as positive, and they could see the short- and long-term benefits of the approach considering their future career paths and were in support of retaining the approach.

From a tutor or instructor perspective, time required to undertake this approach would be an obvious concern, however it was pointed out that given the oral presentation is graded live during and shortly after the presentation that the time may not be that different. In the research project the approach took about 5 hours for ten students, and the researchers point out this time is only slightly longer than the time require to mark and grade a written exam submission. While the researchers do conclude with two areas for development, namely clarity on the grading process, and avoiding bias this approach does appear to be an easy to adopt 'tool' for the right classes and topics.

Conclusion

We have reported on two pieces of research involving methods that work around, and/or work with GAI. These are early examples of the where we should go. We need to continue to carry out academic research, at the experimental level, to investigate approaches that work in the new GAI reality. These research projects should embrace the potential to reinvent academic assessment at a time when the old rules really do not seem as relevant. There has been a call and some movement toward re-thinking academic assessment. GAI provides the impetus to grab this opportunity, to start with the learner, and ask holistically how can we best serve these students during their time with us.

Generative AI (GAI) tools to support assessment

When searching “automated grading ai tools” the fourth response promises “70 Best Automated essay grading software AI tools – 2024”(TopAI.tools, 2024). While this might seem incredible in just the second year of access to GAI tools, essay grading leveraging artificial intelligence (AI) has been in service since at least the mid-90s. Gao, Merzdorf, Anwar, Hipwell, & Srinivasa, (2024), reviewed recent research in this rapidly evolving area. They introduced this author to the term AIEd, an emerging field with four areas of interest, 1) Decision-making tools, 2) Intelligent tutoring systems, 3) Adaptive systems which provide scaffolding and content personalization, and 4) Assessment and evaluation tools. The paper is focussed on the fourth aspect assessment. The researchers deployed a rigorous approach to filtering the papers for review which ultimately filtered 838 search results to 93 papers meeting the criteria.

A meta study

The research was focused on **text-based automated assessment systems (TBAAS) based on input-process-output (IPO) perspectives**. In the cases of written text, most studies were on short-answer exam questions. Five categories were identified as shown in table 3 below:

TBAAS Type	Explanation
Automatic Grading System	Designed for evaluating student learning outcomes to grade student responses; the output is a grade/score (without textual feedback)
Automatic Classifier	Classifies student/teacher textual responses into different labels or categories (the label is not a score/ grade)
Automatic Feedback System	Designed for evaluating student learning outcomes, focusing on providing textual/visual feedback and guidance
Automated Writing Evaluation System	Designed for automated essay evaluation to assess linguistic proficiency and improve students’ quality of writing
Multimodal Evaluation System	Designed for multi-input (e. g., drawing and writing input) or providing multi-output (e.g., both scores and visual feedback)

Table 3. Text-based automated assessment system (TBAAS) categories, (Gao et al., 2024).

In looking at the 93 studies researched in Gao et al's., (2024) paper it is worth noting that only three are from 2023. Therefore, a lot of the studies are leveraging AI or Natural Language Processing (NLP) approaches to AI. Few described conceptual understanding or representational thinking. Therefore, an important output from the paper is as a useful guide to investigate how we might improve the five categories, and add to them, leveraging the power of GAI. In turn, experimental research can look to provide more powerful tools to the educators' assessment toolbox.

A GAI tool for knowledge-grounded answer evaluation

Tobler, (2024) developed and tested 'smartgrading', a GAI tool for knowledge-grounded answer evaluation in educational assessments. Looking at the issues around the time and effort required to assess open questions, and the limitations of GAI related to factual reliability and the amount of data that can be submitted, Tobler developed a web interface to step by step assist educators in surmounting these obstacles. The interface is developed on pseudocode to interact with GAIs. The pseudocode covers the required parameters as outlined in the table 4. The web application and the associated code are both provided for those who wish to use it or undertake further research in the area.

Web application: <https://stobler.shinyapps.io/smartgrading>

Code and software: www.github.com/samueltobler/smartgrading

According to Tobler, (2024) the possible applications for the concept cover; enhancing formative assessments in educational settings, evaluating open answers to test items in behavioural studies, double-checking the grading of coursework, and adaptations to other purposes.

The app and code were tested using an empirical approach, leveraging data from a previous research study. The previously graded data from one open question were chosen. The 29 study participants were undergraduate or post-graduate natural science or technology students from a highly ranked Swiss university. The results from a Krippendorff alpha test indicated high agreement between the original manual grading and the GAI-based grading (0.818, SE 0.061, 95% CI [0.689, 0.926]), where the

GAI was seen to be stricter in any discrepancies. These results are encouraging for the further use and development of this tool and its ease of access and use make it an ideal tool for near-term research.

Parameter or variable	Parameter or variable
<i>GPT-related</i>	
OpenAI API Key	An OpenAI API key is essential to make use of their LLMs, also considering the associated computing costs. The API key can be generated on their website after creating a user account.
Temperature	The temperature parameter describes the model's randomness when generating an answer. Higher values (closer to 1) result in less predictable responses, while lower values (closer to 0) make the model's output more deterministic. We recommend adjusting the temperature parameters based on the specific requirements of the situation. If short answers are evaluated, a lower temperature and, thus, a higher deterministic output might be better suited. A less predictable output might be favourable for longer answers to allow for more diverse responses.
Model	We recommend gpt-4-based models such as gpt-4, gpt-4-0610, or more recent models for higher accuracy. Considering the current speed of development, it might quickly occur that the herein-described models are, at some point, outdated. However, the general structure of the application or function is easily adjustable for more advanced models that are yet to come. Accordingly, the application requires a manual model specification. It is also worth noticing that the prices of individual models differ, whereas, generally speaking, more recent models have higher token costs.
<i>Test-related</i>	
Question	Question or exercise that has been solved. It is essential to acknowledge that well-written and clearly stated questions result in higher accuracy when evaluating the answers. For instance, asking students to explain a concept might involve more possible solutions than reformulating it and asking for three differences between two related ideas.
Sample solution	The perfect solution for the question of interest. Similarly, specifying different answer options and writing a complete sample solution increases the evaluation accuracy, considering that the LLM will be asked to use this information as the basis for evaluation.
Evaluation instructions	Evaluation instructions can be used to specify when and for which aspects to give points, when to take points away, and which answer option counts how much. For instance, when asking for three differences, it might make sense to instruct the LLM to give one point per argument.
Points	Points indicate how many points might be assigned to individual answers.
Answers	The students' answers must be provided as individual elements of a vector, for instance, by saving them as a single column «.csv» file.

Table 4. Overview of the required parameters and variables (Tobler, 2024).

A commercial GAI assessment & feedback platform

There are already many GAI informed tools in the marketplace. In relation to assessment and looking at a research-based solution we looked at GRAIDE (<https://www.graide.co.uk/>). GRAIDE was developed by teaching assistants at the

University of Birmingham who had identified the challenges around the grading of essay answers related to Low Formative Volume, Long Timelines, Consistency & Quality of Feedback, Time Sink & Repetitive Work, Poor Analytics, and No Additional Value Generation (GRAIDE, 2023). The GRAIDE model operates at the institution licence level, if a lecturer has access to GRAIDE they can train the system to work specifically on their assessment including pointing the GAI at the resources of most benefit in answering the assessment. The tool was trained and tested with Oxbridge for three separate English questions and returned 99% accuracy. For lecturers and tutors who correct large amounts of scripts in each semester this tool would appear to add substantial value both in time to dedicate to other activities and in the formative feedback provided to students.

Conclusion

In this section we looked at a meta-analysis, a research tool, and a commercial tool, all in the realm of written script assessments assisted by (G)AI. The tools appear to provide valuable utility as they stand. What may have been noticed reading through this section is the absence of reference to academic dishonesty and plagiarisms. These topics were seen as real challenges in the earlier sections of this report and are conspicuous by their absence here. In the next section we take a quick look at academic cheating and dishonesty.

Academic dishonesty and plagiarism

Considering the abilities of GAI to produce credible and hard to distinguish written responses to assessments, one of the most pressing issues discussed in this report thus far is the requirement to maintain academic integrity. The approaches proposed include in-class assessments, oral vivas, and multimodal presentations. A lot of these approaches are based on preventing cheating and academic dishonesty.

“Cheating (academic) is by nature a motivational issue” (Murdock & Anderman, 2006).

In this section we look at the core issue ‘academic dishonesty’ from a psychological perspective to understand how addressing the tendency to cheat in the first place could be seen as another ‘tool’ toward an intelligence augmentation (IA), (Dede, et.al., 2021) approach.

Academic Dishonesty pre-GAI.

Academic dishonesty is not a new phenomenon, there are many varying reports on the extent to which it occurs which in turn are affected by many demographical and situational aspects. A review of the literature in 2002 showed generally, for college students, it is an upward trend, from 23% in 1941 to between 74 and 90% in the ‘90s, (Jensen, Arnett, Feldman, & Cauffman, 2002).

Following a review of both experimental and non-experimental motivation research related to student cheating, Murdock & Anderman, (2006) devised a framework based on motivational reasons students cheat. The framework is based on three guiding motivations to cheat, 1. What is the purpose?, e.g., to gain an award, 2. Can I do this?, e.g. is there a belief that they are not competent enough to complete an assessment. 3. What are the costs?, e.g. how likely is the outcome of being caught. The framework is shown in Fig.1.

In looking at each of these the authors point to possible approaches to change these motivations. For what is the purpose? mastery approaches of assessment seem to reduce cheating over performance approaches, a point also supported by Baran & Jonason, (2020). Looking at Can I do this? Including near-term, proximal goals, is an

approach that is seen to build self-efficacy, an approach which in turn may reduce cheating. For What are the costs? Educators should clearly indicate the seriousness and severity of punishments associated with cheating. However, it is possibly prophetic that the authors caution back in 2006 “An obvious question to be addressed in future research is whether stricter penalties for engagement in cheating behaviours will lead some students to choose more sophisticated and technologically savvy methods of cheating”, (Murdock & Anderman, 2006).

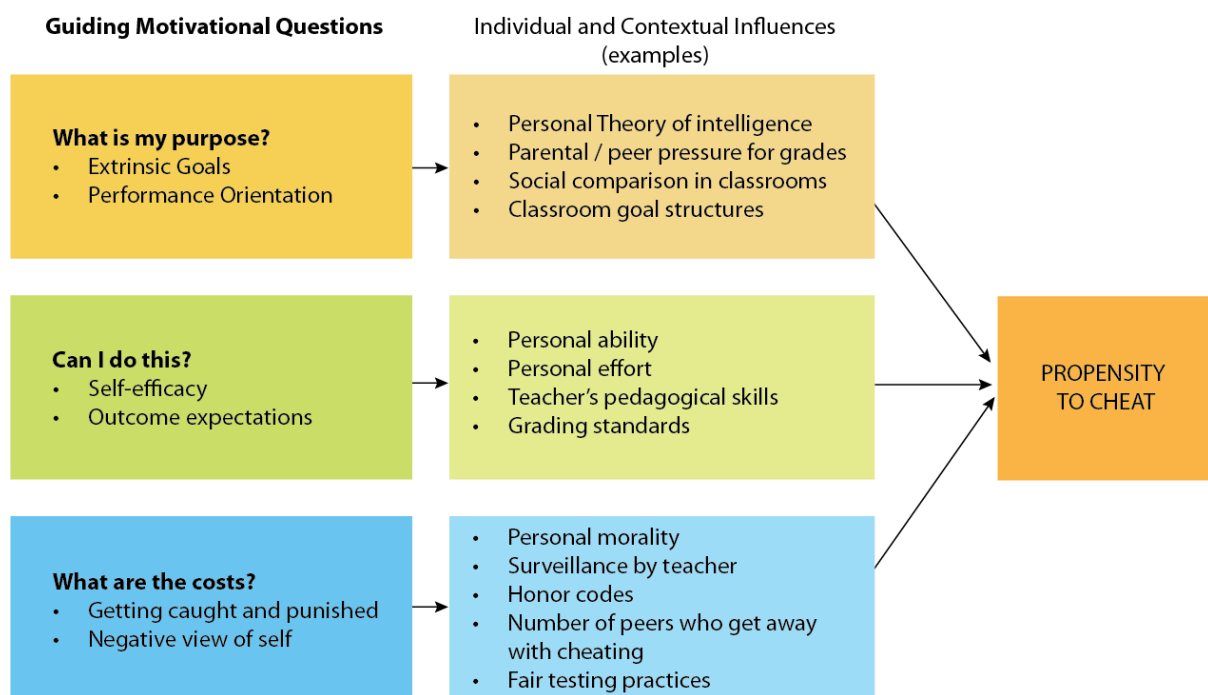


Fig. 1 Proposed motivational framework for integration of the cheating literature (adapted) (Murdock & Anderman, 2006).

Self Determination Theory (SDT) as initially developed by Deci & Ryan has developed to become “one of the most widely researched and applied theories in the field of psychology. Both because of its basic science approach to human motivation and its widely confirmed practical value across multiple domains” (Ryan & Deci, 2019). Self Determination Theory’s view is that we are motivated through ‘need fulfilment’ across the three basic psychological needs for Autonomy, Competence, and Relatedness. In an experiment across two assessments, one group were provided with ‘needs-fulfilment’ approach to the assessment instructions, the other group were given a ‘needs-

frustration' approach where they were instructed exactly what to do, with no choices. The 'needs-frustration' group showed a significantly higher (24.6% and 20.5%) propensity to cheat over those who had their needs supported, (Kanat-Maymon, Benjamin, Stavsky, Shoshani, & Roth, 2015). Most research looks at needs-fulfilment but in taking the opposing view this research shows a link between needs frustration and cheating.

Patall & Leach's, (2015) research lends further support to the concept that choice provision is beneficial in terms of reducing cheating. Experimental assessments looked at the provision of both choice and opportunity to cheat in a 2X2 experiment. Support was shown that even where there was opportunity to cheat the provision of choice mitigated the cheating "because it provides psychological benefits in the form of positive self-beliefs and affect that are equal to those accrued by cheating" (Patall & Leach, 2015).

Academic Dishonesty in the GAI era

As might be expected, there is limited research available related specifically to cheating and academic dishonesty in the current GAI era. Krienert, Walsh, & Cannon, (2022) did survey 1,182 college students in a US university, and like pre-GAI statistics, reported above 71% reported cheating in college assessments. One surprising finding was women (74%) reported significantly more cheating than men (66%). In terms of cheating via technology, 57% reported they had, again women (60%) reported significantly more cheating through technology than men (51%).

The tools for tech cheating reported included earbuds to play answers (hidden under hair), smart watches to send reminders throughout the exam with prompts, and phones potentially accessed during toilet breaks.

These statistics are based on self-report, they do contradict findings in other research, and could be a factor of 'willingness to report'. Either way the figures indicate cheating is a pervasive issue in assessment exams. In addition, Krienert et al., (2022) cite research that indicates a link between academic cheating and later workplace cheating bringing another significant reason to deal with cheating for the students' sake.

A separate and ongoing research project is investigating both educators' and students' views on the Impact of Generative AI on Assessments in Higher Education. To date, the survey has gathered inputs from two universities, one in Australia and one in the US. The authors recognise the rare opportunity to innovate assessment in the GAI era but seek to build sustainable approaches through input from the stakeholders.

Looking at assessment approaches both educators and students were asked to give their preference choice for two assessment approaches. The first was a traditional essay prompt as might appear in a college assignment.

"Write a 5-page essay on [a given topic in your discipline; e.g., Greek mythology, human rights, sustainable energy, sorting algorithms]. You have 7 days to complete the essay."

The second was an 'adapted essay prompt' an approach where educators assume students will leverage GAI and AI.

"You are given a 5-page essay produced by ChatGPT on [a given topic in your discipline; e.g., Greek mythology, human rights, sustainable energy, sorting algorithms]. You have 7 days to analyse the essay and edit it yourself to improve its quality"

The results showed strong support from educators for the "adapted" approach but the support was not so strong from students (Smolansky et al., 2023). Some comments from the students indicate where the reticence comes from, "it kills creativity. You can't ask humans to be the secretary to machines", "[GAI as] a prompt or inspiration but we wouldn't use it directly as a source." (Smolansky et al., 2023). These responses are very heartening in terms of understanding the student view however they are probably in contradiction to a lot of other evidence seen throughout this report.

Conclusion

This section takes a different turn, looking at the potential to address what is a motivation challenge. We do not underestimate the challenge as identified, with the various high cheating rates reported. But from that perspective we must also conclude that cheating has existed and has been prevalent long before the introduction of GAI. Therefore, as we seek to reevaluate assessment in the GAI era there is potential to research approaches to tackle the systemic root causes. The pre-GAI research provides

strong direction, and frameworks built on tested psychological theories, to work with and exciting opportunities to direct Learnovate research projects moving forward.

Closing Remarks

At Learnovate we are keen to further explore the areas of opportunity that the introduction of GAI raises across the approaches to assessment, and addressing the motivational issues that lead to cheating, which appears at least currently to be one of the major concerns. We should not lose sight of the fact that there also exists an opportunity to completely revise what assessment means when you have a GAI available to help you out. We have provided conclusions at the end of each section in this report, so maybe it is fitting that we leave the final words to ChatGPT.

Chat GPT says:

Generative AI in education assessment refers to the utilization of artificial intelligence systems capable of creating or generating content, such as text, images, or even code, to evaluate students' knowledge, skills, and understanding. Here's how generative AI can be applied in education assessment:

1. **Automated Essay Scoring (AES)**: Generative AI can be used to assess students' written responses to essay prompts. By analysing the content, structure, and language of the essays, AI algorithms can assign scores based on predefined criteria, such as coherence, organization, and relevance of arguments. This helps in providing timely feedback to students and reduces the workload of teachers in grading large numbers of essays.
2. **Question Generation**: Generative AI can be employed to create a variety of assessment questions, including multiple-choice, short-answer, and open-ended questions. These questions can be customized to cover specific topics or learning objectives, ensuring a comprehensive assessment of students' knowledge, and

understanding. Additionally, AI-generated questions can adapt to students' proficiency levels, providing personalized assessment experiences.

3. **Programming and Coding Assessment**: In computer science and programming courses, generative AI can be used to evaluate students' coding skills and solutions to programming problems. AI algorithms can analyse the correctness, efficiency, and readability of code written by students, providing feedback on errors, optimizations, and best practices. This enables instructors to assess students' programming abilities at scale and offer targeted support and intervention when needed.

4. **Language Proficiency Assessment**: Generative AI can assess students' proficiency in a second language by generating prompts or scenarios that require written or spoken responses. AI algorithms can analyse the accuracy, fluency, and complexity of students' language use, providing insights into their linguistic abilities and areas for improvement. This is particularly useful for language learning and assessment programs, where automated scoring can expedite the evaluation process and provide immediate feedback to learners.

5. **Creative Assessment Tasks**: Generative AI can be used to design creative and innovative assessment tasks that go beyond traditional formats. For example, AI algorithms can generate multimedia prompts, such as images, audio clips, or videos, that inspire students to demonstrate their understanding and creativity through various forms of expression, such as storytelling, visual art, or multimedia presentations.

Overall, the use of generative AI in education assessment holds promise for enhancing the efficiency, effectiveness, and personalization of assessment practices, enabling educators to gain deeper insights into students' learning progress and providing

students with timely, meaningful feedback to support their learning journey. However, it's essential to ensure that AI-driven assessment systems are transparent, fair, and aligned with educational goals and standards. Additionally, human oversight and interpretation are crucial to validate AI-generated assessments and maintain the integrity and validity of the evaluation process.

References

- Ahmad, N., Murugesan, S., & Kshetri, N. (2023). Generative Artificial Intelligence and the Education Sector. *Computer*, 56(6), 72–76. <https://doi.org/10.1109/MC.2023.3263576>
- Baran, L., & Jonason, P. K. (2020). Academic dishonesty among university students: The roles of the psychopathy, motivation, and self-efficacy. *PLoS ONE*, 15(8 august), 1–12. <https://doi.org/10.1371/journal.pone.0238141>
- Benuyenah, V. (2023). Commentary: ChatGPT use in higher education assessment: Prospects and epistemic threats. *Journal of Research in Innovative Teaching and Learning*, 16(1), 134–135. <https://doi.org/10.1108/JRIT-03-2023-097>
- Chen, J. (2023). Four directions for assessment redesign in the age of generative AI. Retrieved from Times Higher Education website: <https://www.timeshighereducation.com/campus/four-directions-assessment-redesign-age-generative-ai>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, 20(3). <https://doi.org/10.53761/1.20.3.02>
- Dede, C. Etemadi, A., & Forshaw, T. (2021). *Intelligence augmentation: Upskilling humans to And, complement AI*. Retrieved from [https://pz.harvard.edu/sites/default/files/Intelligence Augmentation- Upskilling Humans to Complement AI.pdf](https://pz.harvard.edu/sites/default/files/Intelligence%20Augmentation-%20Upskilling%20Humans%20to%20Complement%20AI.pdf)
- Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education: Artificial Intelligence*, 6(September 2023), 100206. <https://doi.org/10.1016/j.caeai.2024.100206>
- Gardner, D. E., & Giordano, A. N. (2023). The Challenges and Value of Undergraduate Oral Exams in the Physical Chemistry Classroom: A Useful Tool in the Assessment Toolbox. *Journal of Chemical Education*, 100(5), 1705–1709.

<https://doi.org/10.1021/acs.jchemed.3c00011>

GRAIDE. (2023). *AI Assisted Essay grading is hard*.

Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology, 27*(2), 209–228. <https://doi.org/10.1006/ceps.2001.1088>

Kanat-Maymon, Y., Benjamin, M., Stavsky, A., Shoshani, A., & Roth, G. (2015). The role of basic need fulfillment in academic dishonesty: A self-determination theory perspective. *Contemporary Educational Psychology, 43*, 1–9.

<https://doi.org/10.1016/j.cedpsych.2015.08.002>

Koh, E., & Doroudi, S. (2023). Learning, teaching, and assessment with generative artificial intelligence: towards a plateau of productivity. *Learning: Research and Practice, 9*(2), 109–116. <https://doi.org/10.1080/23735082.2023.2264086>

Krienert, J. L., Walsh, J. A., & Cannon, K. D. (2022). Changes in the Tradecraft of Cheating: Technological Advances in Academic Dishonesty. *College Teaching, 70*(3), 309–318. <https://doi.org/10.1080/87567555.2021.1940813>

Lee, J., & Soylu, M. Y. (2023). *ChatGPT and Assessment in Higher Education*. (March), 1–10.

Lodge, Jason, M. (2023). Assessment redesign for generative AI: A taxonomy of options and their viability. Retrieved from <https://www.linkedin.com/pulse/assessment-redesign-generative-ai-taxonomy-options-viability-lodge/>

MacLaren, I., O'Brien, G., Casey, E., Clinch, G., Jackson, N., Lane, B., & Peck, C. (2023). GenAI Guidelines for Educators. *National Academic Integrity Network Group Publications*. Retrieved from [https://www.qqi.ie/sites/default/files/2023-09/NAIN Generative AI Guidelines for Educators 2023.pdf](https://www.qqi.ie/sites/default/files/2023-09/NAIN%20Generative%20AI%20Guidelines%20for%20Educators%202023.pdf)

Memarian, B., & Doleck, T. (2023). ChatGPT in education: Methods, potentials, and limitations. *Computers in Human Behavior: Artificial Humans, 1*(2), 100022. <https://doi.org/10.1016/j.chbah.2023.100022>

Mintz, J., Holmes, W., Liu, L., & Perez-Ortiz, M. (2023). Artificial Intelligence and K-12 Education: Possibilities, Pedagogies and Risks. *Computers in the Schools, 40*(4), 325–333. <https://doi.org/10.1080/07380569.2023.2279870>

Monash University. (2023). Designing assessment for achievement and demonstration of

- learning outcomes. Retrieved from <https://www.monash.edu/learning-teaching/teachhq/Teaching-practices/artificial-intelligence/ai-and-assessment>
- Monash University. (2024). Assessment policy and process. Retrieved from <https://www.monash.edu/learning-teaching/teachhq/Teaching-practices/artificial-intelligence/assessment-policy-and-process>
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5(July), 100151. <https://doi.org/10.1016/j.caeo.2023.100151>
- Murdock, T. B., & Anderman, E. M. (2006). Motivational perspectives on student cheating: Toward an integrated model of academic dishonesty. *Educational Psychologist*, 41(3), 129–145. https://doi.org/10.1207/s15326985ep4103_1
- Nguyen Thanh, B., Vo, D. T. H., Nguyen Nhat, M., Pham, T. T. T., Thai Trung, H., & Ha Xuan, S. (2023). Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, 39(5), 59–81. <https://doi.org/10.14742/ajet.8902>
- Patall, E. A., & Leach, J. K. (2015). The role of choice provision in academic dishonesty. *Contemporary Educational Psychology*, 42, 97–110. <https://doi.org/10.1016/j.cedpsych.2015.06.004>
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2). <https://doi.org/10.53761/1.20.02.07>
- Practice, T., & Bloom, R. (2008). A Revision of Bloom ' s Taxonomy : An Overview David R . Krathwohl. *ReVision*, 41(4), 212–218.
- Ryan, R. M., & Deci, E. L. (2019). Brick by brick: The origins, development, and future of self-determination theory. In *Advances in motivation science* (Vol. 6, pp. 111–156). Elsevier.
- Sanabria-Z, J., Alfaro-Ponce, B., Argüelles-Cruz, A., & Ramírez-Montoya, M. S. (2023). AI-Based Platform Design for Complex Thinking Assessment: A Case Study of an Ideathon Using the Transition Design Approach. *Computers in the Schools*, 40(4), 391–411. <https://doi.org/10.1080/07380569.2023.2256711>
- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator

and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. *L@S 2023 - Proceedings of the 10th ACM Conference on Learning @ Scale*, 378–382. <https://doi.org/10.1145/3573051.3596191>

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1).

<https://doi.org/10.1186/s40561-023-00237-x>

Tobler, S. (2024). Smart grading: A generative AI-based tool for knowledge-grounded answer evaluation in educational assessments. *MethodsX*, 12(December 2023), 102531.

<https://doi.org/10.1016/j.mex.2023.102531>

TopAI.tools. (2024). 70 Best Automated essay grading software AI tools – 2024. Retrieved from <https://topai.tools/s/Automated-essay-grading-software>

UCL. (2023). A three-tiered categorisation of AI use in an assessment for staff to guide students on expectations. Retrieved from <https://www.ucl.ac.uk/teaching-learning/generative-ai-hub/using-ai-tools-assessment# AI tools cannot be used in this assessment>