Preprints.org

Article

Participatory Co-Design and Evaluation of a Novel Approach to Generative AI-Integrated Coursework Assessment in Higher Education

Alex F Martin [*] , Svitlana Tubaltseva , Anja Harrison , G James Rubin

Article

Participatory Co-Design and Evaluation of a Novel Approach to Generative AI-Integrated Coursework Assessment in Higher Education

Alex F Martin [1,*], Svitlana Tubaltseva [2], Anja Harrison [1] and G James Rubin [1]

[1] King's College London, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, London, UK

[2] Richmond American University London, School of Liberal Arts, London, UK

* Correspondence: alex.f.martin@kcl.ac.uk

Abstract: Generative AI tools offer opportunities for enhancing learning and assessment but raise concerns about equity, academic integrity, and the ability to critically engage with AI-generated content. This study explored these issues within a psychology-oriented postgraduate programme at a UK university. We co-designed and evaluated a novel AI-integrated assessment aimed at improving critical AI literacy among students and teaching staff (pre-registration: https://osf.io/jqpce). Students were randomly allocated to two groups: the 'honest' group used AI tools to assist with writing a blog and critically reflected on the outputs, while the 'cheating' group had free rein to use AI to produce the assessment. Teaching staff, blinded to group allocation, marked the blogs using an adapted rubric. Focus groups, interviews, and workshops were conducted to assess the feasibility, acceptability, and perceived integrity of the approach. Findings suggest that, when carefully scaffolded, integrating AI into assessment can promote both technical fluency and ethical reflection. Students engaged critically with AI outputs, while staff recognised the potential to support critical thinking and maintain academic standards. The approach also supports growing calls for authentic assessment that mirrors real-world professional tasks. However, tensions between preserving academic integrity and supporting skill development must continue to be addressed in future work.

Keywords: generative artificial intelligence; pedagogy; postsecondary; curriculum design; student learning; qualitative

1. Introduction

Recent advances in generative artificial intelligence (AI), powered by large language models, present opportunities and challenges for assessment in higher education. AI is now widely used across sectors including health, industry, and research (McKinsey, 2024; Sun et al., 2023), and is permanently reshaping the nature of academic tasks. In educational settings, AI has already shown potential to support learning by providing personalised feedback, scaffolding writing processes, and automating routine tasks (Kasneci et al., 2023). However, evidence on how best to use it in teaching, especially assessment design, remains limited.

Initial studies suggest that while students may benefit from AI-enhanced feedback, overreliance on these tools may undermine opportunities for deep learning and critical engagement (Kasneci et al., 2023; Zawacki-Richter et al., 2019). The integration of generative AI in education also presents challenges. Equity concerns persist, including unequal access to reliable AI tools and the digital skills needed to use them meaningfully (UNESCO, 2024). Academic integrity is also at risk, as AI can be used to cheat in ways that evade detection (Yusuf et al., 2024). Moreover, the use of AI complicates traditional concepts of authorship and originality, raising questions about what constitutes independent academic work. There are also concerns that critical thinking, a key goal of higher education, could be weakened if students accept AI outputs without careful evaluation (Bittle & El-Gayar, 2025).

In response, there is growing recognition of the need to build critical AI literacy among students and staff. This means not just knowing how to use AI tools, but understanding how they work, the wider impacts they have, and how to assess AI-generated content carefully and ethically (Abdelghani et al., 2023). Developing critical AI literacy is needed to prepare students to be thoughtful, responsible users of AI, and should be built into teaching and assessment strategies.

We aimed to improve the critical AI literacy of postgraduate students and teaching staff through the co-design and evaluation of an AI-integrated written coursework assessment. In this assessment, students used generative AI tools to draft a blog critically summarising an empirical research article and produced a reflective, critical commentary on the AI-generated content.

Specifically, we sought to:

1. Use participatory methods to design and evaluate a novel assessment approach.
2. Assess its acceptability and feasibility for students and teaching staff.
3. Test whether teaching staff could distinguish between assessments completed in accordance with the brief and those generated entirely by AI.

Findings were used to develop practical guidance to support the implementation of the assessment within the postgraduate programme and contribute to the wider pedagogical literature on assessment in higher education.

2. Materials and Methods

This study used a participatory evaluation approach. Participatory evaluation involves contributors not just as participants, but as co-designers and co-evaluators (Fetterman et al., 2017), and has been used previously to explore AI-related resources and curriculum development (Cousin, 2006; Teodorowski et al., 2023). A strength of this approach is its emphasis on different forms of expertise, including lived experience, disciplinary knowledge, and teaching practice, which contribute to the development of assessments that are both grounded and relevant.

The protocol is available at OSF Registries: doi.org/10.17605/OSF.IO/JQPCE. We used the Guidance for Reporting Involvement of Patients and the Public short form (GRIPP2-SF) checklist to report involvement in the study (reported in Table A1 in the Appendix#;(Staniszewska et al., 2017). The study was approved by the Research Ethics Panel of King's College London (LRS/DP-23/24-42387; 27 June 2024).

The study took place within a psychology-based postgraduate programme in the United Kingdom. Participants brought a range of expertise, from digital literacy and academic writing to assessment design. By drawing on the knowledge and experience of those directly involved in teaching and in learning, the participatory methods promoted shared ownership and practical relevance, while still allowing space for innovation.

Twelve participants took part in the study, which included eight students and four members of the teaching team from the same academic cohort (2023–24) of a postgraduate course at the Institute of Psychiatry, Psychology and Neuroscience, King's College London, a Russell Group university in the United Kingdom. The teaching staff included a Teaching Fellow, a Lecturer, and two Research Associates. All participants had recently completed or marked a summative assessment within the course.

2.1. Stage 1: AI integrated Assessment

The research team collaborated with other members of the course teaching team to adapt an existing summative assessment already embedded in the curriculum. This assessment required students to write a blog post summarising and critically appraising an empirical research article on mental health. Framed as an authentic assessment, the task included the potential for selected blogs to be published on science communication platforms.

We used the Transforming Assessment in Higher Education framework developed by AdvanceHE to guide our approach to integrating generative AI tools into this assessment (Healey & Healey, 2019). The framework highlights the need for assessments that are authentic, inclusive, and aligned with learning outcomes, emphasising the importance of involving students in the assessment development process. This emphasis aligned with our approach of integrating AI tools to reflect real-world practices and to develop critical AI literacy.

Under the revised assessment approach, students were asked to use two AI tools to assist with drafting a blog based on an empirical article. The written assessment consisted of three components:

1. Two AI-generated blog drafts, using two AI tools.
2. A final blog that combined the strongest elements of the AI outputs with the student's own revisions and original contributions, assessed for the accurate and critical appraisal of the empirical article.
3. A commentary critically reflecting on the AI-generated content and explaining the rationale for revisions made, assessed for the depth of critical and ethical reflection.

The marking matrix was revised to retain the use of a standard critical appraisal checklist for assessing students' understanding of the empirical article, alongside the programme-wide marking framework (stage 2). New criteria were introduced to evaluate students' critical engagement with AI-generated content (stage 3). The adapted format built on the existing learning outcome of critically appraising empirical research, extending it to assess students' ability to reflect on the role of AI in academic work, apply subject knowledge to evaluate AI outputs, and make informed editorial decisions.

2.2. Stage 2: Assessment Trial

All participants were invited to take part in a trial of the adapted assessment. They first attended a workshop designed to support students in their AI-assisted assessment. Microsoft Copilot, in both balanced and precise modes, was the mandated generative AI tool, selected for its free availability for the participants (ensuring equitable access) and to allow direct comparison between model outputs. While Copilot was used in this instance, the assessment was designed to be transferable to other AI tools.

The workshop was delivered in four stages. The first introduced Copilot's core functions, including its strengths, limitations, and examples of effective prompt writing. In the second stage, students practised drafting prompts and used the AI models to generate and revise a mock blog post. The final two stages drew on Gibbs' Reflective Cycle to guide structured learning (Gibbs, 1988). In stage three, students critically appraised an AI-generated blog, and compared the outputs produced by the two Copilot modes. This exercise supported a deeper understanding and analysis of AI-generated content. In the final stage, students reflected on their use of AI and developed an action plan for how they would apply AI tools in future academic work. This reflection aimed to consolidate learning and promote ethical, informed use of generative AI tools.

Feedback on the workshop was collected through qualitative discussions at the end of the session and a short survey. The survey included a Likert-scale question assessing whether the workshop would help students complete the assessment (responses: yes, somewhat, no) and two free-text questions: "What did you learn from the workshop?" and "What was missing from the workshop that would help you feel more prepared for the pilot assessment?"

Student participants were then randomly allocated to one of two groups. Those in the 'honest' group were instructed to follow the coursework brief precisely, using the designated AI tools as directed. Students in the 'cheat' group were given freedom to complete the assessment by any means, including generating the entire submission using AI tools. They were encouraged to be creative and to push the boundaries of the process. Teaching staff participants were asked to mark the submitted assessments and provide written student feedback using the adapted marking matrix. They also indicated whether they believed the student had completed the assessment as instructed ('honest' group) or had been in the unrestricted ('cheat') condition.

2.3. Stage 3: Evaluation

Participants participated in an iterative process of reviewing and refining the assessment materials, including the workshop content, coursework brief, and marking matrix.

To explore the feasibility, acceptability, and perceived integrity of the AI-integrated assessment approach, we conducted a series of semi-structured focus groups with students and individual interviews with teaching staff. This format was chosen to accommodate participant preferences and availability.

The discussion guides are reported in Appendix 2. They were developed to address the study's research aims and to capture experiences across both groups regarding their engagement with

generative AI in the context of assessment. Both focus groups and interviews lasted approximately 45 to 60 minutes and were structured in two parts: the first explored existing knowledge of generative AI and experiences of completing or marking the assessment; the second addressed reflections on the assessment design and its potential for future implementation. In addition, we explored perceptions of 'cheating' in the assessment, including whether students in the 'honest' and 'cheating' groups felt they had met the intended learning outcomes, and whether staff felt able to distinguish between the two. Particular attention was paid to whether the approach supported intended learning outcomes and provided a fair measure of student performance.

We also asked questions about the initial training workshop as part of the interviews. This feedback was reviewed alongside data from the survey questions completed by participants after the workshop and was used to revise and improve the training content.

Focus groups and interviews were conducted via Microsoft Teams. Thematic analysis was led by one researcher (AFM), following Braun and Clarke's (2006) approach, including: familiarisation with the data, initial coding, theme identification, and iterative theme refinement (Braun & Clarke, 2006). Analysis was performed separately for students and teaching staff. Emerging themes were reviewed and refined through discussion within the research team and with participants who participated in subsequent workshops.

We held co-design workshops with students and teaching staff to further refine the assessment brief and marking matrix, respectively. The think aloud technique was used (Charters, 2003; Someren et al., 1994), whereby each section of the assessment materials was reviewed in turn. Participants took part in a facilitated group discussion, voicing their thoughts, suggestions, and reactions in real time as they engaged with the materials. Data saturation was considered to have occurred when no further substantial changes were proposed by the participants. Two workshops were held with students and one with teaching staff, which likely reflects the fact that more extensive feedback had already been gathered from teaching staff during earlier individual interviews and incorporated into the materials prior to the workshops.

Feedback gathered during these sessions was used to inform revisions to the assessment materials. We documented this process using a Table of Changes (ToC) from the Person-Based Approach (Bradbury et al., 2018). We also used a Custom GPT, which allows creation of a personalised version of ChatGPT tailored to specific tasks or knowledge domains, to review the final materials for accessibility and readability.

3. Results

3.1. Assessment Materials and Learning Outcomes

The modifications made to the assessment materials are summarised in the ToC (Table 1).

Feedback on the co-designed assessment materials produced by the Custom GPT indicated that, while both the assessment brief and marking matrix were generally well-structured and aligned with learning outcomes, several refinements could improve readability and accessibility. These included ensuring consistency of language and tone, using bullet points and clearer formatting to support navigation, and clarifying instructions around AI tool use and submission structure. Minor revisions were recommended to the learning outcomes and the reflection criteria to enhance alignment with marking expectations.

The final versions of the assessment materials (workshop proformas, assessment brief and marking matrix) and the amendments recommended by ChatGPT are included in Appendix C and at the Open Science Framework project: osf.io/ctewk/.

Table 1. Table of Changes: Summary of proposed changes to assessment materials with feedback sources and prioritisation.

Assessment Component	Feedback Source	Feedback	Proposed Change	Rationale	Agreed Change	MoSCoW
Assessment Brief	Staff co-design workshop	Instructions lacked clarity on appraising both the original article and the AI's interpretation	Have a separate section that states expectations for appraising the empirical article.	Ensures students understand and address both components of the assessment	1. Clarification added under Instructions section 2. Critical appraisal learning outcome split into three exclusive sections 3. Additional content added at top of Tips section	Must

					describing the need to critically understand the empirical article	
Assessment Brief	Student co-design workshop	Instructions for the second part of the assessment (AI reflection) somewhat unclear	Add more detailed guidance and reflective models/frameworks	Helps students structure their critique and understand expectations	4. Reflective models are recommended under Instructions section	Must
Assessment Brief	Student co-design workshop	Word count guidance inconsistent	State word count for each section and total limit clearly	Reduces confusion and supports appropriate planning	5. Wordcounts are provided for the total and for each part of the report under Instructions section	Should
Assessment Brief	Student co-design workshop	Referencing format unclear	Specify referencing style (e.g., APA 7) in blog guidance	Promotes consistency and addresses AI limitations in citation generation	6. Clarity of referencing style added to Blog Guidance section	Must
Assessment Brief	Student co-design workshop	Background section was appreciated but some found it overwhelming	Use bullet points instead of paragraphs	Improves accessibility and reduces cognitive load	7. The background and relevance section has been streamlined and uses bullet points to break down heavier text	Could
Assessment Brief	Student co-design workshop	Value of tips section highlighted	Retain and expand guidance on recognising AI hallucinations and faults	Reinforces critical AI literacy and practical assessment skills	8. Added to the learning outcomes 9. Added to 'risks' in the Background section 10. Expanded the Tips section	Should
Assessment Brief	Student co-design workshop	Students appreciated the learning outcomes, but overlooked the importance of critical appraisal	Emphasise critical appraisal more clearly throughout the brief and link it directly to learning outcomes	Reinforces key educational focus and clarifies expectations	11. Critical appraisal learning outcome split into three exclusive sections	Must
Assessment Brief	Student co-design workshop	Understanding the purpose improved engagement	Highlight the brief rationale or real-world relevance in the assessment introduction	Enhances motivation and situates learning in context	12. Extended the 'why blogs' and 'why generative AI' sections of the Background section	Could
Assessment Brief	Staff interviews	Complexity of articles may limit scope for critical appraisal	Use more complex articles with intentional limitations or errors that are not highlighted by the study authors	Allows students to demonstrate deeper critical thinking and analytical skills	13. Added to the Instructions for coursework section	Could
Assessment Brief	Staff co-design workshop	Students may not fully demonstrate critical engagement with AI-generated content	Require students to highlight changes made to AI-generated content and include in an appendix	Encourages transparency and supports assessment of student input	14. Added to the Instructions for coursework section	Should
Marking Matrix	Staff co-design workshop	Difficulty in marking AI vs. student input	Integrate AI evaluation criteria into the main marking matrix	Reduces marker cognitive load and reflects integrated learning outcomes	15. Out of scope as it addresses the programme-wide marking matrix	Won't
Marking Matrix	Staff co-design workshop	Lack of marker guidance for grade boundaries	Provide examples of high and low-quality work	Improves marker confidence and consistency	16. Included examples from high and grade bounds	Should
Marking Matrix	Staff interviews	Time consuming to check against academic articles	Provide checklist for articles of key methods, results, and conclusions	Improves marker confidence and consistency	17. Included checklist for empirical articles	Should
Training Workshop	Student co-design workshop	Ethical use of AI is not well understood	Include declaration forms, videos, and exemplar prompts	Supports transparency and	18. Added an interactive brainstorming activity on the ethical use of AI,	Should

		in the training materials	encourages ethical practice	tailored to the specific context of the assessment. 19. Incorporated group breakout discussions to support peer learning and reflective engagement.		
Training Workshop	Student co-design workshop	Students need more practical support in using AI	Offer two workshops at different levels across the programme	Caters to varying levels of familiarity and ensures accessible skill development	20. Revised workshop structure to offer two scaffolded sessions: - *Foundations 1*: Covers ethics and core AI concepts, with optional homework to reinforce learning. 21. *Foundations 2*: Focuses on practical, assessment-based activities, including a review of the King's AI coversheet.	Could
Training Workshop	Student co-design workshop	Students requested more support with AI prompting	Include dedicated time in the workshop focused on writing effective AI prompts	Builds critical AI literacy and confidence in tool use	22. Integrated into revised two-part workshop design, with optional homework to support independent learning.	Should
Training Workshop	Staff co-design workshop	Markers need guidance on evaluating AI-assisted work	Provide training sessions specifically for staff assessing AI-integrated submissions	Supports consistency and confidence in marking across staff teams, and supports new format adoption	23. Marked as a priority for inclusion before the next phase of the trial implementation.	Won't
Training Workshop	Training session survey	Limited prior understanding of how to use AI; prompting guidance was especially valuable	Expand workshop content on prompt engineering, including examples and guided practice sessions	Builds foundational skills for effective AI use and supports equitable engagement with the assessment format	24. Included in the scaffolded workshop structure, aligned with core learning outcomes and feedback priorities.	Could

Feedback on the learning outcomes was generally positive or neutral, with no negative responses offered. Students and teaching staff appreciated the inclusion of learning outcomes and found them helpful for understanding the purpose of the assessment. Some suggested making the link between the learning outcomes and the specific assessment tasks more explicit, to improve alignment and clarify expectations. The major change that emerged from all feedback sources was the need to communicate that critical appraisal of the original empirical article is as important as appraisal of the ability of AI to generate seemingly useful content. One student noted that engaging with the AI output highlighted inaccuracies, such as fabricated participant details, which prompted them to critically verify the content against the original source. This process, while demanding, was seen as intellectually valuable: "It forces you to actually figure out whether you're critically appraising the critical appraisal." Another contributor reflected on the need to distinguish between assessing AI literacy and assessing critical thinking, suggesting that the learning objectives should clearly indicate which of these skills is being prioritised. This feedback informed revisions to the assessment brief and the revised learning outcomes.

Our revised learning outcomes became:

1. **Critical appraisal**: Students will demonstrate the ability to critically appraise academic content by:
    i. Evaluating an empirical research article using an established critical appraisal checklist.
    ii. Assessing the accuracy, relevance, and limitations of AI-generated content in relation to the original empirical article.
    iii. Comparing outputs from different AI tools, identifying their strengths and weaknesses in academic content generation.

2. **Generative AI literacy**: Students will develop foundational AI literacy by using generative tools to support scientific blog writing. They will demonstrate an understanding of AI capabilities and limitations, including the ability to identify common errors such as fabrication or hallucination.
3. **Editorial and reflective judgment**: Students will apply editorial judgement to revise AI-generated content, integrating critical analysis and original insight. They will reflect on their use of AI tools and articulate the rationale for content modifications in alignment with accuracy, academic standards, and ethical considerations.

3.2. Feasibility, Acceptability, and Integrity

Table 2 presents summaries of the key findings and illustrative quotes from the thematic analysis of the focus groups and interviews.

Student feedback highlighted that while AI tools could streamline aspects of the writing process, they did not reduce workload due to the effort required to refine outputs. Perceptions of feasibility, acceptability, and integrity varied, with students valuing the opportunity to build critical thinking skills, but also expressing concerns about fairness, skill development, and notably, ownership of their work. Some viewed equitable access and thoughtful integration of AI to be particularly important for maintaining academic standards. Teaching staff found the assessment structure clear, though marking was initially time-intensive because of the dual task of evaluating both AI and student contributions. Efficiency improved with familiarity, and staff recognised the assessment's potential to support critical engagement. While challenges remained in distinguishing AI-generated from student-authored content, most staff endorsed transparent and pedagogically grounded use of AI in academic settings.

Students in the cheat group found that using AI to complete the entire assessment was challenging, with outputs, particularly the reflective commentary, requiring substantial oversight and correction. Some spent a similar amount of time on the task as those in the honest group, while others felt they used somewhat less. Most felt they had achieved the intended learning outcomes due to the time spent checking, appraising, and reflecting on the AI-generated content.

Assessment marks ranged from 35 (fail) to 78 (distinction). For most assessments, marks from different markers fell within a ten-point range, but for one assessment scores ranged more widely (58 to 78). Markers were not significantly better at correctly identifying students in the 'honest' group (6/14 accurately identified) compared to those in the 'cheat' group (3/6 accurately identified; $X^2(1, N = 20) = 0.09$, $p = .769$). Markers' views on identifying students in the cheat condition were polarised: some reported having no clear sense, while others felt very confident that they could recognise AI-generated submissions. However, these subjective impressions were not reflected in their actual ability to accurately distinguish between the groups.

Table 2. Key findings and illustrative quotes from students and teaching staff.

Theme	Key finding	Contributor quotes
Students		
Feasibility: efficiency vs effort	Many students found that generative AI tools streamlined the writing process, but did not necessarily change the overall workload, particularly when refining AI outputs.	"[the assessment] took about the same amount of time that it would normally ... that that was a big shock to me!" "It forces you to figure out whether you're critically appraising the critical appraisal, but you think it's given. So it is a bit, it is more work in that sense."
Feasibility: user variability	Students highlighted varying levels of success when using AI, depending on their AI literacy and academic skills.	"I just couldn't get it to be any longer, no matter how many times I kind of prompt it and same with references, I couldn't get it to do and so that's probably user error." "I actually started off using AI, but then I found that it was giving very predictable answers and very basic answers … so I then added a lot to that and like kind of brought in my own experience into it a bit more."

Acceptability: learning enhancement	Some students saw the integration of AI as a valuable tool for building critical thinking and learning how to appraise content more deeply.	"It would improve students' critical thinking abilities if they don't have to focus as much on things like grammar … they can focus more on the topics at hand, gaps in research, and they can understand the structure of scientific papers a bit easier." "It can help them do the basic task, learn from it, but at the same time still require them to give some critical viewpoints, which again, it can also help kind of see what other papers what other people might be doing and kind of get inspiration from that. So I only see positivity."
Acceptability: skill development	Others raised concerns that reliance on AI might compromise the development of core academic skills.	"It would be sad to lose [the skill of writing] an essay with good grammar … losing some of the core skills of sort of being a student at university" "Are we learning anything, or are we just asking [AI]?"
Integrity and fairness: ethics and ownership	Some students expressed discomfort about using AI-generated content, especially when it felt detached from their own intellectual effort.	"It kind of alienated me from the work … I had to keep going back to be like, did I say this?" "I wasn't exactly tired after I'd finished it because it did a lot of the work, but like, it was like, I don't read it and feel like any kind of pride or any kind of like ownership of it in any way."
Integrity and fairness: equity and academic standards	Some responses reflected concerns around integrity and fairness of using AI in assessment generally.	"It's going to start feeling like we're being assessed on how well we can use AI." "It doesn't make me any more tempted to use it in university assignments, because I feel like I'd come out of it with a degree that I didn't really feel like was mine."
Integrity and fairness: equity and academic standards	Whilst some were more optimistic about AI in the mock assessment.	'The problem happens when some people in the classroom are relying on AI, some people are not. So then it creates a kind of disparity. But if everyone is using it then the kind of standard changes and I would feel that would be the most fair updater of ability because then everyone has the same resource.' "I did feel tempted to cheat, but I then thought it wouldn't do a good enough job 'cause I think it's done its part. It's written two different styles and I'm like OK, I can see that. But I think to have that bit of a meta way of thinking about it and to put things together."
Teaching staff		
Feasibility: complexity of marking	Most staff reported that while the structure was clear, marking took longer due to the added components and the unfamiliar task of evaluating AI use.	"The first one took me two hours … I had to read the paper so that I wasn't taking into account how long it takes." "You have a blog, and a commentary, it doubles the work."
Feasibility: familiarity improved efficiency	Once familiar with the marking expectations, staff found the process more manageable.	"Afterwards, it took about an hour…. once I figured out what I was doing." "I take more time in the first in the first or second, but then generally the blog part it was quite easy to mark."

Acceptability: critical thinking	Staff saw potential in the assessment for encouraging deeper student engagement and critical evaluation.	"Incorporating AI into assessments is interesting, because … you still need to think and you still need to produce something." "I think that's good for some students. The work I read [was] really interesting. Some students did read well and add their critics. I'm really impressed about … how they learn from that."
Acceptability: undermining learning	There were worries that integration of AI might limit the development of core academic skills.	"You ask ChatGPT, it removes the painful experience [of writing well], but I think … we need that struggle." "If we just rely on chat GPT writing things for us and then thinking for us, we will lose the ability to be creative ourselves, because I feel like you learn by doing."
Integrity and fairness: evaluation	Staff expressed uncertainty about distinguishing between student-authored and AI-generated content in more formulaic sections but many were confident they could identify inappropriate use of AI in the assessment.	"AI has done a pretty good job describing the methodology." "You can tell when you read, some sections of the blog that can also be impersonal, like the methods and results … The personal perspective that's easier to judge because you can kind of sense the touch human touch." "There's no critical thinking because obviously the blog has been written by AI, so there's automatically no critical thinking there, and they didn't particularly try to add more on their own behalf."
Integrity and fairness: integration	Despite concerns, staff generally recognised the inevitability of AI in education and advocated for proactive adaptation.	"It helps people understand and think critically about [AI] use." "We might as well embrace it and … teach people how to use it properly."

4. Discussion

The findings from this study add to a nascent body of literature that highlights the dual role of AI-integrated assessments, as tools for digital literacy and as mechanisms for reflective, critical pedagogy. The blog format provided a unique opportunity for students to practise public-facing, accessible academic writing, aligning with real-world expectations in health communication. Pilot findings showed that students found the approach feasible and helpful for developing critical skills, although engaging with AI outputs was perceived to increase workload. Teaching staff initially found marking more demanding and had limited success distinguishing AI-generated content, but valued the assessment's potential to promote ethical and critical AI use.

A key success of the project was the development of students' critical AI literacy, with findings suggesting that the blog assessment promoted active engagement with AI outputs. Students were required not only to use AI tools but to critique their outputs, identify inaccuracies, and justify their editorial decisions. This process encouraged deeper critical engagement and helped students to view AI as a tool requiring human oversight rather than as a source of ready-made answers. The requirement to compare outputs across different AI models also supported the development of critical evaluation skills, as students reflected on the variability and limitations of AI-generated content. Importantly, these findings address concerns raised during the qualitative evaluation, and reflected issues highlighted in previous research, that overreliance on AI could undermine opportunities for deep learning and reflective practice (Kasneci et al., 2023; Zawacki-Richter et al., 2019). These findings align with recommendations that AI in education must go beyond functional skills to include AI literacy, as well as active learning skills and metacognition (Abdelghani et al., 2023).

Beyond promoting critical engagement with AI outputs, the study also highlighted strategies for maintaining assessment integrity and supporting academic skill development. Teaching staff expressed concerns that AI use could make it harder to distinguish original work from AI-generated content. This echoes broader challenges in the literature, where AI use may complicate traditional

definitions of authorship and independent academic work (Yusuf et al., 2024). Although markers were sometimes confident, their accuracy in identifying AI-reliant submissions from the 'cheat group' was poor. This is likely to be because students in the cheat group generally described a similar editorial process to those in the honest group. Nevertheless, the assessment's structure, requiring critical appraisal of the empirical article, critique of AI outputs, evidence of revision, and transparency may have helped to mitigate these risks, although this needs further testing.

By embedding critical evaluation and editorial judgement, the assessment addressed concerns that AI could weaken core academic skills such as critical thinking and reflective analysis (Bittle & El-Gayar, 2025). Students also recognised that genuine engagement, not uncritical acceptance of AI outputs, was needed to meet the learning outcomes. However, the extent to which students internalised critical evaluation versus simply complying with task requirements remains unclear. Future studies could explore students' metacognitive strategies and critical reasoning during AI use through longitudinal or think-aloud methodologies (Charters, 2003; Someren et al., 1994). Overall, the findings suggest that carefully designed AI-integrated assessments can uphold academic integrity while supporting the development of essential academic competencies.

Involving students and teaching staff in the co-design and evaluation process was central to developing an assessment that was authentic, feasible, and acceptable. The participatory approach drew on academic, pedagogical, and lived experience to shape the teaching workshop and assessment materials, helping us spot practical challenges early and promote shared ownership of the development of the assessment (Fetterman et al., 2017; Teodorowski et al., 2023). This aligns with broader calls for more inclusive, responsive, and transparent innovation in educational assessment (Bovill et al., 2016; Healey & Healey, 2019). Future research should continue to embed participatory evaluation to ensure AI-integrated assessment remains student-centred and pedagogically sound.

Several limitations of the study should be acknowledged. First, students were not involved in the initial design phase of the assessment, falling short of authentic co-production (Cook-Sather et al., 2014). Although this was partly mitigated through later participatory evaluation, involving students earlier could have strengthened the creativity, relevance, and ethical responsiveness of the assessment. Second, qualitative feedback was collected following the initial pilot rather than after a full module-wide rollout. As such, findings may reflect early impressions rather than longer-term engagement. However, this timing allowed immediate adjustments and iterative revisions of the assessment materials based on student and staff experiences. Third, the study was conducted within a single institutional setting, which may limit generalisability to other universities or international contexts with different AI access, policies, and pedagogical cultures. However, the participatory methods used to co-design and co-evaluate the assessment are transferable and could be adapted to different educational settings to ensure relevance to local needs. Moreover, the assessment design mirrors real-world professional scenarios involving AI, making it relevant to clinical, research, and public health practice, and supporting the wider application of the skills developed.

5. Conclusions

This study explored the participatory development and evaluation of a generative AI-integrated assessment in postgraduate education. Findings suggest that integrating AI into assessment can promote both technical fluency and ethical reflection when scaffolded appropriately. Students engaged critically with AI outputs, while teaching staff recognised the potential for supporting critical thinking and maintaining academic integrity. Our approach supports growing calls for authentic assessment that mirrors real-world tasks, particularly in professions where AI is becoming more common. However, there remains a tension between preserving academic integrity and using AI to support skill development. Future iterations must continue to navigate this balance carefully, ensuring that critical engagement and ethical practice are at the core of AI-integrated learning.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Abbreviations

The following abbreviations are used in this manuscript:

AI   Generative Artificial Intelligence

References

1.   Abdelghani, R., Sauzéon, H., & Oudeyer, P.-Y. (2023). Generative AI in the Classroom: Can Students Remain Active Learners? arXiv preprint arXiv:2310.03192.

2.   Bittle, K., & El-Gayar, O. (2025). Generative AI and Academic Integrity in Higher Education: A Systematic Review and Research Agenda. Information, 16(4), 296.

3.   Bovill, C., Cook-Sather, A., Felten, P., Millard, L., & Moore-Cherry, N. (2016). Addressing potential challenges in co-creating learning and teaching: Overcoming resistance, navigating institutional norms and ensuring inclusivity in student–staff partnerships. Higher Education, 71, 195-208.

4.   Bradbury, K., Morton, K., Band, R., van Woezik, A., Grist, R., McManus, R. J., Little, P., & Yardley, L. (2018). Using the person-based approach to optimise a digital intervention for the management of hypertension. PLoS One, 13(5), e0196868.

5.   Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

6.   Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. Brock Education Journal, 12(2).

7.   Cook-Sather, A., Bovill, C., & Felten, P. (2014). Engaging students as partners in learning and teaching: A guide for faculty. John Wiley & Sons.

8.   Cousin, G. (2006). An introduction to threshold concepts. In (Vol. 17, pp. 4-5): Taylor & Francis.

9.   Fetterman, D. M., Rodríguez-Campos, L., & Zukoski, A. P. (2017). Collaborative, participatory, and empowerment evaluation: Stakeholder involvement approaches. Guilford Publications.

10.   Gibbs, G. (1988). Learning by doing: A guide to teaching and learning methods. Further Education Unit.

11.   Healey, M., & Healey, R. L. (2019). Student engagement through partnership: A guide and update to the AdvanceHE framework. Advance HE, 12, 1-15.

12.   Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.

13.   McKinsey. (2024). The State of AI in Early 2024: Gen AI Adoption Spi-kes and Starts to Generate Value. Quantum Black. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

14.   Someren, M. W. v., Barnard, Y. F., & Sandberg, J. (1994). The think aloud method: A practical guide to modelling cognitive processes. (No Title).

15.   Staniszewska, S., Brett, J., Simera, I., Seers, K., Mockford, C., Goodlad, S., Altman, D., Moher, D., Barber, R., & Denegri, S. (2017). GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. bmj, 358.

16.   Sun, L., Yin, C., Xu, Q., & Zhao, W. (2023). Artificial intelligence for healthcare and medical education: a systematic review. American journal of translational research, 15(7), 4820.

17.   Teodorowski, P., Gleason, K., Gregory, J. J., Martin, M., Punjabi, R., Steer, S., Savasir, S., Vema, P., Murray, K., & Ward, H. (2023). Participatory evaluation of the process of co-producing resources for the public on data science and artificial intelligence. Research Involvement and Engagement, 9(1), 67.

18.   UNESCO. (2024). AI literacy and the new digital divide: A global call for action. https://www.unesco.org/en/articles/ai-literacy-and-new-digital-divide-global-call-action

19.   Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives. International journal of educational technology in higher education, 21(1), 21.

20.   Zawacki-Richter: O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? International journal of educational technology in higher education, 16(1), 1-27.

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.