UK Healthcare Educators' Experiences and Perceptions of Essay Assessments in the Generative AI Era

Benard Ohene-Botwe*, Cletus Amedu, Mark Ngo

CRRA3G Research Group, Radiography Division, Department of Allied Health. School of Health and Medical Sciences, Clerkenwell Campus, City St George's, University of London, Northampton Square, London EC1V 0HB

*Corresponding author: benard.ohene-botwe@city.ac.uk

Abstract

Background: The rapid use of generative AI (GenAI) by students in essay assessments poses unique challenges to academic integrity. With limited research on this topic, questions arise about how UK educators perceive and experience the use of GenAI in traditional essay-type assessments (TETAs). The aim of this study was to explore the experiences and perceptions of UK healthcare educators regarding TETAS in the Era of GenAI.

Methods: An exploratory descriptive qualitative design was adopted. Semi-structured interviews were conducted with 10 healthcare academics from disciplines including Radiography, Midwifery, Speech and Language Therapy, and Nursing. With the support of NVIVO version-14, data were analysed using thematic analysis.

Results: Three themes emerged: (1) *Experiences of suspicion cues and verification challenges*; (2) *Perceptions of appropriate GenAI use in assessments*; and (3) *Relevance of essay-type assessments in the GenAI era*. Participants reported widespread evidence of AI-generated content in student submissions. They also indicated relying largely on intuitive detection methods such as identifying "red flags" including suspicious references, shallow content, unusual language patterns, and fabricated elements, among others. However, their inability to definitively confirm AI use created uncertainty, frustration, and ethical discomfort, undermining trust between faculty and students and threatening academic fairness.

Conclusion

This exploratory qualitative study reveals that healthcare educators continue to observe what they perceive as irresponsible use of GenAI in essay-based assessments, while facing a complex "detection dilemma." While participants acknowledged the continued relevance of essays and recognised the potential for GenAI to be used appropriately for limited tasks, they expressed concern that unregulated use may profoundly affect learning, assessment integrity, and professional competence. It is therefore recommended that higher education institutions develop clear strategies to address the irresponsible use of AI in assessment and empower educators to detect AI use in essay-based assessments.

Keywords: Artificial Intelligence; Generative AI; Essay Assessments; Healthcare Educators; Experiences and Perceptions

Introduction

Generative Artificial Intelligence (GenAI) refers to a class of artificial intelligence technologies, particularly large language models (LLMs), capable of generating novel content such as text, images, video, audio, code, or synthetic data by learning patterns from existing datasets (Ray, 2023). Utilising deep learning techniques and neural networks, GenAI models analyse data to produce outputs that closely resemble human-created content (Feuerriegel et al., 2024). Prominent examples of GenAI include ChatGPT, Gemini, and Claude (Ng et al., 2025). Among these, ChatGPT, developed by OpenAI, is widely recognised for its versatility and broad application across various domains (Liebrenz et al., 2023). Notably, it reached an estimated 100 million monthly active users within just two months of its launch, making it one of the fastest-growing online applications in history (Hu, 2023).

GenAI is increasingly being adopted across sectors such as education, healthcare, finance and software development(Abbas *et al.,* 2024; Gruenhagen *et al.*, 2024). In the educational context, GenAI models have been documented as offering significant opportunities (Chiu et al., 2023). Many studies have highlighted their potential to support self-paced learning, provide revision platforms, aid educators in scenario creation, assessment development, and group collaboration, and enhance educational inclusivity and accessibility (Bahroun *et al.*, 2023; Abbas *et al.,* 2024; Gruenhagen *et al.*, 2024).

However, as these technologies advance, there is also increasing concern among academic staff about their potential to compromise assessment integrity in higher education (Chan, 2023; Chaudhry et al., 2023). Issues such as contract cheating, falsification or fabrication of sources, and plagiarism using GenAI tools have raised alarms (Abbas *et al.,* 2024). Particularly, there is concern about the challenges these AI tools pose to academic integrity in traditional essay-type assessments where students are required to produce original work (Chan, 2023). A traditional essay-based assessment (TETA) in this study's context is a single essay-type assignment that does not include a reflective component.

To provide clarity in examining these concerns, two key conceptual distinctions underpin this study. First, a distinction is made between AI-supported and AI-generated writing. AI-supported writing involves a human-led process in which students retain intellectual ownership, using GenAI tools for limited support (e.g., idea generation or grammar), whereas AI-generated writing refers to work in which substantial or complete text is produced by AI, replacing the student's authorship (Dwivedi et al., 2023). Second, a distinction is made between appropriate use and misuse of GenAI in

assessments. Appropriate use is transparent, limited, and does not substitute the student's own thinking, while misuse involves undisclosed or substitutive use of AI-generated content presented as original work, thereby undermining academic integrity (UNESCO, 2023).

There are reported instances where students have used these tools irresponsibly to complete assignments, and many academics have warned that the continued inappropriate use of GenAI may hinder students' learning and success, erode their motivation to achieve, and damage the credibility of higher education institutions (HEIs) (Gruenhagen et al., 2024; Xia et al., 2024).

Amid these concerns, ongoing discussions within higher education are questioning the future of TETAs. Some empirical studies have begun to explore academic staff perceptions regarding the impact of GenAI on TETAs (Amani *et al.*, 2023; Lee *et al.*, 2024; Almisad and Aleidan, 2025).

However, most of these studies have been conducted in non-UK contexts, which limits their direct applicability to the UK higher education system. This is due to important differences in pedagogical approaches, assessment frameworks, regulatory structures, and professional accreditation standards, all of which shape how academic integrity and assessment practices are implemented and perceived.

Of note, there remains a critical need to expand this research area within the UK context to specifically understand the perceptions and concerns of lecturers in healthcare professional programmes (LHPPs) (Cassidy, 2023; Guo & Wang, 2024). This focus is especially pertinent: while AI offers potential benefits in healthcare education such as personalised learning and simulation-based training, there are growing concerns among educators that irresponsible use of GenAI in assessments may compromise students' development of the professional competencies required for safe and effective practice—concerns that, while not yet empirically established fully, carry particular weight in the context of regulated healthcare professions (Moorhouse *et al.,* 2023; Rajabi *et al.*, 2023). Given these considerations, the aim of this study was to conduct an in-depth investigation to answer the following research question: What are the experiences and perceptions of UK healthcare educators regarding TETAS in the Era of GenAI? This study makes a specific contribution by foregrounding the perspectives of lecturers in UK healthcare professional programmes, a population largely absent from existing literature, and by examining how disciplinary expectations, professional accreditation standards, and the unique stakes of healthcare graduate competence shape their experiences and judgements about essay assessments in the GenAI era. In doing so, it extends the existing evidence base beyond general higher education contexts and provides insights directly relevant to UK healthcare education policy and practice.

Methods

Study design, sampling and recruitment

An exploratory descriptive qualitative study was conducted to examine UK healthcare educators' perceptions and experiences regarding essay-type assessments in the era of generative AI. Snowball sampling was implemented to recruit 10 participants. Social media was used to initiate the snowball using social media platforms including ResearchGate, LinkedIn, X (formerly known as Twitter), and Facebook.

Inclusion criteria

The inclusion criteria required participants to be UK-based healthcare professionals with knowledge of GenAI and experience marking or moderating essay assessments where AI use was suspected. It is important to acknowledge that this criterion means the sample was purposively composed of educators who had already encountered or suspected GenAI use; this shapes the nature of the findings, which are likely to reflect heightened concern, and may not represent the views of educators without such experiences.

Interview

Semi-structured interviews lasting 30 to 45 minutes were conducted via Zoom by the first author (BOB) until sufficient information power was achieved. This was based on the focused nature of the research question, participant characteristics, and the depth and relevance of the data generated (Malterud et al., 2016). An interview guide (Appendix One) was developed and piloted on two academic staff members. Interviews were audio-recorded and auto-transcribed using Zoom's transcription feature, with transcripts carefully reviewed and corrected by the research team to ensure verbatim accuracy.

Interested participants were provided with a detailed information sheet outlining the nature of the study, their rights, and the voluntary nature of their involvement. Written informed consent was obtained from all participants prior to their interview. Confidentiality was strictly maintained throughout the study. All data were de-identified prior to analysis and reporting. Each participant was assigned a unique number (P1 to 10) to protect their identity. Audio recordings were used solely for transcription purposes and were permanently deleted once transcription and verification were completed.

Data Analysis

Transcripts were imported into NVivo v14 for analysis. Thematic analysis following Braun and Clarke's (2006, 2021) six-step framework was employed using an inductive, descriptive coding approach. The process involved: (a) familiarisation with the data through multiple readings, (b) generating initial codes, (c) searching for themes, (d) reviewing themes for coherence and

distinctiveness, (e) defining and naming themes, and (f) reporting findings with illustrative quotes and thematic maps.

Rigour strategies

Several strategies were employed to enhance rigour and ensure credibility, dependability, confirmability, and transferability. Member checking was conducted by returning transcripts to participants within two weeks for verification. Peer debriefing was performed to promote reflexivity, with research team members meeting to discuss emerging themes and to reduce individual researcher bias. To support dependability and confirmability, an audit trail was maintained documenting methodological decisions, analytic processes and to enhance transferability, relevant participant quotes are provided to support the identified themes (Ngo et al., 2025)

Reflexivity and Positionality

The research team acknowledges the importance of reflexivity in qualitative inquiry. The first author (BOB), who conducted all interviews, is an experienced healthcare academic with a background in radiography and active involvement in assessment design and quality assurance in higher education. This positioning may have facilitated rapport with participants and enriched the depth of discussion; however, it also means that prior assumptions and professional perspectives may have shaped how questions were framed and how responses were interpreted. To mitigate potential bias, the team engaged in ongoing reflexive discussion throughout data collection and analysis, maintained an audit trail of analytic decisions, and employed peer debriefing across the authorship team. Participants' own words are foregrounded through the use of direct quotations to support transparency and allow readers to assess interpretive claims.

Ethics approval and consent to participate

Ethical approval for the study was obtained from the LEaD Research Ethics Committee at City St George's, University of London (Ref: ETH2425-0553). Written informed consent was obtained from all participants prior to their interviews, including consent to publish the study findings. All recruitment and study procedures were conducted in accordance with institutional ethical guidelines and the principles of the Declaration of Helsinki.

Results

Demographics of participants

The study involved ten UK healthcare educators based in the UK. Participants came from diverse healthcare professional backgrounds: Radiography (*n*=3), Nursing (*n*=3), Speech and Language Therapy (*n*=2), and Midwifery (*n*=2). The gender distribution included seven females and three males. All participants were employed in academic roles within HEIs at the time of their interview.

Themes

Three themes were identified in relation to: *(1) Experiences of suspicion cues and verification challenges; (2) Perceptions of appropriate GenAI use in assessments; and (3) Relevance of essay-type assessments in the GenAI era*. Each theme is presented below, accompanied by illustrative quotes from participants.

Theme 1: Experiences of suspicion cues and verification challenges

Participants described their experiences with GenAI during moderation and marking, where they suspected its use by students in assessments. These findings are organised into two subthemes: (1) Awareness and *Suspicion Cues and (2) Detection Dilemma*

Sub-theme 1.1: Awareness and suspicion cues

Participants reported increasing familiarity with GenAI-generated content in student submissions, encountering situations ranging from initial suspicion to confirmed use following student admission. Suspicions were informed by academic judgement, pattern recognition, familiarity with students' usual performance, and the use of detection tools, and were followed by probing actions such as questioning students or initiating academic misconduct procedures. A commonly reported indicator of suspected GenAI use was the presence of non-existent, inaccurate, or mismatched references.

> *"Sometimes the references just don't exist, they look credible at first glance but cannot be verified."(P4)*

> *"…sometimes it generates references for them that doesn't exist."(P5)*

Others mentioned that some essays demonstrated repetitive ideas or impressive-sounding but shallow content, lacking critical depth or originality.

> *"The writing sounds polished, but it repeats the same points in slightly different ways."(P2)*

> *"It ticks the structural boxes but there's no depth, just a diluted version of academic writing."(P8)*

Some participants also suspected the use of GenAI when they observed that students' essays occasionally included "*AI footprints*", that is, AI-generated comments, prompts, or explanatory statements about what the AI had done or intended to do. These often served as unintentional indicators of AI involvement in the writing process.

> *"We also detect via AI footprint where the AI has left a statement about itself. For example, after the essay, you might see something like 'Would you like help making adjustments to what I have generated' and so on."(P9)*

> *"... and sometimes you even see the summary left in, like 'Generated by ChatGPT'."(P2)*

One participant reported using institutional AI detection tools, primarily integrated through Turnitin, to flag potential instances of GenAI use in student submissions. These tools generate a report indicating the likelihood that portions of a text were AI-generated. However, the participant emphasised that such tools are not conclusive and must be interpreted alongside academic judgment to initiate a discussion with the student.

> *"When the assessment is submitted, we have an AI detection software within the university… it gives a report where it highlights sections where the AI detection software predicts that someone may have developed their work using a GenAI tool. One of the challenges is that the AI detection software is not 100% accurate… we are meant to look at that score side-by-side with academic judgment."(P9)*

Participants observed a number of stylistic and lexical differences from regular writings, including excessive use of em dashes; uncharacteristically perfect grammar all the time, unusual or rare

vocabulary (alien terminology) that feel out of place in healthcare and sudden improvement in writing quality compared to earlier work. These were their red flags in suspecting the use of GenAI use.

> *"The writing sometimes uses words the student never used before (…) strange, academic-sounding terms that feel out of place in healthcare."(P3)*

> *"Some of the indicators are zero grammar errors, unusual punctuation (like semicolons, many em dashes, nested clauses), and sudden improvement in writing quality compared to earlier work…" (P9)*

Some participants also suspected the use of AI when student work lacked personal voice, creativity, or familiar phrasing. They observed that the tone of AI-generated essays often felt detached or generic, lacking the nuance and authenticity expected in healthcare student writing.

> *"It just doesn't sound like them anymore. There's no personality or creativity."(P3)*

> *"We are dealing with healthcare students – you expect care-focused language. But what you get is cold, robotic."(5)*

Moreover, drawing on their extensive marking experience and familiarity with common student errors, participants described noticing a shift in the nature of mistakes being made. Some suspected the use of GenAI when essays occasionally included inaccurate information or nonsensical statements in case studies that did not align with course content or logical argumentation.

> *"I've been marking these assessments for at least seven, maybe ten years, so I know the kinds of mistakes students usually make. But since the advent of GenAI, I'm seeing what I can only describe as bizarre errors. For example, (...) students rarely used to get it wrong. But now I'm seeing answers that (...) I wouldn't expect any student to make, and they suggest the content has been generated by a tool that doesn't understand the material."(P6)*

> *"Sometimes it brings in things that are not human... Like, it doesn't understand it in a medical or nursing context."(P5)*

A word cloud highlighting the key indicators or *suspicion cues* used by healthcare educators to suspect the involvement of GenAI in student submissions is presented in **Figure 1**.

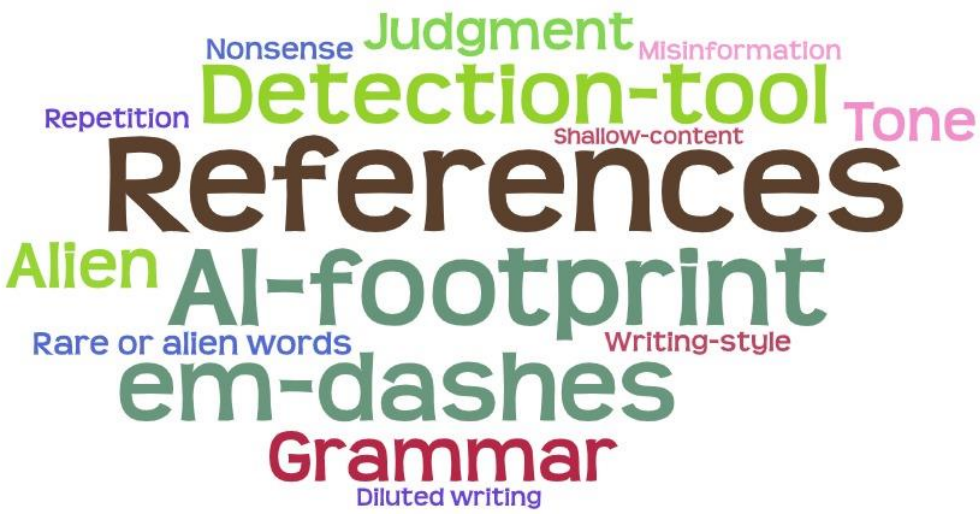

Figure 1: Word cloud on the key indicators or suspicion cues that participants use to suspect GenAI use by students

Sub-theme 1.2: Detection dilemma

Several participants indicated that it is currently very challenging to confidently confirm the use of GenAI in assessments, even when red flags appear. They expressed frustration, noting that when they suspect GenAI was used to write an assignment, they often lack reliable methods to prove it, unless the student admits to it. They emphasised the need for dependable tools to confirm AI use, to avoid the risk of false positives or unjust disciplinary actions. Participants also shared their discomfort with having to question their students' academic integrity, which they felt could damage the trust between staff and students. According to them, this situation contributes to a climate of worry, fear, and uncertainty among both academic staff and students. Overall, they described a mix of emotional strain and ethical concerns when dealing with suspected misuse of GenAI in essay-type assessments.

> *"There are no reliable ways of checking unless students admit… we need standardised frameworks."(P2)*

> *"If someone says 'I didn't use it', you have to believe them… but the one who says 'I used it, I'm sorry' gets penalised." That's unfair."(P5)*

Theme 2: Perceptions of appropriate GenAI use in assessments

Participants described their perceptions of GenAI in relation to its use by students. These findings are organised into two subthemes*: (1) mixed feelings towards GenAI and (2) allowing its use to some extent.*

Sub-theme 2.1: Mixed feelings toward GenAI

Some participants expressed a positive perspective on GenAI, viewing it as acceptable when used appropriately and as a tool that can enhance student learning and provide additional support, particularly for neurodiverse students.

> *"… it can enhance student learning …. I mean, I think I'm not against its use at all."(P2)*

> *"It can support a range of students, those with neurodiversity."(P9)*

Others expressed stronger scepticism, viewing GenAI as a fundamental threat to assessment validity and student development.

> *"(…) has a lot of potential, but without proper regulation and certification… it carries more risks than benefits at the moment."(P10)*

> *"I personally don't like it because one, it doesn't give the room for originality or the intent for the assessment."(P5)*

Sub-theme 2.2: Allowing its use to some extent

In terms of using it for written assessments, participants generally expressed strong reservations about students using GenAI to write their entire assessments. There was a clear consensus that while GenAI should not replace the student's own writing, it could be permitted for limited, supportive functions within the writing process. Acceptable uses identified by participants included scoping, brainstorming or forming ideas, asking questions, assisting with grammar and proofreading, overcoming writer's block, and initiating outlines or structure, provided that AI does not generate full written responses.

"There is a benefit to it, particularly for some very light touch things… like a scoping exercise… it can gather some information … but it has to be exercised with caution..."(P1)

"They may, for example, be guided to use it in sort of like initiating the structure of what the assignment should look like. They can use it to maybe, for example, proofread the work for grammar and stuff like that … But getting it to sort of like write the entire thing for you and not really declaring its usage, then it becomes a problem… AI would help to relieve anxiety … writer's block … and could enhance the overall quality of work if it's used in a responsible manner."(P2)

"It can be helpful for people that English is not their first language when you're trying to construct different things … Maybe essay, the idea, ask questions, and then they can put the ideas together and use the idea to now write their own assignment... If we must, it could be part of a formative assessment … so that people still learn how to write."(P5)

Figure 2 presents a schematic summary of the areas where academics think GenAI can be allowed in essay-type assessments.

Figure 2: Areas where academics think GenAI can be allowed in essay-type assessments

Theme 3: Relevance of Essay-type assessments in the GenAI era

Participants presented mixed positions about the relevance of Essay-type assessments in this era of GenAI proliferation. While most participants defended the continued relevance of essays as a reliable assessment in the era of GenAI and called for some reforms to address academic integrity issues, others argued that they are no longer fit for purpose in a GenAI age, especially in healthcare education, where real-world decision-making is critical.

> *"Personally, I recommend we don't use essays at all… they don't prepare students for real-world scenarios."(P5).*

Among those who support the relevance of essay-type assessments, there was strong consensus that this assessment format enables the evaluation of core competencies, specifically synthesis and critical reflection, which are essential capacities that AI currently struggles to replicate meaningfully.

> *"…essays at Level 7 are designed to assess purposeful synthesis of clinical case data from varied literature and the ability to extract core content, themes, controversies, or similarities or differences. Equally vital is the demonstration of critical reflection, which is the student's voice, a component that AI cannot imitate due to its inability to convey individualised professional insight. While GenAI might offer surface-level support, its tendency to produce inaccuracies, hallucinations, and its failure to meet the nuanced demands of assessment criteria make it an inadequate substitute for authentic student engagement."(P1)*

> *"There are so many skill sets that an essay type of assessment will do to the student. And so I think it's more relevant irrespective of the era that we are in …. The essay in itself... the core knowledge behind their arguments will be their own originality. So that critical thinking is still a good thing for them to have... referencing... appropriate referencing subject-specific contents... they are not there yet. So these type of skills will be really beneficial."(P2)*

> *"AI can't address a case study, which I think is one of its major challenges…" (P4)*

Discussion

Experience with GenAI during moderation and marking

Exploring educators' experiences in education offers critical insights into the practical realities of teaching and learning (Al-Thani and Ahmad, 2025). Consistent with this view, faculty members reported a range of issues encountered during the marking and moderation of essay-type assessments, particularly concerning the use of GenAI tools by students in essay-type assessments.

In particular, participants reported observing an increasing amount of AI-generated content in student essay submissions. Their experiences ranged from mild suspicions of GenAI use to confirmed cases where students openly admitted to relying on such tools. These observations align with prior research (Chan, 2023; Sleator, 2023; Ward *et al.,* 2024), which suggests widespread GenAI use. Sleator (2023), of *The Times* reported that 50% of Cambridge students acknowledged using ChatGPT, while Forbes noted that one in five U.S. college students had used AI in academic work (Nietzel, 2023). The implication for practice is that institutions must develop clear policies and practical strategies to uphold academic integrity in the era of GenAI.

One key strategy involves reliably detecting the irresponsible use of GenAI. Educators reported using a mix of cues and judgment-based strategies to identify suspect submissions, followed by actions such as questioning students or referring cases for misconduct investigations. Their methods were largely intuitive and qualitative, relying on cues such as abrupt shifts in tone, inconsistencies in writing style, unnatural or overly formal language, fabricated references, diluted or repetitive content, and signs of AI output such as unnatural use of several stylistic and lexical writings different from regular writings or the inclusion of "alien" vocabulary. Notably, a set of key codes "referred to as red flags" emerged as indicators of potential AI-generated content. These, as shown in **Figure 1,** include *suspicious references, shallow content lacking critical depth or originality, identifiable AI footprints, odd em-dashes, nonsensical text, misinformation, repetitive patterns, odd writing style, "use of rare or alien words", perfect grammar, and discrepancies in tone.* Other codes included *using detection tools and professional judgment.* These indicators align with previous studies (Cotton *et al.,* 2024; Ward *et al.,* 2024), though this current study also revealed novel cues not previously reported, such as distinctive language patterns, unusual em-dashes, and repeated use of alien words. Conversely, some detection methods identified by Ward *et al.* (2024) were not observed by participants in the present study, suggesting that detection practices may evolve based on context, educator experience, and institutional support. The implication for practice here is the need for staff

training and shared guidelines so that such red flags can be applied consistently, reducing reliance on subjective judgment alone.

Although some participants reported using Turnitin's AI-detection features, they acknowledged these tools are unreliable for confirming AI use. This reflects Ward *et al.*'s (2024) finding that educators increasingly struggle to verify AI-generated writing, as tools like Turnitin and MOSS are less effective against advanced GenAI outputs. The unreliability of these approved plagiarism detection tools for AI use detection, combined with the fact that the subjective cues used by educators cannot consistently confirm AI use, often makes it extremely difficult, if not impossible, for them to take further disciplinary action unless a student admits to using AI. They stressed the need for dependable detection to avoid false positives and unjust penalties, noting that the lack of reliable evidence contributes to a sense of unfairness in addressing suspected cases of academic misconduct. Participants noted that current approaches may inadvertently disadvantage honest students, as allegations often cannot be substantiated when GenAI use is denied, while admissions of use lead to penalties. This imbalance was perceived as unfair, raising a major ethical implications, as unfairness should not be allowed in education (Herr et al., 2024).

The current detection approach was reported to cause growing frustration among educators, echoing Cotton *et al.'s* (2024) findings. Particularly, educators expressed discomfort with questioning their students' academic integrity, an experience they felt could damage the trust between staff and students. They emphasised that this situation contributes to a wider climate of worry, fear, and uncertainty among both academic staff and students. Overall, they described a mix of emotional stress and ethical discomfort in navigating this emerging landscape. Consequently, their experiences imply the need for clearer policies, enhanced detection tools, and supported frameworks that can promote both fairness and confidence in upholding academic integrity.

Perception of GenAI

Educators expressed mixed feelings about GenAI use by students, which is consistent with the literature (Amani *et al.*, 2023; Lee *et al.*, 2024; Almisad and Aleidan, 2025). Some viewed GenAI as a helpful tool when used responsibly, while others were more sceptical, seeing it as a significant threat to assessment validity and student development. There was, however, general agreement that using GenAI to write portions or entire assessments constitutes academic misconduct, a view supported by several authors (Moorhouse *et al.,* 2023; Lee *et al.*, 2024). At the same time, participants agreed that GenAI could be permitted for limited, supportive roles in the writing process without replacing the student's own work. Studies confirm that responsible use can offer benefits such as supporting self-

paced learning, enhancing revision and feedback, and improving accessibility and inclusion *(*Abbas *et al.*, 2024; Levin *et al.*, 2024).

Acceptable uses of GenAI identified by participants for essay assessments included scoping, brainstorming or generating ideas, asking questions, assisting with grammar and proofreading, overcoming writer's block, and initiating outlines or structure, provided that AI does not generate full written assessments (**Figure 2**). These uses were generally not viewed as academic misconduct, provided they were not abused. These findings align with the conceptual distinctions outlined in the Introduction: participants' acceptable uses correspond to AI-supported writing, while their concerns centre on AI-generated writing submitted as the student's own work. Consistent with this framework, disclosure emerged as a critical boundary: participants distinguished not only between the *extent* of AI involvement but also its *transparency*: undisclosed use, regardless of scope, was consistently regarded as misuse. Misuse includes allowing GenAI to write entire or sections of an assignment after using it to generate outlines or content ideas. As advised by the Department of Health (2025), institutions must carefully assess both the benefits and potential risks of limited GenAI use, ensuring it demonstrates clear educational value without compromising safety, ethics, or academic integrity.

Relevance of Essay-type assessments in the GenAI era

When asked about the relevance of essay-type assessments in the GenAI era, participants expressed mixed views. Most supported their continued use but emphasised the need for reforms to safeguard academic integrity. A minority, however, considered essays outdated, particularly in healthcare education, where real-world decision-making is crucial. This view is not strongly supported in the literature. On the contrary, scholars argue that essays remain vital in healthcare education for assessing research, analysis, synthesis, and critical reflection,—skills that are essential for clinical reasoning, professional judgement, and real-world application (Dahl *et al.*, 2023; Lysanets *et al.,* 2024). Theoretical models such as Bloom's taxonomy, Miller's pyramid, and the Dreyfus model also highlight the role of essays within varied assessment strategies (Witheridge *et al.,* 2019; Rosander *et al.*, 2022; Lubbe *et al.*, 2025). The implications for practice suggest that essay-type assessments should be retained but thoughtfully adapted within a broader assessment framework to ensure their continued relevance in the GenAI era. Importantly, there is a clear need to develop and evaluate alternative or complementary assessment approaches that can robustly evidence student learning without being compromised by the irresponsible use of generative AI. For example, assessments could include essays written under invigilated examination conditions, oral vivas, or other forms of authenticated assessment that require students to demonstrate real-time understanding and critical

engagement. Interestingly, a new and closely related strand of work is already underway, focusing on exploring practical strategies to address the assessment challenges identified in this study. Further research in this area is therefore strongly recommended, particularly to evaluate the effectiveness, feasibility, and equity implications of such approaches in diverse educational contexts.

Conclusion

The findings show that healthcare educators face major challenges in upholding academic integrity while dealing with the fast growth of GenAI in healthcare education. The findings demonstrate that educators are experiencing AI use in student submissions. They described relying on specific cues to identify possible AI use and, when necessary, initiating probing steps such as questioning students and referring cases for academic misconduct. However, their inability to definitively confirm AI use creates a climate of uncertainty, frustration, and ethical discomfort that threatens the fundamental trust between faculty and students and the fairness of academic assessment. While participants recognised that GenAI could support the writing process, such as brainstorming, asking questions, providing grammar assistance, overcoming writer's block, and initiating outlines or structures, they expressed serious concerns about students' irresponsible use or overreliance on these tools to produce part or all of an assessment. It is therefore recommended that higher education institutions develop clear strategies to address the irresponsible use of AI in assessment and empower educators to detect AI use in essay-based assessments. Further studies involving larger sample sizes and more balanced gender representation are recommended to explore whether similar findings hold across the broader UK healthcare education community.

Limitation

The study is limited to the views of educators from the professions of radiography, midwifery, speech and language therapy, and nursing, which limits the breadth of perspectives captured across the broader UK healthcare education community. The gender imbalance, while reflective of sector demographics, may underrepresent male perspectives. Some healthcare disciplines had few participants, restricting understanding of discipline-specific nuances. A further important limitation relates to the purposive nature of the sample: all participants were recruited on the basis of having previously suspected or encountered AI use in student submissions. This sampling criterion, while appropriate for the study's exploratory aim, means that the findings are likely to reflect heightened concern and awareness of GenAI challenges. Educators without such experiences may hold different

views, and their perspectives are not captured here. This potential selection bias should be borne in mind when interpreting the findings. The small sample size (n=10), consistent with qualitative exploratory designs, also limits the scope for transferability. While the findings offer valuable insights into the experiences of participants across four healthcare disciplines in the UK, they cannot be assumed to represent the views of all UK healthcare educators, nor to be directly applicable to other educational systems, regulatory frameworks, or disciplinary contexts. Transferability is further limited by the UK-specific pedagogical and accreditation landscape, which shapes how academic integrity and assessment practices are structured and experienced. Readers are encouraged to consider the relevance of these findings to their own contexts through reflexive application rather than direct generalisation.

DECLARATIONS

Ethics approval and consent to participate

Ethical approval for the study was obtained from the LEaD Research Ethics Committee at City St George's, University of London (Ref: ETH2425-0553). All recruitment and study procedures were conducted in accordance with institutional ethical guidelines and the principles of the Declaration of Helsinki.

Consent to participate

Written informed consent was obtained from all participants prior to their interviews, including consent to publish the study findings.

Consent to Publish

Written informed consent was obtained from all participants for the publication of the study findings.

Competing interests

The authors declare that they have no known competing financial or non-financial interests that could have appeared to influence the work reported in this paper.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

None

Acknowledgements

The authors would like to thank all academic staff who participated in this study for their time, insights, and valuable contributions.

References

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, *21*(1), 10. https://doi.org/10.1186/s41239-024-00444-7

Almisad, B., & Aleidan, A. (2025). Faculty perspectives on generative artificial intelligence: Insights into awareness, benefits, concerns, and uses. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1632742

Al-Thani, N. J., & Ahmad, Z. (2025). Teaching Research to Teachers—Traversing from Research-Oriented Education to Research Learning Theory. In N. J. Al-Thani & Z. Ahmad, *Teaching and Learning with Research Cognitive Theory* (pp. 97–112). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87544-1_6

Amani, S., White, L., Balart, T., Arora, L., Shryock, K. J., Brumbelow, K., & Watson, K. L. (2023). *Generative AI Perceptions: A Survey to Measure the Perceptions of Faculty, Staff, and Students on Generative AI Tools in Academia* (arXiv:2304.14415). arXiv. https://doi.org/10.48550/arXiv.2304.14415

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability*, *15*(17), 12983. https://doi.org/10.3390/su151712983

Cassidy, C. (2023). *Australian universities to return to 'pen and paper' exams after students caught using AI to write essays*. https://www.theguardian.com/australia-news/2023/jan/10/universities-to-return-to-pen-and-paper-exams-after-students-caught-using-ai-to-write-essays

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to Revisit Existing Student's Performance Evaluation Approach in Higher Education Sector in a New Era of ChatGPT — A Case Study. *Cogent Education*, *10*(1), 2210461. https://doi.org/10.1080/2331186X.2023.2210461

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

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

Dahl, B. M., Vasset, F., & Frilund, M. (2023). Students' approaches to scientific essay writing as an educational method in higher education: A mixed methods study. *Social Sciences & Humanities Open*, *7*(1), 100389. https://doi.org/10.1016/j.ssaho.2022.100389

Department of Health. (2025). *Generative artificial intelligence (AI) in education*. https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers and Education: Artificial Intelligence*, *7*, 100273. https://doi.org/10.1016/j.caeai.2024.100273

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, *29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Herr, R. M., Deyerl, V. M., Rathmann, K., & Diehl, K. (2024). Perceived Unfairness in University Settings: Findings from a Qualitative Study among Students in Germany. *Education Sciences*, *14*(8), 827. https://doi.org/10.3390/educsci14080827

Hu, K. (2023). *ChatGPT sets record for fastest-growing user base—Analyst note*. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, *6*, 100221. https://doi.org/10.1016/j.caeai.2024.100221

Levin, G., Horesh, N., Brezinov, Y., & Meyer, R. (2024). Performance of CHATGPT in medical examinations: A systematic review and a meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*, *131*(3), 378–380. https://doi.org/10.1111/1471-0528.17641

Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., & Smith, A. (2023). Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *The Lancet Digital Health*, *5*(3), e105–e106. https://doi.org/10.1016/S2589-7500(23)00019-5

Lubbe, A., Marais, E., & Kruger, D. (2025). Cultivating independent thinkers: The triad of artificial intelligence, Bloom's taxonomy and critical thinking in assessment pedagogy. *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13476-x

Lysanets, Yu., Klaus, A., & Bieliaieva, O. (2024). ESSAY WRITING SKILLS AND THEIR SIGNIFICANCE IN MEDICAL AND ECOLOGICAL EDUCATION. *The Medical and Ecological Problems*, *28*(3), 61–67. https://doi.org/10.31718/mep.2024.28.3.07

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

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*, 100151. https://doi.org/10.1016/j.caeo.2023.100151

Ng, D. T. K., Chan, E. K. C., & Lo, C. K. (2025). Opportunities, challenges and school strategies for integrating generative AI in education. *Computers and Education: Artificial Intelligence*, *8*, 100373. https://doi.org/10.1016/j.caeai.2025.100373

Ngo, M., Thorburn, K., Naama, A., & Skelton, E. (2025). Exploring the lived experiences of diagnostic radiographers after transitioning to non-emergency imaging settings. *Radiography*, *31*(2), 102871. https://doi.org/10.1016/j.radi.2025.01.006

Nietzel, M. T. (2023). *More Than Half Of College Students Believe Using ChatGPT To Complete Assignments Is Cheating*. Forbes. https://www.forbes.com/sites/michaeltnietzel/2023/03/20/more-than-half-of-college-students-believe-using-chatgpt-to-complete-assignments-is-cheating/

Rajabi, P., Taghipour, P., Cukierman, D., & Doleck, T. (2023). Exploring ChatGPT's impact on post-secondary education: A qualitative study. *Proceedings of the 25th Western Canadian Conference on Computing Education*, 1–6. https://doi.org/10.1145/3593342.3593360

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Rosander, M., Forslund Frykedal, K., Barimani, M., & Berlin, A. (2022). Experiences from leading parental education groups: Perceived difficulties and rewards as an indication of skill acquisition. *Journal of Child Health Care*, *26*(1), 68–81. https://doi.org/10.1177/13674935211000940

Sleator, L. (2023). *Almost half of Cambridge students admit they have used ChatGPT*. The Times. https://www.thetimes.com/business-money/technology/article/cambridge-university-students-chatgpt-ai-degree-2023-rnsv7mw7z

UNESCO (2023) *Guidance for generative AI in education and research*. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research. Accessed on: 20/03/24

Ward, A., Manoharan, S., & Ye, X. (2024). Exploring Academic Integrity in the Age of Generative AI. *2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–5. https://doi.org/10.1109/ithet61869.2024.10837603

Witheridge, A., Ferns, G., & Scott-Smith, W. (2019). Revisiting Miller's pyramid in medical education: The gap between traditional assessment and diagnostic reasoning. *International Journal of Medical Education*, *10*, 191–192. https://doi.org/10.5116/ijme.5d9b.0c37

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, *21*(1), 40. https://doi.org/10.1186/s41239-024-00468-z

APPENDIX ONE
SEMI-STRUCTURED INTERVIEW GUIDE

Introduction
Welcome and thank the participant for their time.
The purpose of the study and information about the study will be mentioned.
Participants will be assured of confidentiality and anonymity.
Consent will be obtained to record the interview.

INTERVIEW GUIDE

Interview questions.

Share with me any experience where you have suspected the use of GenAI in assessment by student and what happened

What is your general perception of the use of generative AI in educational assessments?

Do you think the use of essay-based assessments is still relevant in the era of generative AI?

How are you able to detect irresponsible use of GenAI?

Share with me some of the challenges of detecting irresponsible use of GenAI?

Closing
Is there anything else you would like to share about your views on generative AI and essay-type assessments?
Do you have any questions or comments about this study?
Thank the participant for their insights and time.