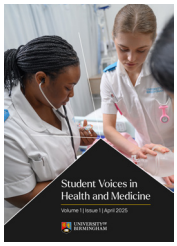
Student Voices in Health and Medicine



Using generative artificial intelligence (GAI) technology as a tool for marking student nursing assignments: A phenomenological qualitative study

Olivia Latheron[1], Daisy Kelly[1]✉, Paula Total-Avioroko[1], Charles Sharp[1], Nutmeg Hallett[1]

[1]School of Nursing and Midwifery, University of Birmingham, Birmingham, United Kingdom

✉ daisy.kelly1@outlook.com

Abstract

Aim: To explore the feasibility and acceptability of using Generative Artificial Intelligence (GAI) to mark and provide feedback on student nursing assignments. **Background**: The recent boom in Artificial Intelligence (AI) has garnered attention with global AI spending reaching an estimated $154 billion by the end of 2023. Nursing is a demanding programme incorporating a variety of assessment methods which are time-consuming to mark. Whilst already used for administrative tasks and plagiarism detection, the use of GAI to mark assignments is an avenue that has not yet been explored. It holds transformative capabilities, potentially serving as virtual tutors, automating tasks and generating educational content. Design: Phenomenology. **Methods**: Semi-structured interviews were conducted with academics teaching at the University of Birmingham and focus groups consisting of nursing students. Participants were asked to evaluate, compare and contrast two feedback samples: one written by an academic and the other generated through GAI. The interviews and focus groups were transcribed and analysed using thematic analysis. **Results**: Evaluation and governance, human touch, development and integration, and time were the key themes identified from the six staff interviews and two focus groups. When comparing examples of feedback, almost all participants favoured that which was GAI generated, stating it was objective and highly detailed. **Conclusions**: Whilst interviewees showed concern regarding how a GAI tool would be governed and the potential reduction of human touch, benefits such as increased grading efficiency and objectivity were acknowledged. Staff believed it would eliminate the narrative that students are treated unfairly and possibly mitigate the need for moderation. Both groups discussed the need for extensive guidance to effectively implement GAI into practice. Although it was apparent that GAI would not completely replace human markers, the possibility of educators using GAI as a feedback tool or as a way of assessing formative assignments was welcomed with positivity.

Keywords: Generative artificial intelligence; Nurse education; Student feedback

Introduction

Generative Artificial Intelligence (GAI) has swiftly transitioned from a theoretical concept to a pivotal tool across numerous sectors, including health, social care and education. The International Data Corporation (IDC) predicted that global GAI spending would reach $154 billion by the end of 2023, an increase of 27% from the year before (IDC, 2023). From identifying cancerous cells to creating visual art, GAI can now perform a huge number of undertakings and is slowly infiltrating all aspects of everyday life (Al-Shamasneh et al., 2017). Despite its expanding role, there are still numerous avenues that are yet to be explored when it comes to GAI, including many aspects of education.

Nursing is a demanding academic course; in the UK, for example, student nurses are expected to spend half their time on placement and dedicate the remaining half to university study. To assess the capabilities of each student, a variety of methods are deployed such as assignments, coursework and Objective Structured Clinical Examinations (OSCEs). Grading these assessments and providing feedback is often a repetitive and time-consuming task, and which can lead to student dissatisfaction with feedback and lecturer burnout (Pitt et al., 2017). While advancements in technology have facilitated aspects of the marking process, the potential for GAI to enhance consistency, quality of feedback and educator productivity is significant but remains unexplored (Kumar, 2023; Zhang, 2023).

The rapid development of GAI has sparked both enthusiasm and debate. Proponents argue that GAI can improve efficiency and objectivity in grading, while critics raise concerns about privacy, bias and the potential hindrance to critical thinking (Gherhes, 2018). These concerns are particularly pertinent in the context of education, where the human touch is often considered indispensable. Moreover, there is a lack of robust evidence for the benefits of using GAI to grade papers (Holmes et al., 2023).

Within the broader GAI landscape, GAI occupies a distinct niche. GAI refers to systems capable of producing new content – e.g. text, images, audio or video – by learning patterns from existing data. This category of GAI gained significant attention with the development of large learning models (LLMs) like ChatGPT, which generate coherent and contextually relevant text by processing and learning from vast amounts of data (MIT News, 2023). This capability makes GAI particularly useful for tasks like generating responses or evaluating written content, as it can produce contextually relevant and coherent output (IBM, 2023). Unlike more advanced forms of GAI, which aim to replicate or surpass human intelligence across a wide range of activities, GAI is specifically designed to excel at these narrower tasks (MIT News, 2023). Its specialisation allows it to be effectively applied in educational settings, where it can assist in grading assignments by generating detailed feedback or assessing the quality of student work.

Within education, GAI has been used primarily to enhance administrative tasks such as plagiarism detection, curriculum development and monitoring student performance to identify effective teaching methods (Chen et al., 2020; Chaudhry et al., 2022). It has also been used in the preparation of lecture materials and plans, with educators increasingly turning to GAI to enhance efficiency and objectivity (Malik and Gangopadhyay, 2023; Nah et al., 2023). Research suggests that GAI could personalise learning experiences and reduce the workload of educators, particularly in the labour-intensive task of providing feedback (Haseski, 2019; Gocen et al., 2021). A survey of 1,685 educators revealed that more than half found data recording and analysing to be the most labour-intensive task, with almost as many stating they were overwhelmed by the sheer volume of marking (Gibson et al., 2015). When asked how their job role could be improved and stress relieved, one third suggested different marking arrangements, whilst a quarter wanted more time for assessment, highlighting the time-consuming and laborious nature of grading assignments.

The potential of GAI to ease the burden of grading is significant, freeing educators to focus on teaching, lesson planning and research. Students could also benefit from this technology, getting timely, detailed and unbiased feedback (Boud and Falchikov, 2007; Hassan et al., 2022). However, the reliance of GAI on human-fed data introduces potential risks such as bias and discrimination (Baker, 2021). Additionally, concerns about the lack of human touch, privacy and the risk of over-reliance on technology persist (Kumar, 2023). The limitations of GAI, such as the potential misinterpretation of nuanced language, could impact the accuracy of grading, further complicating its implementation in education.

Despite the existing literature on the advantages and disadvantages of GAI as a tool for providing feedback, there is a lack of research into the views of students and educators. This project aims to address this deficiency, by exploring these perspectives and comparing GAI-generated feedback with that provided by human lecturers. By doing so, it aims to contribute to a more comprehensive understanding of the role that GAI can play in nursing education, balancing its potential benefits against the ethical and practical challenges it presents.

Aims, objectives and research question

The aim of this study was to explore the feasibility and acceptability of using GAI to provide feedback on nursing assignments.

The objectives were:

- To identify whether GAI is comparable to a lecturer when marking nursing assignments.
- To explore the views of nursing students and educators regarding using GAI as a tool to mark assignments.

Methods

Design

Phenomenology, an inductive, qualitative research method, aims to describe certain phenomena from the perspective of those who have experienced it (Teherani et al., 2015). This approach endeavours to ascertain meaning behind lived experiences and requires the researcher to scrutinise the phenomenon without any predetermined expectations (Neubauer et al., 2019). This qualitative method was selected to accumulate rich data and to gain a deep insight into student and staff opinions on using GAI as a tool for marking assignments (Rodriguez et al., 2018).

Inclusion and exclusion criteria

Given the limited resources and time constraints placed upon the project, convenience sampling was used for practicality (Elfil et al., 2017) - by sending out emails and announcements to staff and students. All lecturers were employed by the nursing and midwifery department at the University of Birmingham. They were diverse in age, experience, gender and ethnicity to obtain a range of perspectives and came from any background of nursing whether, paediatric, adult or mental health. Focus groups, composed of pre-registration nursing students across two universities in Birmingham. Students invited to participate could be of any age, ethnicity and gender and ranged from first to fourth years within any field of nursing. Given the qualitative nature of the study, a sample size calculation was not performed. Data collection stopped when data saturation was achieved, referring to the point at which no new insights are identified, ensuring

validity and robustness (Vasileiou et al., 2018). Data saturation was reached when data began to repeat itself making further exploration redundant (Hennink et al., 2022).

Data collection

Data was gathered through in-person student focus groups and online lecturer interviews, using a series of open-ended questions regarding their views on the use of GAI as a marking tool. Although interview and focus group structures were created to provide guidance, enhance replicability and reduce the likelihood of asking leading questions (Cairns-Lee et al., 2022), a semi-structured approach was taken allowing the interviewee to delve deeper into the topic area and obtain rich, detailed data (DeJonckheere et al., 2019).

One-to-one interviews were conducted with academics due to avoid power dynamics between lecturers that may affect openness (Bullock, 2016). This also allowed for flexibility, as the researcher could change the line of questioning and thoroughly explore the participant's views. Furthermore, this approach gave the opportunity to ask follow-up questions for further elaboration (Alamri, 2019). Focus groups with students were held, to stimulate discussion and debate (Leung et al., 2009). Focus groups bring members of the study population together in a moderated environment, capitalising on communication to generate data about a specific topic (Nyumba et al., 2018). Less intimidating than an interview, focus groups are quick and gather detailed data that allows for a rich blend of perspectives (Tausch et al., 2016).

To generate questions for the interviews and focus groups, it was important to define clear research objectives and understand the target audience, taking factors such as their background and level of education into consideration (George, 2023). A literature review identified gaps in current understanding to ensure questioning was both relevant and contributed towards deepening current knowledge (Müller-Bloch et al., 2015). To avoid yes or no answers, open-ended questions were formulated to encourage richly detailed answers. Each of the researchers independently composed a unique set of questions and through discussion the most favourable were chosen, fine-tuned and organised into a logical sequence to form the basis of the interviews.

At the end of the questioning in interviews and focus groups, participants were presented with two sets of feedback, one written by a lecturer and the other by GAI (see Supplementary data). Participants were asked to comment on both pieces of feedback and state which they would prefer to receive. To obtain the GAI-generated feedback, a 2000-word essay, written by a consenting nursing student, was uploaded to the free version of ChatGPT 3.5. To meet the word limit of ChatGPT 3.5, citations and references were deleted. To ensure a fair comparison, the learning objectives and marking rubric were also provided to ChatGPT, enabling it to generate feedback aligned with the same criteria used by the human assessor. All interviews and focus groups were recorded and transcribed to enable analysis.

Data analysis

Interviews and focus groups were recorded using Microsoft Teams and automatically transcribed into Word documents, then checked for errors. Braun and Clarke's (2022) thematic analysis framework were used to analyse the data because it enables thorough interpretation and collation of the data into descriptive themes, synthesising key aspects from participants. The first step involved becoming familiarised with data, whereby researchers reviewed the interview and focus group transcripts, allowing them to become acquainted with each participant interaction. At this stage, it was decided that findings from interviews and focus groups would be analysed together, given the large cross-over yielded.

Following this, each researcher independently identified transcribed extracts and labelled them with a code.

Researchers met together to compare shared codes which were combined to construct overarching themes, encapsulating key narratives within the dataset. Themes were reviewed and finalised in an ongoing iterative process of merging, creating and disregarding codes and themes. The final four themes were 'evaluation and governance', 'human touch', 'development and integration' and 'time'. Written reporting of themes enabled refinement as an integral part of the analytic process. Data extracts from individual interviews or focus groups were used to illustrate and evidence analytic claims under each theme.

Ethical considerations

Ethical approval was obtained from the School of Nursing and Midwifery Research Ethics Committee. Participant information forms detailing the aims, the benefits of taking part and how confidentiality and anonymity would be maintained were sent to students and lecturers. Before participating, we obtained both verbal and written consent from the focus group and interview participants. We informed them that their data could be removed at any point before data analysis, ensuring they were aware of their right to withdraw.

It is essential that all researchers uphold the guidance within the Data Protection Act (2018) to circumvent unauthorised access, disclosure, destruction or alteration of data. Adhering to these regulations, participants were given pseudonyms, files were only shared amongst the researchers and supervisors, and once data was no longer required it was deleted.

Results

Characteristics of participants

During a four-week data collection period in November 2023, six semi-structured, 30-minute interviews with nursing lecturers were conducted, as well as two focus groups involving eight nursing students from two universities. Students were all in their final year of study, four of whom were from the University of Birmingham and the other four from Birmingham City University. Seven students were studying paediatric nursing and one was in the adult field.

Findings

This section presents a description of the findings including a comparison between GAI- and lecturer-generated feedback under four key themes: Evaluation and Governance; Human Touch; Development and Integration; and Time.

Comparison between GAI and lecturer feedback

Participants found it easy to determine which feedback was GAI-generated. GAI-generated feedback was described as 'repetitive', 'lengthy' and heavily reliant on the language used within the assignment, with its 'verbosity' 'putting people off' reading it. However, it was 'detailed', 'comprehensive' and 'referred back to the learning outcomes' offering suggestions for improvement. Its objectiveness was highlighted with automated feedback making 'direct comments about the essay' as opposed to the student themselves.

Lecturer feedback was 'brief', digestible and linked to certain aspects of the essay but lacked detail. Comments did 'not tell you how to improve' or how to 'be more critical of the literature'. Although the academic made 'positive comments' there was no 'justification of the mark' with many stating this feedback was 'unhelpful' and one participant stating they were 'embarrassed' by the standard. When asked which they would prefer to receive, 100% of students chose GAI generated feedback, compared to 83% of staff.

Evaluation and governance

Given the robust quality assurance process within the university and the importance of academic integrity, a key theme was 'evaluation and governance'. To reach these standards, multiple participants recognised the need for this technology to be 'monitored and evaluated' with staff 'overseeing the work that GAI does' or being 'moderated by members of the academic team'.

In addition to meeting university standards, one participant also stated the need to 'reassure the Nursing and Midwifery Council (NMC) that we are meeting standards for supervision, support and assessment'. Feedback and marking 'standardisation' was deemed essential to maintain the 'quality and governance around assessment'. Whilst some believed there might be outcry from the NMC, others articulated this contemporary method of marking would have to be NMC approved, meeting their guidelines prior to its rollout.

There was concern from staff regarding how GAI would handle 'extenuating circumstances or grievances and undergraduates expressed the fear that they would be 'more likely to fail' if graded by GAI. Challenges may emerge when students fail and 'need to go to an appeal panel', prompting participants to enquire about the capacity of GAI to withstand human scrutiny. There was apprehension regarding the accountability of errors made by GAI, with partakers questioning 'who ultimately has responsibility for the decisions that GAI arrives at?' lecturers, software developers or artificial intelligence itself?

Staff mentioned the moderation process where one individual is allocated to moderate assignments marked by multiple assessors, promoting standardisation. Some remarked that, given the objectivity of GAI, marking standardisation would likely improve, resulting in 'consistent feedback', instead of using the moderation process as a 'second check'. However, one interviewee highlighted that students would still see a disparity between lecturers if GAI was used, as some would go way beyond their expectations, investing more time and effort into using it as a tool. Several ethical issues were discussed, with students' primary concerns centred on the possibility that reflective essays, in which they examine their own poor practice, may be 'held against them' and possibly affect their future employment. Respondents were more willing to have GAI grade academic essays instead of reflective pieces which are based on 'personal experience'. Data security was another ethical issue and lecturers stressed the importance of being 'transparent' with students regarding the use of GAI.

Human touch

A recurring theme was concern that using GAI would result in a loss of human touch as it lacks 'any kind of human intuition or substance'. Lecturers were concerned they would 'lose sight of their students' abilities', whilst students worried there would be 'more of a separation' between staff and students. Students believed that the academics teaching the module would have 'more of an insight' than GAI as well as 'experience in nursing' and therefore be better equipped to mark assignments. Staff conveyed the importance of providing 'personalised and respectful' feedback to students who have put 'time and effort into writing the work', instead of GAI generated feedback which they believed could feel 'soulless' and 'hollow'.

Despite the overwhelming consensus that 'students deserve individualised feedback' 'written uniquely for them', one participant indicated some lecturers simply 'cut and paste in stock phrases.' Therefore, feedback written by academics is not always bespoke and may already lack the human touch.

Whereas a GAI marking tool may work for other university programmes, nursing lecturers emphasised their desire to 'do right by people' and 'make a difference'. Several staff members believed the implementation of GAI may prevent them from making this positive contribution and cultivating the next generation of

healthcare workers.

Assignments at the university are usually marked anonymously, but this can be hard to achieve because lecturers form a rapport with their students and 'can by and large work out who wrote it'. As one lecturer put it: 'As a lecturer, it can be quite hard to mark without personal interest'. Using GAI to mark assessments could enhance anonymity, providing 'more objective feedback' and may remove the narrative that students have been treated unfairly. In addition to reducing 'personal judgement' from lecturers, it was also suggested that GAI could provide students with more 'transparency' regarding the marking process and where exactly 'the marks are being gained and lost'.

Development and integration

Many questioned how a GAI marking tool would be developed and implemented. Most participants agreed that 'using general software like ChatGPT' would be impractical given 'that it is a very generalised programme' and would not 'take into account the NMC'. Therefore, the university would have to develop their own. One lecturer noted that GAI would have to be nursing 'programme specific', allowing for the incorporation of 'professional regulations' and nursing standards. The tool would require a deep insight into 'critical discussion' and 'evaluation of data', however there was hesitation amongst lecturers about how the GAI would respond when students go 'beyond the scope of even the assessor's knowledge'.

Although GAI is already used within the university, to mark 'multiple choice exams' and 'create case studies', there was a lack of knowledge regarding this technology. Students feared that GAI 'would expect things to be worded a certain way' and were unsure how GAI would be able to grade their work. Staff expressed the concern that they could use a GAI tool incorrectly, whilst others were unconvinced that the introduction of such a tool would be welcomed, with many academics being 'stuck in their ways'. One interviewee mentioned that alternate marking methods such as 'VoiceThread [software for providing verbal feedback]' have been trialled, but not adopted and GAI may meet a similar fate.

Though the importance of embracing 'innovation' was conveyed, fourth year students were anxious that they would have to 'consider a different type of writing style' to appease GAI and therefore it would be better introduced during the first year. Students in the focus groups suspected that GAI marking would favour GAI writing and students may be tempted to submit GAI-generated essays in a bid to achieve a higher grade. Staff and students seemed more open-minded at the prospect of using GAI to mark and provide formative feedback, as this often requires greater detail regarding 'syntax, grammar, structure and level of discussion' to assist the student with their summative piece. The use of GAI as a 'second check' was also suggested. However, it was pointed out that if students desired the additional feedback produced by GAI, then they could simply run it through the software themselves.

Time

When asked about possible advantages of using GAI as a tool for feedback and marking, the most prevalent response amongst staff was that it would 'reduce workloads', saving time which could be spent elsewhere. It was also suggested that GAI could be used for 'specific aspects' of the marking process, allowing 'additional time to create more effective feedback'. Although allocated the same number of calendar days to mark assignments, the actual time lecturers spent marking varied widely. Ultimately, marking takes up a 'considerable amount of time', which could be used for teaching, research or supporting students. Additionally, GAI may also reduce disparity between lectures, providing more 'objective feedback' whilst speeding up the moderation process. By increasing the rate at which essays are assessed, feedback can be given to students in a timely fashion, allowing them to improve their academic skills, incorporating

them into the next submission. While it is essential for assessment feedback to be prompt, it becomes inconsequential if the commentary is poor. One lecturer theorised that, if staff are able to mark more efficiently, this may result in an increased number of assessments students are set, giving lecturers more work to mark, creating an endless cycle.

Though most of the lecturer and student participants were convinced that GAI would increase the speed of grading, one participant mentioned that 'it probably wouldn't save an awful lot of time for quite a while, because we'd have to be trained' and furthermore, learning how to use the technology within the strict university guidelines would be labour-intensive. There was also discussion amongst students questioning whether we should pay lecturers if they were to use GAI to mark essays, when students are expected to dedicate time and effort into writing them, with many stating this seems unfair.

Discussion

This phenomenological study provides an insight into staff and student views regarding the use of GAI to mark nursing assignments. Findings were consistent with prior research on the use of GAI in academia yet filled a gap, given the lack of phenomenological research with nursing staff and student views. From the interviews and focus groups, it was clear that there was limited knowledge about GAI and this discourse, as evidenced by the findings, is characterised by a cautious optimism tempered by concerns about governance, fairness and the preservation of the human touch. Despite these concerns, the need to embrace innovation was apparent.

As an essential part of the learning process, feedback informs students about the quality of their performance, supports decision making and enhances professional and educational development. Within nursing, constructive criticism encourages reflective learning, closing the gap between actual and anticipated performance (Burgess et al., 2020). To have the desired effect, feedback must be explicit, descriptive, specific and honest (Bienstock et al., 2007), criteria demonstrated by the GAI-generated feedback in this study, but not that from the lecturer, due to insufficient detail. However, to be acted upon, feedback must also be valued. Hardavella et al (2017) found that advice given by a perceived role model holds greater value, suggesting students may be more likely to disregard GAI-generated feedback. Withey (2013) highlighted that whilst students recurrently criticise the quality of feedback, they continually make inadequate use of it. Our research showed that almost all participants favoured the GAI generated comments. However, participants were presented with only one example of each type of feedback meaning the notion that GAI feedback is more favourable may lack external validity due to potential discrepancies between academics' feedback (Schinske et al., 2014). With factors such as bias, fatigue and stress resulting in unreliable marking, GAI could eliminate these inconsistencies, increasing objectivity (Mumford et al., 2021; Ellis, 2022). Despite GAI feedback being more descriptive than lecturers' feedback, it is not as valued by students. Students may be more likely to accept constructive criticism from GAI if it is used as a tool, not a replacement, and if GAI can increase fairness through reduced marking discrepancy.

The word 'assess' derives from the Latin word 'assidere' which translates 'to sit beside', emphasising the importance of teacher-student relationships (Swaffield, 2011). A recurring theme throughout our study was the lack of human touch that marking with GAI would result in. Nevertheless, GAI is gradually becoming more human-centric due to artificial neural networks, natural language processing and deep learning (Dai et al., 2022). GAI has unique benefits because it is not affected by human limitations or errors and can provide consistent feedback (Mumford et al., 2021). GAI is already used to grade multiple choice questions, given their simplicity, but there is hesitancy regarding how effective GAI would be at marking short answer questions and essays (Ramesh et al., 2022). In this study, there was even greater apprehension concerning

the ability for GAI to mark reflective essays, which are based on experience and self-awareness, instead of academics who can draw on their own practice (Ullmann, 2019). Although GAI receives criticism for a lack of human touch, it is considered a technology that is accurate, reliable and consistent, with Korteling et al. (2021) reporting that human teams could be enhanced with GAI, resulting in fewer cognitive constraints and biases, as these systems excel in selecting and processing large amounts of data. Technological advancements have resulted in the creation of new GAI assisted essay grading platforms, such as Graide (2023), developed by the University of Birmingham. Graide (2023) works alongside academics, offering suggestions for feedback, possibly giving students their preferred amalgamation. This study suggests that participants favour GAI as a supplementary tool, rather than a replacement for lecturers in the marking and feedback process, yet they prefer GAI written feedback. Applications such as Graide (2023) may provide the optimal combination.

During the Covid-19 pandemic, lockdowns resulted in educators having to teach remotely. Despite multiple benefits including increased flexibility and elimination of costs related to commuting, students believed it had an adverse impact on their learning experience and felt less motivated to engage with teaching (Wang et al., 2018). In a study by Serhan (2020), only 9.68% of students felt Zoom improved their learning and 22.58% enjoyed using it. Students felt they were paying too much for the poor quality of education and felt disconnected from their lecturers. These concerns were reflected in our study, with the assumption that the use of GAI for marking and feedback could result in student disengagement and upset. Additionally, to slow the spread of coronavirus, a multitude of human responsibilities were replaced by GAI (Lauri et al., 2023), causing concern that GAI is creating mass unemployment, with the projection that GAI will replace five million jobs in the United States alone (Cerullo, 2023). Our research highlights that whilst participants did not believe human markers would be completely replaced by GAI, it was speculated there may be outcry from students and the public who believe lecturers are not doing the job they are paid to do, especially given the 2012 increase in tuition fees to £9,000 per year (Sá, 2014).

Students had a multitude of concerns regarding the use of GAI to mark their work, believing it would be incapable of marking reflective essays and they would have to alter their style of writing to appease a GAI marker. Throughout their training, nurses are expected to adhere to the NMC Code (NMC, 2018) and those responsible for teaching them are required to do the same. The Code (NMC, 2018) encourages healthcare professionals to respond to people's preferences and concerns, highlighting the need for student apprehensions to be acknowledged. Whilst the NMC would likely ensure that GAI used for assessment aligns with their standards which accentuate integrity, they may question how effectively GAI is able to assess sensitive and subjective topics such as patient care and decision-making.

Considering the cost of university for students in some parts of the UK, it is unsurprising that students demand a high-quality learning experience. According to the 2024 National Student Survey – an annual survey of final-year undergraduate students in the UK - 81% of respondents believed the marking and assessment had been fair, although this ranged from 36% to 100% across institutions (Office for Students, 2024). However, when students were asked how often feedback helped to improve work, only 72% responded positively. Students vocalise the need for clear feedback that facilitates reflection, allowing them to minimise the gap between current and desired educational performance before their next assignment (Blair et al., 2013). When questioned on the advantages of using GAI as a marking tool, the most notable answer amongst staff and students was the increased efficiency of marking, allowing educators to focus on other aspects of education, whilst students would have the more efficient assessment turnaround they want. Alam (2021) found that 40% of teaching time is focused on activities that could be computerised, one of which is marking and staff in this study emphasised that it is time-consuming and labour-intensive. Given

GAI efficiency and accuracy, research participants encourage the idea of lecturers working alongside GAI. A similar conclusion was reached by Mizumoto et al (2023) who highlighted the greatest benefits of using GAI systems may be in conjunction with human markers.

Strengths and limitations of the work

The inclusion of both lecturers and students was a notable strength, as it acknowledged the importance of considering the views of all stakeholders (Nyanchoka et al., 2019). In this study, participants included those who would be directly affected by the possible implementation of GAI in marking. Additionally, the use of Braun and Clarke's (2022) thematic analysis framework ensured a systematic and rigorous approach to data analysis, enhancing the credibility and trustworthiness of the findings.

A significant limitation of the study was that it was conducted by novice researchers who lacked expertise and confidence. Additionally, social desirability bias may have influenced participants' responses, leading them to answer in ways they believed would be viewed favourably by others (Bergen et al., 2020).

External generalisability, the degree to which findings can be applied to a wider context, may have been impeded due to the fact that participants were all studying or teaching within one of two universities and not further afield. The sample size was small with most participants studying children's nursing, further reducing generalisability (Murad et al., 2018). Furthermore, only one example of lecturer feedback was presented, so external validity is limited and views on the quality of academic feedback cannot be generalised due to expected variance between human markers.

Recommendations for further research

Given the contemporary nature of GAI, research into the topic is scarce, requiring further exploration. Whilst this study scrutinised feedback given by GAI in comparison to lecturers, future work could compare the overall grades given by academics to those generated by GAI. Future research could assess the ability of software such as Graide (2023) to provide feedback for nursing assignments by comparing it with general applications, such as ChatGPT and lecturer marking. Further investigation is needed on a more diverse sample to generalise the results to a wider audience (Tiokhin et al., 2019).

The results obtained from GAI models such as ChatGPT heavily rely on the formulated prompts given by a user (Hasse and Hanel, 2023). Due to time constraints, this study only utilised a single prompt to generate feedback. Future research may want to compare how feedback differs when different prompts are given. While we explored students' initial opinions on the topic, it would be informative to introduce this marking method in an educational setting to measure their motivation and engagement with assignments.

Conclusion

Most participants favoured the GAI-generated feedback, but there were concerns about how the software would be regulated and aligned with NMC standards. The potential loss of the human touch in the marking process was a concern, with educators expressing concerns about losing track of student progress. However, the increased objectivity and efficiency that GAI could bring to marking was also acknowledged, potentially eliminating the perception of unfair treatment among students. The enhanced objectivity also raised the possibility of streamlining or even removing the moderation process. Other research suggests that GAI can be influenced by the quality of its training data. While there were doubts that GAI could fully replace human markers, many participants were open to its use as a marking tool to support lecturers or to provide

formative feedback on practice assignments.

References

Al-shamasneh, A. R. M. and Obaidellah, U. H. B. (2017) 'Artificial intelligence techniques for cancer detection and classification: Review study'. *European Scientific Journal*, 13(3), pp. 342-370. https://doi.org/10.19044/esj.2017.v13n3p342

Amisha, Malik, P., Pathania, M. and Rathaur, V. (2019) 'Overview of artificial intelligence in medicine'. *Journal of Family Medicine and Primary Care*, 8(7), pp. 2328-2331. https://doi.org/10.4103/jfmpc.jfmpc_440_19

Alam, A. (2021) 'Possibilities and apprehensions in the landscape of artificial intelligence in education'. In *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, Nagpur, India, pp. 1-8. https://doi.org/10.1109/ICCICA52458.2021.9697272

Alamri, W. A. (2019) 'Effectiveness of qualitative research methods: Interviews and diaries'. *International Journal of English and Cultural Studies*, 2(1), pp. 65-70.

Baker, R.S. and Hawn, A. (2021) 'Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32, pp. 1052-1092. https://doi.org/10.1007/s40593-021-00285-9

Bergen, N. and Labonté, R. (2020) '"Everything is perfect and we have no problems": Detecting and limiting social desirability bias in qualitative research'. *Qualitative Health Research*, 30(5), pp. 783-792. https://doi.org/10.1177/1049732319889354

Bienstock, J. L., Katz, N. T., Cox, S. M., Hueppchen, N., Erickson, S. and Puscheck, E. E. (2007) 'To the point: Medical education reviews - providing feedback'. *American Journal of Obstetrics and Gynecology*, 196(6), pp. 508-513. https://doi.org/10.1016/j.ajog.2006.08.021

Blair, A., Curtis, S., Goodwin, M. and Shields, S. (2013) 'What feedback do students want?'. *Politics*, 33(1), pp. 66-79. https://doi.org/10.1111/j.1467-9256.2012.01446.x

Boud, D. and Falchikov, N. (eds.) (2007) *Rethinking assessment in higher education*. London: Routledge.

Braun, V. and Clarke, V. (2022) 'Conceptual and design thinking for thematic analysis'. *Qualitative Psychology*, 9(1), pp. 3-26. https://doi.org/10.1037/qup0000196

Bullock, A. (2016) 'Conduct one-to-one qualitative interviews for research'. *Education for Primary Care*, 27(4), pp. 330-332. https://doi.org/10.1080/14739879.2016.1176874

Burgess, A., van Diggele, C., Roberts, C. and Mellis, C. (2020) 'Introduction to the Peer Teacher Training in health professional education supplement series'. *BMC Medical Education*, 20, 454. https://doi.org/10.1186/s12909-020-02279-y

Cairns-Lee, H., Lawley, J. and Tosey, P. (2022) 'Enhancing researcher reflexivity about the influence of leading questions in interviews'. *The Journal of Applied Behavioral Science*, 58(1), pp. 164-188. https://doi.org/10.1177/00218863211037446

Cerullo, M. (2023) *Here's how many U.S. workers ChatGPT says it could replace*. MoneyWatch. Available at: https://www.cbsnews.com/news/chatgpt-artificial-intelligence-jobs/

Chaudhry, M. A. and Kazim, E. (2022) Artificial Intelligence in Education (AIEd): A high-level academic and industry note 2021. *AI Ethics*, 2, pp. 157–165. https://doi.org/10.1007/s43681-021-00074-z

Chen, L., Chen, P. and Lin, Z. (2020) Artificial intelligence in education: A review'. *Ieee Access*, 8, pp. 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Dai, C. P. and Ke, F. (2022) 'Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review'. *Computers and Education: Artificial Intelligence*, 3, 100087. https://doi.org/10.1016/j.caeai.2022.100087

Data Protection Act 2018 (c.12), London: The Stationery Office. Available at: https://www.legislation.gov.uk/ukpga/2018/12/contents

DeJonckheere, M. and Vaughn, L. M. (2019) Semistructured interviewing in primary care research: a balance of relationship and rigour. *Family Medicine and Community Health*, 7(2), p. e000057. https://doi.org/10.1136/fmch-2018-000057

Elfil, M. and Negida, A. (2017). Sampling methods in clinical research; an educational review. *Emergency*, 5(1), p. e52. https://doi.org/10.22037/emergency.v5i1.15215

Ellis E. (2022) 'The potential of artificial intelligence in assessment feedback'. *Times Higher Education*. Available at: https://www.timeshighereducation.com/campus/potential-artificial-intelligence-assessment-feedback

George, G. (2023) 'Understanding your target audience by profiling a typical member', In S. Rowland (ed.) *Teaching science students to communicate: A practical guide*. Cham: Springer International Publishing, pp. 129–135.

Gherheș, V. (2018) 'Why are we afraid of artificial intelligence (AI).' *European Review of Applied Sociology*, 11(17), pp. 6-15. https://doi.org/10.1515/eras-2018-0006
Gibson, S., Oliver, L. and Dennison, M. (2015) *Workload challenge: Analysis of teacher consultation responses* (Vol. 355). London: Department for Education.
Gocen, A. and Aydemir, F. (2021) 'Artificial intelligence in education and schools'. *Research on Education and Media*, 12(1), pp. 13-21. https://doi.org/10.2478/rem-2020-0003
Graide (2023) *Improve your grading process with bespoke AI*. Available at: https://www.graide.co.uk/ Accessed: 11 December 2023
Hardavella, G., Aamli-Gaagnat, A., Saad, N., Rousalova, I. and Sreter, K. B. (2017) 'How to give and receive feedback effectively'. *Breathe*, 13(4), pp. 327–333. https://doi.org/10.1183/20734735.009917
Haseski (2019) What do Turkish pre-service teachers think about artificial intelligence, *International Journal of Computer Science Education in Schools,* 3(2). https://doi.org/10.21585/ijcses.v3i2.55
Hassan, A., Ali, S.A. and Abdulkhaleq, N.M.S. (2023) 'Artificial intelligence in educational examinations', In A.M.A. Musleh Al-Sartawi, A. Razzaque and M.M. Kamal (eds) *From the internet of things to the internet of ideas: The role of artificial intelligence*. EAMMIS 2022. Cham: Springer, pp. 73–85. https://doi.org/10.1007/978-3-031-17746-0_7
Haase, J. And Hanel, P. (2023) 'Artificial muses: generative artificial intelligence chatbots have risen to human-level creativity', *Journal of Creativity,* 33(3), 1000066. https://doi.org/10.1016/j.yjoc.2023.100066/
Hennink, M. and Kaiser, B. N. (2022) 'Sample sizes for saturation in qualitative research: A systematic review of empirical tests'. *Social Science and Medicine*, 292, 114523. https://doi.org/10.1016/j.socscimed.2021.114523
Holmes, W., Bialik, M. and Fadel, C. (2019) *Artificial intelligence in education: The social consequences of AI in education.* Boston, MA: The Centre for Curriculum Redesign
IDC (2023) *Worldwide spending on AI-centric systems forecast to reach $154 billion in 2023.* Available at: https://www.idc.com/getdoc.jsp?containerId=prUS50454123 Accessed: 7 August 2023
IBM (2023) *What is Generative AI?* Available at: https://www.ibm.com/topics/generative-ai Accessed: 16 August 2024
Korteling, J. E. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C. and Eikelboom, A. R. (2021) Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4, 622364. https://doi.org/10.3389/frai.2021.622364
Kumar, R. (2023) 'Faculty members' use of artificial intelligence to grade student papers: a case of implications'. *International Journal for Educational Integrity*, 19, 9. https://doi.org/10.1007/s40979-023-00130-7
Lauri, C., Shimpo, F. and Sokołowski, M. M. (2023) 'Artificial intelligence and robotics on the frontlines of the pandemic response: the regulatory models for technology adoption and the development of resilient organisations in smart cities'. *Journal of Ambient Intelligence and Humanized Computing*, 14, pp. 14753-14764. https://doi.org/10.1007/s12652-023-04556-2
Leung, F. H. and Savithiri, R. (2009) 'Spotlight on focus groups'. *Canadian Family Physician Medecin de Famille Canadien*, 55(2), pp. 218–219.
Malik, S. and Gangopadhyay, A. (2023) 'Proactive and reactive engagement of artificial intelligence methods for education: A review'. *Frontiers in Artificial Intelligence*, 6. https://doi.org/10.3389/frai.2023.1151391
MIT News (2023) *Explained: Generative AI.* Massachusetts Institute of Technology. Available at: https://news.mit.edu/2023/explained-generative-ai-1109 Accessed: 13 August 2024
Mizumoto, A. and Eguchi, M. (2023) 'Exploring the potential of using an AI language model for automated essay scoring'. *Research Methods in Applied Linguistics*, 2(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050
Müller-Bloch, C. and Kranz, J. (2015) 'A framework for rigorously identifying research gaps in qualitative literature reviews', In *ICIS 2015 Proceedings: Research Methods*. Available at: https://aisel.aisnet.org/icis2015/proceedings/ResearchMethods/2
Mumford, S. and Atay, D. (2021) 'Teachers' perspectives on the causes of rater discrepancy in an English for academic purposes context'. *Assessing Writing*, 48, 100527. https://doi.org/10.1016/j.asw.2021.100527
Murad, M. H., Katabi, A., Benkhadra, R. and Montori, V. M. (2018) 'External validity, generalisability, applicability and directness: A brief primer'. *BMJ Evidence-Based Medicine*, 23(1), pp. 17-19. https://doi.org/10.1136/ebmed-2017-110800
Nah, F., Zheng, R., Cai, J., Siau, K. and Chen, L. (2023) 'Generative AI and ChatGPT: Applications, challenges and AI-human collaboration'. *Journal of Information Technology Case and Application Research*, 25(3), pp. 277-304. https://doi.org/10.1080/15228053.2023.2233814

Neubauer, B. E., Witkop, C. T. and Varpio, L. (2019) 'How phenomenology can help us learn from the experiences of others'. *Perspectives on Medical Education*, 8(2), pp. 90–97. https://doi.org/10.1007/s40037-019-0509-2

Nursing and Midwifery Council (2018) *The code: Professional standards of practice and behaviour for nurses, midwives and nursing associates.* Available at: https://www.nmc.org.uk/standards/code/ Accessed: 20 March 2023

Nyanchoka, L., Tudur-Smith, C., Porcher, R. and Hren, D. (2019) 'Key stakeholders' perspectives and experiences with defining, identifying and displaying gaps in health research: A qualitative study protocol'. *BMJ Open*, 9(8), p. e027926. https://doi.org/10.1136/bmjopen-2018-027926

Nyumba, T.O., Wilson, K., Derrick, C. J. and Mukherjee, N. (2018) 'The use of focus group discussion methodology: Insights from two decades of application in conservation'. *Methods in Ecology and evolution*, 9(1), 20-32. https://doi.org/10.1111/2041-210X.12860

Office for Students (2024) *Download the NSS data*. Available at: https://www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/download-the-nss-data/ Accessed: 19 August 2024

Pitt, E. and Norton, L. (2017) "Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it'. *Assessment and Evaluation in Higher Education*, 42(4), 499-516. https://doi.org/10.1080/02602938.2016.1142500

Ramesh, D. and Sanampudi, S. K. (2022) 'An automated essay scoring systems: A systematic literature review'. *Artificial Intelligence Review*, 55(3), pp. 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Rodriguez, A. and Smith, J. (2018) 'Phenomenology as a healthcare research method'. *Evidence-Based Nursing*, 21(4), pp. 96–98. https://doi.org/10.1136/eb-2018-102990

Sá, F. (2014) *The effect of tuition fees on university applications and attendance: Evidence from the UK*. Institute for the Study of Labor, (8364). Available at: https://docs.iza.org/dp8364.pdf

Schinske, J. and Tanner, K. (2014) 'Teaching more by grading less (or differently)'. *CBE Life Sciences Education*, 13(2), pp. 159–166. https://doi.org/10.1187/cbe.cbe-14-03-0054

Serhan, D. (2020) 'Transitioning from face-to-face to remote learning: Students' attitudes and perceptions of using zoom during COVID-19 pandemic'. *International Journal of Technology in Education and Science*, 4(4), pp. 335-342. https://doi.org/10.46328/ijtes.v4i4.148

Swaffield S (2011) 'Getting to the heart of authentic assessment for learning'. *Assessment in Education: Principles, Policy & Practice,* 18(4), pp. 433–449. https://doi.org/10.1080/0969594X.2011.582838

Tausch, A. P. and Menold, N. (2016) 'Methodological aspects of focus groups in health research: results of qualitative interviews with focus group moderators'. *Global Qualitative Nursing Research*, 3. https://doi.org/10.1177/2333393616630466

Teherani, A., Martimianakis, T., Stenfors-Hayes, T., Wadhwa, A. and Varpio, L. (2015) 'Choosing a qualitative research approach'. *Journal of Graduate Medical Education*, 7(4), pp. 669-670. https://doi.org/10.4300/JGME-D-15-00414.1

Ullmann, T. D. (2019) 'Automated analysis of reflection in writing: Validating machine learning approaches'. *International Journal of Artificial Intelligence in Education*, 29(2), pp. 217-257. https://doi.org/10.1007/s40593-019-00174-2

Tiokhin, L., Hackman, J., Munira, S., Jesmin, K. and Hruschka, D. (2019) 'Generalizability is not optional: Insights from a cross-cultural study of social discounting'. *Royal Society Open Science*, 6(2), 181386. https://doi.org/10.1098/rsos.181386

Vasileiou, K., Barnett, J., Thorpe, S. and Young, T. (2018) 'Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period'. *BMC Medical Research Methodology*, 18, 148. https://doi.org/10.1186/s12874-018-0594-7

Wang, Q., Huang, C. and Quek, C. L. (2018) 'Students' perspectives on the design and implementation of a blended synchronous learning environment'. *Australasian Journal of Educational Technology*, 34(1). https://doi.org/10.14742/ajet.3404

Withey, C. (2013) 'Feedback engagement: Forcing feed-forward amongst law students'. *The Law Teacher*, 47(3), 319-344. https://doi.org/10.1080/03069400.2013.851336

Zhang, T. (2023) 'The contributions of AI in the development of ideological and political perspectives in education'. *Heliyon*, 9(3), pp. e13403–e13403. https://doi.org/10.1016/j.heliyon.2023.e13403