Chapter

Exploring AI Applications in Essay-Based Assignments: Affordances and Risks

Ahmad Alzahrani and Ying Zheng

Abstract

This study examined the feasibility of employing artificial intelligence (AI) for feedback provision on essay-based assignments in a UK Higher Education setting. Although the critical role of feedback in enhancing students' learning experiences is widely recognised, resource limitations and large student numbers often hinder its quality and timely delivery. Through in-depth interviews with four participants from a university in the UK, this research investigated AI applications in essay evaluation, utilising data from 12 AI-generated essays and their corresponding feedback. The aims of the study are to evaluate tutors' abilities in discerning human and AI-generated essays, as well as evaluating the quality of AI-generated feedback from their perspectives. Findings showed that assessors could detect certain characteristics consistent with AI generation and noted ethical concerns regarding deviations from academic standards. Participants also acknowledged AI's capacity for swift feedback delivery as compared to human. The results of this study help enhance our understanding of AI's affordances and risks in assessment and feedback, particularly in the less explored university essay assignments.

Keywords: artificial intelligence (AI), essay-based assignment, feedback, assessment, ethics

1. Introduction

Quality feedback is recognised as a key factor in improving students' learning experiences and achievements. However, due to resource limitations, providing timely and constructive feedback to many students is a challenging task. Automated feedback systems (AFS) are increasingly seen as a potential solution, but they have been less commonly applied to open-ended writing tasks, such as essay assignments and project proposals [1]. Recent advancements in Generative Pre-trained Transformer (GPT) models, particularly ChatGPT, offer new possibilities for enhancing AFS by providing more natural and context-specific individualised responses.

Additionally, ethical discussions are taking place within educational and scholarly realms. While taking advantages of the methodological advancements in AI, scholars emphasise the need for increased methodological rigour and ethical scrutiny

in practice [2]. The emergence of Large Language Model (LLM) powered by AI chatbots, such as OpenAI's, raises scholarly and practical concerns regarding their potential applications, ethical implications, and the distinction between AI- and human-produced texts [3].

Moreover, the widespread use of AI models among students and academics prompts questions about how tutors can adapt course contents and assessment methods to mitigate the impact of students' extensive reliance on them [4]. An alternative perspective suggests empowering educators to incorporate AI for various educational purposes, such as generating lecture topics, demonstrations, exam questions, assignments, content explanations of contents, ideation exercises, and grading essays or programming assignments [5–8].

Similarly, *tutoring*, recognised as a highly individualised and efficient method to improve student learning, faces a shortage of adequately trained tutors [9]. Although tutor training programmes have been developed, a significant gap exists as most programmes lack specific formative feedback, leaving a research void in tutors receiving feedback on their assessment methods. Researchers now advocate emphasising utilising pre-trained Large Language Models to give the tutors precise formative feedback on their tutoring practices, emphasising the assessment of the accuracy of AI-generated feedback in enhancing tutor learning and performance [10].

Integrating AI use as discussed above has been seen as a promising opportunity to improve feedback. This is because creating individualised feedback for assignments is an intricate task [11]. Additionally, recent research recognises the above-mentioned shortcomings of current evaluation approaches and suggests investigating the use of Large Language Models (LLMs) as a potential remedy for automating error identification and facilitating teacher assessments in classroom-based second language (L2) learners' writing assessment [12]. Moreover, research regarding the use of ChatGPT in providing scoring information acknowledges the limitation of directly employing pre-trained models like GPT-3.5 for tasks involving student language and underscores the necessity of fine-tuning on domain-specific data [13].

The study posits that integrating GPT models, like ChatGPT, into automated feedback systems can significantly improve the quality and relevance of feedback for open-ended writing tasks. The research design evaluates this by comparing GPT-based feedback with traditional methods, assessing its impact on accuracy, timeliness, and contextual relevance, while also exploring tutors' views on how AI can enhance their feedback practices. The chapter introduces the challenges of current feedback systems, presents GPT models as a potential solution by reviewing the literature on AI model use in feedback and assessment, and discusses their practical applications and ethical implications, guiding the reader through problem identification, solution evaluation, and future research directions.

2. Literature review

The literature review identified several gaps in understanding how humans detect AI-generated essays and assess the quality of AI-generated feedback from their viewpoints. First, existing studies focused mainly on identifying AI-generated abstracts (e.g. not full essays), which limited understanding due to the greater complexity of full essays. Second, ethical considerations and acceptance of AI tools in education, though gaining more and more attention, have been underexplored empirically so far. Existing opinions vary and lack depth on issues like bias and transparency. Lastly,

most studies offered short-term views of AI feedback integration without considering long-term impacts on student learning, teacher practices, and educational quality. Longitudinal studies are needed to evaluate tools' sustainability and evolving effectiveness in education. Addressing these gaps can provide a more comprehensive understanding of AI's role in educational settings, enhancing efficacy of both AI and human contributions to learning and assessment.

In this regard, research revealed a low identification rate among reviewers of research abstracts while assessing scholars' capacity to differentiate between AI- and human-produced abstracts [14]. In a similar vein, these challenges were acknowledged as more significant in identifying AI use in student submissions. In this context, although introducing a novel keyword analysis revealed the potential of detecting ChatGPT's influence on student writings, the output was described as vague, calling for the need for more specific prompting in the detection process [4]. However, the focus on abstracts instead of full essays in these two previous studies limited their scopes, as abstracts did not capture the depth present in full essay assignments.

Similarly, de Winter et al. [4] addressed the challenges of conclusively identifying ChatGPT use in student submissions. The study concluded that although introducing a novel keyword analysis revealed the potential of detecting ChatGPT's influence on student abstracts and academic publications, they acknowledged challenges, such as vague outputs and the need for specific prompting. Moreover, the study by Dai et al. [1] explored the feasibility of using ChatGPT for providing written feedback on a data science project assignment in an Australian university. Their investigation focused on the clarity of the generated feedback, its alignment with instructor-provided feedback, and the inclusion of effective feedback elements. The evaluation included readability, agreement with human instructors using a marking rubric, and application of a theoretical feedback model. Their findings indicated that ChatGPT's feedback readability scores fell within the 3.75–4.0 range, outperforming over 75% of instructor feedback.

Examining assessors' capacity to identify AI is crucial, yet equally significant is evaluating the quality of feedback they provide. In this regard, investigating ChatGPT's feedback clarity, alignment with instructor feedback, and effectiveness revealed its superior scores, especially in providing process-focused feedback, surpassing instructor feedback [1]. Hirunyasiri et al. [10] looked into the ability of GPT-4 to precisely evaluate elements within effective praise given by human tutors to students. Their focus was on comparing the accuracy of GPT-4 assessments using zero-shot and few-shot chain of thought prompting approaches. Results showed that zero-shot and few-shot chain of thought methods produced similar outcomes in which GPT-4 moderately identified specific and immediate praise but struggled to recognise tutors' ability to deliver genuine praise, especially in the zero-shot prompting scenario.

A few studies looked into human voices in this matter. For example, Nguyen [15] investigated the perspectives of English as a Foreign Language (EFL) teacher at Van Lang University in Vietnam on integrating ChatGPT-4 for generating feedback in writing sessions. The study involved 20 EFL teachers who incorporated ChatGPT into language education, collecting quantitative and qualitative data through online surveys and structured interviews. The findings indicated a positive attitude among EFL teachers towards ChatGPT integration, emphasising the importance of professional training, enhancing user understanding of ChatGPT's limitations, and ensuring responsible chatbot usage for effective implementation in language classes. Cao and Zhong [16] investigated the effectiveness of ChatGPT-based feedback compared to traditional teacher feedback

and self-feedback in improving Chinese to English translation skills among Master of Translation and Interpretation students. The findings suggested that while traditional feedback methods outperformed ChatGPT in overall translation quality, ChatGPT-based feedback showed strengths in enhancing lexical proficiency and referential cohesion. Therefore, the potential of integrating ChatGPT as an additional resource in translation practice alongside traditional teacher-led methods is worth the effort.

Similarly, Pankiewicz and Baker [11] employed GPT-3.5 model to automate feedback generation for programming assignments, assessing its impact on students. The study compared an experimental group, receiving GPT hints, with a control group. Results indicated that students valued GPT-generated hints, leading to reduced reliance on regular feedback and improved performance in tasks with GPT hints. The experimental group also completed assignments more quickly for tasks with GPT hints.

Another important study, Bewersdorff et al. [12], explored the linguistic analysis of feedback, highlighting challenges in identifying logical errors in complex student experiment protocols. The research investigated the potential of LLMs to automatically identify errors in these protocols. The primary objective was to establish a foundation for generating personalised feedback, evaluating the AI system's accuracy in discerning both fundamental and complex errors, and its practical usability in education. Using a dataset of 65 student protocols, the study built an AI system based on GPT-3.5 and GPT-4, comparing its accuracy to those of human raters. The findings revealed varying levels of accuracy in error detection, with the AI system excelling in identifying fundamental errors but facing challenges with more complex errors. This study provided insights into LLMs' potential applications in education as well as LLMs' capabilities in detecting errors in enquiry-based learning.

3. Method

The four tutors in this study had varying levels of teaching experience, ranging from one year to over 15 years, with two being native English speakers and two non-native. Despite the small sample size, their diverse perspectives and substantial experience in essay evaluation made them suitable for in-depth qualitative analysis. All participants had experience marking essays from the specific postgraduate module under review. The group consisted of one female and three males. They were not informed of the nature of the 12 essays and feedback they were to assess prior to the study. Their task was to evaluate the quality of the essays, assign marks using the marking criteria, and then assess the corresponding AI-generated feedback.

In this study, several ethical considerations were carefully considered, particularly around consent processes and data privacy. First, informed consent was obtained from all tutor participants before conducting the interviews and essay evaluations. The participants were fully briefed about the purpose of the study, the nature of the tasks, and the use of AI-generated essays and feedback without revealing their source until after the assessment to ensure unbiased evaluations. Regarding data privacy, the participants' identities were anonymised to protect their confidentiality, with no personal identifying information disclosed in the reporting of the results. All data collected, including interview transcripts, essay evaluations, and AI feedback assessments, were securely stored, ensuring compliance with data protection regulations. Additionally, participants were given the option to withdraw from the study at any point, guaranteeing their autonomy throughout the research process. These ethical safeguards ensured the protection of participants' rights and privacy during the study.

Phase	Details
Research Design	• Qualitative case-study methodology • Focus on AI (GPT-3.5) application in essay evaluation • Semi-structured interviews with university tutors in the UK
Participants' selection	• 4 university tutors (1 female, 3 males) • Experience in essay marking from 1 to 15+ years • 2 Native English speakers, 2 Non-native speakers
Generating Essays & Feedback	• 12 essays generated using GPT-3.5 based on module guidelines • Essays divided into 3 levels: Excellent, Merit, Pass • AI-generated feedback for each essay
Evaluation Process	• Tutors tasked with evaluating 12 essays across 3 levels and topics • Compare AI feedback with traditional feedback methods • Assess the quality of AI-generated essays and feedback
Data Collection	• Semi-structured interviews • Annotated essays from tutor evaluations collected for comparative analysis

Table 1.
Summary of the research methods.

The study aimed to answer the two research questions, each intertwined with the exploration of AI in academic assessment. First, the study aimed to unveil the perception of human raters, seeking to discern their proficiency in comparing essays organically crafted by students and those generated by AI. Second, the study aimed to contrast traditional assessment modalities with AI-assisted paradigms, particularly focusing on marking and feedback provision within academic modules. A summary of the study methodology is presented in **Table 1**.

4. Results

Nine aspects of the findings are summarised in **Table 2**, which are elaborated in answering the three lines of enquiries in this study, they are, (1) tutors' ability

Aspects	Details	Examples	Contribution to outcome
AI Characteristics	AI-generated essays showed odd language and style.	Flowery language noted by Carl; poetic styles by Helen and Kyle.	Demonstrated the difficulty in distinguishing AI content due to non-academic language and style.
Unreliable Contents	Issues with accuracy and relevance of references.	Questionable references criticised by Omar and Kyle.	Revealed AI's limitations in producing credible, well-supported academic content.
Superficial Analysis	Essays lacked deep insights and meaningful analysis.	Superficial and template-like analysis noted by Helen and Omar.	Showed that AI essays often fail to engage critically with the material.
Not Academic Essays	Essays did not meet academic tone and style standards.	Non-academic language noted by Carl, Helen, and Kyle.	Highlighted the mismatch between AI-generated content and academic writing norms.

Aspects	Details	Examples	Contribution to outcome
Lack of Coherence	Poor logical structure and organisation in some essays.	Disjointed arguments observed by Carl and Omar.	Identified structural deficiencies, impacting readability and argument flow.
Lack of Criticality	Essays lacked critical analysis and had repetitive points.	Repetitive content noted by Kyle and Omar.	Demonstrated AI's failure to provide the critical engagement and original insights needed for high-quality writing.
Efficiency and Identifying Inconsistency	AI feedback was efficient and identified inconsistencies quickly.	Timely feedback for international students; inconsistency highlighted by Helen.	Showed AI's strength in providing rapid feedback and improving workflow efficiency.
Language Support	AI supported language acquisition and provided feedback.	Kyle approved AI tools for enhancing language skills.	Emphasised AI's role in supporting language learning and supplementing traditional methods.
Lack of Constructive Feedback	AI feedback lacked depth and formative guidance.	Inadequate feedback noted by Kyle.	Highlighted the need for human evaluators to provide nuanced, constructive feedback.

Table 2.
Summary of the study results.

to detect AI-generated essays; (2) tutors' evaluation of AI-generated essays against marking rubrics; and (3) tutors' evaluation of AI-generated feedback.

Details, examples, and contributions to the research outcome are provided in the summary, with the first three aspects addressing the first enquiry regarding human raters' ability to detect AI-generated essays, i.e., "AI characteristics", "unreliable contents", and "superficial analysis". "Not academic essays", "lack of coherence", and "lack of criticality" inform the second enquiry, followed by "efficiency and identifying inconsistency", "language support for students and teachers", and "lack of constructive feedback" addresses the third enquiry.

4.1 Tutors' ability to detect AI-generated essays

4.1.1 AI generated

Tutors agreed that some essays displayed characteristics consistent with AI generation, which raised important questions about the role of technology in academic writing. Carl, for example, showed scepticism about the essays' authenticity and authorship as he noted several features that could indicate AI involvement, including "odd language", "flowery expressions", and "inappropriate references".

> *Carl: Well, it's both kind of self-congratulatory saying how wonderful I am, and it's also a ridiculously flowery language*

Carl indicated that the language used in the essay is overly focused on praising the author or the subject matter. It could also imply that the author is more concerned with showcasing their own achievements or opinions rather than providing objective analysis or valuable insights. The term "flowery language" typically refers to writing

that is overly exaggerated with excessive use of metaphors, similes, adjectives, and other literary devices. Carl commented that while flowery language can sometimes enhance the beauty of prose, when used excessively, it may obscure the intended message or come across as pretentious, especially in academic writings.

Similarly, Helen expressed her suspicion about the essay's origins, lack of a clear line of argument, and critical evaluation as potential indicators of AI involvement. She suggested that the essay's poetic and idiomatic writing style may be characteristic of AI-generated content, especially when combined with the absence of coherent analysis. She commented that

> *I like this essay, in a good way. However, in terms of the language, it uses a lot of poetic and idiomatic writing. I think that's why it made me feel like it's machine generated, for me as a non-native speaker, I feel I am more sensitive to this type of language.*

Helen highlighted a specific aspect of the essay that she appreciates—its use of poetic and idiomatic language, which suggests if the essay employs creative and expressive language, they might enhance its appeal and make it more engaging to read to some readers. However, this part is interesting because it introduces a contrast. Despite appreciating the poetic and idiomatic writing, the tutor also felt that this style somehow gave the impression of being machine-generated. Furthermore, this tutor added a personal perspective here, indicating that her status as a non-native English speaker might influence her perception of the essay's language. This suggests that she may be more accustomed to certain linguistic features that could be indicative of machine generation, especially if they deviate from typical patterns of second language learners' language use.

Additionally, Kyle argued that the essay's characteristics aligned with those of AI-generated contents and pointed out further that the bland content, inappropriate language, and structural issues as potential indicators of AI involvement.

> *Researcher: So you seriously think it's AI generated? It's likely to be AI generated essay according to you.*
> *Kyle: Yeah, basically because of the sort of blandness of the writing.*

He commented that the "blandness" in writing, to him, means the texts lack depth, creativity, or personal touch. The use of repetitive phrases, generic language, or lack of clear expression could also contribute to this perception of blandness. The use of the word "blandness" implies a lack of distinctive or unique qualities in the writing, which could be interpreted as a deficiency in creativity or originality.

Furthermore, Kyle's use of the phrase "sort of" before "blandness" suggests a degree of hesitation or uncertainty in his assessment, indicating that he may not be entirely confident in his conclusion. This hesitation could stem from the challenge of accurately distinguishing between AI-generated and human-generated content, particularly as AI technologies become more and more sophisticated and capable of mimicking human writing styles.

4.1.2 Unreliable content

The tutors' observations regarding the reliability of the essay's content shed light on the importance of thorough research and accurate referencing in academic writing.

In this regard, Omar, in his comment below, focused on issues related to the accuracy and relevance of the essay's content. He noticed the presence of irrelevant sections, factual inaccuracies, and outdated references, suggesting lack of credibility and academic rigour in the essays that he deems AI generated.

> *Omar: reading through the reference list, some references, for example, "the promise of assessment engineering", didn't sound like a real one to me.*

Omar's doubt about the legitimacy of the reference suggests his concern about the essay's authenticity and reliability. He further commented that in academic writing, references should be precise and credible. The possibility that the reference may have been generated without rigorous academic scrutiny hints at one of the issues with AI-generated text. Omar's critique, therefore, highlights the potential pitfalls of relying on AI-generated content for academic purposes. It serves as a reminder that while AI may be able to assist in generating text, the "hallucinations" it provides often fail to meet the rigorous standards of authenticity and reliability required in scholarly work.

Similarly, Kyle echoed concern about the reliability of the essay's content, emphasising the need for accurate referencing and evidence-based argumentation. Kyle's remarks below indicated his scepticism regarding its reliability and authenticity.

> *Kyle: I'll agree it's well-structured, I think the key points are the sources though. I don't think the arguments presented in the essay are supported by sources.*

His critique centres on three main points: the essay's lack of genuine source support, the coherence of the arguments and their link to the sources. First, Kyle acknowledges the essay's structural soundness, which can be a symbol of strength of AI-generated content. Advanced AI models are adept at creating well-organised texts that mimic human writing to certain extent. However, structural integrity alone does not equate to reliability. The absence of authentic sources is a significant flaw Kyle highlights. AI-generated essays sometimes lack solid citations, which challenge the essay's credibility, suggesting it is not based on genuine research or information.

Second, Kyle's hesitation and repeated phrases, such as "I don't think" and "it's not an argument supported by sources", indicate his uncertainty about the essay's argumentative strength. AI-generated content, while structurally sound, often fails to present compelling, evidence-backed arguments. This deficiency is crucial, as persuasive writing relies heavily on the ability to substantiate claims with credible sources. Without this foundation, the essay's arguments could appear superficial and unconvincing.

4.1.3 Superficial analysis

All four tutors remarked on the superficial nature of the essay's analysis and described the essays as merely scratching the surface of the topic without going into deeper insights or offering meaningful interpretations of the subject matter. Superficial analysis may result from a lack of critical thinking, insufficient research, or a failure to engage with complex ideas or perspectives. For example, Helen's observation below indicated that the essay failed to provide a thorough analysis, suggesting it might be AI-generated.

> *Helen: However, the problem is, although they mentioned the validity, they didn't go further, or go deeper into the validity, or go to different aspects and with more specific and relevant literature.*

Helen noted in one of the essays she evaluated that while the essay briefly mentions the concept of validity, it does not go into deeper aspects of the topic. A comprehensive analysis would typically explore various dimensions of validity, such as construct, content, and criterion validity, and reference-specific relevant literature to support its points. The lack of this depth suggests a superficial treatment of the subject. Similar point was also discussed by Omar who asserted that the essay he referred to was characterised by a superficial analysis.

> *Omar: I understand you cannot evidence every point in your essay due to word count, but you need to evidence a few critical points, and this one I evaluated reads more like a template.*

This template-like nature is indicative of AI-generated text, which often uses a generic structure and adjusts minor details to fit different prompts or topics. The superficiality is further highlighted by the need for only minimal adjustments to make the essay suitable for different tests or bands, suggesting a lack of depth and specificity. Rather than offering deep insights or robust evidence, the AI-generated essay tended to skim the surface, providing just enough to appear coherent without delving into substantial or original thought.

Additionally, Omar's mention of the essay being "more like a template" and the suggestion to "just need to change a few words" reveal a reliance on a pre-structured format that can be easily adapted to various contexts. This template-like nature is indicative of AI-generated text, which often uses a generic structure and adjusts minor details to fit different prompts or topics. The superficiality is further highlighted by the need for only minimal adjustments to make the essay suitable for different tests or bands, suggesting a lack of depth and specificity.

4.2 Tutors' evaluation of AI-generated essays against the rubrics

4.2.1 Not academic essays

In light of the original rubrics of the module under investigation, there is a consensus among the participants regarding the essays' lack of adherence to academic standards, which is indicative of several underlying issues, highlighting various aspects of academic writing that the AI-generated essays failed to meet. For example, Carl's evaluation provided insights into language usage and stylistic elements of the essay. He noted the presence of odd language and expressions, suggesting a departure from the formal tone expected in academic writing.

> *Carl: some wordings were rather odd. It started with things like "this essay embarks on an ambitious journey. It will surpass conventional boundaries, reflecting excellence and original thought". I mean, this is not typical language of a student essay, or even in any kind of academic.*

The above quote provides a critique of the language used in a student essay, noting its atypical phrasing. The critique implies that such ambitious language is unusual and

potentially inappropriate for the context, suggesting a mismatch between the essay's language and the expected tone and style of academic work. This analysis underscores the importance of aligning writing style with audience expectations, particularly in academic settings, where clarity and appropriateness are crucial.

Similarly, Helen's assessment below emphasised the absence of a clear argument and critical evaluation in the essay. She pointed out that the essay failed to articulate a coherent thesis or engage critically with the topic. This critique underscored the importance of developing a well-defined argument supported by evidence and analysis in academic writing. Without a clear argument, the essay lacked direction and failed to fulfil the fundamental requirements of scholarly discourse.

> *Researcher: What features help you recognise? ...Can you summarise it please?*
> *Helen: OK, so the first one is from the text itself. It's not written by an L2 student, and there's a really poetic and informatic writing style.*
> *Researcher: When you say poetic, do you mean it is like a poem?*
> *Helen: Yes, a lot of metaphors.... it's not that academic essay to me.*

In her evaluation, Helen provides a clear rationale for why the text in question does not qualify as an academic essay. Her main points revolve around the writing style and the presence of certain literary features that are atypical for academic writing. She notes that the essay is "poetic and informatic", indicating that it employs a style more characteristic of creative writing than of scholarly analysis. Specifically, Helen points out the frequent use of metaphors, which she finds unsuitable for an academic context.

In a similar vein, Kyle echoed concerns about the essay's failure to meet academic standards. He said some of the language would be expected from a storybook, not academic writing.

> *Kyle: The language is articulate and engaging. I mean, I'll certainly describe it as articulate and engaging... But in a story book, you know. It's not an academic style at all.*

His evaluation revealed a critical perspective on its academic validity. Initially, Kyle acknowledges the essay's effective use of language, stating, "The language is articulate and engaging". This suggests that while the essay is well written and likely captivating for readers, its style is more suited to storytelling rather than academic discourse.

He explicitly differentiated between the qualities of effective narrative writing and the requirements of academic writing. Academic essays typically prioritise clarity, objectivity, and evidence-based arguments, adhering to specific structural and stylistic conventions. This evaluation highlights the fundamental difference between engaging narrative and academic rigour. Kyle's critique suggests that while the essay may excel in creativity and readability, it lacks the formal tone, structured argumentation, and scholarly depth, which are the typical characteristics of academic writing.

4.2.2 Lack of coherence

The tutors observed that some essays lacked coherence, which underscored the importance of logical structure and organisation in academic writing. For example, Carl highlighted several structural issues that contributed to the essay's lack of coherence. He noted the presence of random sentences, inappropriate references, and a disjointed structure that hindered reader comprehension.

> *Carl: In some cases, it's almost like a random collection of sentences, all of which were OK and on topic, but didn't join together properly.*
> *Researcher: That sounds like AI generated it to you?*
> *Carl: Well, if it was generated by AI, it wasn't a very good AI.*

Carl's evaluation of the essay highlights a fundamental issue of coherence, indicating that while the sentences were individually acceptable and relevant to the topic, they failed to form a unified, coherent piece. This lack of connection among sentences suggests that the essay lacked a logical flow, making it difficult for readers to follow the argument or narrative. Carl's comment implies that the essay's sentences were disjointed, preventing the text from conveying a clear and cohesive message. It seems the primary issue lies not in the relevance or correctness of the individual sentences but in the essay's inability to weave these sentences into a coherent narrative or argument. Effective writing requires more than just relevant content; it demands a logical progression of ideas, which was evidently lacking in the essay Carl evaluated.

Also, Omar mentioned the essay's lack of coherence and flow, citing issues with the order of presentation and disconnected ideas. In this sense, the essay failed to provide a cohesive narrative or develop ideas in a logical progression, which emphasised the importance of structuring the essay in a way that facilitated smooth transitions between paragraphs and sections. This disjointedness makes it difficult for readers to follow the essay's argument, diminishing its overall effectiveness.

> *Omar: No coherence, and arguments are not talking about what it promised to talk about in the introduction.*

Omar's observation that the arguments do not align with what was promised in the introduction is a critical weakness. The introduction of an essay sets expectations for the reader by outlining the main points or arguments that will be explored. If the body of the essay swings away from these points, it not only breaks the reader's trust but also undermines the purpose of the introduction. This misalignment suggests that the student either did not plan their essay effectively or failed to stay on topic, both of which are detrimental to the essay's overall quality.

4.2.3 Lack of criticality

The tutors' observations about the essay's lack of critical evaluation underscore the importance of analytical thinking and engagement with scholarly literature in academic writing. For example, Kyle's evaluation emphasises the essay's failure to engage critically with the topic or present a coherent argument.

> *Kyle: I mean, while the argumentation covers the transformative impact on timelines, it may lack in-depth critical analysis, occasionally veering towards a descriptive approach.*

Kyle's evaluation highlights that while the essay addresses the transformative impact on timelines, it fails to go into a deep critical analysis. Instead, it tends to adopt a descriptive approach. Criticality in academic writing involves more than just describing or summarising information; it requires engaging deeply with the subject matter, questioning assumptions, evaluating evidence, and considering alternative perspectives. In the absence of criticality, an essay may fall short in several aspects. It

may lack originality, merely repeating existing knowledge without adding anything substantial to the academic discourse. Moreover, it may overlook contradictions or biases inherent in the arguments presented, thus failing to provide an understanding of the topic.

Kyle's critique suggests that the student's essay may be superficial in its treatment of the subject matter. While it may acknowledge the transformative impact on timelines, it fails to interrogate the underlying assumptions or implications critically. As a result, the essay may not fulfil the expectations of academic rigour and intellectual depth. To address this deficiency, the student needs to cultivate a more critical mindset, actively questioning assumptions, engaging with conflicting viewpoints, and offering insightful interpretations. By doing so, they can elevate their analysis beyond mere description and contribute meaningfully to the scholarly conversation.

Similarly, Omar noted the presence of superficial analysis and factual inaccuracies, indicating a failure to engage critically with the topic or evaluate the evidence presented. His assessment of the student's essay shows that the essay seems to suffer from a repetitive nature where the same point is repeated without deeper analysis or exploration of alternative perspectives. Criticality in writing involves the ability to evaluate, question, and engage with the material being discussed. It demands a thoughtful examination of ideas, considering various angles, and offering deep insights. Such an absence of criticality in the essay suggests a superficial engagement with the topic, failing to go into its complexities or challenge prevailing assumptions.

Moreover, Omar's observation about the lack of criticality aligning with typical features of AI-generated text underscores the nature of the essay. AI-generated content often lacks the human capacity for critical thinking, relying on algorithms to generate text based on patterns and data inputs rather than genuine analysis. In essence, Omar's evaluation suggests that the AI-generated essay under evaluation falls short of demonstrating critical thinking skills essential for academic discourse. Without criticality, the essay fails to offer meaningful contributions to the conversation, resembling more of a product of automation than genuine intellectual enquiry.

4.3 Tutors' evaluation of AI-generated feedback

4.3.1 Efficiency and identifying inconsistency

The four tutor participants emphasised that AI feedback systems offered unparalleled efficiency and speed in providing feedback on academic writing, taking Omar for example.

> *Omar: Timely feedback is important for international students.*

Timely feedback is crucial for international students, and AI algorithms significantly enhance this process. By rapidly analysing essays, AI can identify grammar errors and provide constructive suggestions much faster than human graders. This speed improves the feedback turnaround time, enabling students to receive prompt guidance on their writing. Timely feedback is particularly important for international students who may face language barriers and cultural differences in academic writing.

Quick and constructive feedback allows them to understand their mistakes and learn how to improve their skills more efficiently. It helps them adapt to academic expectations and standards, reducing the time and stress associated with waiting for tutor grading. Consequently, the ability of AI to deliver immediate, detailed feedback

supports international students in making necessary revisions and progressing in their studies more effectively.

Similarly, Helen highlighted the potential benefits of AI in providing feedback, especially in translation work. The conversation begins with the researcher steering the discussion towards exploring how AI might be advantageous in offering feedback. She prompts the participant to consider whether AI could provide any significant affordances or advantages that could be beneficial, specifically asking if AI could offer any "good thing that we can actually benefit from".

> *Researcher: do you see any advantage that AI could offer in providing feedback? Either in this module or in your translation module, is there any affordances? Is there any advantage? Is there any good thing that we can actually benefit from?*
> *Helen: Yes. They can highlight the inconsistency. I need to read through the translation pieces and find there are some kind of inconsistency, but machine can highlight much more quickly.*

Helen responds by pointing out a specific advantage of using AI in translation tasks. She explains that AI can quickly highlight inconsistencies within translation pieces, a task that would typically require a thorough and time-consuming manual review. By automating this process, AI can identify discrepancies and errors rapidly, allowing for a more efficient workflow. This capability is particularly valuable because it helps ensure the accuracy and consistency of translations, which are critical aspects of quality in this field.

The interview emphasises the practical application of AI in enhancing the efficiency and effectiveness of various tasks. In the context of translation, AI's ability to pinpoint issues means human reviewers can focus their efforts on more complex aspects of the translation process. This not only improves the overall quality of the work but also significantly reduces the time and effort required for manual reviews.

Moreover, this discussion illustrates the broader implications of AI in different fields. By automating routine yet essential tasks, AI can optimise workflows and enhance productivity. It allows professionals to allocate their time and skills to more strategic and creative endeavours, thus maximising the value of human input while leveraging the strengths of AI technology. This synergy between human expertise and AI capabilities represents a transformative potential, making processes more efficient and outcomes more reliable.

4.3.2 Language support for students and educators

Tutor interviewees, particularly Kyle, emphasised the valuable language support provided by AI feedback systems.

> *Kyle: I'd be very happy for students to use AI to help them with their language.*

His statement above reflects a positive stance on the use of AI in language learning. He expresses positive attitude towards the idea of students using AI tools to aid their language learning. This perspective aligns with the growing acceptance of technology in education, where AI can offer personalised learning experiences, immediate feedback, and access to a vast array of resources. His approval suggests that he views AI as a beneficial tool to traditional learning methods, enhancing students' ability to practice and improve their language skills. By advocating for AI use, he acknowledges

its potential to address individual learning needs and accelerate proficiency. This endorsement highlights a shift towards integrating advanced technologies in educational settings, aiming to make learning more efficient and accessible.

While AI feedback offered many advantages, participants emphasised that it should be used as a supplementary tool for educators rather than a replacement for human feedback. For example, Omar emphasises the potential role of AI in the creative and evaluative processes of brainstorming and writing. He suggests that AI can be employed as a tool to enhance the initial stages of idea generation by providing a critical perspective. The phrase "a pair of eyes" can be interpreted metaphorically as a fresh or unbiased viewpoint that AI brings to the table. This notion highlights AI's capability to assist in refining ideas by offering feedback that is detached from human biases or preconceptions.

> *Omar: you can use AI in the brainstorming stage as a pair of eyes that gives some criticality to your thoughts because you sometimes generate your text, and then you can ask for feedback.*

Omar's mention of using AI for generating text and subsequently seeking feedback underscores a cyclical, iterative process. AI's role in this cycle can be twofold: first, as a collaborator in producing content, and second, as an evaluator that helps to improve and polish the output. This dual functionality allows for continuous refinement and enhancement of ideas and written material. It reflects a growing recognition of AI as a valuable tool, where it can act as a supportive partner rather than a replacement for human creativity.

In summary, Omar advocates for integrating AI into the creative process to provide critical feedback and to foster continuous improvement. This approach not only enhances the quality of the output but also facilitates a more dynamic and interactive creative workflow. By leveraging AI in this manner, users can benefit from an additional layer of critical analysis that complements their own insights and expertise.

4.3.3 Lack of constructive feedback

Although tutor participants argued for the importance of AI in feedback, some tutors, particularly Kyle, noted the absence of constructive feedback provided.

> *Researcher: So, feedback is good, we give human marks, give feedback for formative purposes to help them to learn. But the AI generated feedback won't be able to do that job, at least the current stage of AI generated feedback can't do that.*
> *Kyle: It doesn't. It didn't in the three that you gave me. So that way, yeah.*
> *Researcher: right, right, right.*

This interview brings into focus the current limitations of AI in the realm of educational feedback. The researcher begins by affirming the critical role of human feedback in the learning process, particularly for formative purposes. Formative feedback is essential because it provides students with detailed, personalised insights that help them understand their strengths and areas for improvement. This type of feedback is not just about correcting mistakes but also about guiding students in their learning journey, fostering their development in a supportive manner.

The researcher asserts that AI-generated feedback, at its current stage, cannot perform this function effectively. This claim highlights a significant gap between what

AI can offer and the depth of feedback that humans provide. Human feedback is rich in context, empathy, and understanding, which are crucial elements in education that AI has yet to master.

Kyle's response, noting that the AI feedback he reviewed did not meet necessary standards, corroborates the researcher's point. His statement, "It didn't in the three that you gave me", serves as a concrete example of AI's shortcomings in this area. This practical observation adds weight to the researcher's argument and underscores the importance of maintaining a human touch in educational feedback. The researcher's repeated affirmation, "Right, right. Right. Right", indicates her strong agreement and possibly a sense of urgency about this issue. It suggests that they see this as a significant concern that needs addressing. This repetitive affirmation could also reflect their recognition of the challenges and complexities involved in integrating AI into educational settings.

The conversation between the researcher and Kyle shows a broader debate within educational technology. While AI has shown great promise in various domains, its application in delivering formative feedback remains limited. The dialogue suggests a need for a cautious approach to integrating AI in education. It emphasises that while AI can support certain tasks, the irreplaceable value of human interaction, judgement, and insight into teaching and learning processes must be preserved. This balance is crucial to ensure that the adoption of AI enhances rather than diminishes the quality of education.

5. Discussion

5.1 AI detectability abilities of human raters

The analysis of the four tutor participants' data demonstrated their keen attention to detail and deep understanding of academic writing dynamics. Their evaluations revealed a multifaceted approach, highlighting the complexities involved in differentiating between AI and human content. One key area of focus was the tone of the language used in the essays. Tutors analysed the essays for signs of artificiality, identifying abnormalities such as overly ornate phrasing, disjointed sentence structures, and the presence of unusual idioms or expressions that hinted at non-human origins. These linguistic irregularities served as red flags, prompting further scrutiny to determine the authenticity of the content. This aligns with the findings of Floridi and Chiriatti [17], who discussed common linguistic irregularities in AI-generated text are indicators of non-human authorship.

Beyond linguistic analysis, the assessors examined the essays, searching for signs of AI involvement. They identified important issues, such as a lack of coherent argumentation, superficial analysis, and the absence of original insights, as potential indications of automated generation. Their critiques extended beyond surface-level assessments, exploring the scholarly discourse and the intellectual rigour expected in academic writing. This comprehensive evaluation included examining the depth of analysis, originality of thought, and the overall intellectual engagement demonstrated in the essays. Desaire et al. [18] underscore the importance of these indicators in evaluating academic writing quality, highlighting how the depth of analysis can differentiate human-authored from AI-generated content.

Additionally, the tutor participants paid close attention to the structural coherence and organisation of the essays. They noted deficiencies, such as disjointed arguments,

inadequate transitions between ideas, and a lack of logical flow, which detracted from the overall coherence of the work [4]. These observations highlighted the human-like qualities of organisation and coherence that are often lacking in AI-generated content. Crossley and McNamara [19] also emphasise these qualities as hallmarks of skilled academic writing, noting that the absence of well-organised and logically coherent arguments is a significant indicator of AI involvement.

Overall, the human assessors in this study demonstrated an understanding of both linguistic and substantive elements of academic writing, allowing them to identify potential indicators of AI-generated content. This multifaceted approach underscores the importance of a detailed and comprehensive assessment process in distinguishing between human and AI-generated essays and feedback. In this regard, McNamara et al. [20] discussed the role of natural language processing in evaluating writing quality, which can aid in distinguishing human-written content from AI-generated text, reinforcing the need for detailed and thorough evaluation criteria.

The above discussion highlights the need for a detailed approach in detecting AI-generated essays, focusing on linguistic irregularities, such as odd language and disjointed structures, as well as a lack of coherent argumentation and original insights. They also identified structural deficiencies, like poor transitions and logical flow, as indicators of AI involvement. This comprehensive evaluation underscores the importance of analysing both linguistic and intellectual depth to distinguish between human and AI-generated content, aligning with previous research on writing quality assessment.

5.2 AI affordances to feedback and evaluation

In the assessment of academic writing, human evaluators demonstrate a depth of understanding that stems from their ability to recognise distinctions and contextualise their evaluations within the broader landscape of academic standards and expectations. They scrutinise essays, identifying specific shortcomings such as a lack of coherence in argumentation, deficiencies in critical analysis, and the presence of language that falls short of the formal genre expected in scholarly discourse. Drawing upon their individual expertise and experience, human assessors offer personalised feedback tailored to the unique strengths and weaknesses of each piece of writing [21, 22]. This personalised approach enables students to receive targeted guidance for improvement, addressing their specific areas of concern and fostering a deeper understanding of academic writing conventions. For example, human raters can pinpoint nuanced issues in student writing and provide context-specific feedback that AI systems might miss [21]. This level of detailed, contextually rich feedback is crucial in helping students understand and meet the complex demands of academic writing [22].

On the other hand, artificial intelligence (AI) feedback systems offer distinct advantages in terms of efficiency and consistency. Using algorithms to rapidly analyse essays, these systems can provide feedback at a pace unmatched by manual grading processes [23]. This rapid turnaround can significantly enhance the learning process, as students receive timely insights that allow them to quickly address and rectify their mistakes [24]. Moreover, AI feedback ensures a level of consistency and standardisation in evaluation, as it applies predefined criteria uniformly across different student submissions [23]. This standardised approach promotes fairness and transparency in the assessment process, as all students are evaluated according to the same set of guidelines [24].

Furthermore, AI feedback systems can offer personalised support tailored to individual student needs. By analysing writing proficiency levels, learning styles, and specific areas requiring improvement, AI algorithms can adapt feedback to address each student's unique requirements [23]. This personalised feedback enhances the relevance and effectiveness of the guidance provided, ultimately contributing to improved learning outcomes [24]. This capability can be particularly beneficial in large classes where individualised attention from human instructors is limited [20]. Thus, while both human and AI feedback systems have unique strengths, integrating human evaluators' contextual understanding with AI's efficiency and consistency could offer a more comprehensive and effective approach to academic writing assessment [20].

5.3 Risks and ethical issues

Participants evaluating AI-generated feedback compared to human understanding identified several key risks that could impact students' learning and improvement. One major concern raised was the lack of constructive criticism provided by AI systems. Without actionable insights, students may struggle to identify areas for growth and develop their writing skills effectively [25]. Additionally, participants noted issues with the clarity of AI-generated feedback. Clear feedback is essential for students to understand where to improve and how to address those areas. Unclear feedback can lead to confusion and frustration, ultimately hindering students' ability to make meaningful revisions to their work.

Consistency in feedback provision was another area of concern highlighted by the participants. While AI systems offer standardised criteria for evaluation, inconsistencies were still observed in the feedback provided. Inconsistent feedback may confuse students and undermine their confidence in the assessment process, potentially leading to dissatisfaction and mistrust [26]. Moreover, participants expressed worries about the lack of individualisation in AI-generated feedback. Individualised feedback considers the unique strengths and weaknesses of each student's work, providing tailored guidance for improvement. Without this personalised approach, students may feel that their specific needs are not met, which could result in disengagement and frustration [27].

Also, ethical concerns regarding the integrity of academic standards in the face of AI technology were raised, prompting reflections on the potential consequences of relying on AI-generated content without transparent guidelines. Ethical considerations loomed large in their evaluations as they grappled with the implications of AI technology in academic integrity. Participants raised concerns about the reliability of AI-generated content, emphasising the need for clear ethical frameworks to govern its use in educational settings. Their reflections underscored the importance of maintaining the credibility and standards of academic scholarship amidst technological advancements.

6. Conclusions

Overall, tutor participants' multifaceted approach to detecting AI-generated content showcased their holistic understanding of academic writing. Their insights transcended surface-level analysis, exploring linguistic, substantive, ethical, and structural dimensions. In doing so, they underscored the indispensable

role of human judgement and critical evaluation in preserving the integrity and quality of academic discourse in an era of advancing AI technology. Additionally, combining human understanding and AI feedback offers a comprehensive approach to evaluating academic writing. Human assessors bring a depth of insights, while AI systems could offer surprising efficiency and consistency in many aspects. By integrating these approaches, educators can optimise the feedback process, providing students with timely, relevant, and effective guidance for improving their academic writing skills.

Future research should focus on advancing AI systems to provide more personalised and constructive feedback, addressing current limitations in critical analysis and individualised guidance (see **Table 2**). Exploring the ethical implications, including academic integrity and bias, is crucial, as is investigating hybrid assessment models that combine AI's efficiency with the depth of human judgement. Research should expand beyond small qualitative case studies, incorporating larger, more diverse samples across institutions and disciplines to enhance generalisability. Additionally, mixed-method approaches and longitudinal studies could offer deeper insights into AI's long-term impact on learning outcomes. Examining AI's role in varied educational contexts, including primary and secondary schools, vocational training, and different cultural and linguistic environments, could help push the agenda of inclusivity and adaptability in global education further.

Acknowledgements

The authors would like to thank The Centre for Higher Education Practice (CHEP) at the University of Southampton for funding this research.

Conflict of interest

The authors declare no conflict of interest.

Exploring AI Applications in Essay-Based Assignments: Affordances and Risks
DOI: http://dx.doi.org/10.5772/intechopen.1008230

Author details

Ahmad Alzahrani[1*] and Ying Zheng[2]

1 Faculty of Arts and Humanities, King Abdulaziz University, Jeddah, Saudi Arabia

2 Faculty of Arts and Humanities, University of Southampton, Southampton, United Kingdom

*Address all correspondence to: aksalzahrani@kau.edu.sa

References

[1] Dai W, Lin J, Jin H, Li T, Tsai Y-S, Gašević D, et al. Can large language models provide feedback to students? A case study on ChatGPT. In: Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT'23); 10-13 July 2023; Orem. Utah, United States: IEEE; 2023. pp. 323-325

[2] De Costa PI, Sterling S, Lee J, Li W, Rawal H. Research tasks on ethics in applied linguistics. Language Teaching. 2021;**54**:58-70. DOI: 10.1017/S0261444820000257

[3] Tai AMY, Meyer M, Varidel M, Prodan A, Vogel M, Iorfino F et al. Exploring the Potential and Limitations of ChatGPT for Academic Peer-Reviewed Writing: Addressing Linguistic Injustice and Ethical Concerns. 2023. Available from: https://journal.aall.org.au/index.php/jall/article/view/903 [Accessed: 6 August 2024]

[4] de Winter J, Dodou D, Stienen A. ChatGPT in education: Empowering educators through methods for recognition and assessment. Informatics. 2023;**10**:87. DOI: 10.3390/informatics10040087

[5] Atlas S. ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI. 2023. Available from: https://digitalcommons.uri.edu/cba_facpubs/548/ [Accessed: 3 December 2023]

[6] Holmes J, Liu Z, Zhang L, Ding Y, Sio TT, McGee LA, et al. Evaluating large language models on a highly-specialized topic, radiation oncology physics. Frontiers in Oncology. 2023:**13**. DOI: 10.3389/fonc.2023.1219326

[7] McNichols H, Feng W, Lee J, Scarlatos A, Smith D, Woodhead S, Lan A. Exploring Automated Distractor and Feedback Generation for Math Multiple-Choice Questions Via in-Context Learning. 2023. Available from: arXiv.2308.03234 [Accessed: 6 August 2024]

[8] Mondal H, Marndi G, Behera JK, Mondal S. ChatGPT for teachers: Practical examples for utilizing artificial intelligence for educational purposes. Indian Journal of Vascular and Endovascular Surgery. 2023;**10**:200-205. DOI: 10.4103/ijves.ijves_37_23

[9] Kraft MA, Falken GT. A blueprint for scaling tutoring and mentoring across public schools. AERA Open. 2021;**7**. DOI: 10.1177/23328584211042858

[10] Hirunyasiri D, Thomas DR, Lin J, Koedinger KR, Aleven V. Comparative analysis of GPT-4 and human graders in evaluating praise given to students in synthetic dialogues. In: Proceedings of the 24th International Conference on Artificial Intelligence in Education (AIED); 3-7 July 2023. Tokyo, Japan: AIED; 2023. 12 pages workshop paper

[11] Pankiewicz M, Baker RS. Large language models (GPT) for automating feedback on programming assignments. In: Proceedings of the 31st International Conference on Computers in Education (APSCE); 4-8 December 2023. Matsue, Shimane, Japan: ICCE; 2023

[12] Bewersdorff A, Seßler K, Baur A, Kasneci E, Nerdel C. Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. Computers and education. Artificial Intelligence. 2023;**5**. DOI: 10.1016/j.caeai.2023.100177

[13] Latif E, Zhai X. Fine-tuning ChatGPT for automatic scoring. Computers and Education: Artificial Intelligence. 2023. Available from: arXiv:2310.10072 [Accessed: 6 August 2024]

[14] Casal JE, Kessler M. Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. Research Methods in Applied Linguistics. 2023;**2**:3. DOI: 10.1016/j.rmal.2023.100068

[15] Thu NT, H. EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. International Journal of Language Instruction. 2023;**2**:1-47. DOI: 10.54855/ijli.23231

[16] Cao S, Zhong L. Exploring the Effectiveness of ChatGPT-Based Feedback Compared with Teacher Feedback and Self-Feedback: Evidence from Chinese to English Translation 2023. Available from: arXiv:2309.01645 [Accessed: 6 August 2024]

[17] Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. Minds and Machines. 2020;**30**:681-694. DOI: 10.1007/s11023-020-09548-1

[18] Desaire H, Chua AE, Isom M, Jarosova R, Hua D. Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. Cell Reports Physical Science. 2023;**4**:6. DOI: 10.1016/j.xcrp.2023.101426

[19] Crossley SA, McNamara DS. Say more and be more coherent: How text elaboration and cohesion can increase writing quality. Journal of Writing Research. 2016;**7**:351-370. DOI: 10.17239/jowr-2016.07.03.02

[20] McNamara DS, Crossley SA, Roscoe R. Natural language processing in an intelligent writing strategy tutoring system. Behavior Research Methods. 2013;**45**:499-515. DOI: 10.3758/s13428-012-0258-1

[21] Sadler DR. Beyond feedback: Developing student capability in complex appraisal. Assessment & Evaluation in Higher Education. 2010;**35**:535-550. DOI: 10.1080/02602930903541015

[22] Carless D, Salter D, Yang M, Lam J. Developing sustainable feedback practices. Studies in Higher Education. 2011;**36**:395-407. DOI: 10.1080/03075071003642449

[23] Wilson J, Roscoe RD. Automated writing evaluation and feedback: Multiple metrics of efficacy. Journal of Educational Computing Research. 2019;**58**:87-125. DOI: 10.1177/0735633119830764

[24] Shermis M, Garvan C, Diao Y. The impact of automated essay scoring on writing outcomes. Online Submission [thesis]. University of Florida; 2010

[25] Nguyen A, Gardner L, Sheridan D. Data analytics in higher education: An integrated view. Journal of Information Systems Education. 2020. Available from: https://aisel.aisnet.org/jise/vol31/iss1/5 [Accessed: 6 August 2024]

[26] Zawacki-Richter O, Marín V, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education -where are the educators? International Journal of Educational Technology in Higher Education. 2019;**16**:1-27. DOI: 10.1186/s41239-019-0171-0

[27] Boud D, Molloy E. Rethinking models of feedback for learning: The challenge of design. Assessment & Evaluation in Higher Education. 2013;**38**:698-712. DOI: 10.1080/02602938.2012.691462