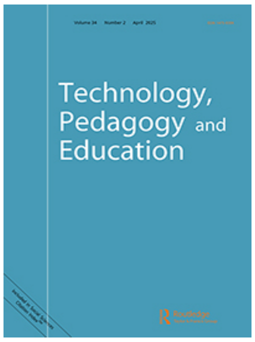

Technology, Pedagogy and Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rtpe20

AI and peer reviews in higher education: Students’ multimodal views on benefits, differences and limitations

Gabriela C. Zapata, Bill Cope, Mary Kalantzis, Anastasia Olga (Olnancy) Tzirides, Akash K. Saini, Duane Searsmith, Jennifer Whiting, Nikoleta Polyxeni Kastania, Vania Castro, Theodora Kourkoulou, John W. Jones & Rodrigo Abrantes da Silva

TECHNOLOGY, PEDAGOGY AND EDUCATION
https://doi.org/10.1080/1475939X.2025.2480807

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

AI and peer reviews in higher education: students' multimodal views on benefits, differences and limitations

Gabriela C. Zapata [a], Bill Cope [b], Mary Kalantzis [b], Anastasia Olga (Olnancy) Tzirides [b], Akash K. Saini [b], Duane Searsmith [b], Jennifer Whiting [b], Nikoleta Polyxeni Kastania [b], Vania Castro [b], Theodora Kourkoulou [c], John W. Jones [d] and Rodrigo Abrantes da Silva [e]

[a]School of Education, University of Nottingham, Nottingham, UK; [b]Department of Education Policy, Organization & Leadership, University of Illinois, Urbana-Champaign and Chicago, IL, USA; [c]Department of Teaching and Learning, Elementary Education Program, University of Iowa, Iowa City, IA, USA; [d]Childhood/Early Childhood Education Department, SUNY Cortland, Cortland, NY, USA; [e]Departamento de Letras Modernas, Universidade de São Paulo, São Paulo, Brazil

ABSTRACT

Since the launch of ChatGPT in November 2022, educational researchers and practitioners have sought to understand the ways in which these new generative AI technologies might influence education. This article describes one such effort. The focus of the investigation was chatbots responding from large language models to the review of open-ended student work. Specifically, the authors examined university students' multimodal views of the benefits and limitations of AI reviews as compared to human feedback. The participants were postgraduate students in a public American university. The students' opinions of their experiences with both types of reviews were expressed linguistically, visually and gesturally, and they were submitted to discursive and socio-semiotic analyses. The results revealed a preference for human reviews. Nevertheless, the participants also identified several benefits for AI feedback, as well as ways in which it had complemented human reviews, overwhelmingly welcoming its addition as part of their educational experience.

ARTICLE HISTORY
Received 29 June 2023
Accepted 29 October 2024

KEYWORDS
Generative AI; AI feedback; peer reviews; higher education

Introduction

The landscape of education is evolving and, with it, the methodologies for providing constructive feedback are transforming. Peer feedback, a time-honoured tradition in academia, involves the evaluation of one's work by peers within the same educational cohort (Nicol et al., 2014). This collaborative assessment is grounded in the belief that constructive criticism from peers can foster a deeper understanding of the subject matter, refine critical thinking skills and nurture a sense of academic community (Double et al., 2020). Peer-to-peer feedback is an excellent example of crowdsourcing (a general practice of obtaining information or input into a task soliciting the efforts of a large number of people; see Surowiecki, 2004) that reduces the time and workload of teachers (Li et al., 2022) and allows them to concentrate their efforts on other learning and pedagogical issues (Double et al., 2020; Gamage et al., 2021).

CONTACT Gabriela C. Zapata gabriela.zapata@nottingham.ac.uk School of Education, University of Nottingham, Nottingham, UK

As we embrace the digital era and leverage the capabilities of generative artificial intelligence (GenAI) models, there emerges a unique opportunity to further revolutionise the feedback process, offering more comprehensive, timely, and tailored evaluations (Otaki, 2023; Wongvorachan & Bulut, 2022). AI feedback, particularly in the form of chatbots leveraging large language models, represents a modern technological approach to assessment (Tzirides et al., 2023, 2024; Tzirides, Zapata, Bolger, et al., 2024; Zapata et al., 2024). Driven by advancements in natural language processing, AI feedback offers automated personalised evaluations with the potential for scalability based on predefined criteria (Clarizia et al., 2018; Wongvorachan & Bulut, 2022). These systems draw upon vast linguistic databases to generate responses that simulate human-like engagement with student submissions. AI-generated feedback has ignited significant discussions and debates within academic circles (Baidoo-Anu & Ansah, 2023; Cope & Kalantzis, 2023b; Dai et al., 2023; Otaki, 2023) with the potential to reshape the way in which we teach and learn (Božić, 2023).

The purpose of this study is to contribute to these ongoing discussions by investigating chatbots responding from large language models to the review and assessment of university students' complex, open-ended work. This investigation was part of a comprehensive project examining GenAI and human feedback from different perspectives. In this work, we explore participants' multimodal views of the benefits and limitations of AI reviews as compared to peer feedback. In the first section of the article, we provide an overview of existing work on student feedback, and we also touch on AI educational research in connection with this area. The next part of the article introduces the study. This is followed by the results of the investigation and their discussion. The final sections focus on the limitations of GenAI in educational contexts, as well as ways of addressing them, and conclude this work.

Peer and AI feedback

Since the late 1990s, peer feedback[1] has been one of the most researched topics in educational research, particularly in connection with higher education (Ryan et al., 2021). Several systematic literature reviews have highlighted its many pedagogical and social benefits. For example, in one of the first comprehensive examinations of existing studies, Dochy et al. (1999) posited that through the assessment of classmates' assignments, university students could acquire academic skills that they could apply to their own work as well as develop 'high levels of responsibility [to ensure they are] fair and accurate with the judgments they make of their peers' (p. 338). Additionally, the articles reviewed by the authors revealed overall positive student attitudes towards peer reviews and the information received from this type of feedback. A more recent review carried out by Van Zundert and her colleagues (Van Zundert et al., 2010) rendered similar results, confirming both learners' openness towards peer feedback and its several academic and social benefits.

Nevertheless, some of the studies discussed in these two works have also shown that not all types of feedback are beneficial, and peer assessment can vary in its reliability and effectiveness. This is important because learner 'output … is dependent on the nature and quality of the … comments that students receive of their work or learning [as they] are an important prerequisite for the impact of feedback process' (Winstone & Carless, 2020, p. 7). Hattie and Timperley's work (2007) has focused specifically on this process, furthering our understanding of the ways in which it can advance or hinder learning. Based on their examination of existing evidence, these scholars have proposed a model of feedback that addresses the prerequisites needed for it to be effective for student use.

This model revolves around three specific questions that offer guidance on different aspects of the task to be completed and/or a product to be developed (Hattie & Timperley, 2007). The first question helps learners understand *where they are going*, as the feedback given is connected to the objectives to be achieved based on specific criteria and/or expectations. The second question, *how students are going*, entails 'information relative to a task or performance goal, often in relation to some expected standard' (Hattie & Timperley, 2007, p. 89). To answer the third question, *where to next*, feedback needs to be oriented towards specific actions that can lead to further ways of learning.

These three questions can be embedded in four different levels of feedback – *task*, *process*, *self-regulation,* and *self*. Feedback given at the *task* level focuses on the achievement of a specific task or product, and it involves establishing connections with expectations or criteria (e.g. how students' work might reflect a specific criterion or what might be needed to get closer to it). Feedback at the *process* level 'is more specific to the process underlying tasks or relating and extending tasks . . . [and it] appears to be effective . . . for enhancing deeper, [more comprehensive] learning' (Hattie & Timperley, 2007, p. 93). The remaining levels, *self-regulation* and *self*, are not considered as academically effective for task completion as the other two levels, but they do play a social and personal role. That is, in these two levels, feedback is focused on students' ability (*self-regulation*) and effort (*self*), and therefore, it can act as a motivating factor, enhancing learners' 'commitment, control, and confidence . . . , [as well as their] engagement or feelings of efficacy in relation to the[ir] learning' (Hattie & Timperley, 2007, p. 93 and 96). Both the questions and levels included in this model make it a comprehensive tool for the analysis of peer assessment, and it could also be applied in the exploration of AI feedback, an area of interest in AI-supported education.

In their recent analysis of the possible impacts of AI in education, Nguyen et al. (2022, pp. 4221 and 4223) characterised AI as 'one of the most pivotal developments of the century', positing that it can be 'seen as an influential tool to empower new paradigms of instruction . . . [including] computer-assisted collaborative learning . . . [and] the use of automatic assessment'. Lane et al. (2016) echoed these beliefs, pointing to the enormous possibilities for AI in areas such as collaborative, immersive, affective, and exploratory learning. Clearly, there is an increased demand for developing AI technologies that can be utilised by educators for more effective and inclusive support of learners' educational needs (McGrath et al., 2023), and two areas of interest are assessment and formative feedback (Dai et al., 2023; Tzirides, Zapata, Bolger, et al., 2024; Tzirides, Zapata, Kastania, et al., 2024; Zapata et al., 2024).

Specifically, recent work has shown that GenAI can be a valuable tool for the automatic grading of assignments. For example, Mizumoto and Eguchi's (2023) study, which analysed 12,100 English essays authored by individuals from 11 distinct linguistic backgrounds between 2006 and 2007, revealed that GenAI significantly reduced the time required for grading, ensured consistency in scoring, and was able to provide immediate scores and feedback. Dai et al. (2023) have reported similar results. This work compared ChatGPT-based reviews and instructor feedback on 85 open-ended student-written assignments in a postgraduate university programme. The findings showed that AI feedback (based on five different criteria) was more readable and consistent than that offered by the students' instructor. Additionally, the AI was able to provide specific guidance, albeit limited, akin to the process-level feedback in Hattie and Timperley's (2007) model. In spite of these strengths, the AI reviews also exhibited limitations, as they 'could not offer a reliable assessment of student performance compared with the instructor' (Dai et al., 2023, p. 5).

Steiss et al. (2024) reported similar results in a recent paper that examined the quality of formative feedback provided by GenAI (GPT 3.5) compared to human evaluators for high school students' writing. The study compared these two types of feedback in connection to 200 student essays based on five criteria: criteria-based, clarity of directions for improvement, accuracy, prioritisation of essential features, and supportive tone. Descriptive statistics and effect sizes were employed to determine whether there were differences in feedback quality for the whole sample, for essays of different overall quality, and for English-speaking students and English learners. The results showed that human evaluators provided higher quality feedback than the AI in all categories except for criteria-based feedback. Additionally, the analyses suggested there was no difference in feedback quality based on language status (English-speaking vs. English learners) for both GPT and human evaluators. Based on these results, Steiss et al. concluded that while well-trained evaluators provide higher quality feedback, AI feedback can still be useful in certain contexts, particularly for formative early drafts or when well-trained educators are unavailable.

Even though the studies presented in this section have offered data on GenAI's possible contributions to both the assessment of student work and the provision of formative feedback, one crucial aspect missing from this body of work is students' voice. That is, if we are to embrace GenAI as part of the learning experience, it is important to explore how this innovation is regarded by students. The objective of this article is to achieve this goal by investigating university students' views on AI and peer feedback. As previously discussed, peer reviews offer a variety of academic and social benefits to students, and complementing the peer-to-peer review process with AI-enhanced technology as an additional form of feedback could result in more constructive learning and deeper understanding of the given subject-matter. In what follows, we present our examination of learners' multimodal views of peer and AI reviews after experiencing both types of feedback.

The present study

Research questions

This work sought to answer the following research questions:

(1) What benefits and drawbacks of peer and AI reviews are reported by postgraduate students after experiencing both types of feedback?
(2) How might these two types of feedback complement each other?

Participants

The participants in this work were recruited during Spring 2023 from two online higher education courses in a college of education at a university in the Midwestern region of the United States. Even though the study was based on curricular content and activities, participation in it was voluntary and not a requirement for course enrolment. Ninety-three students of mixed genders and ethnic/racial backgrounds and ages ranging from 25 to 45+ years, pursuing Certificate, Master's, and Doctorate degrees, participated in the study. All were working professional educators undertaking their graduate degrees part-time. Additionally, the students were highly experienced in areas ranging from school to higher education, workplace, and community education, crossing diverse discipline areas.

Educational context

This study is part of a comprehensive research project for which both quantitative and qualitative data were collected digitally through a suite of interconnected apps in an established social knowledge and learning platform developed by two of this article's authors in 2009 (Tzirides et al., 2023). An intervention leveraging a large language model for dialogue applications was designed to provide postgraduate students in the programmes of focus with AI feedback on their multimodal texts, in addition to anonymised peer reviews. OpenAI's GPT-3 was connected to an app within the digital platform used via application programming interface (API) to offer machine feedback to learners on their works.[2] This feedback relied on the same explicitly stated assessment measures and rubric used by students to review their and their peers' work. The rubric items and GPT prompts were drawn from the multiliteracies framework *Learning by Design* (Cope & Kalantzis, 2023a; New London Group, 1996). This pedagogy takes an epistemological approach to learning, focusing not solely on cognition but more broadly on knowledge-making activities which in addition to cognition involve material practices, embodied activity, and socio-emotional engagement (Cope & Kalantzis, 2015). These are high-level, abstract review criteria.

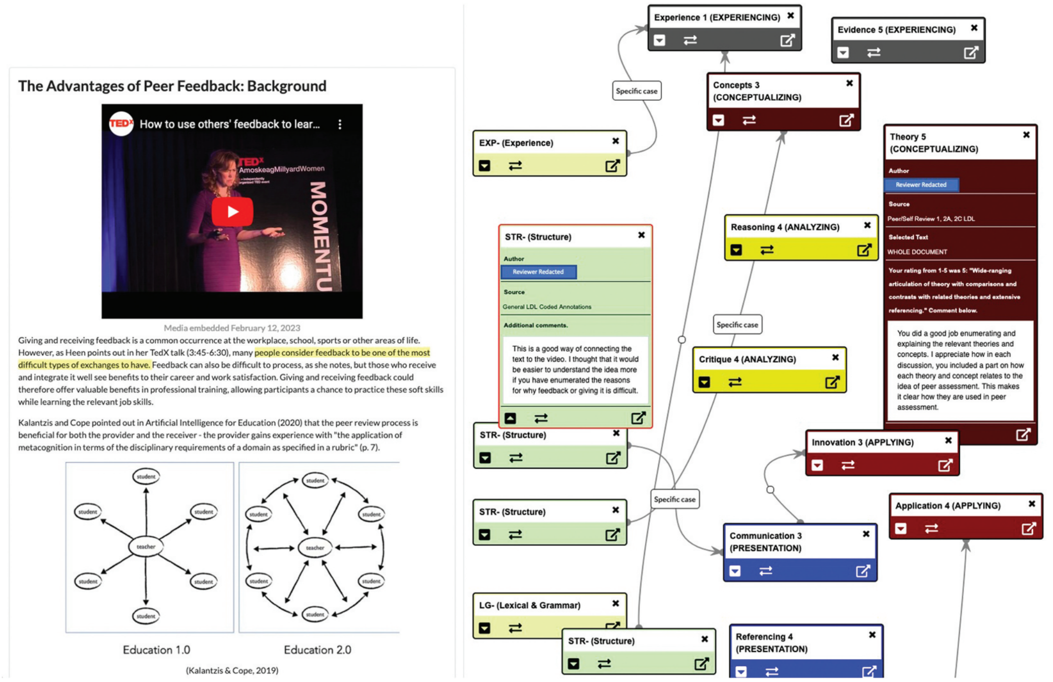

Figure 1. Sample peer review of first draft of a participant's work. The work under review is on the left, and the rubric-based peer review map is on the right. Two nodes are opened in this screenshot; the rest have been folded closed by the reviewer for visual clarity while building the review map. The green node highlighted in red is currently active, with an annotation coded as STR (a structural issue with the text that requires improvement). This refers to the yellow highlighted block of text in the reviewed work. Mousing over the node highlights the text. The brown node is a whole-text review against the theory review criterion. The reviewer's rating is recorded in the top bar, and their narrative explanation has been opened out for the writer.

The peer and AI reviews were carried out in connection with the participating students' major project for one of the courses in their programmes of study. Participants chose their own topic to examine educational theory and practice, and they developed their work throughout the term. Upon submission of their first complete draft, they digitally reviewed two other students' works against the rubric criteria within the learning platform, and they also received the same type of peer feedback on their own products. An example from one of these works and its peer review are shown in Figure 1.

After feedback in anonymised peer reviews, the participants revised their texts. At this point, a new step had been added to the existing workflow – the AI review. First, the learning platform's API took the extended student text and broke it into sections. This was helped by the platform's strictly structural and semantic markup of sections and subsections, as well as HTML paragraphing at a more granular level. These 'sections' were then fed to the GPT-3 'text-curie-001'. This model is smaller and less powerful than the model subsequently used for review generation, 'text-davinci-003', but it is less expensive, faster, and does a competent job on the summarisation task (which would not necessarily benefit from the more powerful model). Each section was summarised, and then sections were concatenated. After that, criterion by criterion, the summarised text was sent to GPT-3 multiple times, each time with a different rubric criterion as a prompt. The review was displayed as a node graph in the platform's tool where the AI rubric criterion nodes were generated, one for each criterion on the rubric. Participants then compared the human and machine reviews before the final revision and publication of their work into the course community and personal portfolios. A sample AI review is presented in Figure 2.

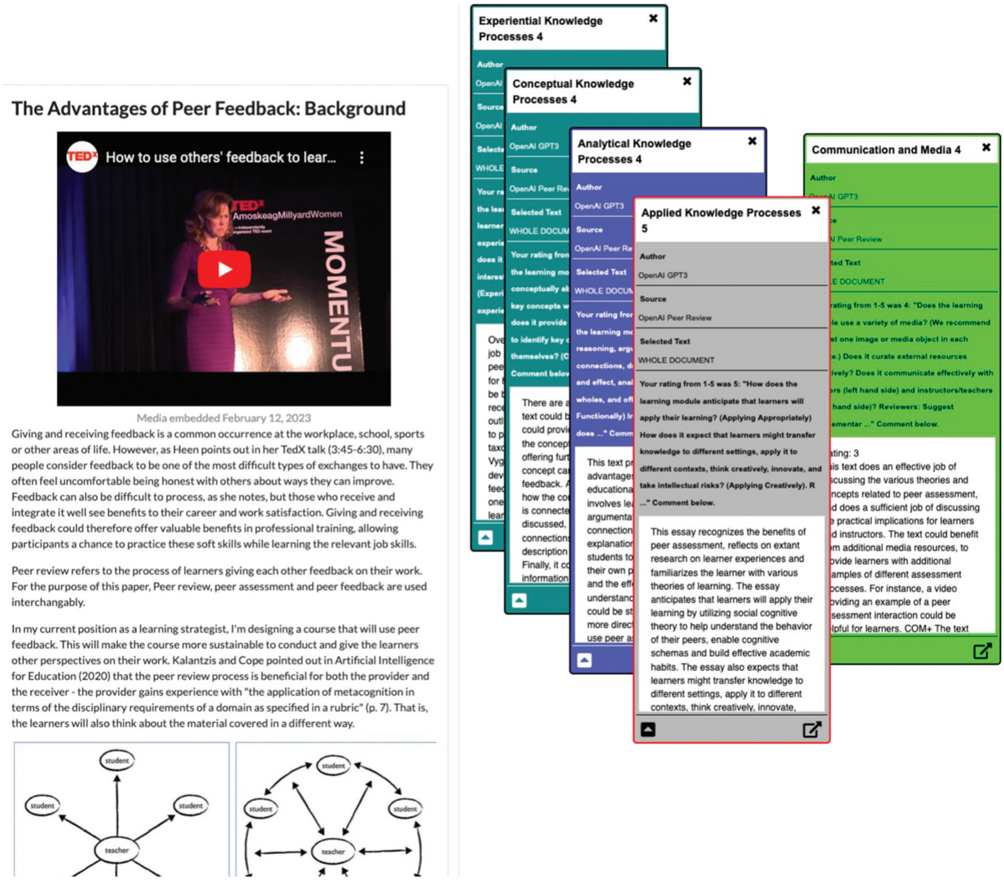

Figure 2. AI review of the revised, post-peer review version of a participant's work.

Instruments

After their work had been reviewed by their peers and the AI, the participants were invited to express their opinions of both review types by sharing an image and/or screenshot to multimodally depict their experience. Since work in the courses in which the students were enrolled required them to create multimodal artefacts (i.e. combining a variety of semiotic resources such as text and images), it was deemed appropriate to ask them to comment on their experiences using more than just language. This choice was also grounded in our belief, guided by Van Leeuwen's (2011), that the analysis of the participants' multimodal communication would offer a richer, more holistic and nuanced understanding of their experiences with the two types of reviews than could be achieved solely through text analysis. In addition to their artefacts, the learners submitted three words to summarise their perceptions of peer and AI reviews, and they completed written reflections (with a 150-word limit) comparing them. All these became the sources of data for this study.

Procedures

The students' linguistic responses were analysed qualitatively with the software *MAXQDA 2022* (VERBI Software, 2021) using thematic analysis. This type of analysis was chosen because it has been used in a myriad of works that have focused on participants' opinions and have employed similar instruments for data collection (Braun & Clarke, 2006). The first step involved the careful reading of the students' responses and the recording of their overall, general impressions of peer and AI reviews. This was followed by the specific identification of themes and exemplifying statements. In the final stage of the analysis, themes were cross-examined employing Glaser's (1965) constant comparative method to ensure that there were no discrepancies in the initial analysis, and the percentage of themes in connection with coded documents (i.e. each student's responses) was calculated. Additionally, word clouds were created to offer visual compilations of the words that had been chosen by the participants to summarise their overall experiences with both review types.

The analysis of students' multimodal responses was grounded in the tenets of social semiotics (Cope & Kalantzis, 2020; Kalantzis & Cope, 2020; Kress & van Leeuwen, 2021; Van Leeuwen, 2005). This methodology allows for the identification and categorisation of the semiotic resources (e.g. textual [language and typography], visual, gestural) used by meaning-makers, as well as the examination of their motivation, sociocultural context, and intended audience. The analysis in this work first involved the identification and description of each of the semiotic resources embedded in the artefacts submitted by the participants (e.g. the textual elements were considered in terms of both the linguistic message they conveyed and their typography). In the next step, the relationship among different semiotic elements was established, and the overall message expressed multimodally was unveiled in connection with the prompt to which it had been attached. The findings resulting from these analyses are presented and discussed in the following two sections of the article.

Results

Analysis of textual data

The results of the thematic analysis of the participants' textual responses revealed an overall preference for human feedback, with only three participants choosing AI reviews over peer ones. Nonetheless, themes also point to both benefits and drawbacks for both types of reviews, with most participants' regarding them as complementary and welcome sources of feedback. A summary of the themes resulting from the qualitative analysis and their percentage with respect to the responses submitted (i.e. the coded documents) is presented in Figure 3. The specific results connected with each review type are discussed below, in separate sections.

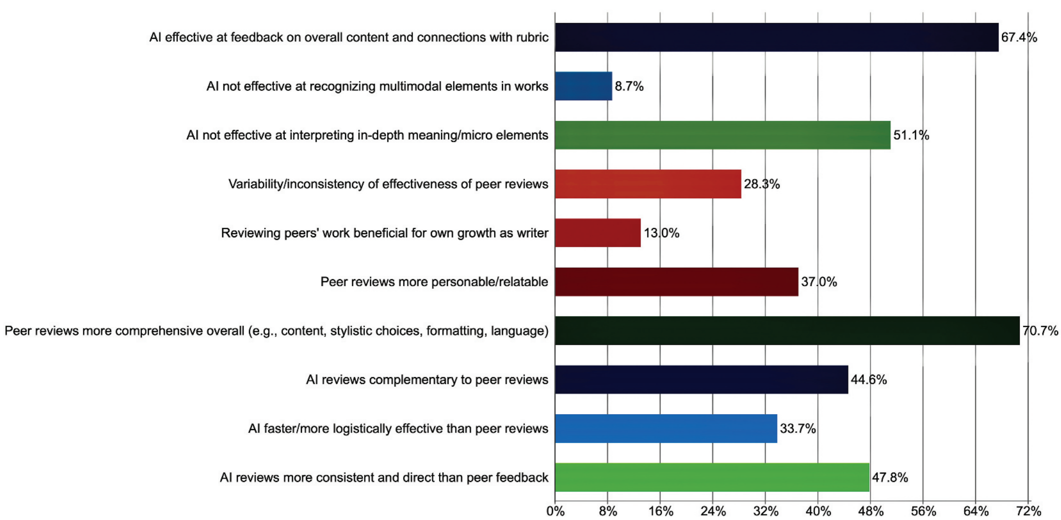

Figure 3. Percentage of themes in coded documents.

Peer reviews

The participating students believed peer reviews had been more beneficial than the AI feedback in terms of the comprehensive and in-depth feedback they had provided with respect to essay content, formatting and stylistic choices:

> [This peer review] is not only specific but demonstrates that the reviewer has the entirety of the work and its purpose in mind. I think this is a key distinction between human peer reviews and their current AI counterparts: a human reviewer has a thesis, a guiding idea behind their critique that they form over the course of reviewing the work, while an AI does not. (Participant #4)

> The peer review helps identify several . . . issues . . . that the AI didn't. The peer review provided specific examples and suggestions that felt a bit more helpful as a writer as they were more relevant to the work. (Participant #14)

> There were multiple annotations, instead of overall feedback per category, allowing . . . commentary to be understood with ease. (Participant #53)

> Peer review comments are more specific, although both reviewers also provide comments and ratings for the entire article. Because peer feedback is more specific, such as pertaining to a specific point of view, structural or analytical issue, or even just a formatting problem, peer feedback is generally easier to implement. (Participant #70)

Also, the participants characterised human reviews as being more personable and relatable:

> Human peer review can give more reassuring and encouraging feedback to authors than the AI. (Participant #3)

> [A peer review] is truly what human interaction can be at its best – two writers communicating and making-meaning together to share a common goal. It is yet another way to actively participate in the growth of a peer and makes the online learning experience much warmer because of this sort of human interaction. (Participant #20)

> Peer review is the personal experience the review brings. Not only does the reviewer offer constructive feedback, but commentary on what has resonated with them in the work and why. I enjoy the connection created through the work. And that is something that an AI review cannot replicate. (Participant #83)

Another benefit noted was the growth the students had seen in their own writing as a result of reviewing their classmates' work:

> I have learned so much reviewing others' work. For starters, reading the broad topics have made me a better learner. I am now a more professional writer. (Participant #13)

> Reviewing other students' papers gave me insight into how I could better structure my own paper. (Participant #38)

> Primarily, I get a lot out of reviewing the work of others . . . Seeing the assignment through the interpretive lens of another writer allows me to analyze my own writing and how I can organize, communicate, and structure my paper to be better. In this way, reviewing the works of others . . . greatly improves my own writing. (Participant #61)

As evinced in these quotes, the participants appear to have benefitted from their peers' comments. Nevertheless, a limited number of students ($n = 8$) also expressed their dissatisfaction with reviews that took too long, were not completed on time and/or only offered general or succinct remarks without specific suggestions for improvement.

The opinions expressed in the textual prompts were also reflected in the words chosen by the participants to summarise their experiences with peer reviews. For example, as shown in the visual compilation presented in Figure 4, the crucial aspect highlighted was the social, personal connection that transpired in the feedback received from peers as well as its subjective nature, which manifested in both the positive and negative experiences described by the participating students. Some of the words had a clear positive connotation (e.g. 'supportive', 'meaningful', 'thoughtful', 'encouraging', 'honest' and 'helpful') and seemed to focus on the emotional, more personal aspects of the reviews. Other words had a more negative connotation (e.g. 'disappointing', 'spotty' and 'lacking') and appear

Figure 4. Visual compilation of the words used by participants to describe peer reviews.

to signal the lack of quality and time delay in the comments received by some students. Some terms also emphasised the comprehensiveness and usefulness of human reviews (e.g. 'thorough', 'specific', 'constructive', 'targeted', 'enlightening', 'actionable' and 'purposeful'). Overall, the tone of the words chosen was more positive than negative, confirming the participants' preference for this type of review.

AI reviews

When considering the effectiveness of the AI reviews, the students in this work noted that the comments received had been useful for identifying 'big picture' revisions and for establishing connections between content and the rubric categories on which their work would be assessed. For example, Participant #54 stated that 'the AI review ... brings an interesting perspective, [as] it is very rubric-based, which is good because sometimes [students] get caught up and forget portions of the rubric'. Also, some participants highlighted the usefulness of AI feedback for later stages of the writing process. For instance, Participant #86 believed the AI review had offered 'a solid foundation as a stepping stone for the beginning stages of revision', while Participant #96 felt it was a valuable tool 'in regard to referencing'.

Most participants, however, regarded the AI feedback as too limited in terms of specific content and stylistic suggestions due to what they believed was the AI's lack of understanding of the context and nuances of academic writing:

> I thought the feedback was really vague and too generalized ... It could not tell me where I needed to add more details, better structure, or deeper analysis. (Participant #8)

> The feedback I received on my work seemed general and could [have] been applied to any number of works. It did a great job summarizing, but not interpreting, my writing. (Participant #12)

> While I remain impressed by the level at which the AI performs, it still isn't able to parse my work quite right. What I mean is that it doesn't always register my intent how a human reader would, and so some responses are a bit ... off. (Participant #55)

Another limitation reported was that the AI had suggested the incorporation of multimodal elements that were already part of the document reviewed. Despite these drawbacks, the AI reviews were praised for their speed and overall efficiency as well as for the consistency of the feedback provided, which contrasted with the variability found in the quality of some peer comments and their lack of expediency.

The words used by the participants to describe the AI feedback seem to focus on its objective and impersonal, data-driven nature (in comparison to peer reviews) and, like the textual responses, they also pointed to both its advantages and shortcomings (Figure 5). For instance, words such as 'fast', 'straight-forward', 'immediate', 'practical', 'satisfying' and 'instant' suggest that the AI comments were easy to understand, and they provided clear and actionable information as well as instilled a sense of progress or momentum in participants with respect to the development of their work. In contrast, terms like 'superficial', 'decontextualised', 'general/generic', 'rough', 'a-bit-hollow', 'disjointed' and 'hit-or-miss' convey a sense of insufficiency and lack of substance or detail, and they imply that there was a lack of focus, sophistication or coherence in the reviews. Despite these weaknesses, most participants appear to welcome and enjoy this type of reviews, as long as they do not replace peer comments, but complement them.

Figure 5. Visual compilation of the words used by participants to describe AI reviews.

Analysis of multimodal data

Like the textual responses, the 44 multimodal elements submitted by the participants to communicate their experiences with peer and AI reviews highlighted advantages and drawbacks. Additionally, some of the artefacts conveyed students' uneasiness, anxiety, and/or ambivalent emotions towards the use of AI for educational purposes, while others expressed the opposite, celebrating instead the potential for human–AI collaboration. The semiotic resources in all products were combined in single, non-segregated frames (i.e. all the semiotic resources were presented together in one semiotic space). The prevailing forms of communication were the visual, spatial, and gestural (though a limited number included some textual features). The social semiotic analysis of artefacts connected with each review type is presented separately, in the next two sections.

Peer reviews

The main advantage of peer reviews highlighted in eight of the 14 multimodal representations submitted by the participants was their diverse and collaborative nature. This was achieved through the use of different colours, photos or illustrations of diverse students, spatial and gestural features denoting proximity and equality, and facial expressions and body language conveying dialogue, harmony, and a friendly exchange of ideas. An example of these artefacts (submitted by Participant #40) is presented in Figure 6, where a group of diverse human figures can be seen working together. The main visual resource is colour, which is not only employed to express ethnic diversity but also diversity of ideas, embodied in the varied, colourful dialogue balloons positioned above each of the human representations. Also, these balloons communicate a sense of inclusion and equality (all figures have a voice). This is complemented by the spatial and gestural modes of communication: Each person is the same size, can be fully seen, and has been placed in equal proximity to others. Their posture indicates friendly collaboration and interaction (e.g. the male figure in the green sweater appears to have his arm on the back of the figure with glasses, suggesting support and friendship). Another important element that acts as a visual and gestural metaphor of the collaborative, active thinking taking place in the scene depicted is the bright, shining lightbulb positioned centrally, above the human figures.

Another interesting example is the artefact submitted by Participant #78, who used their multimodal product to highlight another advantage of peer reviews mentioned in the textual data – its comprehensive nature. As seen in Figure 7, this student chose an abstract representation of a human eye extending beyond its shape to the area surrounding it to convey the idea that their classmates' feedback was all-encompassing, perhaps going above and beyond what could be seen and entailing both meaning comprehension and interpretation. This idea was supported by the textual legend the participant added to the image: 'Peer/Human Review – Comprehensive [and] Helpful'.

Figure 6. Sample multimodal representation of collaborative nature of peer reviews (participant #40). Image in the public domain.

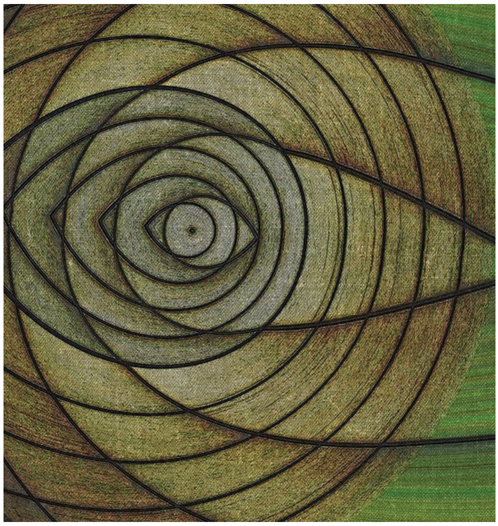

Figure 7. Sample multimodal representation of comprehensive nature of peer reviews (participant #78).Image in the public domain.

While the diversity and comprehensiveness of peer comments appear to have been welcomed by these students, the six remaining representations conveyed more negative emotions, as the participants who submitted them seem to have found the diversity of opinions negatively subjective and overwhelming (i.e. it was difficult to reconcile comments that differed quite drastically). For example, one of the images submitted depicts an arm and a hand holding a pen that act as a visual and gestural synecdoche for the writer, who appears to be buried in paper reviews. The writer, nevertheless, is shown as breaking through the reviews, emerging triumphally. This is conveyed gesturally by the fist gripping the pen in a gesture symbolising triumph, despite the overwhelming quantity of the paper comments. Another drawback highlighted multimodally by these students was the reviewers' unreliability and/or delay in terms of the submission of their feedback.

AI reviews

Of the 30 multimodal artefacts submitted by the participants to convey their experiences with non-human reviews, 15 presented the AI tool as isolated robots with humanoid features, located in the metaverse (represented as an abstract, dark, and/or vacuous space) or empty classrooms. Most of these robots exhibit neutral facial expressions and are either looking at a particular point within the setting where they are or are staring blankly into the metaverse/classroom: There is no visual contact between the figures and viewer. These visual and gestural elements appear to highlight the impersonal, decontextualised 'cold' tone of the AI reviews, which contrasts with the collaboration and communication emphasised in the artefacts depicting peer feedback.

Four of these participants also made reference to the AI tool's inability to offer creative and/or comprehensive suggestions for improvement beyond those connected to the categories in the rubric on which the reviews were based. For example, Participant #28 textually equated the AI feedback to that of 'a brilliant child, [who] lacks the life experience and context of other related fields'. This message is embedded in the multimodal artefact submitted by this student (Figure 8). Meaning is conveyed visually and gesturally through a photo of a robot's upper torso and head (with humanoid features), carrying what appears to be a tablet and looking up at the viewer, implying it is not tall. Both the expression in the robot's eyes and the shape of its mouth suggest eagerness to please. That is, the figure is looking up at a taller viewer, like a child would do with an adult, and its

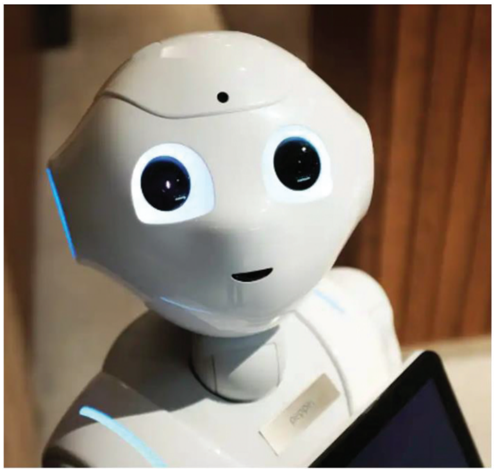

Figure 8. Sample multimodal representation of the AI tool as 'a brilliant child' (participant #28). Image in the public domain.

gestures and posture instil an air of expectancy either for contact or approval. Clearly, this is a friendly tool, but its knowledge is limited, like a child's.

While these 15 participants chose to focus on the limitations of AI reviews, eight of their classmates viewed them as complementary to peer feedback, submitting multimodal artefacts that emphasised AI–human communication and collaboration. These artefacts include photos or illustrations of robots shaking hands with humans or making physical contact with them (e.g. fingers touching or heads creating a common, Venn-diagram-like space). The representations rely on visual and gestural synecdoche, as only one arm or an arm and a hand or the head for each figure is shown (i.e. no full-bodied robots or human beings are included in the images chosen), which could be interpreted as a generalised view of this relationship. The AI – human contact either takes place in the metaverse or in a computer lab, conveying the idea that it is digital, as were the AI and peer reviews received by the participating students. Also important is the fact that in all images, the technology and human elements are the same size, and though they might originate from opposite sides (e.g. left vs. right), they converge in a central point. This could mean that, for these students, AI and peer reviews could complement each other and could equally contribute to the improvement of their work.

In contrast with the eight participants who highlighted AI–human communication, the remaining seven students expressed their anxiety and/or ambivalent feelings towards the use of AI in educational contexts. To convey their views, they resorted to images from movies that were either created with live-action and motion-capture computer-animated animation or featured villainous AI (e.g. *2001 Space Odyssey* and *Blade Runner*). For example, Participant #4 used an image from the film *The Polar Express*, whose type of animation has been associated with Mori's (2012) concept of *the uncanny valley* (Russell, 2021). Masahiro Mori was a leading Japanese professor of robotics who posited that 'a person's response to a humanlike robot would abruptly shift from empathy to revulsion as it approached, but failed to attain, a lifelike appearance' (Mori, 2012, p. 98). By choosing this representation, the student clearly drew parallels between the uneasiness produced by the uncanny valley and their experience with AI.

Discussion

The results of the analysis of the participants' textual and multimodal responses suggest that they were able to identify benefits and drawbacks in both peer and AI feedback (research question #1).

For example, even though a small number of students expressed their disappointment at the comments they had received from their classmates for being too succinct, delayed, or not having been submitted at all, most participants believed peer reviews were more comprehensive than AI feedback with respect to all aspects of their writing. Not only were peers able to provide specific suggestions both in terms of content and form, but they also offered emotional support, with the interaction among classmates resulting in collaboration and meaningful communication. Additionally, carrying out reviews of other students' work was seen as beneficial for the development of the reviewer's own writing skills. Like the participants in the studies discussed by Dochy et al. (1999) and Van Zundert et al. (2010), the learners in this work displayed positive attitudes towards peer reviews, overwhelmingly expressing their preference for this type of feedback.

In contrast, the AI reviews were not regarded as effective as the peer comments, and for some participants, they felt impersonal, superficial, and cold. Nevertheless, most students believed that, though the AI could not replace human reviews, nor could it offer feedback of the same quality, it could work well as a complementary tool because of its expediency, logistical effectiveness, and consistency. These opinions mirrored the GenAI advantages highlighted by Mizumoto and Eguchi (2023) and Steiss et al. (2024). Some participants clearly embraced this idea through their choice for multimodal representations that seem to celebrate AI–human collaboration and interaction. Other students, however, expressed their uneasiness and ambivalent feelings towards the inclusion of AI in educational contexts, drawing parallels with negative, fictional representations of technology. Nonetheless, the participants' overall experiences with AI reviews appear to suggest that their incorporation to instruction would be welcomed.

Indeed, the findings suggest that the peer and AI reviews complemented each other effectively. Together, they seem to have encompassed the four feedback levels included in Hattie and Timperley's (2007) feedback model (Figure 9), which resulted in a comprehensive review process that would not been possible if only one type of feedback had been implemented (research question #2). For example, the AI feedback can be mostly equated with the *task* level, as the guidance provided to the participants was connected to their work and its reflection of the rubric categories

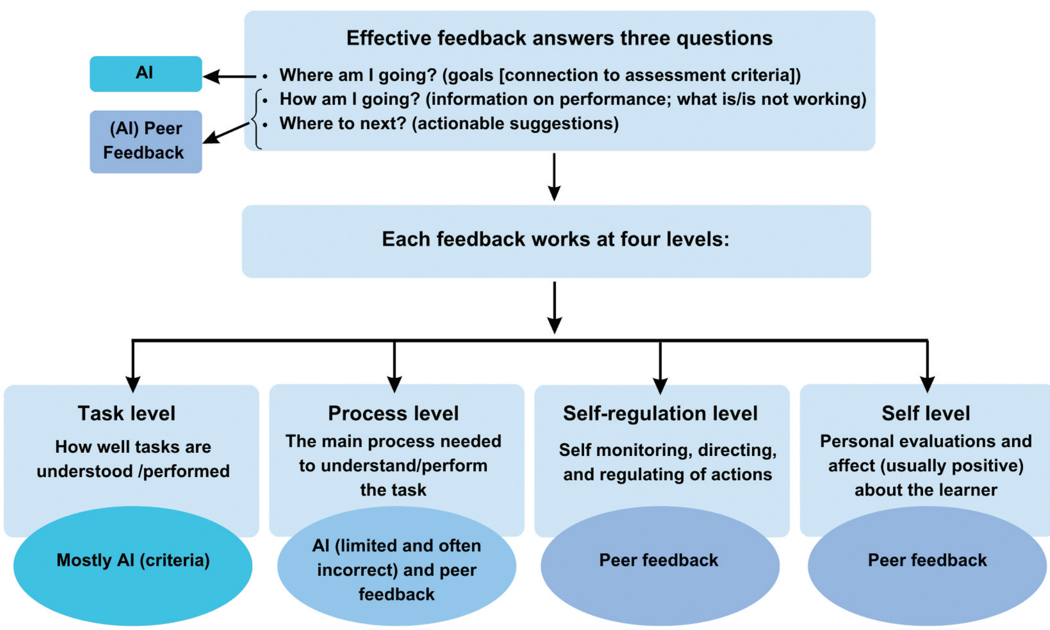

Figure 9. Correspondence between AI and peer reviews and Hattie and Timperley's (2007) feedback model. Adapted from Hattie and Timperley (2007, p. 87).

on which it would be assessed and the overall objective to be achieved (*where* question in the model). In contrast, the specific and comprehensive feedback as well as personal comments resulting from peer-to-peer reviews reflected the type of guidance characteristic of the model's *process*, *self-regulation,* and *self* levels and the questions *how* and *where to next*. Even though, as in the studies by Dai et al. (2023) and Steiss et al. (2024), the AI reviews in this work offered some guidance at the *process* level, the participants felt the information given was too limited, and therefore not as useful as the feedback received from their peers. These results imply that the adoption of both AI and peer reviews could offer a rich educational experience, which is reflected in Participant #97's words: 'Coupled with the peer review, I think this is the best overall experience I've had with others reviewing my work'.

Limitations of AI reviews and suggestions for implementation

When considering the limitations of AI reviews highlighted by the participants in this study, it is important to consider that large language models, like the GPT employed in this work, are still *just* language models – they have no understanding of language (Church & Mercer, 1993; Cope & Kalantzis, 2023b). Machine intelligence records characters in binary notation and performs statistical analyses of patterns in these characters. Limited though this is as mechanised intellectual work, the sheer force of numerical modelling and calculation can today achieve feats that are superhuman. In addition to responding to prompts, they perform second-order operations including summarisation, translation, application of instructions, and planning. The power of this unmeaning brute force was evident to the participants in this article, as most of them embraced the addition of AI feedback to peer reviews, and they deemed it helpfully different from human feedback in a number of ways, even against identical prompts.

Our results suggest that one key to the implementation of AI reviews lies in a new domain of human–computer interaction termed 'prompt engineering'. GPT technology has two related components. One is the corpus of text that constitutes the large language model upon which the word-after-word statistical analysis is based. The second is the chatbot that queries the large language model and prompts the generation of a response. The generated response is only as good as the prompt (Cope & Kalantzis, 2023b). The key, we would therefore argue, is to use well-formed prompts based on a balanced range of analytical perspectives as well as customise the AI to encompass a knowledge storage directly connected to students' needs. These changes would allow AI feedback to be as comprehensive and accurate as possible. In our project, this goal was achieved through the development of an API which carefully calibrated the student's interaction with the GPT using as the basis for the prompts rubrics created in our specific learning platform. These rubrics were theoretically grounded in the established pedagogical framework *Learning by Design* (Cope & Kalantzis, 2015, 2023a). This feature of our work, is, to the best of our knowledge, unique, and, as noticed by the participants, resulted in AI feedback that, though not as specific as peers' comments, was still deemed useful for the development of learners' work.

When considering some of the participants' reticence, anxiety, and/or lack of confidence towards AI feedback, we feel one way to address these feelings would be to provide students with more information about the way in which these reviews work, focusing both on the technology involved and on issues of privacy and data safety (Borenstein & Howard, 2021). Other ways to expand learners' understanding of the technologies involved and their capabilities in connection with feedback would be the incorporation of collaborative classroom opportunities for the critical analysis of examples of both AI and peer feedback, to identify strengths and weaknesses as well as the affordances offered by the combination of both review types. This in-depth understanding of AI and peer assessment processes could mitigate feelings of discomfort and also result in students' embracement of AI in educational contexts.

Study limitations

While this work offered data on university students' views on peer and GenAI reviews, it exhibits the limitations often found in teaching and learning scholarship, such as the description of a short instructional intervention in a specific educational context. Also, the study relies solely on self-reported data, which introduces the possibility of response bias. Furthermore, since this study was carried out shortly after ChatGPT became publicly available, the implementation relied on a now outdated version of GenAI. Since the completion of the project described, LLMs have become more sophisticated, which has not only extended their capabilities, but has also simplified their customisation to better answer specific learner needs.

Despite the limitations described, we believe that this article offers valuable data for our understanding of students' preferences and the ways in which the combination of GenAI and human feedback can enrich learning experiences. We also feel that this work could constitute a departure point for more in-depth explorations of both social and academic aspects of GenAI in education, such as the changes to classroom interactions and teacher and student roles resulting from this innovation and the effectiveness of machine and human feedback on students' performance.[3]

Conclusion

The results of this study suggest that, purposefully calibrated, new generative AI technologies can be put to good use supporting learning. Clearly, the participants in this work identified weaknesses in the non-human feedback they received and felt it could not compete with the detailed and comprehensive information offered by their peers, nor could it substitute the personal and communicative aspects of peer-to-peer collaboration. Nevertheless, most students also believed the incorporation of AI reviews into their educational context had enriched their learning experiences, not only at times compensating for human subjectivity and lack of expediency but also providing effective connections to criteria and the overall objectives of the task. The experiences detailed in this article thus point to possible, fruitful human–machine collaborations, where AI does not *replace*, but effectively *complements*.

Notes

1. In this article, we employ Huisman et al.'s (2020) conceptualisation of peer feedback, defined as 'all task-related information that a learner communicates to a peer of similar status which can be used to modify [their] thinking or behavior' (p. 328).
2. The study discussed in this article was carried out in Spring 2023. At that time, the only publicly available GenAI version was GPT-3, which limited the technology employed.
3. We have expanded the findings detailed in this article in our recent work, which examines, among other areas, the development of students' AI literacy as well as their views on the effects of customised (i.e. calibrated) GenAI feedback on their writing (see Tzirides, Zapata, Bolger, et al., 2024; Tzirides, Zapata, Kastania, et al., 2024; Zapata et al., 2024).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Gabriela C. Zapata is Associate Professor of Education at the University of Nottingham. She also serves as the editor of the book series Multiliteracies and Second Language Education (Routledge) and co-editor-in-chief of the journal *Diversity & Inclusion Research* (Wiley). Her main research areas are Learning by Design and second and heritage language pedagogy, Generative AI in higher education and teacher training, and multimodal social semiotics.

Bill Cope is a Professor of Education at the University of Illinois, Urbana-Champaign. His research interests are literacy, multimodal meaning, digital learning ecologies, and the applications of artificial intelligence in education.

Mary Kalantzis is a Professor of Education and former Dean of the College of Education at the University of Illinois, Urbana-Champaign. Her research interests include theories and practices of pedagogy, approaches learner differences, literacy, and e-learning ecologies.

Anastasia-Olga (Olnancy) Tzirides is a Visiting Lecturer of Modern Greek at the University of Illinois Chicago and a Research Scientist at the University of Illinois, Urbana-Champaign. Her research interests include advanced digital technologies, artificial intelligence, user experience research, learning design, online education, and language and culture learning.

Akash Kumar Saini is a PhD student at the University of Illinois Urbana-Champaign. He specialises in designing and testing digital tools to enhance learning and integrate technology into education. Currently, his research explores how AI can support students in both higher education and K-12 settings, focusing on its strengths and limitations.

Duane Searsmith is an E-Learning Technical Specialist. He has developed writing analytics with the support of a series of IES and NSF grants. As well as developing the CGScholar AI Helper and CGAnalytics software, he has worked with advanced AI and natural language processing techniques and systems for over two decades. Mr Searsmith has recently built prompt engineering software that creates peer review feedback, using generative AI, from user-designed rubrics.

Jennifer Whiting is a PhD student at the University of Illinois Urbana-Champaign. She is an academic and educator with a career spanning education, curriculum design and information technology. Her teaching experience covers a range of institutions and programmes, including the Police Training Institute, the University of Illinois Urbana-Champaign.

Nikoleta Polyxeni Kastania is a PhD student at the University of Illinois Urbana-Champaign. She is also the Outreach Coordinator for the Illinois Center for Transportation at the Grainger College of Engineering at the University of Illinois, Urbana-Champaign.

Vania Castro is a Teaching Assistant Professor in the Department of Education Policy, Organization and Leadership at the University of Illinois Urbana-Champaign. Her current research interests include emerging technologies in education, mobile learning, multiliteracies, teacher education, and artificial intelligence.

Theodora Kourkoulou is a Clinical Assistant Professor in the College of Education at the University of Iowa. Her expertise and interests lie in the educational applications of informal and playful technologies, educational technologies, and AI in education.

John Jones holds a PhD in Global Studies in Education from the University of Illinois Urbana-Champaign and is currently an Assistant Professor in the Childhood/Early Childhood Education Department at SUNY Cortland. His research interests include philosophy of education, particularly American Pragmatism and the work of William James, media studies, and education technology.

Rodrigo Abrantes da Silva holds a PhD in Digital Literacies from the University of São Paulo. He specialises in designing digital technology experiences aligned with pedagogical principles to empower students. His work spans secondary and higher education, as well as teacher training, focusing on technology-driven learning practices.

ORCID

Gabriela C. Zapata http://orcid.org/0000-0002-3069-9866
Bill Cope http://orcid.org/0000-0003-2129-395X
Mary Kalantzis http://orcid.org/0000-0002-6552-3780
Anastasia Olga (Olnancy) Tzirides http://orcid.org/0000-0002-8641-206X
Akash K. Saini http://orcid.org/0000-0001-5452-0413
Duane Searsmith http://orcid.org/0000-0002-1155-0716
Nikoleta Polyxeni Kastania http://orcid.org/0009-0009-5677-9424
Vania Castro http://orcid.org/0000-0002-4930-5345
Theodora Kourkoulou http://orcid.org/0000-0002-8254-9253
John W. Jones http://orcid.org/0009-0004-2951-7684
Rodrigo Abrantes da Silva http://orcid.org/0000-0002-6631-8017

References

Baidoo-Anu, D., Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4337484

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, *1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Božić, V. (2023). *Artificial intelligence in education* [Unpublished manuscript]. https://doi.org/10.13140/RG.2.2.33278.25925

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, *19*(1), 1–24. https://aclanthology.org/J93-1001.pdf

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An education support system for student. In A. Castiglione, F. Pop, M. Ficco, & F. Palmieri (Eds.), *Cyberspace safety and security. CSS 2018. Lecture notes in computer science* (Vol. 11161, pp. 291–302). Springer. https://doi.org/10.1007/978-3-030-01689-0_23

Cope, B., & Kalantzis, M. (2015). The things you do to know: An introduction to the pedagogy of multiliteracies. In B. Cope & M. Kalantzis (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 1–36). Palgrave Macmillan.

Cope, B., & Kalantzis, M. (2020). *Making sense: Reference, agency, and structure in a grammar of multimodal meaning*. Cambridge University Press.

Cope, B., & Kalantzis, M. (2023a). Towards education justice: Multiliteracies revisited. In G. C. Zapata, M. Kalantzis, & B. Cope (Eds.), *Multiliteracies in international educational contexts: Towards education justice* (pp. 1–33). Routledge.

Cope, B., & Kalantzis, M. (2023b). A multimodal grammar of artificial intelligence: Measuring the gains and losses in generative AI. *Multimodality & Society*, *4*(2), 123–152. https://doi.org/10.1177/26349795231221699

Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y., Gasevic, D., & Chen, G. (2023, April 13). Can large language models provide feedback to students? A case study on ChatGPT. *EdArXiv*. https://doi.org/10.35542/osf.io/hcgzj

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and coassessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331–350. https://doi.org/10.1080/03075079912331379935

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Gamage, D., Staubitz, T., & Whiting, M. (2021). Peer assessment in MOOCs: Systematic literature review. *Distance Education*, *42*(2), 268–289. https://doi.org/10.1080/01587919.2021.1911626

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445. https://doi.org/10.2307/798843

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2020). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International*, *57*(3), 328–338. https://doi.org/10.1080/14703297.2019.1630294

Kalantzis, M., & Cope, B. (2020). *Adding sense: Context and interest in a grammar of multimodal meaning*. Cambridge University Press.

Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge.

Lane, H. C., McCalla, G., Looi, C.-K., & Bull, S. (2016). Preface to the IJAIED 25th anniversary issue, part 2. *International Journal of Artificial Intelligence in Education*, *26*(2), 539–543. https://doi.org/10.1007/s40593-016-0109-9

Li, J., Huang, J., & Cheng, S. (2022). The reliability, effectiveness, and benefits of peer assessment in college EFL speaking classrooms: Student and teacher perspectives. *Studies in Educational Evaluation*, *72*, Article 101120. https://doi.org/10.1016/j.stueduc.2021.101120

McGrath, C., Cerratto Pargman, T., Juth, N., & Palmgren, P. J. (2023). University teachers' perceptions of responsibility and artificial intelligence in higher education – An experimental philosophical study. *Computers and Education: Artificial Intelligence*, *4*, Article 100139. https://doi.org/10.1016/j.caeai.2023.100139

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *SSNR*. https://doi.org/10.2139/ssrn.4373111

Mori, M. (2012). The uncanny valley (K. F MacDorman & N. Kageki, Trans.). *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. Original work published 1970. https://doi.org/10.1109/MRA.2012.2192811

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2022). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, *28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Otaki, B. (2023). *Feedback in the era of generative AI: An explorative focus group study of student and educator perspectives in higher education* [Master's Thesis]. University of Gothenburg. https://gupea.ub.gu.se/handle/2077/77610

Russell, C. (2021, December 3). The disturbing uncanny valley of Robert Zemeckis film 'Polar Express'. *Far Out*. https://faroutmagazine.co.uk/the-disturbing-valley-robert-zemeckis-polar-express/

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2021). Identifying the components of effective learner-centered feedback information. *Teaching in Higher Education*, *28*(7), 1565–1582. https://doi.org/10.1080/13562517.2021.1913723

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Booth Olson, C. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning & Instruction*, *91*, Article 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Doubleday.

Tzirides, A. O., Saini, A., Zapata, G. C., Searsmith, D., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, T., Jones, J., da Silva, R. A., Whiting, J., & Kastania, N. P. (2023). Generative AI: Implications and applications for education. *arXiv*. https://arxiv.org/abs/2305.07605

Tzirides, A. O., Zapata, G. C., Bolger, P. A., Cope, B., Kalantzis, M., & Searsmith, D. (2024). Exploring instructors' views on fine-tuned generative AI feedback in higher education. *International Journal on E-Learning*, *23*(3), 319–334. https://www.learntechlib.org/primary/p/225173/

Tzirides, A. O., Zapata, G. C., Kastania, N. P., Saini, A. K., Castro, V., Abdul Rahman Ismael, S., You, Y., Afonso dos, S. T., Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, *6*, Article 100184. https://doi.org/10.1016/j.caeo.2024.100184

Van Leeuwen, T. (2005). *Introducing social semiotics*. Routledge.

Van Leeuwen, T. (2011). Multimodality and multimodal research. In D. Mannay & L. Pauwels (Eds.), *The SAGE handbook of visual research methods* (pp. 549–569). Sage. https://doi.org/10.4135/9781446268278

Van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning & Instruction*, *20*(4), 270–279. https://doi.org/10.1016/j.learninstruc.2009.08.004

VERBI Software. (2021). *MAXQDA 2022*. *Computer software*. maxqda.com

Winstone, N., Carless, D. (2020). *Designing effective feedback processes in higher education: A learning focused approach*. Routledge.

Wongvorachan, T., & Bulut, O. (2022). Feedback generation through artificial intelligence. *Open/Technology in Education, Society, and Scholarship Association Conference Proceedings*, *2*(1), 1–9. https://doi.org/10.18357/otessac.2022.2.1.125

Zapata, G. C., Saini, A. K., Tzirides, A. O., Cope, B., & Kalantzis, M. (2024). The role of AI feedback in university students' learning experiences: An exploration grounded in activity theory. *Ubiquitous Learning: An International Journal*, *18*(2), 1–30. https://doi.org/10.18848/1835-9795/CGP/v18i02/1-30