Ubiquitous Learning: An International Journal
ISSN: 1835-9795 (Print), ISSN: 2475-9686 (Online)
Volume 18, Issue 2, 2025
https://doi.org/10.18848/1835-9795/CGP/v18i02/1-30

Original Research

The Role of AI Feedback in University Students' Learning Experiences: An Exploration Grounded in Activity Theory

Gabriela C. Zapata, University of Nottingham, UK
Akash K. Saini, University of Illinois, Urbana-Champaign, USA
Anastasia Olga (Olnancy) Tzirides, University of Illinois, Urbana-Champaign, USA
Bill Cope, University of Illinois, Urbana-Champaign, USA
Mary Kalantzis, University of Illinois, Urbana-Champaign, USA

Received: 08/12/2024; ***Accepted:*** 09/16/2024; ***Published:*** 11/04/2024

Abstract: This study employs Engeström's second-generation activity theory (AT) to examine the transformative potential of Generative AI (GenAI) in providing formative feedback in higher education. Specifically, this research focuses on the experiences of fifty students with generic and calibrated GenAI feedback in a graduate program in the US. Through the analysis of participants' multimodal views and textual reflections after their experiences with these two types of AI reviews, we uncover the role that the AI reviewer played in the peer and AI review activity system of which students were part. The examination of the semiotic elements embedded in the participants' artifacts points to clear opinion differences in connection with the effectiveness and role of generic and calibrated AI feedback in the review activity system. The results show that while the generic AI reviewer was deemed an imperfect, limited tool, its calibrated counterpart was welcomed by students for its effectiveness and was even regarded by some participants as a new member of the community of practice within the activity system. Based on the findings resulting from this investigation, we suggest effective strategies for AI-human collaboration in higher education, aiming to enhance teaching and learning practices through advanced AI applications.

Keywords: *Generative AI, Generic vs. Calibrated AI Feedback, Higher Education, Activity System*

Introduction

The label "artificial intelligence" (AI) was coined in 1955 by computer scientist John McCarthy as a hook to attract funding for a workshop held at Dartmouth College for a small group of computer luminaries. In the words of the proposal, AI was "making a machine behave in ways that would be called intelligent if a human were so behaving" (McCarthy et al. 2006, 11). For many decades, the promise of AI was not realized, or at least not on a practicable or widely applicable scale. A first generation of rule-based, expert, and symbolic logic systems were not able to address the complexities of the empirical world, leading to the "AI winter," when the project was all but abandoned (Nilsson 2009). There were, however, promising yet limited applications of the expert systems paradigm in education beginning with the PLATO computer learning system of 1949 (Cope and Kalantzis 2023c; Dear 2017) and intelligent tutoring systems (Graesser et al. 2001).

From the turn of the twenty-first century, a second generation of AI development centered on empirical datasets and supervised machine learning (annotating data for analysis of statistical patterns) or unsupervised machine learning (labeling patterns that present themselves in the data). In education, perhaps the best example of this was automated essay assessments where a sample of student texts was graded by expert human examiners and new texts were graded by machines on the basis of their similarities, statistically determined by natural language processing (Cope et al. 2011; Shermis 2014; Warschauer and Grimes 2008).

Recently, with the advent of Generative AI (GenAI) and the wide accessibility of platforms such as OpenAI's *ChatGPT* or Google's *Gemini*, we have seen yet another paradigm shift—a third-generation development in AI, which retains the statistical bent of the second generation but now uses a technique called "self-supervised learning" to measure the proximity of words. The rapid development of this new AI iteration has raised deep concerns for educators, including the inability of detectors to determine whether student work is their own or AI generated; automation that may undermine the professional role of teachers; the privacy of teachers and students; "hallucination" or the creation of false facts; failure to acknowledge sources and the possibility of inventing fictional sources; the biases inherent to source texts from which GenAI draws; the quality of the filters required to moderate these sources; as well as fundamental concerns with respect to education ethics (Cope and Kalantzis 2023b).

Despite these challenges and concerns, GenAI also presents great promise for education (Johnson 2023; Mollick and Mollick 2023). For instance, many of the teaching routines that have been a staple part of teachers' professional practice when reviewing written assignments can now be supported by GenAI. This includes support and feedback during and after writing, calibrated to specific subject areas.

This article explores the possible benefits of calibrated GenAI in education. Specifically, this study employs Engeström's (1999) second-generation activity theory (AT) to examine the transformative potential that GenAI might have when providing formative feedback in higher education. This research focuses on the experiences of university students with generic and customized GenAI feedback in a graduate program in the US. Through the analysis of participants' multimodal views and textual reflections after their experiences with these two types of AI reviews, we seek to uncover the role that the AI reviewer played in the feedback activity.

This work is organized as follows. The first part offers information on existing work on GenAI feedback and introduces the study's theoretical framework. This is followed by the presentation of the research questions and methodology. In the final sections of the article, we discuss our results and provide suggestions for GenAI-supported instruction.

Generative AI Feedback

Feedback has long been recognized as a critical component of the learning process, essential for guiding students toward academic improvement (Black and Wiliam 2009; Nicol and Macfarlane-Dick 2006; Shepard 2009). Traditional feedback models have typically been conceptualized as unidirectional processes, wherein instructors deliver summative comments or grades aimed at correcting errors or reinforcing desired behaviors in students. While these feedback models have historically played a crucial role in providing students with essential performance feedback, Garrison and Ehringhaus (2011) argue that they encounter significant limitations, particularly in terms of scalability, time efficiency, and the ability to offer personalized guidance—concerns also noted by Shepard (2009). Moreover, contemporary research advocates for a more dynamic and reciprocal approach to feedback, emphasizing the importance of engaging students as active participants within the feedback process (Nicol and Macfarlane-Dick 2006; Winstone et al. 2016).

Recent advancements in feedback mechanisms have emphasized the transformative role of automated feedback tools, particularly in online learning environments, where the absence of face-to-face interaction necessitates alternative feedback methods. For instance, in a randomized controlled trial involving 1,136 instructors, Demszky et al. (2023) found that the use of an automated feedback tool enhanced instructor engagement with student contributions by 13 percent, leading to higher assignment completion rates and greater overall course satisfaction in an online computer science course. These tools can range from simple automated quizzes that provide immediate responses to students' answers, to more complex systems like intelligent tutoring systems. The integration of AI into these systems has further enhanced their ability to provide personalized, timely, and actionable feedback (Bulut and Wongvorachan 2022) that can be used to cultivate learner agency and productive action (Buckingham Shum et al. 2023) across different subjects without needing specific training (Jürgensmeier and Skiera 2024). Seo et al. (2021) found that there was a unanimous appreciation among instructors for the immediate feedback provided by AI, particularly during times when they were unavailable.

More recently, GenAI has helped significantly advance the application of AI to educational feedback. Unlike traditional AI, which relies on rigid algorithms and pre-set rules, GenAI uses advanced natural language processing techniques to generate more dynamic and contextually relevant feedback that provides scalable and automated evaluations based on predefined criteria (Bulut and Wongvorachan 2022). These systems draw on extensive linguistic databases to generate responses that simulate human-like interactions with student submissions, ensuring that the outputs are "grammatically correct and semantically meaningful" (Brynjolfsson et al. 2023, 4). Cotton et al. (2023) underscore the ability of GenAI to efficiently grade assignments while providing personalized feedback, thereby enhancing both the efficiency and customization of the learning experience. Meyer et al. (2024) further affirm that GenAI can achieve accuracy levels comparable to human assessors and can complement reviews and ratings by offering additional insights.

Zhai (2023), for example, has demonstrated that these systems can significantly address challenges in science education by creating performance-based assessments, evaluating student progress, providing personalized feedback, and recommending supplementary materials. Similarly, Mizumoto and Eguchi's (2023) analysis of 12,100 English essays from individuals of eleven different linguistic backgrounds revealed that GenAI significantly reduced grading time, ensured consistent scoring, and provided immediate scores and feedback. Moreover, Wan and Chen (2024) posit that GenAI, with minimal examples, can effectively assist in generating feedback that is highly consistent in the level of detail regardless of the amount of grading workload, potentially reducing the time required for grading student responses.

Steiss et al. (2024) reported similar results in a recent article that examined the quality of formative feedback provided by generative AI (GPT 3.5) compared to human evaluators for high school students' writing. The study compared these two types of feedback in connection with two hundred student essays based on five criteria: being criteria-based, clarity of directions for improvement, accuracy, prioritization 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.

Despite the apparent beneficial effects of GenAI feedback reported in the studies discussed previously, our work (e.g., Saini et al. 2024; Tzirides et al. 2023; Zapata et al. 2024), like Steiss et al.'s (2024), has also shown that when generic AI feedback was compared with that offered by humans, university students preferred the latter, citing quality, usefulness, and actionability as major factors for their preference. Specifically, our research revealed that participants felt human feedback (provided by peers) was more specific and comprehensive than that offered by the GenAI reviewer with respect to all aspects of their writing. Human reviewers were not only able to identify and pinpoint discrepancies between their peers' text and expected results based on assessment rubrics but also provided constructive suggestions for improvement with regard to content as well as stylistic, linguistic, and multimodal features. Additionally, the students in our work believed that the AI could not equal, let alone replace, the warmth, empathy, and support embedded in peer comments, which they felt highlighted strengths in their work, encouraging them to continue the revision process, and positively influencing their investment in their work.

In order to address the GenAI shortcomings identified by our participants and in previous studies, in Spring 2024, we added RAG (Retrieval Augmented Generation) processes to the GenAI reviewer used by our students, with a vector database consisting of 35 m tokens—the final, post peer- and instructor-reviewed drafts of all our graduate students' work for the past five years as well as instructors' writings (e.g., articles published in peer-reviewed journals). The objective of this work is therefore to explore whether the modifications we introduced to the GenAI exerted a change in our students' views of its feedback as well as the overall review activity. In the next section, we introduce the theoretical framework that grounded our work.

Theoretical Framework

Activity Theory

This study seeks to understand the role that GenAI feedback played in the formative assessment of participants' multimodal course projects before and after the AI was calibrated to align with the disciplinary context of their graduate programs. The theoretical basis of this work is Engeström's (1994, 1999) second-generation version of AT, first proposed in the late 1980s (see Engeström 2015, for a detailed account of AT's evolution). This theory is grounded in a cultural–historical approach to human cognition and behavior (Cole 1996; Vygotsky 1978, 1986). The major premise of this approach is that "the structure and development of human psychological processes emerge through culturally mediated [through material and symbolic artifacts such as language], historically developing, practical activity" (Cole 1996, 108). Vygotsky (1978) proposed that cognitive development results from the merging of fundamentally biological lower mental functions with socially originated higher mental functions. That is, higher mental functions originate in an interpersonal plane through the interaction of human beings with the outside world and then move to the intrapersonal plane, through a process of appropriation or internalization in a non-linear but recurring developmental process. This implies that a person's social relationships and the set of cultural values, principles, and procedures with which they are in contact shape their cognition.

Another important notion in this theory is that of mediated action. Vygotsky (1978, 1986) believed that, through the use of tools, human beings interact with their social environment and that in their action, they affect that environment and are affected by it. When discussing the function of tools in this theory, Vygotsky not only referred to material tools but also to the use of psychological tools (or cultural signs), which include abstract "artifacts" such as language, diagrams, art, music, and math. While the use of material tools is oriented toward outside objects and through these tools the individual exercises their influence on them, the use of signs can be directed toward the subject and other human beings, and their use results in differing patterns of behavior in the individual using signs and/or in other individuals. Psychological tools also play a fundamental role in allowing people to internalize the socio-cultural experiences that will determine the way they view and interact with the world. Also of importance is the concept of "context," since "to give an account of culturally mediated thinking, it is necessary to specify not only the artifacts through which behavior is mediated but also the circumstances in which the thinking occurs" (Cole 1996, 131).

Sharing the same foundations as the cultural–historical approach to human cognition and behavior, Engeström's (2015) AT offers a conceptual model that can be applied for the understanding of the social and dialogic nature of human activity, where knowledge and meaning are collectively constructed and distributed (Cole and Engeström 1993). The model "is deeply contextual and oriented at understanding historically specific local practices, their objects, mediating artifacts, and social organization, [while] seek[ing] to explain…qualitative changes in human practices over time" (Engeström 1999, 378). Engeström (1994, 1999) represents his second-generation AT model as a triangle with six interconnected components: the subject, the instruments (or tools), the object, the rules of engagement, the community of practice, and the division of labor.

The three main components—the subject, the instruments (or tools), and the object—are located in the upper part of the triangle, which represents "the level of mediated action through which the subject transforms the object in the process of acting upon it" (Cole 1996, 140). The action represented in the upper segment of the triangle "exists 'as such' only in relation to the bottom of the triangle" (Cole 1996, 140), where the rest of the components—the community of practice, the rules of engagement, and the division of labor—are located. A community of practice can be defined as a group of people who, being engaged in activities of similar nature, might "share the same general object" even if its members' actions are different; "the rules refer to explicit norms and conventions that constrain actions within the activity system; [and] the division of labor refers to the division of object-oriented actions among members of the community" (Cole 1996, 140–141). Lastly, an activity system also has an outcome, defined by Engeström (1995) as the "action response."

Since its development, Engeström's different iterations of AT have been applied in countless social contexts (e.g., hospitals and businesses; see Engeström 2008, 2015; Wiser et al. 2019), including educational institutions. Also, AT has been employed in the examination of human–computer interactions (e.g., see review by Clemmensen et al. 2016). In the next section, AT will be examined in connection with education, with a specific focus on those activities that have involved the use of technology.

Activity Theory in Technology-Supported Learning

In education, a myriad of studies "ha[ve] proven AT to be a very useful framework…[as] it acknowledges that learning occurs most naturally and meaningful[ly] in the context of an activity, in which cognition is distributed across people [e.g., students and teachers] and artifacts [e.g., textbooks, educational apps, etc.]" (Dolata et al. 2023, 61–62). That is, because of its comprehensive conceptual nature, AT offers scholars the tools to unveil the role that not only personal but also cultural and institutional factors might play in a specific learning activity. The interplay of these factors is perhaps best captured in Engeström's (1994) second-generation model, particularly when the activity involved is bounded in time and space. This model (presented in Figure 1) is ideal for examining both the mediated and social nature of learning experiences by allowing for a focus on individual students and the tools (including those relying on digital and AI technologies) they use to develop their knowledge as well as the ways in which they interact with peers and teachers. Evidence for the effectiveness of this model for the examination of technology-supported learning has been provided in two recent works—Chung et al. (2019) and Zheng et al. (2019).

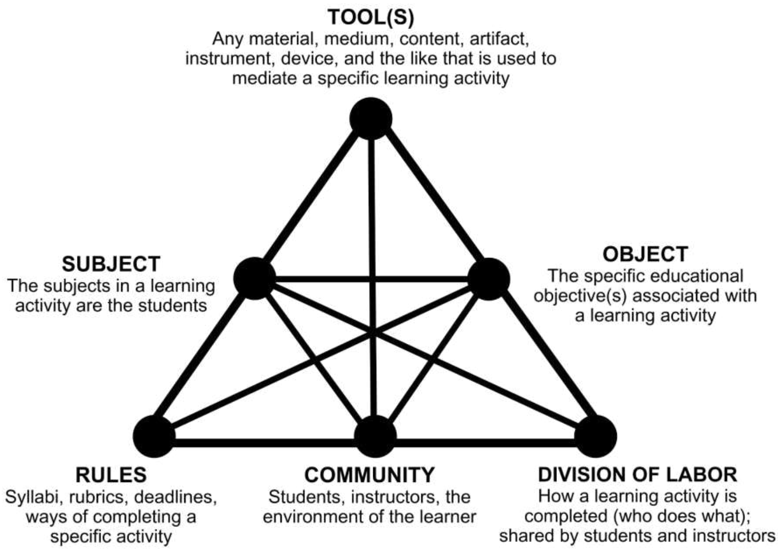

Figure 1: Engeström's (1994) Second Generation AT Applied to Learning
Source: Adapted from Chung, Hwang, and Lai 2019, 3

Chung et al. (2019) applied AT's second-generation model to understand the effectiveness of mobile learning in situating students in real-world contexts and activities. The study involved the examination of sixty-three articles published between 2010 and 2016. The researchers categorized the studies based on dimensions such as context, tools, control, communication, subject, and objective and examined the chosen works' distribution across these categories. The results revealed that AT has been a useful tool to identify the effectiveness of mobile learning in connection with various aspects of the learning process, such as the enhancement of learning outcomes, the promotion of active learning, and the development of personalized and situated learning tasks connected to real-life experiences. Based on the resulting AT analysis, Chung et al. recommend the ongoing exploration of these types of technologies, with a focus on specific subject domains (e.g., math), which would lead to further understanding of their application in situated learning activities.

Zheng et al. (2019) also employed second-generation AT to review studies investigating the incorporation of technology in educational contexts. The focus of this work was 134 articles on technology-supported peer assessment published between 2006 and 2017. The scholars developed an analysis framework based on AT, which included six components: subjects, objects, tools, rules, criteria used for peer assessment, and division of labor. The studies were coded based on this framework, and a correlation analysis using adjusted residual values was also carried out. This combination allowed for a comprehensive understanding of how technology-supported peer assessment was designed and implemented in the works examined. The results revealed that technology-supported peer assessment had been predominantly implemented in higher education and within social science, in mostly unstructured ways (i.e., lacking scaffolding). Additionally, general learning management systems were found to be more commonly used than dedicated peer assessment tools, which limited the affordances of the review activity. Also absent from the findings were students' perceptions and emotions, which led to Zheng et al.'s recommendation for more research in these areas.

More recently, AT has been proposed as an ideal approach for the examination of human–computer interaction involving GenAI in different social contexts, including education. For example, Vartiainen et al. (2023; see also Clemmensen et al. 2016 and Vartiainen and Tedre 2024) believe that the framework can help us examine human–AI machine ecosystems (Rahwan et al. 2019) to uncover the influence that humans exercise on technology and the ways in which GenAI technologies can, in turn, affect human behavior and agency. More importantly, AT's analytical concepts can unveil the role that GenAI technologies play in human activity—i.e., whether humans' relationship with GenAI "is just interactional…[or] can evolve into an intangible connection involving sensing…, emotional resonance, and even a sense of partnership" (Vartiainen and Tedre 2024, 20).

Dolata et al. (2023) have focused on this type of relationship by exploring the role of pedagogical agents (PAs) in various activity systems, with a particular emphasis on learning activity systems. The scholars defined PAs as "digital agents," such as the AI reviewer in our article, "capable of communication in natural language…designed to help a human learner improve their knowledge or skills" (Dolata et al. 2023, 57). Dolata et al. reviewed prior research on PAs to analyze how various characteristics of a learning activity featuring interactions between students and PAs can influence learning outcomes. Based on their analysis, the scholars proposed enhancing the agency assumption of AT (i.e., the active role of the subject in shaping their own development and the world around them) by regarding it as a perceived agency assumption (PAA). That is, the premise is to consider individuals (i.e., subjects) as not only engaging in and experiencing their own activities but also perceiving and understanding the actions of others (both in terms of intention and agency) (see also Engeström 2008). A focus on students' PAA within a learning activity involving a GenAI PA could thus provide insights into the role students assigned to PAs (i.e., a tool or a community member), which, in turn, would allow for a better understanding of the impact of GenAI technologies on their behavior and cognition and its influence on overall learning experiences.

As GenAI becomes more ingrained in education, understanding the relationship between PAs and students becomes crucial. The purpose of this work is to delve deeper into this relationship as well as contribute to the existing literature on GenAI assessment. To achieve this goal, we employ Engeström's (1994) second-generation AT model as well as Dolata et al.'s (2023) concept of PAA with regard to the students' participation in the AI review activity system. We employ these theoretical conceptualizations in the examination of the participants' multimodal representations of their experience with the GenAI reviewer before and after its calibration. In the following sections of the article, we introduce the research questions guiding this work and its methodology.

Research Questions

This work sought to answer the following research questions:

1. Based on its perceived effectiveness and usefulness, what role did the participants assign to the AI reviewer in their graduate course's paper review activity system when a generic GenAI interface was used?
2. Did the role the participants assigned to the AI reviewer in their graduate course's paper review activity system change when a specifically calibrated GenAI interface was used?
3. Did the paper review activity system change when a specifically calibrated GenAI interface was used?

The Present Study

Participants

Fifty graduate students of mixed genders and ethnic/racial backgrounds and ages ranging from 25 to 45+ years participated in this work. They were all enrolled in certificate, master's, and doctorate programs in education at a university in Midwestern US and were recruited for this work during two academic terms—Fall 2023 and Spring 2024—from four online graduate courses. All the participants were professional educators completing their graduate degrees part-time, and they were all experienced in areas ranging from school to higher education, workplace, and community education, crossing diverse discipline areas.

Educational Context

The focus of the four graduate classes in which the participants were enrolled was the relationship among learning, technology, and pedagogy, including the critical assessment of social, cultural, and historical factors as well as the examination of relevant theories and frameworks. The participants both accessed the materials for their courses and completed their course tasks through the digital tools offered by the *CGScholar* platform (Cope and Kalantzis 2023a). This platform was first developed in 2000, and since then, experimental online writing and assessment spaces have continued to be developed in a number of loosely linked applications under the overall platform name *CGScholar.com* (Cope and Kalantzis 2023d). Students' work in each of the graduate classes offered on CGScholar was assessed through the development of multimodal critical pieces examining technology, educational theory, and practice. Students chose their topics and then incrementally worked on these projects throughout the semester, receiving both GenAI and peer feedback at different points of the development process. This feedback resulted in a collaborative review activity the goal of which was the provision of formative suggestions for the improvement of students' work before final submission.

Both peer and AI feedback relied on the same explicitly stated assessment measures and rubric, which were grounded in the multiliteracies framework *Learning by Design* (Cope and Kalantzis 2023d; New London Group 1996). This pedagogy takes an epistemological approach to learning, focusing both on cognition and knowledge-making activities involving material practices, embodied activity, and socio-emotional engagement (Cope and Kalantzis 2015). These are high-level, abstract review criteria. The components of the activity system connected to the review of the participants' work and the relationship among them are visually presented in Figure 2.

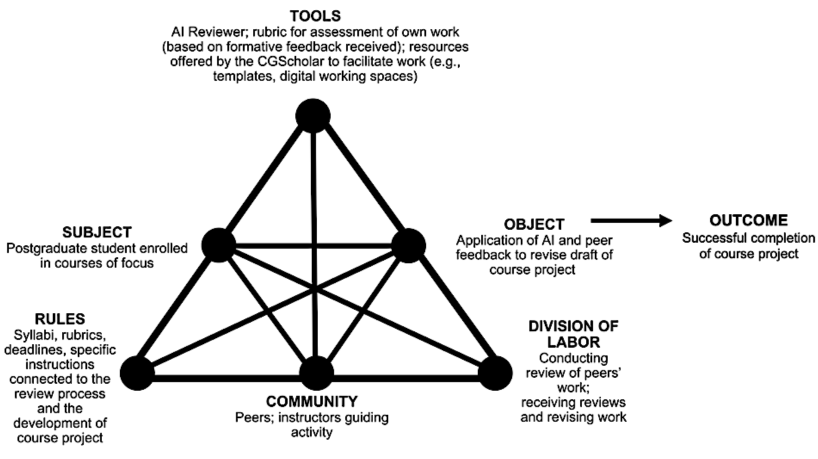

Figure 2: Engeström's (1994) Second-Generation AT Applied to the Overall Review Activity

Generative AI Feedback in Activity of Focus

GenAI reviews were first incorporated into the review activity in Spring 2023, and they constituted a novel addition to the feedback offered to students on their course projects, as they complemented the formative suggestions provided by peers and instructors. From January to

December 2023, this new intervention leveraged a large language model (LLM) for dialogue applications (Tzirides et al. 2023; Zapata et al. 2024). OpenAI's GPT-3 was connected to an app within *CGScholar* via application programming interface (API). When called for, the AI review would loop through students' entire work once for each rubric criterion, using it as the prompt (i.e., the rubric criteria read like prompts), and it would generate qualitative feedback. This feedback would also be appended with an overall rating for each criterion.

In January 2024, in order to address the weaknesses and limitations identified by students in connection with the GenAI feedback they had received in their graduate courses (see Tzirides et al. 2023; Saini et al. 2024; Zapata et al. 2024), the AI reviewer was recalibrated. The calibration was aimed not only at addressing these drawbacks but also at aligning AI feedback more closely with the disciplinary context of the graduate programs so that its suggestions would be more specific and actionable for students to revise their work. While still relying on an LLM (Open AI's GPT4), the newly calibrated AI encompassed a customized knowledge storage of approximately 35 million words deriving from relevant research articles and students' works. Additionally, each student's full submission was now included in the prompt created for the AI review. Once the calibrated AI was implemented, we sought to investigate whether this change had affected the review activity and, if so, in what ways it had influenced the effectiveness of the feedback offered to students as well as the role it played in the review activity system.

Instruments

After their course papers had been reviewed by both peers and the AI, the participating students were asked to create images employing their preferred GenAI image generator platform to convey their experiences with AI reviews. These visual artifacts were complemented with textual comments in which the participants expanded their views on the effectiveness of AI feedback. Both the GenAI images and the textual reflections were collected in connection with both the generic (Fall 2023) and specifically calibrated AI feedback (Spring 2024). These two sets of visual artifacts and reflections became the data for this article. By collecting these data, our objective was to focus on the subjects' (i.e., students) PAA within the review activity system in connection with the AI reviewer (Dolata et al. 2023).

The decision to focus on these two sources of data was driven by both our participants' classroom experiences and the richness we felt the analysis of multimodal artifacts would add to our investigation. That is, since work in the courses in which the students were enrolled required them to create multimodal artifacts (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) work, that the examination of the participants' multimodal communication would offer a more holistic and nuanced understanding of their experiences with AI reviews than could be achieved solely through text analysis.

Procedures

Once all participants had submitted their AI-generated images, they were analyzed following a similar approach to the one employed by Putland et al. (2023) for the exploration of GenAI images. Since these scholars' work was based on the same types of images as the present study, we deemed this choice appropriate. Also, we felt that adopting an analysis grounded in "Barthian notions of denotation (what is depicted?) and connotation (what is meant or implied?)" (Putland et al. 2023, 10) would result in an in-depth understanding of the message expressed multimodally by our participants. In this work, the analysis first involved determining who the participants in the artifacts were (i.e., who was depicted) and the setting in which they were placed. In the next stage, we examined the relationship among participants and setting, based on the semiotic elements (e.g., gestures, spatial organization, color, size) used to convey how each element related to others. The artifacts' semiotic elements were also analyzed following the tenets of social semiotics (Cope and Kalantzis 2020; Kalantzis and Cope 2020; Kress and van Leeuwen 2021).

These analyses led to the interpretation of the overall message expressed multimodally in connection with the participants' view of AI feedback. Specifically, we focused on the role that the AI reviewer had in students' PAA (Dolata et al. 2023) within the review activity system. The participants' linguistic reflections were employed as further sources of information in the interpretation of the multimodal artifacts. Once the analyses had been finalized, the data before and after the AI calibration were compared. The findings resulting from these processes were then interpreted in connection with Engeström's (1994, 1999) second-generation version of AT and are presented and discussed in the following two sections of the article.

Results

The analysis of the multimodal, AI-generated artifacts the participants developed before and after the GenAI reviewer calibration revealed that meaning had been conveyed through the visual, spatial, and gestural modes of communication. An examination of the semiotic elements embedded in these products suggests that most students sought to convey their experiences with the AI reviewer metaphorically, employing colors, object position, size, and gestures/body language to express emotions and relationships. Intended figurative meanings were also described and extended in the participants' textual reflections. The findings point to clear opinion differences in connection with the effectiveness of AI feedback and the students' PAA (Dolata et al. 2023) with regard to the role the AI reviewer played in the review activity system before and after the calibration. These views are presented in the next two sections.

Students' Views Before the AI Calibration

The AI-generated multimodal artifacts developed by the participants after receiving generic AI feedback, as well as the textual reflections accompanying them, convey overall feelings of dissatisfaction and sometimes uneasiness toward the adoption of these types of reviews.

Even though the AI reviewer employed the same rubric as its human counterparts in the review activity, most students regarded AI feedback as too limited in terms of specific content and stylistic suggestions due to what they believed was its lack of understanding of the context and nuances of academic writing. This feedback was also characterized as being too general—i.e., unable to offer the kind of useful examples and detailed, in-depth actionable points found in most peers' comments. A further negative aspect of the AI reviewer was highlighted in connection with its inability to detect multimodal elements incorporated into the students' essays, which resulted in mistaken feedback on already present information that the AI had failed to identify. These types of errors were significant considering that the participants were expected to produce multimodal course projects. A final, important drawback of the AI noted by all participants was its lack of "humanity," evident from the tone and language choices embedded in its comments. The following exemplify these opinions:

> I found the AI reviews to be stark…while the peer reviews have…a sense of empathy, and a general human quality (obviously) that AI cannot replicate. (Participant 5)

> The AI review…can sometimes fall short on the human aspects of the project. By this, I mean…emotions and reasoning….A technological construct does not have these lived experiences. (Participant 25)

> There is a certain human element [that] an AI cannot quite attain, especially regarding the more personal benchmarks such as experience and purpose. (Participant 47)

Multimodally, these views were communicated through varied, metaphoric representations that relied on colors, object size and position, and gestures/body postures for meaning making. In most artifacts, the AI was represented as a machine or robot lacking any features that would lead to being identified as a real human being. Additionally, in some representations, the students expressed their preference for human reviews over AI feedback through blunt light/dark color contrasts and the juxtaposition of representations of doom-like scenes and positive human interactions. Figures 3 to 5 offer examples of these types of representations.

For instance, in the multimodal artifact in Figure 3, Participant 17 depicted the AI reviewer as a melting computer with a lifeless human face (communicated, for example, by its empty eyes), in a vacuous space, surrounded by disorganized wires, seemingly disconnected from reality and/or the social context of the learning community. Through this depiction, this student sought to emphasize the fact that, although "posing as [an] intelligent and human-like [reviewer, the AI was] just a fake, hollow representation of one," characterized by the " 'melting down' [of] the [review] process, [in which] each AI review [was] less useful, accurate, and complete than the previous one." The disorganized array of disconnected wires appears to have been introduced by the participant to symbolically convey the idea of "generative AI as a prototype in the infancy of its development," while the blank space in which the computer was placed represented "the empty promises of AI, yet to be realized" (Participant 17). These negative views were further expressed visually by the dark, ominous shades of black, blue, and gray that prevail in the image.

Figure 3: AI-Generated Multimodal Artifact Created by Participant 17

The artifact in Figure 4, developed by Participant 14, offers a similar negative view. In this representation, the student is sitting at her desk while working on her project. She is in a dark room, with a large window, in what appears to be an apartment in a big city. The AI reviewer is represented as a large, menacing robotic structure with wires connected to different parts of the room, facing the student. There appears to be no social connection between the two entities (i.e., there is no physical or eye contact), except for the wires attached to the student's computer and the rest of the environment. The robotic structure also seems to be blocking the participant's view from the outside world. Through this visual isolation and lack of communication, as well as the choice of predominantly dark shades of red and lack of light, the student conveyed feelings of isolation and doom (even the plant on the student's desk appears to be dead) and her negativity with regard to the AI. These emotions were also expressed textually: "I felt removed from my peers while learning in an online environment [and] after I completed my AI review, I felt discouraged by the feedback that I deemed harsh and contradictory.…The almost ominous size of the robot represents how I felt after reading my AI review" (Participant 14).

Figure 4: AI-Generated Multimodal Artifact Created by Participant 14

Similar emotions were expressed by Participant 24 in their multimodal artifact (Figure 5), which contrasted AI and peer reviews in side-by-side representations, in separate semiotic frames (AI reviews on the left, human feedback on the right). Like Participants 17 and 14, this student chose to portray their experience with AI reviews in dark, blue, and gray monotones, depicting the AI as an isolated computer in what appears to be a sea of papers (representative of students' written works). No human beings are seen in the image, which constitutes a figurative view of the AI reviewer as "an abstract, chaotic landscape…, symbolizing [its] often confusing [,] impersonal…and repetitive nature" (Participant 24). In contrast, in the peer review representation, the participant's emphasis was on personal, solely human communication, expressed through the presentation of a female student and her male peer sitting next to each other. The figures' body position, spatial proximity, and gestures signal that they are engaged in conversation involving their written work, represented by the notebooks and papers on their desk and in the female student's hands. Both figures are dressed in bright colors and seem to be situated in a brightly painted classroom. In their textual reflection, Participant 24 offered more information on their intended meaning, stating their wish "to show the human peer review side in warmer tones of orange and green, with figures in a discussion representing the…more personalized nature of human feedback." Clearly, the semiotic separation resulting from the presentation of the two review experiences in different frames was also embedded in the differing semiotic elements chosen by this student to convey their views of the two types of feedback.

Figure 5: AI-Generated Multimodal Artifact Created by Participant 24

The application of Engeström's (1994) second-generation AT model to the review activity based on the analysis of the participants' multimodal depictions, as well as the textual reflections accompanying them, suggests that they regarded the generic AI reviewer solely as a tool. The artifacts are also reflective of the students' characterization of AI reviews as not organized, limited, at times erroneous, and, thus, mostly not effective for the achievement of their objective—i.e., the incorporation of feedback for the improvement of their written work. In contrast, the participants considered their peers as effective collaborators within the review activity's community, sharing with them the division of labor by mutually reviewing their written works and co-constructing knowledge through the provision of specific suggestions and actionable points for the improvement of drafts. These relationships and the role attributed to the AI reviewer and its perceived effectiveness is visually represented in the AT system in Figure 6. While the lower aspects of the system did not appear to have been affected in the review activity, the AI review seems to have disrupted the upper part of the system during the review process, since the feedback offered was not effective enough to aid subjects in reaching their objective. This disruption is graphically represented by the dotted lines connecting the subject, tools, object, and outcome.

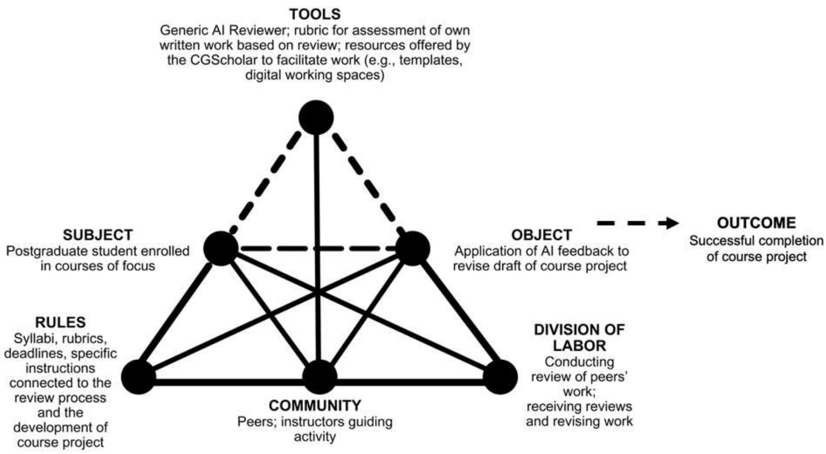

Figure 6: Engeström's (1994) Second-Generation AT Applied to the Review Activity Pre-AI-Calibration, with a Focus on the Generic AI Reviewer

Results After the Calibration of the AI Reviewer

The analysis of the artifacts submitted by the participants after the GenAI reviewer had been recalibrated points to a dramatic change in their opinions on its effectiveness as well as their PAAs with regard to its role within the review activity system. Both the multimodal representations and the textual reflections accompanying them suggest that students felt that the calibrated AI's comments were much more specific and relevant for the development of the topics on which they had chosen to focus. For example, Participant 10 characterized the feedback received as "excellent, [having] allowed [them] to make significant changes," while Participant 26 felt the AI had "efficiently analyze[d] and enhance[d] the clarity of [their] written content…provid[ing] what [they] consider to be *transformative insight*" (emphasis added). Participant 34 appeared to echo these views, describing the AI reviewer as "a *potent* tool for refining your work" (emphasis added). Participant 38 offered a comprehensive reflection that summarized some of the instructional benefits identified by other students as well as shed more light on the nature of the information given by the calibrated reviewer:

> I was absolutely astonished by the type of feedback that the AI review was able to generate. The feedback was not only specific and clear, but also based on my research and other research within the field. For instance, when commenting on my need for a larger counterargument section, the AI suggested multiple references I could use to provide arguments for that section…I have plenty of actionable feedback to go off of. (Participant 38)

The positive opinions expressed linguistically were mirrored in participants' multimodal representations. Again, students resorted to the use of colors, object size and position, and gestures/body postures to figuratively convey their experiences with the calibrated AI reviewer. In the case of some participants (e.g., 14 and 17), the comparative analysis of the semiotic elements in the artifacts created pre- and post-calibration revealed a stark multimodal contrast. For example, the emotions expressed visually and gesturally by the image generated by Participant 14 post-calibration (Figure 7) are clearly in opposition to those expressed in Figure 4 (pre-calibration). While the AI review is still represented as a computer in Figure 7, it now exhibits more human-like features and emotions, and it seems to be communicating with the student (e.g., there is eye contact), who is still sitting at her desk, but now appears to be relaxed and enjoying her interaction with the AI (both the student and AI are smiling, and they are closely positioned). Additionally, unlike Figure 4, the image is bright, with predominantly light colors, which contributes to its overall positive tone. Details such as the now thriving plant on the student's desk further support this message. The participant expanded on their multimodal viewpoint by stating that they were "blown away by the thoroughness and helpful specific suggestions from [the] AI review, and [that's] why the woman in this image is so happy."

Figure 7: AI-Generated Multimodal Artifact Created by Participant 14 to Convey Their Experience with the Calibrated GenAI Reviewer

The artifact generated by Participant 17 after the AI calibration also evinces contrasting opinions with regard to the AI's effectiveness and, more importantly, its perceived agency and role in the review activity system. This student characterized their experience in very positive terms, stating that

> the AI review…was improved by leaps and bounds…This time around I was impressed by the capabilities of AI, [as its feedback] made sense, was well-organized in the same format for each rubric category,…articulated clear strengths and weaknesses in my work, and—most importantly—made appropriate, useful suggestions for revision that drew from a knowledge base impossible for any individual peer to have. (Participant 17)

In this participant's post-calibration representation, the AI is no longer a "melting down," ineffective computer (Figure 3); instead, it has assumed a human, female shape exuding light and digital connections (Figure 8). This female AI is the most salient image in the picture in terms of both size and visual details, and this prominence appears to convey the idea of her power over the small human shapes that have been placed next to her, on the left. The setting is also more "human" since the location appears to be a beach at dawn, which might also imply that the AI is no longer in a digital vacuum and has now become part of the real world. The human figures are walking toward the new "day," while the AI remains in place, overlooking this journey. Participant 17 explained this symbolic representation in these terms:

> Th[is] image…reflects my experience with a robotic, but realistic looking woman representing AI waking up to the dawn of a new day in generative AI application. Her presence dominates the image, overshadowing the small human figures walking

> away into the horizon, just as the AI review I received dominates my revision process, overshadowing peer comments in its timeliness and robustness. (Participant 17)

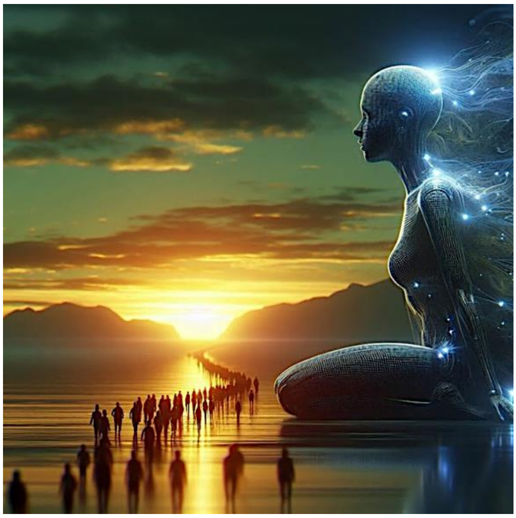

Figure 8: AI-Generated Multimodal Artifact Created by Participant 17 to Convey Their Experience with the Calibrated GenAI Reviewer

Clearly, the experience that Participants 14 and 17 had with the calibrated AI reviewer positively influenced their views on its effectiveness. Additionally, both their artifacts and reflections point to a change in the way they started to perceive the AI with regard to its agency within the activity system. That is, these students' choice of human features to represent the AI reviewer might signal a shift in the AI's perceived role, conveying its transformation from an imperfect tool into a collaborative, effective "partner," whose suggestions might surpass those offered by human peers.

The perception of the AI reviewer as a collaborative "peer" was also found in other participants' representations and textual reflections. For example, in the artifact generated by Participant 11 (Figure 9), the AI reviewer is presented as a humanoid robot sitting at a table with the peer reviewer and the student whose paper is being reviewed. The three figures are in a physical classroom, and both their body postures and proximity seem to indicate they are involved in a conversation regarding the student's paper, which is on the desk. The facial features of the robotic AI are expressive, and they suggest it is fully interacting with the human figures. These multimodal choices were further clarified by Participant 11, who stated that their objective was to create a "visual representation showing how all three, student, AI, and…peer (human), could work together.…All three characters in the image are sitting at the same eye level, which helps to illustrate the power dynamic between the three." This student's reflection and their AI visualization, as well as those generated by Participants 14 and 17 post calibration, suggest that, although still not fully human, the AI reviewer has become much more than a tool.

Figure 9: AI-Generated Multimodal Artifact Created by Participant 11 to Convey Their Experience with the Calibrated GenAI Reviewer

Similar attitudes toward and perceptions of the AI reviewer were also conveyed linguistically. For example, Participant 49 felt that the calibrated AI "was far better at providing feedback than [themselves]." Additionally, this student highlighted how difficult it had now become to distinguish between human and AI feedback: "The strangest part about the AI reviews is that if you told me that they were one of my peers', I would believe it. The AI writes in a very well-balanced way, both offering critiques and a soft landing of compliments." The perceived "humanity" of the AI post calibration was also embedded in the reflections offered by Participant 26, who referred to the AI reviewer as distilling "clarity of thought…brilliance, [and] wisdom," as well as Participant 23, who believed that "the revision provided by the AI was much *kinder* and *softer* than peer feedback" (emphasis added). Participant 30 also focused on social aspects when describing the AI review process as "similar to classmates convening to discuss our class and its assignments…enjoying their time together and actively participating in the discussion." Even though in some of these students' views the calibrated AI's feedback was deemed more effective than the suggestions provided by classmates, it is important to emphasize that none of the participants felt that the AI should replace peer reviews. Instead, both review types were welcomed by all students, who felt that "when used together, [they] are the most well-rounded" (Participant 25) and "they are both helpful in their own ways" (Participant 4).

The participants' multimodal and linguistic depictions of their experiences with AI feedback post-calibration suggest changes in the review activity system, particularly in connection with the students' PAA with regard to the AI reviewer. When Engeström's (1994) second-generation AT model is applied to this system (Figure 10), the analysis reveals changes in its tools, community, and division of labor. That is, since the participants' (subjects) appeared to perceive the AI reviewer as an agentive peer/collaborator more than a tool, the AI could be seen as having (or on its way to) become part of the community of practice. Also, since in some participants' views, the AI's suggestions started to surpass those provided by human peers, the division of labor exhibited changes, as the responsibility of offering actionable feedback now relied on both classmates and the AI reviewer. Clearly, the calibration modified the existing review system, at least as perceived by some participants.

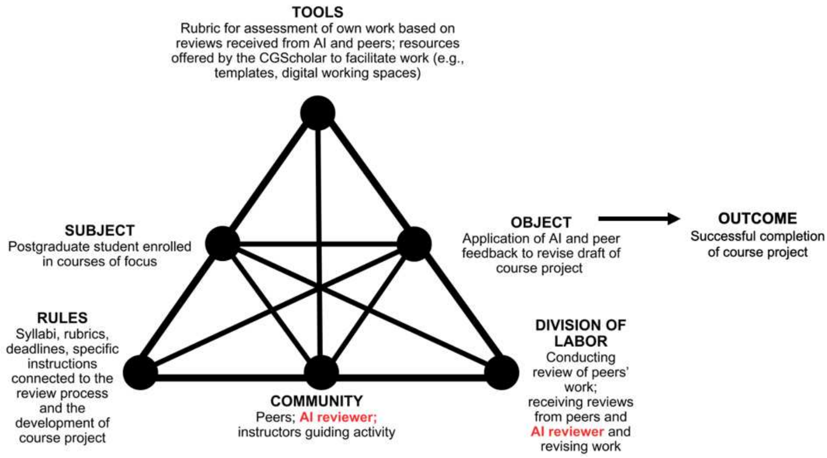

Figure 10: Engeström's (1994) Second-Generation AT Applied to the Review Activity After the GenAI Reviewer Calibration

Discussion

The first two research questions this work sought to answer probed into the role the participants assigned to the AI reviewer before and after calibration. The application of Engeström's (1994) second-generation AT model to the review activity based on the analysis of students' multimodal artifacts and textual reflections suggest that there was a change in the way the AI was perceived by most of them. Before calibration, the AI seems to have been regarded solely as a tool—another component (albeit imperfect) offered by *CGScholar*, the digital environment in which the learning process was taking place, with no other role in the review activity system. This perception appears to have been rooted in the contrast participants identified between AI and peer feedback. The differences noted originated in the nature of the comments offered both in terms of their effectiveness and social/communicative value. For example, the AI reviewer's suggestions were characterized as being less specific and actionable than those provided by peers, and, more importantly, they appeared to have lacked the emotional, human quality the participants welcomed and enjoyed when interacting with their classmates. These results mirror those reported in previous comparisons between generic GenAI and human feedback (e.g., Saini et al. 2024; Steiss et al. 2024; Tzirides et al. 2023; Zapata et al. 2024), where the quality of human suggestions was deemed higher and more academically and socially relevant than those given by GenAI.

The calibration of the GenAI reviewer in January 2024 seems to have affected the review activity dramatically, particularly with regard to the students' (the system's subjects) PAA with regard to both the place of the AI and its role within the activity system. The analysis of the semiotic elements in the participants' multimodal representations and the textual reflections point to their view of AI as being more like a collaborative partner and, in some cases, a member of the community of practice, than a tool. Clearly, the changes made to the AI to answer the specific needs of the students in the program influenced their perceptions, as the detailed, actionable, and thoughtful feedback it offered was very similar, and at times even more effective than peers' comments. Since the participants could now see a connection between their work and the AI's words and could, therefore, identify with the suggestions offered, they seemed to have grown closer to the AI, bestowing it with human qualities and incorporating it into the community of practice. Interestingly, this change did not diminish the relevance of peer reviews; instead, all participants seemed to have embraced both human and AI feedback, highlighting the richness that this combination had brought to their learning process.

The third research question in this study focused on changes in the review activity system that might have taken place as a result of the GenAI calibration. The findings not only revealed a re-structuring of the community of practice and division of labor within the system but also offered evidence for the symbiotic relationship between humans and GenAI, reflective of mediated human activity (Cole 1996; Cole and Engeström 1993). For example, the AI calibration originated in the weaknesses and drawbacks that students had identified in the generic reviewer, pointing to transformations in the tool as a result of human use. Additionally, the modifications introduced evince human control over the AI's content and actions, and they constitute a clear example of the training and application levels within the GenAI's life-cycle model proposed by Abdelghani et al. (2023) for the effective use of GenAI in education.

On the other hand, the calibrated AI appeared to have shaped human action and thinking, transforming both individual and collective agency within the review system. The effective and socially positive interactions between the participants and the AI seemed to have resulted in a new community of practice, where humans (both the students themselves and the peers reviewing their work) and the AI shared ideas, resources, and support, leading to the emergence of a novel collective intelligence and combined expertise, enriching and expanding the review process, and, in turn, the participants' graduate studies through personalized, interactive, and empowering learning experiences. These findings highlight the transformative potential of responsible and mediated GenAI in human activities previously discussed by Cope and Kalantzis (2023e, 2024) and Vartiainen and Tedre (2024), including educational contexts such as the one described in this work. This study also offers further support for the application of Engeström's (1994) AT for the examination of the role that GenAI might play in education, contributing to Dolata et al.'s (2023) and Vartiainen et al.'s (2023) work. In the next section of the article, we analyze the pedagogical implications resulting from our investigation.

Conclusion: Pedagogical Implications

Based on the findings presented in this article, we offer pedagogical recommendations for educators and educational professionals in higher education seeking to integrate GenAI for formative feedback. These implications aim to enhance the educational process, foster a more collaborative learning environment, and address potential challenges. By leveraging the insights gained from this study, educators can better understand how to effectively use GenAI to support and enrich student learning experiences.

- Enhancement of Educational Process
    - Improved Feedback Quality: The study demonstrates that calibrated GenAI can deliver feedback that is specific, clear, and actionable, often surpassing generic AI and enriching peer feedback, resulting in higher overall feedback effectiveness. This indicates that integrating calibrated GenAI into the feedback process can substantially enhance the quality of formative assessments. Educators should consider utilizing such tools to provide more detailed and context-specific feedback, thereby facilitating deeper learning and improved academic outcomes.
    - Personalized Learning: The ability of calibrated GenAI to offer tailored feedback based on individual student submissions supports personalized learning. This approach aligns with modern educational theories (Kalantzis and Cope 2015; Cope and Kalantzis 2016) that emphasize addressing individual learning needs and differences. Institutions should leverage GenAI to create more personalized learning experiences that can cater to diverse student populations and learning styles.
- Fostering Collaborative Learning Environments
    - AI as a Collaborative Partner: The transformation of the AI reviewer from a mere tool to a collaborative partner within the learning community, as perceived by students, suggests that GenAI can play a significant role in collaborative learning environments. Educators should promote the use of GenAI not only as a feedback tool but as a co-learner that participates in the learning process alongside students and peers. This can enhance the sense of community and shared purpose in educational settings.
    - Balancing Human and AI Feedback: The study indicates that while GenAI feedback is highly valued, it does not replace the need for human peer feedback. Instead, the combination of both AI and human feedback is seen as providing the most comprehensive support for student learning. Educators should strive to balance the use of AI with human interactions to ensure that students benefit from both the precision of AI and the empathy of human feedback.

- Addressing Challenges
    - Calibration and Context-Specificity: The effectiveness of GenAI is heavily dependent on its calibration to specific disciplinary contexts. This necessitates ongoing efforts to update and refine AI systems to ensure they remain relevant and effective. Educational institutions must invest in the development and maintenance of calibrated GenAI systems to ensure they provide meaningful and contextually appropriate feedback.
    - Training and Familiarization: Both students and educators need to be adequately trained to use GenAI tools effectively. AI literacy, which encompasses understanding the capabilities and limitations of GenAI, is becoming an essential attribute for all educational professionals. This literacy enables them to grasp both the opportunities and threats posed by AI technologies (Tzirides et al. 2024). Consequently, professional development programs should be designed to help educators integrate GenAI into their teaching practices effectively.
    - Ethical Considerations: The integration of GenAI in education raises important ethical considerations, particularly regarding data privacy and the transparency of AI processes. Institutions must establish clear guidelines and policies to address these concerns, ensuring that the use of GenAI is ethical and aligned with broader educational values.

In conclusion, this work demonstrates the transformative potential of GenAI in providing formative feedback in higher education. By enhancing feedback quality, supporting personalized learning, fostering collaborative environments, and addressing associated challenges, educators can harness the power of GenAI to significantly improve teaching and learning practices. The ongoing refinement and ethical implementation of these technologies will be critical to their successful integration into educational systems.

Acknowledgment

Funding acknowledgment: US Department of Education, Institute of Education Sciences: “The Assess-as-You-Go Writing Assistant” (R305A090394); “Assessing Complex Performance” (R305B110008); “u-Learn.net: An Anywhere/Anytime Formative Assessment and Learning Feedback Environment” (ED-IES-10-C-0018); “The Learning Element” (ED-IES-lO-C-0021); “InfoWriter: A Student Feedback and Formative Assessment Environment” (ED-IES-13-C-0039); and “The CGScholar AI Helper: A Generative AI Writing and Learning Assistant” (91990024C0010). Bill and Melinda Gates Foundation: “Scholar Literacy Courseware.” National Science Foundation: “Assessing ‘Complex Epistemic Performance’ in Online Learning Environments” (Award 1629161).

AI Acknowledgment

The authors declare that generative AI or AI-assisted technologies were not used in any way to prepare, write, or complete essential authoring tasks in this manuscript.

Informed Consent

The authors have obtained informed consent from all participants.

Conflict of Interest

The authors declare that there is no conflict of interest.

REFERENCES

Abdelghani, Rania, Hélène Sauzéon, and Pierre-Yves Oudeyer. 2023. "Generative AI in the Classroom: Can Students Remain Active Learners?" *arXiv:2310.03192v2* [cs.CY]. https://doi.org/10.48550/arXiv.2310.03192.

Black, Paul, and Dylan Wiliam. 2009. "Developing the Theory of Formative Assessment." *Educational Assessment and Accountability* 21 (1): 5–31. https://doi.org/10.1007/s11092-008-9068-5.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. *Generative AI at Work*. Working Paper 31161. National Bureau of Economic Research.

Buckingham Shum, Simon, Lisa-Angelique Lim, David Boud, Margaret Bearman, and Phillip Dawson. 2023. "A Comparative Analysis of the Skilled Use of Automated Feedback Tools Through the Lens of Teacher Feedback Literacy." *International Journal of Educational Technology in Higher Education* 20: 1–42. https://doi.org/10.1186/s41239-023-00410-9.

Bulut, Okan, and Tarid Wongvorachan. 2022. "Feedback Generation Through Artificial Intelligence." *OTESSA Conference Proceedings* 2 (1): 1–9. https://doi.org/10.18357/otessac.2022.2.1.125.

Chung, Ching-Jung, Gwo-Jen Hwang, and Chiu-Lin Lai. 2019. "A Review of Experimental Mobile Learning Research in 2010–2016 Based on the Activity Theory Framework." *Computers & Education* 129: 1–13. https://doi.org/10.1016/j.compedu.2018.10.010.

Clemmensen, Torkil, Victor Kaptelinin, and Bonnie Nardi. 2016. "Making HCI Theory Work: An Analysis of the Use of Activity Theory in HCI Research." *Behaviour & Information Technology* 35 (8): 608–627. https://doi.org/10.1080/0144929X.2016.1175507.

Cole, Michael. 1996. *Cultural Psychology: A Once and Future Discipline*. Belknap Press of Harvard University Press.

Cole, Michael, and Yrjö Engeström. 1993. "A Cultural-Historical Approach to Distributed Cognition." In *Distributed Cognitions: Psychological and Educational Considerations*, edited by Gavriel Salomon. Cambridge University Press.

Cope, Bill, and Mary Kalantzis. 2015. "The Things You Do to Know: An Introduction to the Pedagogy of Multiliteracies." In *A Pedagogy of Multiliteracies: Learning by Design*, edited by Bill Cope and Mary Kalantzis. Palgrave.

Cope, Bill, and Mary Kalantzis. 2016. "Conceptualizing e-Learning." In *e-Learning Ecologies: Principles for New Learning and Assessment*, edited by Bill Cope and Mary Kalantzis. Routledge.

Cope, Bill, and Mary Kalantzis. 2020. *Making Sense: Reference, Agency and Structure in a Grammar of Multimodal Meaning*. Cambridge University Press.

Cope, Bill, and Mary Kalantzis. 2023a. "Creating a Different Kind of Learning Management System: The CGScholar Experiment." In *Promoting Next-Generation Learning Environments Through CGScholar*, edited by Matthew Montebello. IGI Global.

Cope, Bill, and Mary Kalantzis. 2023b. "Generative AI Comes to School (GPT and All That Fuss): What Now?" *Educational Philosophy and Theory* 59 (6): 840–844. https://doi.org/10.1080/00131857.2023.2213437.

Cope, Bill, and Mary Kalantzis. 2023c. "A Little History of e-Learning: Finding New Ways to Learn in the PLATO Computer Education System, 1959–1976." *History of Education* 52 (6): 905–936. https://doi.org/10.1080/0046760X.2022.2141353.

Cope, Bill, and Mary Kalantzis. 2023d. "Towards Education Justice: Multiliteracies Revisited." In *Multiliteracies in International Educational Contexts: Towards Education Justice*, edited by Gabriela C. Zapata, Mary Kalantzis, and Bill Cope. Routledge.

Cope, Bill, and Mary Kalantzis. 2023e. "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.

Cope, Bill, and Mary Kalantzis. 2024. "Platformed Learning: Reshaping Education in the Era of Learning Management Systems." In *Varieties of Platformisation: Critical Perspectives on EdTech in Higher Education*, edited by Duncan A. Thomas and Vito Laterza. Palgrave Macmillan.

Cope, Bill, Mary Kalantzis, Sarah J. McCarthey, Colleen Vojak, and Sonia Kline. 2011. "Technology-Mediated Writing Assessments: Principles and Processes." *Computers and Composition* 28 (2): 79–96. https://doi.org/10.1016/j.compcom.2011.04.007.

Cotton, Debby R. E., Peter A. Cotton, and J. Reuben Shipway. 2023. "Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT." *Innovations in Education and Teaching International* 61 (2): 228–239. https://doi.org/10.1080/14703297.2023.2190148.

Dear, Brian. 2017. *The Friendly Orange Glow: The Untold Story of the PLATO System and the Dawn of Cyberculture*. Pantheon Books.

Demszky, Dorottya, Jing Liu, Heather C. Hill, Dan Jurafsky, and Chris Piech. 2023. "Can Automated Feedback Improve Teachers' Uptake of Student Ideas? Evidence from a Randomized Controlled Trial in a Large-Scale Online Course." *Educational Evaluation and Policy Analysis* 46 (3): 483–505. https://doi.org/10.3102/01623737231169270.

Dolata, Mateusz, Dzmitry Katsiuba, Natalie Wellnhammer, and Gerhard Schwabe. 2023. "Learning with Digital Agents: An Analysis Based on the Activity Theory." *Journal of Management Information Systems* 40 (1): 56–95. https://doi.org/10.1080/07421222.2023.2172775.

Engeström, Yrjö. 1994. *Training for Change: New Approach to Instruction and Learning in Working Life*. International Labor Office.

Engeström, Yrjö. 1995. "Innovative Organizational Learning in Medical and Legal Settings." In *Sociocultural Psychology: Theory and Practice of Doing and Knowing*, edited by Laura M. W. Martin, Katherine Nelson, and Ethel Tobach. Cambridge University Press.

Engeström, Yrjö. 1999. "Innovative Learning in Work Teams: Analyzing Cycles of Knowledge Creation in Practice." In *Perspectives on Activity Theory*, edited by Yrjö Engeström, Reijo Miettinen, and Raija-Leena Punamäki. Cambridge University Press.

Engeström, Yrjö. 2008. *From Teams to Knots: Activity-Theoretical Studies of Collaboration and Learning at Work*. Cambridge University Press.

Engeström, Yrjö. 2015. *Learning by Expanding: An Activity Theoretical Approach to Developmental Research*. 2nd ed. Cambridge University Press.

Garrison, Catherine, and Michael Erskine Ehringhaus. 2011. "Formative and Summative Assessments in the Classroom." https://www.amle.org/wp-content/uploads/2020/05/Formative_Assessment_Article_Aug2013.pdf.

Graesser, Arthur C., Kurt VanLehn, Carolyn P. Rosé, Pamela W. Jordan, and Derek Harter. 2001. "Intelligent Tutoring Systems with Conversational Dialogue." *AI Magazine* 22 (4): 39. https://doi.org/10.1609/aimag.v22i4.1591.

Johnson, W. Lewis. 2023. "How to Harness Generative AI to Accelerate Human Learning." *International Journal of Artificial Intelligence in Education*, 1-5. https://doi.org/10.1007/s40593-023-00367-w.

Jürgensmeier, Lukas, and Bernd Skiera. 2024. "Generative AI for Scalable Feedback to Multimodal Exercises." *International Journal of Research in Marketing* 41 (3): 468–488. https://doi.org/10.1016/j.ijresmar.2024.05.005.

Kalantzis, Mary, and Bill Cope. 2015. "Learning and New Media." In *The SAGE Handbook of Learning*, edited by D. Scott and E. Hargreaves. SAGE Publications.

Kalantzis, Mary, and Bill Cope. 2020. *Adding Sense: Context and Interest in a Grammar of Multimodal Meaning*. Cambridge University Press.

Kress, Gunther, and Theo van Leeuwen. 2021. *Reading Images: The Grammar of Visual Design*. 3rd ed. Routledge.

Meyer, Jennifer, Thorben Jansen, Ronja Schiller, Lucas W. Liebenow, Marlene Steinbach, Andrea Horbach, and Johanna Fleckenstein. 2024. "Using LLMs to Bring Evidence-Based Feedback into the Classroom: AI-Generated Feedback Increases Secondary Students' Text Revision, Motivation, and Positive Emotions." *Computers and Education: Artificial Intelligence* 6: 1–10. https://doi.org/10.1016/j.caeai.2023.100199.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. 2006. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955." *AI Magazine* 27 (4): 12. https://doi.org/10.1609/aimag.v27i4.1904.

Mizumoto, Atsushi, and Masaki Eguchi. 2023. "Exploring the Potential of Using an Ai Language Model for Automated Essay Scoring." *SSRN*. https://doi.org/10.2139/ssrn.4373111.

Mollick, Ethan R., and Lilach Mollick. 2023. "Assigning AI: Seven Approaches for Students, with Prompts." The Wharton School Research Paper. https://doi.org/10.2139/ssrn.4475995.

New London Group. 1996. "A Pedagogy of Multiliteracies: Designing Social Futures." *Harvard Educational Review* 66: 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u.

Nicol, David J., and Debra Macfarlane-Dick. 2006. "Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice." *Studies in Higher Education* 31 (2): 199–218. https://doi.org/10.1080/03075070600572090.

Nilsson, Nils J. 2009. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press.

Putland, Emma, Chris Chikodzore-Paterson, and Gavin Brookes. 2023. "Artificial Intelligence and Visual Discourse: A Multimodal Critical Discourse Analysis of AI-Generated Images of 'Dementia.'" *Social Semiotics*, 1–26. https://doi.org/10.1080/10350330.2023.2290555.

Rahwan, Iyad, Manuel Cebrian, Nick Obradovich, et al. 2019. "Machine Behaviour." *Nature* 568: 477–486. https://doi.org/10.1038/s41586-019-1138-y.

Saini, Akash K., Bill Cope, Mary Kalantzis, and Gabriela C. Zapata. 2024. "The Future of Feedback: Integrating Peer and Generative AI Reviews to Support Student Work." *EdArXiv*, May 15. https://doi.org/10.35542/osf.io/x3dct.

Seo, Kyoungwon, Joice Tang, Ido Roll, Sidney Fels, and Dongwook Yoon. 2021. "The Impact of Artificial Intelligence on Learner–Instructor Interaction in Online Learning." *International Journal of Educational Technology in Higher Education* 18: 1–23. https://doi.org/10.1186/s41239-021-00292-9.

Shepard, Lorrie A. 2009. "Commentary: Evaluating the Validity of Formative and Interim Assessment." *Educational Measurement: Issues and Practice* 28 (3): 32–37. https://doi.org/10.1111/j.1745-3992.2009.00152.x.

Shermis, Mark D. 2014. "State-of-the-Art Automated Essay Scoring: Competition, Results, and Future Directions from a United States Demonstration." *Assessing Writing* 20: 53–76. https://doi.org/10.1016/j.asw.2013.04.001.

Steiss, Jacob, Tamara Tate, Steve Graham, et al. 2024. "Comparing the Quality of Human and ChatGPT Feedback of Students' Writing." *Learning and Instruction* 91: 101894. https://doi.org/10.1016/j.learninstruc.2024.101894.

Tzirides, Anastasia Olga [Olnancy], Akash Saini, Gabriela C. Zapata, et al. 2023. "Generative AI: Implications and Applications for Education." *arXiv:2305.07605 [cs.CY]*. https://doi.org/10.48550/arXiv.2305.07605.

Tzirides, Anastasia Olga [Olnancy], Gabriela Zapata, Nikoleta Polyxeni Kastania, et al. 2024. "Combining Human and Artificial Intelligence for Enhanced AI Literacy in Higher Education." *Computers and Education Open* 6: 100184. https://doi.org/10.1016/j.caeo.2024.100184.

van Leeuwen, Theo. 2011. "Multimodality and Multimodal Research." In *The SAGE Handbook of Visual Research Methods*, edited by Eric Margolis and Luc Pauwels. SAGE Publications.

Vartiainen, Henriikka, Päivikki Liukkonen, and Matti Tedre. 2023. "The Mosaic of Human-AI Co-Creation: Emerging Human-Technology Relationships in a Co-Design Process with Generative AI." *EdArXiv*, November 2. https://doi.org/10.35542/osf.io/g5fb8.

Vartiainen, Henriikka, and Matti Tedre. 2024. "How Text-to-Image Generative AI Is Transforming Mediated Action." *IEEE Computer Graphics and Applications* 44 (2): 12–22. https://doi.org/10.1109/MCG.2024.3355808.

Vygotsky, Lev S. 1978. *Mind in Society*. Harvard University Press.

Vygotsky, Lev S. 1986. *Thought and Language*. MIT Press.

Wan, Tong, and Zhongzhou Chen. 2024. "Exploring Generative AI Assisted Feedback Writing for Students' Written Responses to a Physics Conceptual Question with Prompt Engineering and Few-Shot Learning." *Physical Review Physics Education Research* 20 (1): 1–13. https://doi.org/10.1103/PhysRevPhysEducRes.20.010152.

Warschauer, Mark, and Douglas Grimes. 2008. "Automated Writing Assessment in the Classroom." *Pedagogies: An International Journal* 3 (1): 22–36. https://doi.org/10.1080/15544800701771580.

Winstone, Naomi E., Robert A. Nash, Michael Parker, and James Rowntree. 2016. "Supporting Learners' Agentic Engagement with Feedback: A Systematic Review and a Taxonomy of Recipience Processes." *Educational Psychologist* 52 (1): 17–37. https://doi.org/10.1080/00461520.2016.1207538.

Wiser, Fabian, Carolin Durst, and Nilmini Wickramasinghe. 2019. "Using Activity Theory Successfully in Healthcare: A Systematic Review of the Theory's Key Challenges to Date." Presented at the Proceedings of the 52nd Hawaii International Conference on System Sciences, Grand Wailea, Hawaii, January 8–11, 2019. https://hdl.handle.net/10125/59528.

Zapata, Gabriela C., Akash K. Saini, Bill Cope, Mary Kalantzis, and Duane Searsmith. 2024. "Peer-Reviews and AI Feedback Compared: University Students' Preferences." *EdArXiv*, March 29. https://doi.org/10.35542/osf.io/uy8qp.

Zhai, Xiaoming. 2023. "ChatGPT for Next Generation Science Learning." https://doi.org/10.2139/ssrn.4331313.

Zheng, Lanqin, Nian-Shing Chen, Panpan Cui, and Xuan Zhang. 2019. "A Systematic Review of Technology-Supported Peer Assessment Research: An Activity Theory Approach." *International Review of Research in Open and Distributed Learning* 20 (5): 168–191. https://doi.org/10.19173/irrodl.v20i5.4333.

ABOUT THE AUTHORS

Gabriela C. Zapata: Associate Professor, School of Education, University of Nottingham, Nottingham, UK
Corresponding Author's Email: Gabriela.Zapata@nottingham.ac.uk

Akash K. Saini: Graduate Student, Learning Design and Leadership Program, Department of Education Policy, Organization & Leadership, University of Illinois Urbana-Champaign, Illinois, USA
Email: aksaini2@illinois.edu

Anastasia Olga (Olnancy) Tzirides: User Experience Researcher, Learning Design and Leadership Program, Department of Education Policy, Organization & Leadership, University of Illinois, Urbana-Champaign, Illinois, USA
Email: aot2@illinois.edu

Bill Cope: Professor, Department of Education Policy, Organization & Leadership, University of Illinois, Urbana-Champaign, Illinois, USA
Email: billcope@illinois.edu

Mary Kalantzis: Professor, Department of Education Policy, Organization & Leadership, University of Illinois, Urbana-Champaign, Illinois, USA
Email: marykalantzis@illinois.edu