A Systemic Functional Linguistics Discourse Analysis of Learner-Centered, Generative AI Feedback in Higher Education

Gabriela C. Zapata[1], Anastasia Olga (Olnancy) Tzirides[2], Bill Cope[2], Yu-ling You[3], Mary Kalantzis[2], and Duane Searsmith[2]

[1]*School of Education, University of Nottingham, Nottingham, United Kingdom*

[2]*Department of Education Policy, Organization & Leadership, University of Illinois, Urbana-Champaign and Chicago, United States*

[3]*National Changhua University of Education, Changhua City, Taiwan*

Corresponding author: Gabriela C. Zapata, Gabriela.Zapata@nottingham.ac.uk

A Systemic Functional Linguistics Discourse Analysis of Learner-Centered, Generative AI Feedback in Higher Education

The rapid development of Generative AI (GenAI) has opened new possibilities for its use in higher education, particularly in assessment and formative feedback. This study investigates the pedagogical effectiveness of GenAI-generated feedback using Systemic Functional Linguistics and Appraisal Theory to analyze the language used by GenAI reviewers. We compared two sets of GenAI-generated reviews on student writing from a graduate program in an American university. The first set came from a platform connected to OpenAI's GPT-3, while the second used GPT-4, customized with a 35-million-word disciplinary corpus. The second version aimed to align more closely with the program's academic context and provide more relevant, theoretically grounded feedback to the students enrolled in it. Through discourse analysis, we identified linguistic features that made the calibrated AI reviewer more pedagogically effective. Our findings highlight how tailoring GenAI systems to disciplinary language and feedback frameworks can improve the quality of support offered to university students. Based on our results, we also discuss pedagogical implications and offer recommendations for further research.

Keywords: Generative AI; higher education instruction; formative feedback; Systemic Functional Linguistics; Appraisal Theory; discourse analysis

Introduction

Artificial intelligence has rapidly transitioned from a futuristic possibility to a present-day reality across multiple domains, including education (Cope & Kalantzis, 2023c; Kalantzis & Cope, 2024; Luckin, 2025). Among the most impactful developments is Generative AI (GenAI), particularly large language models (LLMs) like OpenAI's *ChatGPT* and Google's *Gemini*, which have shown remarkable capabilities in generating human-like text. This unique aspect of GenAI technologies has resulted in valuable opportunities in higher education, particularly in the area of assessment. For instance, existing studies (e.g., Clarizia et al., 2018; Wongvorachan & Bulut, 2022; Zapata et al., 2024b) have shown that these technologies can provide more comprehensive, timely, and tailored evaluations, with the potential for scalability based on predefined criteria. Nevertheless, our previous work (e.g., Saini et al., 2024; Tzirides et al., 2023, 2024a; Zapata et al., 2024a, 2025) and other studies, such as those conducted by Dai et al. (2023) and Steiss et al. (2024), have also posited that individualized, high-quality GenAI feedback might still be difficult to achieve in comparison with that offered by instructors and peers. It is therefore crucial to continue exploring GenAI feedback tools to investigate how their effectiveness can be enhanced.

In this paper, we seek to contribute to the understanding of GenAI tools for formative feedback by focusing on the language used by GenAI reviewers through the lens of Systemic Functional Linguistics (Eggins, 1994; Halliday & Matthiessen, 2014) and Appraisal Theory (Martin & Rose, 2007; Martin & White, 2005). Specifically, we investigated two implementations of GenAI formative feedback on student work in a graduate program in a US university: A non-calibrated version based on a generic LLM and a fine-tuned version enhanced with disciplinary knowledge. Our goal was to evaluate how each version constructed meaning, enacted evaluative stance, and supported student learning. By situating our findings within contemporary feedback and educational technology scholarship, we aim to contribute to a more in-depth understanding of how GenAI can support, rather than replace, human-centered learning.

In the first section of the paper, we discuss the key characteristics of human feedback, and we present our research on GenAI formative assessment in higher education as well as the questions that resulted from this exploration. In the next part of the paper, we introduce the theoretical framework that grounded our analysis in this study and the ways in which the GenAI feedback was analyzed. This is followed by the results of the investigation and their discussion, where we address the differences between the non-calibrated and fine-tuned GenAI reviews. In the final part of the paper, we consider pedagogical implications and offer suggestions for future research.

Characteristics of Human, Formative Feedback

Formative feedback is a central pedagogical practice in higher education, particularly in the development of student writing, as it can serve not only as a tool for guiding revisions but also as a vehicle for fostering motivation, self-efficacy, and metacognitive reflection (Holmeier et al., 2018; Nicol & Macfarlane-Dick, 2006; Wiliam, 2011). Existing research on instructor and peer formative feedback (e.g., Hyland & Hyland, 2001; Morris et al., 2021; Nicol et al., 2014; Pearson, 2022; Van Zundert et al., 2010) has highlighted its multidimensional nature, which includes informational, affective, and interpersonal components. These components operate in tandem to both assess and nurture learning.

For instance, a foundational study carried out by Hyland and Hyland (2001) showed that praise, criticism, and actionable advice were the primary discursive functions used in written instructor feedback. Nevertheless, these functions were rarely delivered in isolation. Rather, for example, praise was often used as a softening strategy to mitigate the potentially face-threatening nature of criticism. Additionally, the authors reported the frequent use of hedging devices, such as modal verbs (e.g., “might” and “could”), interrogative syntax (“Is there a better way to express this?”), and personal attribution (“I think”) to frame negative evaluations in a more palatable form. These strategies appeared to serve both cognitive and emotional purposes, clarifying the nature of the feedback while maintaining rapport with students.

In his comprehensive review of existing research on instructor formative feedback in the last 30 years, Pearson (2022) further reinforced Hyland and Hyland’s (2001) findings by identifying paired-act patterns (i.e., praise and critique offered in the same sentence) as a hallmark of effective formative feedback. In the studies reviewed, these patterns, as well as the discursive functions reported in Hyland and Hyland’s work, were shown to preserve student motivation and minimize defensive responses.

Additionally, to further describe instructor feedback, Pearson (2022) compiled a typology of characteristics commonly found in written comments, including their tone (e.g., advisory, descriptive), syntactic structure, text specificity, and degree of explicitness. Pearson also posited that feedback that includes specific suggestions (e.g., "Consider elaborating on this idea by...") is more actionable and more likely to be revised by students than vague or purely evaluative statements (see also Ferris, 1997 and Nurmukhamedov & Kim, 2010). That is, effective advisory feedback often combines content critique with revision strategies, making clear not just what is problematic but also how to improve it.

The feedback characteristics and strategies identified by Hyland and Hyland (2001) as well as in the studies reviewed by Pearson (2022) have also been extensively documented in corpus-based studies. For example, Lee (2013) analyzed 126 feedback reports from UK universities and found that modal verbs such as "could," "might," and "would" were the most frequently employed linguistic markers of hedging. These forms, which carry lower degrees of certainty, were predominantly used in criticism and suggestions. This aligns with broader findings in linguistic pragmatics, which argue that mitigation helps reduce the social tension that arises from evaluative discourse (Crismore & Vande Kopple, 1988; Hyland, 1996). In assessment, the use of modal verbs reflects an effort from instructors to negotiate the dual role of both assessor and facilitator, a theme echoed across other relevant feedback literature (e.g., see Anson, 1989).

Equally important is the structural organization of feedback. In her work on instructor comments on graduate students' writing in the UK, Mirador (2000) identified a specific rhetorical pattern recurring across all the feedback texts analyzed. This pattern, which this scholar defined as *the clinching pattern*, comprises six typical moves: (1) A summary with general impressions, (2) the recapitulation of the student's ideas, (3) the description of strengths, (4) the identification of weaknesses, (5) suggestions for improvements, and (6) an overall judgment. These moves, Mirador posited, appear to create a coherent narrative that can support both reflection and action. Moreover, this structure seems to be not only pedagogically functional but also culturally situated, reflecting a genre-specific way of communicating evaluative content (Swales, 1990; Yelland, 2011).

The interpersonal dimension of feedback has also received increased attention, particularly in relation to how it engages students as active participants in learning. For instance, Ädel (2018) highlighted the dialogic function of metadiscursive "you" in teacher feedback, showing how such language can foster a conversational tone that clarifies expectations and helps resolve misunderstandings. This use of direct address and reflexivity positions feedback as an interaction rather than a one-sided critique, cultivating a relational ethos that supports student agency (Hyland, 2000/2013; Nicol & Macfarlane-Dick, 2006).

Importantly, this interpersonal quality is not limited to teacher-student exchanges. A review on existing research on peer feedback by Van Zundert et al. (2010) similarly underscored students' openness to engaging with one another's feedback and identified benefits such as enhanced academic performance, greater motivation, and improved self-regulatory practices (see also Bargh & Schul, 1980; Black & Wiliam, 1998; Falchikov, 2001). Together, these findings point to the value of dialogic and relational approaches to feedback, whether from teachers or peers, in promoting deeper learning and student empowerment.

In sum, existing studies on human formative feedback such as the ones presented in this section have shown that it constitutes a multi-layered communicative practice involving far more than highlighting errors. It is characterized by clear rhetorical conventions (Mirador, 2000), linguistic strategies of mitigation (Hyland & Hyland, 2001), a dialogic nature (Ädel, 2018; Van Zundert et al., 2010), and an evolving understanding of pedagogical best practices (Pearson, 2022). Effective feedback is timely, specific, dialogical, and motivational, aligning with learning goals while maintaining a supportive tone (Holmeier et al., 2018). As such, teachers (and peers) offering effective, actionable guidance appear to be not only conscious of what they say, but also of how and why they say it.

GenAI Formative Feedback

While formative feedback has long been recognized as a cornerstone of effective instruction and student learning in higher education, the findings presented in the previous section suggest that its quality depends on more than the delivery of information: It is a tailored, relational, and rhetorical practice (Hyland & Hyland, 2001; Morris et al., 2021; Pearson, 2022). The multifaceted nature of formative feedback might present both opportunities and challenges when it is mediated by GenAI technologies. This has been the focus of our work since January 2023. Through a series of studies, we have explored how GenAI feedback may serve to enhance conventional, human-based approaches.

Our research has involved the participation of graduate-level students in online courses at a Midwestern US university. These participants were part of Master's, doctoral, and certificate programs in Education, enrolled in classes that examined the interplay between learning, technology, and pedagogy. These courses critically analyzed sociocultural and historical dimensions and explored various theoretical and practical frameworks. Instructional content and assignments were delivered and completed through a digital platform that has been in development since 2000 (Cope &

Kalantzis, 2023a). Over time, this platform has evolved to include various experimental tools for writing and assessment. In these courses, students are assessed through critical, mostly written projects (though multimodal elements are encouraged) focused on educational theory, technology, and practice. Projects are student-selected and developed progressively over each academic semester with iterative feedback from peers and instructors before final submission.

In January 2023, a novel component was introduced into the feedback process: GenAI-generated reviews. These AI-generated comments were designed to complement existing formative feedback from both peers and instructors. Over the course of the year, this system utilized OpenAI's GPT-3, integrated into the platform via an application programming interface (API). The feedback from instructors, peers, and the AI was unified under a shared set of ten assessment criteria, grounded in the multiliteracies pedagogy *Learning by Design* (LbyD). [1] This educational approach emphasizes learning as an epistemic process, incorporating cognition, hands-on engagement, and emotional-social interaction (Cope & Kalantzis, 2023b).

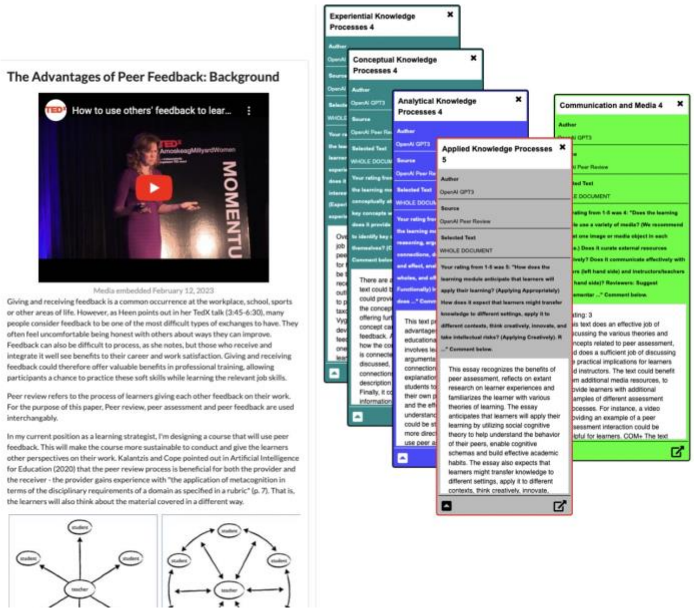
In the revised formative feedback system, the GenAI tool assessed each student submission by looping through the entire text once for each of the ten rubric criteria, effectively treating each criterion as a prompt. For each pass, the AI generated qualitative feedback specific to that criterion and appended it with an overall rating (see Tzirides et al., 2023). This process resulted in a set of structured comments, aiming to mirror instructor and peer feedback. By using the assessment rubric as a guiding scaffold, the system ensured consistency in evaluation while offering targeted

[1] A schematic view of these criteria can be found at https://tinyurl.com/FeedCriteria.

observations on multiple dimensions of students' work. A visual overview of the GenAI feedback is presented in Figure 1.

Figure 1

Sample GenAI feedback



To assess this AI component's efficacy in the feedback process, we conducted a series of convergent, mixed-methods studies that gathered students' reflections on both AI and peer feedback. Our dataset included written and multimodal reflections as well as Likert-scale ratings. Our thematic, socio-semiotic, and statistical analyses (Saini et al., 2024; Tzirides et al., 2023, 2024a; Zapata et al., 2024a, 2025) consistently showed that students favored peer feedback, noting its empathetic tone, detailed interpretation, and relevance to the course context. Peer assessments were seen as more insightful and tailored, addressing not only the content but also the stylistic and multimodal aspects of student work. This contextual awareness, informed by shared learning experiences, made peer feedback feel more meaningful and supportive than that offered by the AI reviewer. Additionally, our participants believed that the human element of warmth and empathy in peer reviews provided emotional encouragement and helped sustain their motivation.

These findings seemed to highlight specific weaknesses in the GenAI review process that warranted further improvement. In response, beginning in January 2024, we incorporated a RAG (Retrieval Augmented Generation) approach into the AI system. This update introduced a vector database still relying on an LLM (OpenAI's GPT4) via an API, but now containing 35 million tokens from previous students' final, reviewed work as well as scholarly, published peer-reviewed writings by the program instructors. Our goal was to provide AI-generated feedback with greater alignment to disciplinary knowledge and pedagogical standards.

These technological updates were evaluated through another round of mixed-methods research, using textual, multimodal, and numerical data sources (Tzirides et al., 2024b; Zapata et al., 2024b). The aim was to determine whether students' experiences with the updated AI system influenced their perceptions of the GenAI reviewer. The findings showed a clear shift: Students rated the recalibrated AI feedback as more relevant, precise, and transformative. Moreover, this improvement appeared to alter students' views of the AI itself, with some now describing it as a collaborative partner or even as part of their academic community. And, in certain cases, students even perceived the AI's feedback to be on par with or better than peer comments (Zapata et al., 2024b).

The results of these studies led us to reconsider the role of language in the perceived effectiveness of GenAI-generated feedback. While recalibration clearly seemed to improve the relevance and quality of the AI responses, what remained unclear was how this improvement was being realized linguistically. In particular, we became interested in the discursive features that made the feedback from the fine-tuned system feel more supportive, specific, and pedagogically meaningful in students' eyes. This realization prompted us to conduct a detailed linguistic analysis of the reviews to identify the language patterns that contributed to students' improved reception of the AI feedback. In the next section, we present the study that emerged from this line of inquiry. First, we introduce the theoretical framework that grounded our analysis. This is followed by a description of the categories we considered when analyzing the generic and calibrated AI reviews.

The Present Study

Theoretical Framework

This study is grounded in Systemic Functional Linguistics (SFL), a model of language developed by M.A.K. Halliday (1975, 2009) that views language as a social semiotic system; i.e., a resource for making meaning shaped by and shaping social contexts. Unlike traditional or generative approaches to grammar, which emphasize syntactic form or universal rules, SFL focuses on how language choices function in context to enact *ideational*, *interpersonal*, and *textual* meanings simultaneously (Halliday & Hasan, 1985; Halliday & Matthiessen, 2014). These three *metafunctions* offer complementary perspectives on every utterance, making SFL especially suitable for analyzing discourse in education, where meaning-making is both intentional and situated (Coffin, 2013; Gibbons, 2006; Moore et al., 2018).

The *ideational metafunction* is concerned with the representation of experience. It captures how language encodes actions, participants, and circumstances through what is termed *the transitivity system* (Eggins, 1994; Halliday & Matthiessen, 2014). This system allows us to explore how ideas are constructed, agency is distributed, and phenomena are made knowable through linguistic choices. The *interpersonal metafunction* focuses on the enactment of social relationships and evaluative stance. It includes grammatical resources that reflect speaker authority, affect, obligation, and negotiation (Eggins, 1994). The *textual metafunction*, meanwhile, attends to the internal organization of discourse, realized through *thematic progression* and *cohesive ties* that contribute to information flow and textual coherence (Eggins, 1994; Halliday & Matthiessen, 2014).

To examine interpersonal meaning in more depth, we drew on the Appraisal framework developed by Martin and White (2005) and further elaborated by Martin and Rose (2007). As a key component of the interpersonal metafunction, Appraisal Theory focuses on how speakers and writers express emotions, make judgments, and evaluate things and events. It consists of the following three subsystems:

- **Attitude**, which includes *affect* (feelings), *judgment* (moral evaluations), and appreciation (aesthetic evaluations);
- **Engagement**, which accounts for how speakers acknowledge or exclude alternative perspectives (monoglossic vs. heteroglossic expressions); and
- **Graduation**, which modulates the intensity or preciseness of evaluative language through scalar adjustments.

Together, SFL and Appraisal Theory offer a rich set of tools for analyzing how GenAI-generated feedback communicates meaning. SFL enables a systemic account of clause-level choices and their relation to broader discourse functions, while Appraisal Theory helps us evaluate the interpersonal tone and persuasive quality of the feedback.

We believe this combined framework is particularly well suited to our research aims because, as previously discussed, feedback is inherently multifunctional: It conveys content, offers evaluations, structures discourse, and fosters pedagogical relationships. In the context of GenAI, these functions must be realized without the benefit of shared context or human intuition. Understanding how these functions are (or are not) successfully performed through linguistic choices is central to assessing the educational value of AI feedback. Thus, SFL and Appraisal Theory not only offer a principled basis for analysis but also align directly with the pedagogical and ethical concerns raised by the deployment of AI in educational contexts (Cope & Kalantzis, 2023c, 2024; Kalantzis & Cope, 2024).

Data Collection and Analysis

To investigate the linguistic qualities of GenAI-generated feedback, both generic and calibrated, we conducted a discourse analysis informed by SFL and Appraisal Theory. The dataset consisted of AI-generated formative comments from two iterations of our feedback system: One using GPT-3 with no calibration (Appendix A), and one using GPT-4, fine-tuned with a 35-million-word disciplinary corpus (Appendix B). We selected a focused sample of 12 reviews (6 from the non-calibrated output and 6 from the fine-tuned version) for our detailed discourse analysis for both practical and methodological reasons. Although our broader study engaged with 91 non-calibrated and 50 calibrated reviews in various ways, the purpose of this particular investigation was to perform a qualitative, in-depth linguistic analysis grounded in SFL and Appraisal Theory. These frameworks require fine-grained attention to clause-level linguistic features, such as transitivity, modality, and theme-rheme structure, which makes large-scale analysis impractical without compromising analytical depth.

We intentionally balanced the sample across both review types and rubric categories to ensure comparability and to capture a representative cross-section of feedback discourse. The selected comments were strategically chosen to reflect a range of performance levels and included both strengths and areas for improvement, allowing us to analyze how language patterns varied across functions (i.e., ideational, interpersonal, and textual). This approach enabled us to probe into the discursive mechanisms that students had previously described as more or less pedagogically meaningful (see Saini et al., 2024; Tzirides et al., 2023, 2024a, 2024b; Zapata et al., 2024a, 2024b, 2025). In other words, the sample size reflected a methodologically appropriate subset for the qualitative goals of this study, prioritizing depth of analysis over breadth of coverage.

Our analysis of the 12 GenAI reviews centered around three primary dimensions, each corresponding to a metafunction in SFL. The ideational analysis was conducted through the transitivity system, focusing on the kinds of processes (e.g., material, mental, verbal) used by the AI, as well as participant roles and circumstantial elements (e.g., location, cause, manner) (Eggins, 1994). That is, this dimension assessed how the feedback represented student work, whether agency was attributed to the student writer, and whether the feedback offered concrete, content-relevant suggestions. Emphasis was placed on examining whether process types reflected a dynamic, agentive view of students' role in knowledge production (Halliday & Matthiessen, 2014).

The interpersonal analysis explored how the AI established rapport, communicated stance, and positioned students in relation to the feedback. Drawing on Appraisal Theory (Martin & Rose, 2007; Martin & White, 2005), we focused on attitude (e.g., expressions of praise or critique), engagement (e.g., whether the feedback acknowledged alternative perspectives), and graduation (e.g., how intensity or certainty was scaled). This allowed us to determine whether feedback communicated encouragement, caution, authority, or empathy. These features were identified in previous research as essential to effective human feedback (e.g., Ädel, 2018; Holmeier et al., 2018; Van Zundert et al., 2010).

Finally, the textual analysis investigated the organization of information, particularly theme–rheme structure and cohesion. We examined how feedback comments introduced topics, structured propositions, and linked ideas across sentences using lexical ties, reference chains, and conjunctions (Eggins, 1994). This allowed us to evaluate the coherence of AI reviews and the extent to which they mirrored the progressive rhetorical development found in human feedback, such as Mirador's (2000) clinching pattern.

By triangulating these three dimensions, we were able to generate a holistic picture of how GenAI performs as a feedback provider and how improvements in calibration might have impacted the pedagogical quality of the AI reviewer's language. This multilayered analysis not only illuminated the linguistic mechanisms behind students' changing perceptions of AI feedback, but also provided concrete evidence of how alignment with disciplinary discourse conventions might have enhanced both clarity and engagement.

Results

Our linguistic analysis revealed significant differences in ideational, interpersonal, and textual realizations between non-calibrated and calibrated GenAI-generated feedback. The findings suggest that calibration not only improved the specificity and coherence of AI feedback but also reshaped how student agency, rhetorical structure, and evaluative stance were linguistically encoded. In what follows, in separate sections, we present detailed results across the three metafunctional dimensions for both generic and calibrated GenAI reviews.

Generic AI Reviews

Ideational Meaning

The non-calibrated reviews were characterized by a narrow deployment of process types, which rendered them generic, formulaic, and categorical. For example, material processes (i.e., those representing actions or events in the physical world [what people or things do]) appeared occasionally and were typically vague (e.g., "More sources should be included to support the different claims made in the text."), suggesting revision without acknowledging what the student had done or projecting possibilities instead of analyzing what was actually written.

Also, mental (expressing internal experiences like thinking, feeling, or perceiving) and verbal (involving saying, stating, or claiming) processes were almost entirely absent. For instance, the reviews rarely referenced what the student thought or argued, which undermined recognition of their reflective or argumentative labor (e.g., no comments like “The writer appears to suggest...” or “The author questions...”). The absence of these two process types, as a result, impaired interpretive depth and led to flattened discourse (i.e., student writing was reviewed as a static object, not a voice in a feedback dialogue).

Moreover, human agency was underrealized or generalized, with the AI reviewer frequently attributing actions and evaluations to non-human entities such as “the text” or “the essay.” The student was often absent or distanced as a participant (e.g., “The paper discusses…” vs. “The writer adeptly weaves together …” [calibrated review]). Although the AI occasionally referred to “the writer,” such instances were largely confined to critiques or suggestions for improvement (e.g., “The writer does not define the concepts of metaverse or pre-service teacher education”). The tendency to obscure the writer’s presence limited the depth of feedback as well as the opportunity to engage the student as an active agent in the revision process.

Circumstantial elements were scarce and vague, leaving feedback unanchored from disciplinary context (i.e., there were few or no references to educational settings, theoretical frameworks, or audience considerations). For example, statements such as “more research data should be included” were not supplemented with reasons why, where, or in what context such data would be effective. The lack of situational detail thus limited the richness of the experiential context and weakened the reviews' explanatory or evaluative value.

Interpersonal Meaning

In terms of appraisal, overall, the non-calibrated feedback felt impersonal, as the reviews exhibited consistent lack of emotional and interpersonal depth, across all categories of affect, judgment, and appreciation. For instance, the tone remained emotionally flat, with virtually no expressions of warmth, empathy, or encouragement. Comments such as "This was compelling" or "You've made a thoughtful point" were notably absent, replaced instead by dry, procedural remarks like "The essay text provided shows a good understanding of the potential applications of the innovative technologies…" This emotionally generic tone might have contributed to a sense of detachment; i.e., without affective language, students might have felt that the GenAI had not engaged meaningfully with or understood their work.

Judgments about student performance were primarily communicated through abstract, generic language, often referencing rubric-based standards rather than specific, individualized evaluations. Phrases like "more explanation… should be included" or "the essay does not provide any clear reasoning" dominated the commentaries, offering little insight into students' particular efforts or achievements. The absence of personalized or detailed feedback might have limited the AI reviews' usefulness for learning and development. Additionally, this feature might also have reinforced a sense of being assessed by an automated system, a single-voiced monoglossic authority rather than a dialogic instructor or peer who recognizes the complexity and intention behind the work.

Appreciation of student writing was similarly vague and unspecific, lacking aesthetic or intellectual depth. Praise, when offered, tended to be formulaic (e.g., through the overuse of the adjective “good,” as in “does a good job”) seldomly indicating what exactly was done well or why it mattered. This lack of detailed recognition failed to convey intellectual interest in students’ ideas or writing style. As a result, the feedback might have come across as transactional, giving students the impression that their work was merely processed rather than thoughtfully considered. This, in turn, might have led to students’ feeling unseen or undervalued, diminishing the motivational impact of positive feedback.

Additionally, the AI’s feedback style relied heavily on definitive, overly absolute statements, lacking hedging and offering little room for alternative interpretations or partial success. That is, evaluations such as “The writer fails to provide a clear argument” were delivered with finality, excluding language that might soften or open up the commentary. GenAI comments therefore felt rigid and categorical, with rare use of expressions that suggested degrees of achievement or progress (i.e., limited use of hedging). This all-or-nothing approach can prevent students from seeing how their work might evolve or where it partially meets expectations, limiting its value as a tool for learning and revision.

Textual Meaning

Overall, the analysis revealed that, textually, non-calibrated reviews felt generic, repetitive, and incomplete, characterized by unmarked themes as well as limited theme development and cohesion. For instance, repetitive theme-rheme structures manifested through lack of variety in clause openings, with the perpetual use of “The text/essay...” or, occasionally, “The writer...” This redundancy resulted in mechanical cadence and prevented the accumulation of rhetorical complexity. Thematic progression was also minimal, and sentences lacked cohesion beyond basic additive and adversative conjunctions like “additionally” and “however.” As a result, reviews read as disjointed checklists, instead of cumulative or integrative commentaries on students’ work.

Likewise, cohesive ties were weak. Lexical cohesion was impaired by the overuse of generic terms like “the text,” “the concepts,” or “the technology” with little synonymic variation or expansion (achieved, for example, through the employment of semantically related terminology). This aspect of the reviews prevented lexical chaining, which made the feedback sound mechanical. Also, the text often failed to make semantic relationships (e.g., cause-effect, elaboration) between ideas explicit (a key element in effective formative guidance). A final weakness with regards to the textual organization of the non-calibrated AI feedback was the partial, limited reflection of Mirador’s (2000) clinching pattern. As can be seen in the sample in Appendix A, the feedback offered a general impression and recapitulation, identified weaknesses, and provided general recommendations. However, the content was disjointed and generic, very seldom leading to actionable guidance. Overall, therefore, the non-calibrated reviews resembled feedback templates more than dialogic responses.

Calibrated AI Reviews

Ideational Meaning

In contrast to the generic reviews, the calibrated GenAI feedback exhibited a specific and academically appropriate use of diverse process types. For instance, the deployment of material processes through the use of verbs such as “applies,” “integrates,” and “weaves” positioned the student writer as an active participant in academic inquiry. That is, these verbs conveyed actions that were clearly aligned with scholarly activity, such as integrating theories or applying frameworks. A clear example of this feature can be seen in the comment “The writer integrates Dewey’s and Shusterman’s theories with the practice of mural making,” which not only identifies what the student has done but also how theoretical understanding is operationalized within a pedagogical practice. This contrasts sharply with generic reviews, which flattened the writer's role into a passive subject rather than an active knowledge constructor. By leveraging a broader transitivity repertoire, the calibrated feedback highlighted intellectual labor and made student agency visible in meaningful ways.

Mental processes were equally prominent, reflecting internal states such as cognition, reflection, and evaluation. Clauses like "demonstrates understanding," "reflects on pedagogical beliefs," or "considers the implications of" serve a dual function: They affirm the student's capacity for critical thinking while also offering suggestions for deeper engagement. This attention to the internal world of the student writer signaled an appreciation of their intellectual trajectory, not just the outcome of their writing. In the example, "The author could reflect more on personal beliefs that shaped their interpretation of Dewey," the feedback not only suggested a revision but also validated the value of self-reflection as part of scholarly inquiry. Such mental process clauses add subtle evaluative depth, acknowledging student intentionality while inviting greater introspection and elaboration.

Additionally, the use of verbal processes such as "argues," "articulates," and "claims" was notable. These types of verbs positioned the student not just as a thinker but as a speaker embedded within a scholarly dialogue. This dialogic framing, seen, for example, in the comment "The writer articulates a clear position on digital equity by referencing current empirical research," present the student as a participant in an ongoing academic discourse. Also, alongside well-placed circumstantial elements such as "in the context of inclusive classroom design" or "within the framework of community engagement," these reviews grounded arguments in relevant sociocultural settings, anchoring feedback within authentic educational contexts. As a result, the calibrated feedback did not merely evaluate student work abstractly but tied it to meaningful academic and real-world applications, enhancing relevance and pedagogical value. Overall, therefore, the ideational choices in the calibrated AI feedback constructed a richly textured account of student agency, knowledge creation, and educational practice.

Interpersonal Meaning

The calibrated AI feedback demonstrated a notable richness in the use of attitude resources, particularly in the dimensions of judgment, appreciation, and subtle expressions of affect. Unlike the generic reviews that often relied on vague descriptors like “good” or “could further be improved,” calibrated responses delivered judgments with contextual precision, addressing specific aspects of academic and intellectual performance. For instance, rather than stating that a section “lack[ed] depth,” a calibrated review observed, “There is a missed opportunity to delve deeper into specific personal anecdotes that could enhance the argument’s emotional resonance.” Such feedback not only evaluated but explained the basis of the critique, supporting the development of metacognitive awareness.

Appreciation was similarly enriched and moved beyond surface-level praise to engage with conceptual and aesthetic dimensions of student work. A comment like “The integration of fieldwork examples with critical pedagogy is particularly compelling” not only praised the student’s technique but also positioned their work within scholarly discourse. While affect remained generally understated, which we might argue is appropriate for academic contexts, there were discernible moments of encouragement and recognition. A particularly illustrative example is the comment, “The reference to the author’s professional standpoint in… adds credibility and a personal touch, suggesting a deep-seated interest in the subject matter.” This kind of feedback not only signals the reader's engagement but also affirms the significance of the student's perspective. These types of observations might have offered subtle emotional support, reinforcing the student's role as a valued academic contributor and aligning with Pearson's (2022) emphasis on feedback that sustains motivation.

In terms of engagement, the calibrated AI feedback was distinctly more dialogic and inclusive than its generic counterpart. The use of heteroglossic constructions such as "some scholars argue" or "consider[ing] alternative theories or practices in art education that challenge or complement…" opened up space for student agency and interpretation, rather than shutting it down with definitive claims. For example, comments such as the following acknowledged the legitimacy of the student's stance while simultaneously inviting further critical engagement:

> "While the masking tape example is strong, additional case studies or varied applications (e.g., different age groups, different materials) would strengthen the argument. For instance, how does intra-active pedagogy function in non-art classrooms or in different cultural contexts? Expanding the scope of examples would make the application argument more robust."

These kinds of suggestions might have created a feedback space that could be perceived as both supportive and dialogically open, modeling the kind of academic reasoning expected at the graduate level.

Moreover, calibrated reviews frequently referenced external sources or disciplinary figures/frameworks (e.g., Dewey, TPACK, specific empirical studies), thereby not only grounding the evaluation in scholarly discourse but also positioning the AI as a peer interlocutor rather than an unquestionable authority. This constituted a meaningful shift from monologic judgment to collaborative meaning-making. The feedback therefore became not only evaluative but instructional, scaffolding students into academic discourse practices and reinforcing their role in ongoing theoretical conversations.

Finally, the calibrated reviews made strategic and pedagogically effective use of graduation resources (i.e., those linguistic features that scale meaning through intensity and focus). Rather than flattening performance into binary categories like "strong" or "weak," these reviews modulated evaluations with scalar descriptors such as "somewhat effective," "reasonably clear," "well-articulated," and "particularly strong." This graded feedback offered students a clearer sense of where their work fell on a continuum and suggested the degree of improvement required. For instance, a statement like "The critique of digital equity is somewhat limited and would benefit from further contextualization" signals both the limitation and a concrete path for enhancement. Softening, hedging devices such as "could be strengthened by…" or "might benefit from…" preserved student motivation while maintaining academic rigor. Importantly, these strategies aligned with the face-saving conventions of effective human feedback, as identified by Hyland & Hyland (2001) and echoed in Pearson's typology. This calibration of modality, alternating between assertive and tentative tones depending on context, helped maintain a professional yet encouraging voice. By offering feedback that was both formative and affirming, the AI reviews might have cultivated a sense of growth and development, rather than judgment or finality, mirroring the relational ethos often associated with the most impactful human feedback.

Textual Meaning

The calibrated AI feedback also demonstrated marked improvement in textual organization, particularly in how it handled thematic progression and information flow. Unlike generic reviews that often began each sentence with repetitive structures, calibrated feedback employed a varied thematic structure to emphasize reasoning, evaluation, or pedagogical function. Examples include linguistic choices such as "To enhance the connection to personal experience…" and "By drawing on Dewey's concept of experiential learning…" These marked themes served rhetorical purposes, signposting evaluative moves, shifting focus, and foregrounding key reasoning, and, as a result, created a more dynamic and instructive feedback narrative. Additionally, thematic progression across clauses showed strong cohesion. For instance, the progression from "The writer integrates theoretical perspectives…" to "This integration reveals…" and later followed by, "Such application suggests…" resulted in a pattern that mirrored how ideas develop in scholarly writing.

Cohesion was further reinforced through logical connectors and lexical chains, demonstrating an understanding of how to build semantic continuity. For example, words like "framework," "application," "construct," and "model" appeared in cohesive clusters, building conceptual bridges throughout the feedback. Connectors such as "Moreover," "Although…," and "Overall" were not only present but used strategically, enabling smooth transitions and making evaluative reasoning transparent. Also, referential cohesion was carefully managed, with demonstrative pronouns and determiners like "this" or "these theories…" clearly pointing to antecedents and previously stated ideas. This, in turn, eliminated ambiguity, supporting readability and instructional clarity. These textual features not only mirror academic discourse norms but also make feedback more actionable by helping students trace ideas logically through the feedback narrative.

Furthermore, the *Given → New* information structure was consistently respected, a hallmark of effective academic writing. Statements in the calibrated reviews often began with information familiar to the student (e.g., "The essay presents *Dewey's theory* clearly") and then expanded into new insights (e.g., "*This theory* allows for a nuanced view of embodied learning in urban classrooms”). Such structuring supports both comprehension and revision by aligning with cognitive processing patterns. The feedback also employed interpersonal textual features such as modality (e.g., imperative constructions such as “Include more…,” “Provide empirical evidence…,” and “Explore more…”) and hedging (e.g., paired-act patterns such as “While the essay effectively outlines…, incorporating critical perspectives on… could provide a more balanced view”) with purpose and care. By doing so, the calibrated AI reviewer avoided either authoritarian or overly tentative tones, instead achieving balance with phrases like “could benefit from…” or “may be strengthened by…” when offering critique, and “This clearly demonstrates…” or “The essay effectively…” when offering praise. These choices enabled feedback to be assertive yet supportive, mirroring the tone of a thoughtful instructor or peer.

The level of rhetorical sophistication seen in the calibrated reviews suggests that the AI was tuned not only to linguistic form but also to the pedagogical function of feedback in academic settings. This is also reinforced by the comprehensive presence of all moves in Mirador’s (2000) clinching pattern. As seen in the sample calibrated review in Appendix B, the AI feedback included a general impression and recapitulation of students’ work, it highlighted strengths, but also called attention to weaknesses, offering actionable suggestions, and it provided an overall judgment.

Discussion

The linguistic distinctions between non-calibrated and calibrated GenAI feedback highlight a transformative shift in how AI can be configured to support learning. When analyzed through the lenses of SFL (Eggins, 1994; Halliday & Matthiessen, 2014) and Appraisal Theory (Martin & Rose, 2007; Martin & White, 2005), it becomes evident that calibrated feedback moves significantly closer to the discursive and pedagogical standards expected of effective human commentary, be it offered by instructors or peers. Our recalibration process enabled the AI to model practices long associated with formative feedback excellence, such as rhetorical sensitivity, dialogic engagement, and constructive alignment with academic norms. That is, rather than functioning as a detached evaluator, the calibrated GenAI model assumed the role of a semi-pedagogical agent, providing feedback that was not only more useful but also more attuned to specific student needs.

A central area of improvement emerged in the ideational metafunction, where calibrated reviews employed a wider and more pedagogically relevant range of process types. Material and mental processes were particularly prominent, supporting the representation of students as agentive thinkers and doers rather than passive recipients of critique. In contrast to the repetitive relational clauses seen in the generic feedback, the fine-tuned model integrated process types that enacted scholarly work, signaling intellectual engagement and agency. These linguistic choices align with Halliday's (1994) view of language as a semiotic resource for enacting social and experiential realities. More importantly, these choices support formative practices described by Nicol and Macfarlane-Dick (2006), where effective feedback contributes to students' identities as knowledge-makers. By explicitly positioning students as participants in academic discourse, rather than merely as objects of evaluation, the calibrated AI helped normalize student voice within scholarly genres, a critical step in democratizing academic literacy (Cope & Kalantzis, 2023a).

The interpersonal metafunction also saw notable gains. Calibrated feedback was dialogic rather than monologic, exploratory rather than prescriptive, and attuned to differences rather than binary. Through controlled modality and specific judgment/appreciation resources, the feedback performed both evaluative and relational work. This rhetorical positioning aligns closely with Pearson's (2022) synthesis of effective feedback strategies, which emphasizes the use of hedging, personal attribution, and affectively supportive tone. Our findings also reinforce Zapata et al.'s (2024b) argument that calibrated AI can support a more "relational" feedback ecology, in which students and GenAI are treated as co-constructors of meaning.

Nevertheless, it is also important to recognize that affective resonance remains limited: While encouragement was present, the calibrated reviews lacked the kind of empathetic or emotionally attuned commentary that often characterizes high-impact human feedback (Ädel, 2018; Hyland & Hyland, 2001; Pearson, 2022). For example, there were almost no instances of expressive appreciation for creativity or intellectual risk-taking, what might be described as the "goosebump effect." Thus, while interpersonal tone was respectful and motivating, it did not yet reach the level of relational warmth that human feedback can sometimes offer.

The calibrated feedback also exhibited a heightened awareness of textual organization, especially in its use of thematic progression, cohesion, and information flow. Rather than defaulting to static openings like "The writer...," the AI employed marked themes to signal evaluative stances, guiding the student through coherent argument chains. This kind of thematic structuring mirrors the "given-new" information patterning described by Eggins (1994), a feature associated with high-quality academic writing. Additionally, lexical cohesion and rhetorical connectors were used with clarity and purpose, making the feedback not only more readable but also more instructive as a model of scholarly discourse (Swales, 1990; Yelland, 2011). These textual improvements were not merely stylistic: They enhanced the pedagogical impact of the feedback by aligning it with academic writing conventions students are expected to master themselves. In this sense, the feedback might have served a dual role, as both response and scaffold for disciplinary discourse development.

Despite these notable advancements, limitations persist. Even in its calibrated form, GenAI lacks the full dialogic and affective richness of human feedback. It cannot draw on personal histories with students or shared classroom/lived experiences, resources that human educators and peers routinely leverage in feedback practices (Ferris, 1995; Treglia, 2008). Moreover, the AI struggled with in-depth interpretations of students' ideas, such as identifying implicit ideological tensions or responding emotionally to the aesthetic or ethical dimensions of a student's work. This mirrors Cope and Kalantzis's (2023c, 2024) broader critique of AI in education: While generative systems can approximate surface-level academic genres, they remain limited in their ability to engage in the transformative, affective, and context-sensitive labor of teaching. Thus, even as we recognize the capacity of calibrated GenAI to enhance feedback quantity and consistency, we must also remain attentive to its epistemological and relational constraints.

In sum, our results have shown that calibrated GenAI feedback does represent a significant step forward in the development of AI-enhanced pedagogy. Through careful tuning grounded in educator-authored data, our AI system was able to adopt linguistic and rhetorical patterns that support dialogic engagement, knowledge growth, and, possibly, academic literacy development. Its ability to model effective feedback moves, such as those identified by Mirador (2000), Pearson (2022), and others, suggests it can be meaningfully integrated into formative assessment practices. Nevertheless, this integration should be framed as complementary rather than substitutive. GenAI is most effective when paired with human facilitation, peer dialogue, and feedback literacy instruction. As our findings and the student perspectives reported in our previous work show, the promise of GenAI in education lies not in its autonomy but in its capacity to augment the relational, rhetorical, and reflective practices that define powerful learning environments.

Pedagogical Implications

Our findings strongly point to the possibility that calibrated GenAI feedback can be pedagogically meaningful when purposefully designed and integrated into formative feedback cycles. In its calibrated form, GenAI can mirror core principles of effective human feedback: It can be dialogic, specific, structurally coherent, and intellectually respectful. Additionally, as the results showed, the fine-tuned reviews comprehensively modeled Mirador's (2000) clinching pattern, delivering feedback that not only identified strengths and weaknesses but also offered concrete, pedagogically grounded suggestions for improvement. This suggests a key implication: Instructors can deploy calibrated AI feedback as a scalable way to model high-quality academic discourse and critical thinking practices, particularly in writing-intensive or large-enrollment environments where personalized human feedback is not always feasible. Crucially, however, the AI must be calibrated with contextually rich, educator-chosen data. Clearly, as we have posited in this work, uncalibrated feedback remains too generic and misaligned with pedagogical intent.

Second, the results point toward a shift in how AI can support identity-building and agency-enhancing feedback, particularly in programs emphasizing student-centered and socially situated learning. When properly tuned, GenAI models can reinforce the idea that student authors are knowledge-makers embedded in academic and sociocultural contexts. The use of agentive language, heteroglossic stance-taking, and disciplinary references can help position students as legitimate contributors to scholarly conversations. This aligns with the kind of feedback that nurtures epistemic agency, supports metacognitive development, and affirms student voice, particularly in diverse and globalized learning environments. For instructors, this means AI feedback can serve not just as assessment, but as affirmation of scholarly identity, a critical component of assessment that might support motivation and retention, especially for historically minoritized learners.

Nevertheless, this study also underscores that even the most sophisticated GenAI models should not be seen as replacements for human feedback, but rather as pedagogical partners in a cyber-social learning ecology. While AI can provide consistent, linguistically coherent, and theoretically anchored commentary, it lacks the affective resonance, context sensitivity, and reciprocal engagement that characterize transformative human feedback. This implies that the design of AI feedback systems must be situated within relational pedagogies that prioritize connection, care, and context. In practice, this means embedding AI feedback in a broader feedback ecosystem, supplemented by peer and instructor input, and framing it explicitly as a tool to support learning, not judge performance.

Suggestions for Further Research

While this study points to the promise of calibrated GenAI feedback in enhancing the quality and pedagogical alignment of machine-generated comments, further research is needed to explore how students engage with and revise based on this feedback. Future studies should employ longitudinal, mixed-methods designs that track how learners interpret, trust, and act upon GenAI comments across multiple drafts and assignments. Particular attention should be paid to differences in uptake across demographic and sociocultural variables, language backgrounds, and academic experience levels. These studies could illuminate the complex socio-affective dynamics at play in AI-supported learning environments and contribute to a more in-depth understanding of AI's role in writing development.

Additionally, further research should investigate how calibrated GenAI feedback functions across genres, disciplines, and linguistic registers. While this study focused on academic writing within an education context, the rhetorical expectations and feedback norms vary significantly across fields such as STEM, the humanities, and multilingual writing programs. Comparative research that calibrates GenAI to discipline-specific discourse practices could uncover both affordances and limitations in diverse contexts. Moreover, exploring co-authorship models, in which students and AI iteratively co-construct drafts, may help move beyond feedback into more dialogic and generative pedagogical applications. Such work could contribute to broader conversations about AI literacy, academic integrity, and the evolving nature of authorship in the age of generative tools.

Conclusion

This study illustrates that when GenAI feedback is pedagogically calibrated, it can approximate the rhetorical and relational qualities of high-quality human feedback. Through enriched ideational, interpersonal, and textual meaning-making, calibrated AI feedback systems can not only assess student writing but also model academic discourse and foster identity-building. However, their value lie in their integration into human-centered, cyber-social feedback and learning ecosystems, where AI can amplify rather than replace the transformative power of teacher and peer interaction. The challenge ahead is not whether to use AI in education, but how to design or configure it in ways that honor pedagogical purpose, context, and humanity.

References

Ädel, A. (2018). Variation in metadiscursive "you" across genres: From research articles to teacher feedback. *Educational Sciences: Theory & Practice, 18*(4), 777–796. https://files.eric.ed.gov/fulltext/EJ1202069.pdf

Anson, C. M. (1989). Response styles and ways of knowing. In C. M. Anson (Ed.), *Writing and response: Theory, practice, and research* (pp. 332–366). National Council of Teachers of English. https://www.researchgate.net/publication/313182977_Response_Styles_and_Ways_of_Knowing

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*(5), 593–604. https://doi.org/10.1037/0022-0663.72.5.593

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

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

Coffin, C. (2013). Using systemic functional linguistics to explore digital technologies in educational contexts. *Text & Talk, 33*(4-5), 497-522. https://oro.open.ac.uk/35410/1/SFL%20special%20edition%20proofs2.pdf

Cope, B., & Kalantzis, M. (2023a). Creating a different kind of learning management system: The CGScholar experiment. In M. Montebello (Ed.), *Promoting next-generation learning environments through CGScholar* (pp. 1–18). IGI Global.

Cope, B., & Kalantzis, M. (2023b). 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. (2023c). A multimodal grammar of artificial intelligence: Measuring the gains and losses in generative AI. *Multimodality & Society, 0*(0) 1–30. https://doi.org/10.1177/26349795231221699

Cope, B., & Kalantzis, M. (2024). On cyber-social learning: A critique of artificial intelligence in education. In D. Kourkoulou, A. O. Tzirides, B. Cope, & M. Kalantzis (Eds.), *Trust and inclusion in AI-mediated education: Where human learning meets learning machines* (pp. 3–34). Springer. https://doi.org/10.1007/978-3-031-64487-0_1

Crismore, A., & Vande Kopple, W. J. (1988). Readers' learning from prose: The effects of hedges. *Written Communication, 5*(2), 184-202. https://doi.org/10.1177/0741088388005002004

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

Eggins, S. (1994). *An introduction to systemic functional linguistics*. Pinter Publishers.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. Psychology Press.

Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*(1), 33–53. https://doi.org/10.2307/3587804

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339. https://doi.org/10.2307/3588049

Gibbons, P. (2006). *Bridging discourses in the ESL classroom: Students, teachers and researchers*. Bloomsbury Publishing.

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. Edward Arnold.

Halliday, M. A. K. (2009). Context of culture and of situation. In J. J. Webster (Ed.), *The essential Halliday* (pp. 55–84). Continuum.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Deakin University Press.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar* (4th ed.). Arnold.

Holmeir, M., Grob, R., Nielsen, J. A., Rönnebeck, S., & Ropohl, M. (2018). Written teacher feedback: Aspects of quality, benefits and challenges. In J. Dolin & R. Evans (Eds.), *Transforming assessment: Contributions from science education research* (pp. 175-208). Springer.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics, 17*(4), 433–454. https://doi.org/10.1093/applin/17.4.433

Hyland, K. (2000/2013). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press. https://doi.org/10.3998/mpub.6719

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212. https://doi.org/10.1016/S1060-3743(01)00038-8

Kalantzis, M., & Cope, B. (2024). Literacy in the time of artificial intelligence. *EdArXiv*. https://doi.org/10.35542/osf.io/es5kb

Lee, K. Y. (2013). Hedging expressions used in academic written feedback: A study on the use of modal verbs. *Research in Corpus Linguistics, 1*, 33–45. https://ricl.aelinco.es/index.php/ricl/article/view/11

Luckin, R. (2025). Nurturing human intelligence in the age of AI: Rethinking education for the future. *Development and Learning in Organizations: An International Journal, 39*(1), 1-4. https://doi.org/10.1108/DLO-04-2024-0108

Martin, J. R., & Rose, D. (2007). *Working with discourse. Meaning beyond the clause* (2nd ed.). Continuum.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.

Mirador, J. F. (2000). A move analysis of written feedback in higher education. *RELC Journal, 31*(1), 45-60. https://doi.org/10.1177/003368820003100103

Moore, J., Schleppegrell, M., & Palincsar, A. S. (2018). Discovering disciplinary linguistic knowledge with English learners and their teachers: Applying systemic functional linguistics concepts through design-based research. *TESOL Quarterly, 52*(4),1022–1049. https://doi.org/10.1002/tesq.472

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education, 9*(3), 1-26. https://doi.org/10.1002/rev3.3292

Nicol, D. J., & Macfarlane-Dick, D. (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

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

Nurmukhamedov, U., & Kim, S. H. (2010) “Would you perhaps consider …”: Hedged comments in ESL writing. *ELT Journal, 64*(3), 272–282. https://doi.org/10.1093/elt/ccp063

Pearson, W. S. (2022). A typology of the characteristics of teachers’ written feedback comments on second language writing. *Cogent Education, 9*, Article 2024937. https://doi.org/10.1080/2331186X.2021.2024937

Saini, A. K., Cope, B., Kalantzis, M., & Zapata, G. C. (2024). The future of feedback: Integrating peer and generative AI reviews to support student work. *EdArXiv*. https://doi.org/10.35542/osf.io/x3dct

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

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Treglia, M. O. (2008). Feedback on feedback: Exploring student responses to teachers’ written commentary. *The Journal of Basic Writing, 27*(1), 105–137. https://doi.org/10.37514/JBW-J.2008.27.1.06

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

Tzirides, A. O., Saini, A. K., Zapata, G. C., Searsmith, D., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, D., Jones, J. W., Abrantes da Silva, R., Whiting, J., & Kastania, N. P. (2024a). Generative AI in education: Reflections from application with student work. In D. Kourkoulou, A. O. Tzirides, B. Cope, & M. Kalantzis (Eds.), *Trust and inclusion in AI-mediated education: Where human learning meets learning machines* (pp. 287–301). Springer. https://doi.org/10.1007/978-3-031-64487-0_13

Tzirides, A. O., Zapata, G. C., Kastania, N. P., Saini, A. K., Castro, V., Abdul Rahman Ismael, S., You, Y., Afonso dos Santos, T., Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024b). 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 Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270-279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.

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

Yelland, C. (2011). A genre and move analysis of written feedback in higher education. *Language and Literature, 20*(3), 218-235. https://doi.org/10.1177/0963947011413563

Zapata, G. C., Cope, B., Kalantzis, M., Tzirides, A. O, Saini, A. K., Searsmith, D., Whiting, J., Kastania, N. P., Castro, V., Kourkoulou, T, Jones, J., & Abrantes da Silva, R. (2025). AI and peer reviews in higher education: Students' multimodal views on benefits, differences and limitations. *Technology, Pedagogy and Education*, 1-19. https://doi.org/10.1080/1475939X.2025.2480807

Zapata, G. C., Saini, A. K., Cope, B., Kalantzis, M., & Searsmith, D. (2024a). Peer-reviews and AI feedback compared: University students' preferences. *EdArXiv*. https://doi.org/10.35542/osf.io/uy8qp

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

Appendix A

Excerpt from Sample Generic AI Review

"The paper discusses the importance of using multiple modes of communication in order to facilitate learning. It argues that this is essential for effective online learning in the future, and that teachers must be involved in order to create a meaningful experience for students. The article also discusses the potential benefits of holographic technology, which could enable students to more fully engage in multimodal learning. The text could use more specific examples of empirical evidence regarding the effectiveness of technology in online learning. For example, the text could provide survey or interview results from students and teachers who have experienced online learning with technology. Additionally, the text could provide case studies of successful or unsuccessful implementations of technology in online learning. Rating: 2"

Appendix B

Excerpt from Sample Calibrated AI Review

"The essay "[title of essay]" provides an insightful exploration of the transition to one-to-one technology in K-12 education, focusing on the impact of ubiquitous learning and the integration of technology, pedagogy, and content knowledge (TPACK framework). The writer adeptly weaves together personal anecdotes, theoretical frameworks, and empirical data to argue for the benefits and potential drawbacks of this educational shift. However, the essay could benefit from a deeper engagement with a broader range of theories and more explicit connections between the selected theories and the empirical evidence presented.

The writer introduces the concept of ubiquitous learning, drawing on the work of Mary Kalantzis and Bill Cope, and further explores the TPACK framework as proposed by Punya Mishra and Matthew J. Koehler. These frameworks are crucial for understanding the shift towards one-to-one technology in education and its potential impact on teaching and learning. However, the essay could be enriched by incorporating additional theoretical perspectives that address the digital divide, students' digital literacy, and the socio-cultural implications of technology in education. For instance, referencing the work of Harris, Al-Bataineh, and Al-Bataineh (2016) on the effects of one-to-one technology on student academic achievement and motivation could provide a more nuanced understanding of the educational outcomes of this shift… The use of figures and references to empirical studies helps to illustrate the points made. However, the connections between the theoretical frameworks (ubiquitous learning and TPACK) and the empirical evidence could be made more explicit. For example, discussing how the findings from the study on student engagement (Elizondo, 2018) or the benefits for English Language Learners (Liu et al., 2014) specifically relate to the principles of ubiquitous learning or the domains of the TPACK framework would strengthen the conceptual coherence of the essay…

Suggestions for Improvement:

1. Incorporate a discussion of digital equity and the digital divide to critically examine the potential disparities in access and outcomes associated with one-to-one technology initiatives.

2. Make more explicit connections between the empirical evidence presented (e.g., studies on engagement, achievement, and benefits for ELL students) and the theoretical frameworks of ubiquitous learning and TPACK…

Overall, the essay provides a comprehensive and engaging exploration of the shift towards one-to-one technology in K-12 education. By deepening the engagement with a broader range of theoretical perspectives and making more explicit connections between theory and practice, the essay could offer a more nuanced understanding of this educational transformation. Rating: 3”