INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 14, Issue 4, (2025) E-ISSN: 2226-6348

# Mentor or Examiner? A Critical Discourse Analysis of Ai-Generated Feedback in EFL Writing Education

Liang Ting[1], Charanjit Kaur Swaran Singh[2], Tee Tze Kiong[3], Farkhodjon Rakhimjonov[4], Tarsame Singh Masa Singh[5]

[1,2]Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris, Perak, Malaysia, [3]Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia, [4]Faculty of Philology, Alfraganus University, Uzbekistan, [5]Faculty of Education, Languages and Psychology, SEGi University, Selangor, Malaysia

Corresponding Author Email: charanjit@fbk.upsi.edu.my[2]

**DOI Link:** http://dx.doi.org/10.6007/IJARPED/v14-i4/26651

***Published Online:*** 19 October 2025

**Abstract**

This study addresses a critical gap in understanding how artificial intelligence (AI) systems construct evaluative discourse in language education contexts. While AI-powered feedback tools increasingly supplement or replace human assessment in English as a Foreign Language (EFL) writing instruction, limited attention has been paid to the discursive mechanisms through which these systems position learners, construct authority, and shape pedagogical relationships. Drawing on Critical Discourse Analysis (CDA), this article develops a comprehensive framework for analyzing AI feedback discourse through a comparative analysis of two Large Language Models' (LLMs') responses to 63 undergraduate EFL descriptive essays. Employing Fairclough's three-dimensional CDA framework, the analysis reveals distinct patterns in how AI systems construct their evaluative stance, distribute agency, and enact pedagogical authority in feedback. An emergent six-part taxonomy of discourse moves is identified: diagnostic positioning, prescriptive directives, facilitative suggestions, affective engagement, metalinguistic explanation, and comparative benchmarking. Findings indicate that the two LLMs employ contrasting discursive strategies—akin to a mentor versus an examiner—with significant implications for student positioning, learning autonomy, and the nature of pedagogical relationships in digitally mediated contexts. The proposed framework extends CDA methodology to AI-generated educational discourse and offers educators practical tools for critically evaluating AI feedback systems. As educational institutions rapidly adopt AI assessment tools, this taxonomy enables informed decisions about which discursive practices align with desired pedagogical values. The study concludes by discussing implications for student agency, pedagogical authority, and AI literacy in teacher education, and by recommending the development of more pedagogically-aligned AI feedback systems.

**Keywords:** Critical Discourse Analysis, AI-Generated Feedback, EFL Writing Instruction, Pedagogical Authority, Large Language Models

## Introduction

The accelerating integration of artificial intelligence into educational assessment represents a fundamental reconfiguration of pedagogical relationships and evaluative practices. Within this broader technological transformation, generative AI technologies—particularly Large Language Models (LLMs)—are being rapidly deployed as automated feedback systems in writing instruction contexts globally. This phenomenon reflects what Selwyn (2019) characterizes as the "digital rationalization" of education, wherein algorithmic systems increasingly mediate the traditionally human-centred processes of teaching, assessment, and feedback provision. In English as a Foreign Language (EFL) writing instruction specifically, tools such as ChatGPT and other LLM-based platforms are being positioned as scalable solutions to persistent resource constraints, promising immediate, detailed feedback to large student populations while potentially alleviating teacher workload and offering timely formative assistance (Alnemrat et al., 2025; Warschauer et al., 2023).

This technological shift occurs within a critical moment for educational discourse studies. As Knox (2020) argues, the materialization of AI in educational spaces necessitates renewed attention to how algorithmic systems construct, reproduce, or transform existing power relations through language. The discourse produced by AI feedback systems is not merely technical output but constitutes what van Dijk (2014) terms "technologically-mediated social practice"—language that actively shapes educational identities, relationships, and epistemologies. Yet despite the profound implications of AI assuming roles traditionally occupied by human educators, systematic investigation into *how* these systems communicate with learners remains remarkably limited.

Recent empirical studies have begun validating the quantitative effectiveness of AI-generated feedback in improving writing outcomes. Alnemrat et al. (2025) demonstrated that undergraduate EFL students who revised essays using LLM feedback achieved writing gains comparable to those receiving traditional teacher feedback, with no significant performance differences between AI and human feedback groups. Similarly, Poláková and Ivenz (2024) reported measurable improvements in university students' writing quality—including conciseness and grammatical accuracy—following ChatGPT-generated feedback interventions, alongside generally positive student perceptions of AI's pedagogical utility. These findings underscore the technical viability of large language models as scalable complements to teacher feedback in resource-constrained EFL contexts (Alnemrat et al., 2025).

However, the current research trajectory reveals a critical epistemic gap. Existing studies predominantly adopt what Biesta (2010) critiques as "effectiveness paradigms", focusing on *what* measurable outcomes AI feedback produces while neglecting fundamental questions about *how* feedback is discursively constructed and what this communicative mode implies for teaching and learning dynamics. In other words, there exists a paucity of inquiry into the discourse of AI-generated feedback itself—the linguistic mechanisms, evaluative stances, and power relations embedded within automated assessment communications. This gap is particularly concerning given that feedback constitutes far more than information transmission; as Carless and Boud (2018) argue, feedback represents a fundamentally interactional and dialogic process through which learners develop identity, agency, and epistemic authority. The discursive characteristics of AI feedback—how it positions student writers, what pedagogical identity the AI assessor assumes, and how closely its communicative patterns mirror or diverge from human teacher discourse—remain largely unexplored despite their potentially profound influence on learning relationships.

Scholars in applied linguistics and educational assessment have long recognized that teacher-written feedback involves intricate discursive negotiations to balance critique with encouragement, maintain social rapport, and scaffold learning while preserving student agency (Hyland & Hyland, 2006; Winstone & Carless, 2020). This understanding draws from sociocultural theories of assessment that conceptualize feedback as co-constructed meaning-making within specific power geometries (Pryor & Crossouard, 2008). An AI system, operating without genuine social cognition or relational awareness, may deploy fundamentally different linguistic registers and evaluative moves when delivering critique. Emerging evidence from institutional discourse studies suggests significant variation in how AI-related educational communications construct authority. For instance, recent critical analyses of university AI policy documents reveal divergent rhetorical strategies correlating with institutional status: more prestigious institutions adopt authoritative, compliance-oriented language in framing AI use, whereas less selective universities employ inclusive, explanatory rhetoric (Baker & Hübner, 2024). Such findings indicate that AI-mediated discourse can vary substantially in authoritative stance, potentially reflecting and reproducing educational hierarchies through subtle linguistic choices.

Within the specific context of writing feedback, these discursive variations raise theoretically and pedagogically significant questions: Does an AI feedback system enact a voice of hierarchical authority—functioning as an "examiner" imposing evaluative judgments—or does it adopt a more facilitative stance—operating as a "mentor" guiding developmental improvement? How might these communicative positions affect students' epistemic agency, their reception of feedback, and their identity formation as writers? Furthermore, what ideologies of teaching, learning, and assessment become embedded—whether intentionally or inadvertently—within the linguistic architecture of automated feedback systems? To address these critical gaps, the present study conducts a systematic Critical Discourse Analysis of AI-generated feedback provided to EFL student essays, examining how two different LLMs discursively construct their pedagogical identities and position learners within evaluative relationships. Specifically, we compare feedback generated by two distinct LLM systems on the same corpus of student texts to develop an empirically-grounded taxonomy of AI feedback discourse strategies and to examine how each system enacts pedagogical authority through language. Employing Fairclough's (1992) three-dimensional CDA framework, we analyze the textual features, discursive practices, and broader sociopolitical implications of AI feedback, thereby illuminating what we term the "language of algorithmic assessment" in EFL writing contexts.

This research addresses three interrelated objectives: (1) to identify and categorize the primary discourse moves employed in AI feedback—that is, to map *what* communicative actions the AI performs through its commentary (e.g., diagnosing, directing, suggesting, explaining); (2) to reveal systematic differences in feedback style between the two AI systems—specifically, to examine *how* one system may function as a mentor-like coach while another operates as an examiner-like evaluator, and what linguistic mechanisms produce these distinct pedagogical personas; (3) to discuss the educational, ethical, and practical implications of these discursive differences for student agency, teacher authority, and the broader integration of AI assessment tools in language education. This investigation contributes to several scholarly conversations simultaneously. Methodologically, it extends CDA approaches into the emerging domain of AI-generated educational discourse, demonstrating how critical language analysis can reveal power dynamics and ideological positioning within algorithmically-produced texts. Theoretically, it advances understanding of how pedagogical relationships and assessment functions are being reconfigured through technological mediation. Practically, it offers educators and institutional decision-makers a critically informed analytical framework for evaluating and selecting AI feedback tools in alignment with specific pedagogical values and educational goals. As educational institutions accelerate AI adoption often without sufficient critical reflection (Selwyn & Facer, 2022), this research provides essential tools for interrogating not merely *whether* AI feedback works, but *how* it communicates and what educational relationships it constructs in the process.

## Literature Review

### *CDA, Power and AI in Education*

Critical discourse analysis offers a theoretical and methodological lens for examining how language both reflects and shapes social power relations (Fairclough, 1992). CDA approaches such as Fairclough's three-dimensional model posit that any instance of text (here, AI feedback comments) must be analyzed not only at the textual level (vocabulary, grammar, structure), but also in terms of discursive practice (how the text is produced, circulated, and consumed) and social practice (the wider social and institutional contexts and power structures that give the text meaning). This framework is well-suited to studying AI in education, where issues of authority, agency, and ideology are increasingly evident in discourse. A particularly relevant insight comes from a systematic CDA of AI in higher education by Bearman et al. (2023). Analyzing the rhetoric around AI's role in academia, they identified two dominant discourses: a "discourse of imperative response" portraying AI integration as inevitable and non-negotiable, and a "discourse of altering authority" focusing on how AI challenges traditional teacher-student roles and power dynamics. The latter discourse is especially pertinent to EFL writing instruction, where feedback has traditionally been a cornerstone of the teacher's authority. If AI systems begin to assume some of this feedback function, the discourse they use may either reinforce or disrupt established authority relations. The *voice* of the AI—whether it speaks as an authoritative judge or as a collaborative peer—could influence how students perceive the feedback and their own agency in the learning process. As CDA scholars argue, discourse not only reflects social relationships but actively constructs them (Fairclough, 1992; Luke, 2002). Therefore, examining the discourse of AI-generated feedback can reveal underlying ideologies of teaching and learning being embedded (consciously or not) into these tools. In summary, prior literature establishes that discourse analysis is a powerful means to uncover the subtle ways AI systems may perpetuate or transform power relations in educational settings. Building on this foundation, we apply CDA to a new, granular context—the feedback comments given by AI to student writing—to explore how pedagogical authority and learner positioning are constructed in this emerging form of educational discourse.

### *AI-Generated Feedback in EFL Writing*

Research on AI-provided feedback for second language writing has accelerated in recent years, reflecting both technological progress in Natural Language Processing and a pressing need to support writing instruction in large EFL classes. A 2024 systematic review by Shi and Aryadoust identified 31 different AI-based automated writing evaluation and feedback systems studied in the literature, underscoring the diversity of tools and contexts being explored. These range from specialized grammar-correction programs to advanced LLM-based platforms. The review noted that most studies to date have evaluated such systems in terms of accuracy, student improvement, and usability, often reporting generally positive outcomes (e.g., improved linguistic accuracy in revisions) but also calling for deeper analysis of how these systems function pedagogically (Shi & Aryadoust, 2024). A complementary integrative framework is provided by Panadero and Lipnevich (2022), who developed the MISCA model to classify feedback elements by Message content, Implementation method, Student characteristics, Context, and Agent (source of feedback). Using MISCA, researchers have begun to compare human versus AI feedback: for example, certain studies have found that automated feedback tends to focus on local issues like grammar and vocabulary (message content), altering the traditional balance where human teachers might prioritize global issues like content and organization. These emerging frameworks highlight that AI feedback can differ from teacher feedback not only in what is addressed but in how it is delivered and contextualized (e.g., the tone and directives used by the "Agent"). Such differences in the agent's discourse may significantly affect how students interpret and act on feedback. Empirical studies in EFL contexts are beginning to reveal how learners interact with AI feedback tools. Yang et al. (2024) conducted an exploratory study in China where university EFL students used an AI writing evaluation system (*Pigai*) for iterative essay revisions. Their detailed analysis of system-student interaction logs showed "sophisticated engagement patterns" over multiple drafts. Initially, students responded to the AI's corrective feedback on grammar and spelling in a rather mechanical, surface-level way, but over time some learners engaged more critically, especially when the AI provided non-error-related suggestions (e.g., alternative vocabulary or style improvements). Interestingly, the AI feedback lacked explanatory depth—it would indicate errors or offer rephrased sentences without much contextual rationale—and students often ignored or misinterpreted these less explicit suggestions. This finding aligns with other reports that while AI feedback is thorough on form-focused issues, it may provide fewer explicit explanations or examples compared to human teachers, potentially limiting students' uptake of more complex advice (Zhang & Hyland, 2022). In practice, EFL learners might need guidance to fully benefit from AI-generated comments beyond simple corrections. Comparative research on AI vs. human feedback also informs our work. As noted earlier, studies like Alnemrat et al. (2025) and Tran (2025) have quantitatively demonstrated that AI feedback can be as effective as teacher feedback in improving certain aspects of writing. What these studies also highlight, indirectly, is that AI feedback is often delivered in a markedly different manner. For instance, in Tran's (2025) study on Vietnamese EFL learners, the AI feedback was provided immediately and in written form on a platform, whereas teacher feedback was given with some delay and sometimes orally. The immediacy and written, impersonal nature of AI comments could influence how students respond.

One concern that emerges in the literature is how students use the AI feedback they receive. Preliminary evidence suggests that learners may often take AI feedback at face value, implementing suggestions without much reflection. This passive uptake could lead to shallow revisions—students apply fixes but may not truly internalize the underlying writing principles—a phenomenon also warned by Zhang and Hyland (2022) in the context of grammar checkers. The risk of uncritical acceptance reinforces the importance of understanding the manner in which AI feedback is delivered: if the AI's discourse comes across as authoritative and definitive ("correct this, do that"), students might be even less inclined to question it, whereas a more dialogic and explanatory feedback style might invite students to think and make choices. Hence, there is a pedagogical imperative to examine and possibly shape the discourse of AI feedback to foster deeper learning. In summary, prior research shows that AI-generated feedback can effectively address many writing issues and is becoming a viable supplement to teacher feedback in EFL writing classrooms. Yet, how AI feedback is communicated—its discourse moves, tone, and implicit positioning of the learner—remains underexplored. This study builds on the literature by using a CDA approach to fill that gap. We extend the focus from *what* improvements AI feedback yields to *how* the feedback itself is constructed and what it might mean for educational interactions. The next sections outline our methodology for analyzing AI feedback discourse and present a taxonomy of discourse strategies observed in two different LLMs' feedback to student writers.

## Methodology

### *Context and Data*

This research was conducted in the context of an EFL writing program at the undergraduate level. Sixty-three (N=63) students wrote a descriptive essay (approximately 300–400 words each) as part of a course assignment. The prompt asked students to *"describe a place you know very well, giving a clear general impression and organized supporting details".* The essays, written in English by non-native speakers (primarily L1 Chinese), covered various familiar places (e.g., hometowns, tourist sites) and served as the baseline student texts for analysis. To generate AI feedback on these essays, we employed two different Large Language Models, referred to here as LLM-A and LLM-B for anonymity. Each student essay was fed to both LLM-A and LLM-B, yielding two separate feedback outputs per essay. In total, 126 AI feedback responses (63 per LLM) were collected for analysis.

The two LLM systems were chosen to represent distinct approaches to AI feedback. LLM-A is a widely used generative model integrated into a writing assistant platform. It was prompted to provide formative feedback including strengths, weaknesses, and suggestions, and to assign an indicative score out of 15 points (as per the course rubric). In practice, LLM-A's feedback followed a structured template: it typically opened with a brief overall evaluation, then listed Strengths (focusing on content, structure, and language positives) and Areas for Improvement, often subdivided by categories (Content, Structure, Language, Mechanics). LLM-A's comments were concise, bullet-like in places, and included explicit directives (e.g., *"Remove the meta-comment in the conclusion"*) and corrections for specific errors (often formatted as "Issue – Correction" pairs). An example excerpt from LLM-A on one essay illustrates this format: *"Language (Precision): Issue: Missing article in 'largest freshwater lake in China.' Correction: 'the largest freshwater lake in China.'"* In general, LLM-A's feedback averaged 200 words per essay and was highly organized, resembling the style of an experienced examiner systematically marking a script.

LLM-B, in contrast, is a newer LLM-based assistant known for a more conversational style. We prompted LLM-B with a less rigid instruction: to give helpful feedback and a holistic score (on a 15-point scale) with reasoning. LLM-B's responses were more narrative in structure. They usually began with a short paragraph summarizing the essay's overall performance in a conversational tone (often addressing the student as "you" or the essay as "this essay"), followed by separate paragraphs for Content, Structure, Language, and Mechanics aspects of the writing. LLM-B did not explicitly label "strengths" vs "weaknesses", but it interwove praise and critique in each category paragraph. It also ended with a set of *"Specific suggestions"* for revision, written as a series of recommendations (frequently using verbs like *"consider...", "expand...", "add..."*). For instance, LLM-B's feedback on one essay noted: *"Content: While you provide a general impression of Nanchang as a 'vibrant city,' the supporting details lack depth and personal connection... Consider adding specific memories or sensory details to demonstrate your familiarity with the city."* This illustrates LLM-B's tendency to couch critiques within a helpful advisory tone. LLM-B also assigned a score (often phrased as "Score: X points (Y–Z range)"), situating the essay in a performance band. On average, LLM-B's feedback was slightly longer (220–250 words) and read more like prose, in contrast to LLM-A's bullet-point style.

It is important to note that these differences in format and tone between LLM-A and LLM-B were not explicitly hard-coded by us but emerged from the inherent design and default style of the models (and possibly differences in prompt interpretation). This natural variation provided a rich basis for comparative discourse analysis. We treated each set of an essay with its two feedback responses as a mini-case, allowing side-by-side comparison of how two AI systems "responded" to the same student text. All student essays and AI feedback outputs were imported into a qualitative data analysis software for coding.

*Analytical Approach*

Our analysis was informed by Fairclough's (1992) three-dimensional CDA framework, examining each AI feedback document on three levels: textual, discursive practice, and social practice. At the textual level, we conducted a close reading of the language and structure of the feedback comments. We identified linguistic features such as speech acts (e.g., advising vs. commanding), pronoun use (direct address "you" vs. impersonal constructions), modality (hedges like *could, maybe* vs. firm statements), evaluative vocabulary (praise terms like *excellent, clear* vs. critical terms like *lacks, weak*), and the overall organization of the feedback text. We also noted any meta-linguistic terminology (e.g., references to grammar terms, rubric criteria) used by the AI. At the level of discursive practice, we considered how the feedback was produced and for whom. This involved examining the genre conventions the AI appeared to follow (e.g., teacher grading memo, peer tutoring comment, etc.) and the implied interaction—even though the feedback is a one-direction text, we asked: What role is the AI writer adopting? Who is the implied audience (the student, clearly, but addressed as a learner or as a peer)? We also reflected on the intertextual context: the AI models likely drew on vast training data including educational texts, which may shape their style. Finally, at the social practice level, we interpreted what the discourse strategies mean in the context of pedagogical relationships and norms. Here we connected the textual findings to broader concepts like pedagogical authority, learner agency, and cultural expectations in EFL education (for example, how an authoritative tone might align with or challenge the teacher-centred norms in some educational cultures).

To systematically categorize the discourse features of the AI feedback, we employed an inductive coding process akin to qualitative content analysis, embedded within the CDA perspective. Two researchers (the authors) first independently reviewed a subset of 20 feedback samples (10 from each LLM) and noted recurring "moves" or elements in the feedback. We then discussed and compared these initial codes, iteratively refining our codebook. Through this process, six salient categories of feedback discourse emerged (e.g., instances where the AI was diagnosing a problem, giving a direct order, offering a suggestion, etc.). We then coded the entire dataset of 126 feedback responses according to these categories. Segments of text (ranging from a phrase to a couple of sentences) were marked with one or more category labels as appropriate. We achieved a high intercoder agreement on the classification (Cohen's $\kappa = 0.88$ after two rounds of reconciliation), indicating consistency in identifying the discourse features. Throughout, we remained attentive to how these codes map onto Fairclough's dimensions: for instance, a "prescriptive directive" is a textual feature (an imperative form) that suggests a certain discursive practice (an authoritative stance) and potentially reflects a social practice (teacher-centred pedagogy). In presenting our findings, we use illustrative excerpts from the AI feedback (with minor surface edits for brevity or anonymity if needed). All excerpts are labelled with the source LLM (A or B) and the context if relevant. Because our aim is to analyze *how* the AI communicates, we focus on the AI's wording rather than the specific content of the student essay. However, we occasionally reference the student's writing (e.g., what the essay lacked or did well) insofar as it is mentioned in the feedback, to provide context for the AI's comments. The six-part taxonomy of discourse moves is described in detail in the next section, followed by a comparative analysis of how LLM-A and LLM-B differ in their use of these moves and the pedagogical persona they project.

**Findings**

*A Taxonomy of AI Feedback Discourse Moves*

Through CDA-guided analysis, we identified six primary discourse categories in the AI-generated feedback. These categories form a taxonomy of the communicative moves that the LLMs used when responding to student essays. The categories, along with their defining characteristics and examples, are presented as follows.

*Diagnostic Positioning*

In many feedback instances, the AI assumes the role of a diagnostician, evaluating the student's work and positioning it in terms of achievement or shortcomings. This involves statements that summarize the essay's overall quality or the presence/absence of required elements. For example, LLM-B often opened with an evaluative synopsis such as: *"This essay demonstrates basic relevance to the topic but falls short of meeting several task requirements, affecting both content and structure."* Here the AI identifies what the student did ("basic relevance") and did not do ("falls short...requirements"), effectively *positioning* the student's performance on a scale of adequacy. LLM-A likewise engaged in diagnostic positioning, though sometimes this came at the end of the feedback in a summary statement. For instance, LLM-A concluded one essay's comments with: *"A well-organized, insightful essay. Refine minor language errors and enrich descriptions with concrete details to elevate clarity and engagement."* This sentence first positions the essay positively (well-organized, insightful), then diagnoses what would elevate it further (more detail, minor error correction). Diagnostic positioning serves to frame the rest of the feedback; it tells the student where they stand. We observed that LLM-A's diagnostic remarks were often succinct and either wholly positive or a balanced mix, whereas LLM-B sometimes provided a more nuanced or critical diagnosis up front (potentially to justify the score given). In either case, these statements set an overall evaluative tone. By explicitly stating an appraisal, the AI situates itself as an assessor—implicitly asserting authority to judge the work.

*Prescriptive Directives*

A prominent feature of the AI feedback (especially from LLM-A) was the use of direct instructions telling the student exactly what to do to improve. We label these instances prescriptive directives. Linguistically, they are typically imperatives or explicit advice phrased as commands. For example, LLM-A frequently issued bullet-point fixes: *"Remove the meta-comment ('This essay highlights... modifications!') in the conclusion"* or *"Add a concluding sentence summarizing the market's significance (e.g., 'This market encapsulated Beijing's vibrant culture')."* These are unambiguous prescriptions—the AI is not merely suggesting but instructing changes to the text. LLM-B also produced prescriptive directives, though often embedded in its suggestions section. In one case, after its narrative feedback, LLM-B listed: *"Specific suggestions: Remove the meta-commentary in the final paragraph, add personal experiences with specific locations mentioned, and reorganize content using spatial or chronological order to improve coherence."* The bolded verbs (emphasized here) illustrate the imperative tone. Such directives mirror the behavior of a human examiner or editor who marks errors and tells the writer how to fix them. The presence of prescriptive directives in AI feedback is double-edged: on one hand, it provides clear, actionable guidance (which students often appreciate); on the other, it can position the student as a relatively passive recipient of "orders" for improvement. From a discourse perspective, the AI in these moments adopts a powerful speaker role, assuming the right to dictate revisions. Notably, LLM-A's structured format lent itself to a higher frequency of these directives—it would itemize issues and corrections extensively (grammar, word choice, punctuation, etc.). LLM-B, while also giving commands, tended to wrap them in softer language when outside the explicit "Specific suggestions" list. Nonetheless, both AIs at times wrote as strict proofreaders, using prescriptive language that leaves little room for negotiation (e.g., "should be...", "replace X with Y"). This category is central to the "examiner" identity enacted by the AI.

*Facilitative Suggestions*

In contrast to the above, we also observed many instances of the AI offering facilitative suggestions—advice or prompts phrased in a non-authoritative, encouraging manner. These often took the form of hedged recommendations, open-ended questions, or options for the student to consider. LLM-B was notably inclined to use this style. For example, LLM-B advised one student: *"Consider adding more sensory details about sounds, smells, or atmosphere to enhance the vivid description you've established."* The use of "consider" softens the directive, inviting the student to reflect on an idea rather than commanding it. Similarly, LLM-B asked another student, *"Perhaps think about how this market experience helped you understand Beijing's character?"*, effectively suggesting a reflection to include in the conclusion (this was paraphrased in feedback as *"expand the conclusion to reflect on how this experience helped you understand the city"*). LLM-A, while more prescriptive overall, did occasionally employ facilitative language, particularly when addressing higher-order content improvements. For instance, it might say *"Try adding a sensory detail for immersion (e.g., describe the bell's echo in the tower)"*, which, though still an imperative ("try adding"), feels like a gentle nudge to enrich the content creatively. Facilitative suggestions are characteristic of a coaching or mentoring discourse: the AI provides guidance while leaving space for the student's agency in implementing it. This style aligns with formative feedback best practices that human teachers use, such as mitigated criticism and reader-responsible comments (e.g., "you could improve this by..." instead of "this is wrong, do that"). In our data, LLM-B clearly embraced this approach more than LLM-A. The discursive effect is that the AI positions itself more as a collaborator or guide. Such language can encourage students to engage more thoughtfully with the feedback, as it prompts consideration rather than dictating solutions. We will later see how this difference plays into the mentor vs. examiner personas of the two LLMs.

*Affective Engagement*

Both AI models at times engaged in what we term affective engagement—using language that is supportive, encouraging, or attends to the writer's affective needs (confidence, motivation). This category includes praise for what was done well, empathetic remarks, or motivational encouragement. LLM-A, due to its "Strengths" section, consistently provided praise for each essay. It would highlight positives like *"excellent organization (history → architecture → cultural role) and rich vocabulary ('exquisite', 'enduring legacy')"* or *"Clear focus on historical/cultural duality, with precise vocabulary..."* Such comments not only identify strengths but also convey approval, which can boost a learner's confidence. LLM-B also included praise, often interwoven with critique. For example, *"Your personal experience exploring the night market creates an engaging narrative that demonstrates genuine familiarity with the place"* is clearly validating the student's effort and storytelling. Beyond praise, we looked for any socio-emotional cues. Direct expressions of empathy (e.g., "I understand..." or "It's great that you...") were rare, as expected from an AI, but there was an element of encouraging tone in many comments. Phrases like "effectively fulfils the task requirements" or "successfully describes your visit" serve to affirm the student's capability.

Even criticisms were sometimes buffered with positive language (a classic "feedback sandwich" approach). For instance, LLM-B in one case noted the essay was under length and lacking depth but prefaced with *"This essay demonstrates basic topic relevance..."* before delivering the critique. The presence of affective engagement indicates the AI is not solely a cold evaluator; it attempts to emulate the motivational aspect of teacher feedback. This is crucial because research in L2 writing feedback emphasizes the role of praise and a friendly tone in maintaining student receptiveness (Hyland & Hyland, 2001). In our comparison, LLM-A's affective engagement was a bit formulaic (explicit strengths listed for every essay, possibly due to prompt instructions), whereas LLM-B's felt more integrated into the narrative flow of feedback. Nonetheless, both provided encouragement, contributing to a mentor-like dimension in the discourse, particularly when balanced with criticism.

*Metalinguistic Explanation*

A significant portion of the feedback fell into a category of metalinguistic explanation, where the AI stepped back from the specific text at hand and offered general language-related commentary or rules. This often happened in the context of grammar or mechanics corrections. LLM-A, for example, didn't just correct errors but frequently named the rule or issue: *"Passive voice weakens impact – change 'being carefully repaired' to active voice 'experts carefully repairing...'"* or *"Use commas after introductory clauses: 'The night before I left Beijing, I...'"* In doing so, it was providing a mini lesson on language form. LLM-A also used terminology like "article error", "redundancy", "hyphenate compound adjectives", explicitly referencing grammatical concepts and correct usage. LLM-B also engaged in metalinguistic talk, though slightly less systematically. It pointed out, for instance, *"'must - visit' should be* *'must-visit' (hyphenation issue)"* or explained that *"'Let me know if you'd like any modifications!' is clearly instructional text that should be removed"*—implicitly teaching the student about genre-appropriate language. These explanations function on the textual level by clarifying the *why* behind a suggested change, thus educating the student on language rules or writing conventions. This aligns with good practice in feedback, as merely giving the correct answer is less instructive than explaining the rationale. By including metalinguistic explanations, the AI adopts a somewhat didactic teacher role, as opposed to just editor. It's attempting to increase the student's understanding of language. In discourse terms, the AI is positioning itself as an expert in language knowledge (which, in fairness, it is programmed to be).

The tone can sometimes come off as pedantic—e.g., a strict teacher emphasizing rules—but it can also be seen as the AI trying to be helpful by imparting knowledge. Between the two LLMs, LLM-A delivered more of these explicit rule statements (likely owing to its structured approach that parsed feedback by language issues), whereas LLM-B embedded them within its content and language paragraphs as needed. Metalinguistic content in the feedback contributes to the educational value of the comments, shifting the focus from just *fixing this essay* to *improving writing skills* more generally, which is an important goal in EFL writing education.

*Comparative Benchmarking*

The final category we identified is comparative benchmarking, where the AI feedback makes an explicit or implicit comparison to a standard or reference point. This can manifest as references to rubrics, exemplar performances, or even genre exemplars. One clear form of this was LLM-B's practice of situating the score in a range, e.g., *"Score: 8 points (7–9 points range)"*. By doing so, LLM-B was benchmarking the essay against a scoring band (implying, for instance, this essay is a low-average performance if 7–9 is a certain tier). LLM-A gave raw scores (e.g., 12/15) without ranges, but it occasionally alluded to expectations from the prompt, which is another kind of comparison (essay vs. task requirements). For example, LLM-B explicitly contrasted the student's approach with the task expectation: *"The task asks you to describe 'a place you know very well,' but your description reads more like a tourism brochure than personal experience."* Here, the AI is benchmarking the student's writing against the expected personal narrative genre. Another instance from LLM-A: *"Use stronger adjectives ('iconic' Terracotta Army vs. 'amazing')."*—the AI is comparing the student's word choice ("amazing") with a more appropriate or higher-level word ("iconic"), effectively benchmarking lexical choices against a more advanced vocabulary standard.

Comparative benchmarking also appeared when feedback referenced how the essay could be improved by reaching a certain quality: LLM-A wrote *"focus solely on describing Nanchang"* after pointing out a meta-comment—implicitly comparing the essay to an ideal descriptive essay that would have no meta-commentary. In sum, this category captures moments where the AI sets a bar or invokes a comparison (whether to an external standard or between what the student did and what could be). This discourse move situates the student's work in a broader context and can encourage the student to self-evaluate against exemplars. However, it also reinforces the AI's assessor identity, as benchmarking often comes from a position of authority (the one who knows the standard). Both LLMs engaged in this, though LLM-B did so more explicitly via score ranges and direct mentions of task guidelines, whereas LLM-A did so through suggested "upgrades" to meet expected academic writing norms.

Collectively, these six discourse move categories form a descriptive profile of AI feedback communication. They often occurred in combination—for example, a single comment from the AI could involve diagnostic positioning (*"underdeveloped conclusion"*—diagnosing a weakness) followed by a prescriptive directive (*"add a concluding sentence summarizing the main insight"*), perhaps with comparative benchmarking (*"to better mirror the introduction's promise"*). Nonetheless, distinguishing these categories is useful for understanding the balance and style of the feedback. We found that the frequency and emphasis on each category varied between LLM-A and LLM-B, which in turn correlates with the different pedagogical identities each AI seemed to enact. We explore this comparison in the next subsection.

*Contrasting Pedagogical Identities: Mentor vs. Examiner*

A key finding of our analysis is that the two AI systems, despite addressing the same student texts, adopted contrasting pedagogical personas through their discourse. LLM-A's feedback style is akin to an examiner or authoritative instructor, whereas LLM-B often comes across more as a mentor or coach. This distinction emerged from differences in how the above discourse moves were realized and balanced by each model.

*LLM-A – The Examiner*

LLM-A's feedback discourse was characterized by an authoritative, assessment-oriented tone. It heavily utilized prescriptive directives and metalinguistic explanations, signalling a focus on correctness and adherence to rules. For instance, LLM-A would systematically list errors (diagnosing them) and immediately provide corrections, as seen in its treatment of grammar and mechanics issues (articles, hyphenation, etc.). This mirrors the approach of a traditional examiner or proofreader who marks mistakes and tells the student exactly how to fix them. Moreover, LLM-A gave each essay a numeric score out of 15 without additional qualification, much as a teacher might do when grading. The language was often impersonal and declarative (e.g., "Missing article in '...'. Correction: 'the ...'"), which positions the AI as a neutral authority on language usage, not inviting debate or personal engagement. Even compliments in LLM-A's feedback, while present, were phrased in a formal register ("Exceptional organization...precise vocabulary"). There was little use of "I" or direct address; instead of saying "I liked your description of X," LLM-A would state "Content: Vivid details about X support the impression well". This impersonal style is typical of an examiner's report, emphasizing objectivity.

Another indicator of LLM-A's examiner identity is how it adhered to a rubric-like structure. By explicitly separating content, structure, language, etc., and commenting on each, the AI mimicked institutional feedback forms. It also tended to enforce the assignment criteria strictly—for example, repeatedly noting the requirement of a "general impression" and deducting points if not explicit, or pointing out if the essay length was insufficient. This shows LLM-A aligning itself with the *assessor's expectations* set by the task. In essence, LLM-A speaks *to* the rubric and *for* the institution. The student is cast in a relatively submissive role: the one being evaluated and corrected. LLM-A's frequent use of imperatives ("Add this, Replace that, Use X, Remove Y") reinforces a hierarchical dynamic wherein the AI, like a strict teacher, directs the learning. The benefit of this approach is clarity—students know exactly what to do. However, the potential downside is reduced student agency; the feedback can be taken as a checklist of orders to implement, possibly without much reflection (as noted earlier, a concern if students become passive). In terms of social practice, LLM-A's discourse could be seen as reproducing a teacher-centred model of education, where authority lies with the feedback-giver and compliance is expected from the learner.

*LLM-B – The Mentor*

In contrast, LLM-B's discourse exhibited a more dialogic and supportive tone, aligning with a mentoring role. LLM-B often addressed the student directly ("you") and framed its feedback as advice for improvement rather than judgment. It extensively used facilitative suggestions—for example, *"consider adding..."*, *"you could expand by..."*—which inherently give the student a choice and encourage engagement with the suggestion. This approach tends to preserve the student's agency, as it implies that the student is ultimately responsible for deciding how to revise. LLM-B also integrated affective engagement more seamlessly, acknowledging positive aspects in the midst of critique (e.g., praising vivid details and genuine familiarity before pointing out the abrupt ending in the Beijing market essay). The tone is reminiscent of a human tutor who aims to boost confidence while guiding improvements. One hallmark of LLM-B's mentor-like identity is its effort to explain and contextualize feedback in relation to the student's intentions or the task's purpose. For instance, rather than simply saying "the essay lacks personal experience", LLM-B elaborated that the student's description *"reads more like a tourism brochure than personal experience"*, explicitly connecting to the task ("describe a place you know well"). This not only benchmarks the performance (as discussed in comparative benchmarking) but also treats the student as a partner in understanding the gap—a strategy a mentor would use to raise the student's awareness.

LLM-B's narrative style, writing feedback in full sentences and cohesive paragraphs, creates an impression of a conversation or a letter to the student, rather than a checklist. It sometimes even used first person plural or hypothetical phrasing ("you might want to...", "we could imagine more details about..."), which softens the authority and makes the feedback feel like a collaboration in improving the text. Moreover, LLM-B showed *flexibility* and *acknowledgement of the student's perspective* in ways LLM-A did not. For example, where a student's essay had a particular strength, LLM-B would sometimes mention how that strength contributed to the overall effect (e.g., how the student's personal narrative added engagement), almost as if empathizing with the student's effort. It also avoided an overly formal tone; its language was professional but approachable, occasionally using transitional phrases in feedback like "however", "while", which mimic how a teacher might explain something in person. All these traits make LLM-B's feedback sound less like an authoritative decree and more like a constructive critique from a mentor invested in the student's growth. Importantly, LLM-B still gave a score and identified issues—it was not lenient in substance (indeed, LLM-B often assigned slightly lower scores than LLM-A for the same essay). The difference lies in delivery. The discursive practice here is one of *simulated teacher-student dialogue*: LLM-B anticipates the student's needs (e.g., providing rationale, maintaining encouragement) much as a human teacher would when trying to keep a student motivated while pointing out flaws.

It is illuminating to compare how the two LLMs handled the same feedback point in a specific case. In one essay, the student had included a sentence like "Let me know if you'd like any modifications!" at the end—obviously a leftover instructional text not meant to be in the final draft. LLM-A's feedback likely would simply instruct removal as a factual correction. LLM-B's actual comment was: *"The phrase 'Let me know if you'd like any modifications!' is clearly instructional text that should be removed."* While it is a directive ("should be removed"), notice LLM-B added context ("clearly instructional text") explaining why it doesn't belong. LLM-A might have said "Remove 'Let me know...' (not part of the essay)". Both convey the same correction, but LLM-B's phrasing gently teaches genre awareness (the student learns *why* it's inappropriate), whereas LLM-A's hypothetical phrasing would focus on the act of removal.

This micro-difference exemplifies the mentor vs. examiner vibe: the mentor teaches and guides, the examiner corrects and expects compliance. In terms of the taxonomy categories: LLM-A leaned more on prescriptive directives, diagnostic positioning, and metalinguistic explanations, and used affective engagement in a formulaic but present way; facilitative suggestions were less frequent for LLM-A. LLM-B, on the other hand, used more facilitative suggestions and affective language, with plenty of diagnostic positioning as well, but it balanced criticism with a coaching tone. Both used comparative benchmarking, but LLM-B did so to frame the student's work against objectives (mentor's strategy), whereas LLM-A's comparisons (through scores or ideal solutions) felt like strict standards being imposed (examiner's strategy).

It is noteworthy that these differences in discourse have practical implications. A student receiving LLM-A's feedback might perceive it as formal evaluation—something to "fix" to get a better score—essentially treating the AI feedback as they would a teacher's red pen comments. In contrast, a student receiving LLM-B's feedback might feel they are in a process of improvement and learning, since the feedback speaks to them more and explains more. Of course, these perceptions would also depend on the student's own approach and the classroom context (e.g., whether the AI feedback is mandatory, how the teacher frames it). But purely in textual terms, LLM-A *enacts* a more top-down pedagogical role, while LLM-B *enacts* a more collaborative one.

In summary, our comparative discourse analysis reveals that the design and communicative style of an AI feedback system can align with different educational roles. One system (LLM-A) behaved like an examiner—providing clear, criteria-focused, authoritative feedback—and another (LLM-B) behaved more like a mentor—providing supportive, explanatory, and choice-oriented feedback. These contrasting identities were not explicitly labelled as such by the systems but emerged through the language choices they made. The findings raise important questions: How do these differing styles affect student revision behavior and learning? Do students respond better to the friendly guide or the strict critic? And what are the implications for teachers who might incorporate such AI feedback into their practice? We turn to these questions in the Discussion, linking our findings to broader issues of student agency, teacher authority, and AI literacy.

## Discussion

The above findings demonstrate that AI-generated feedback is not a monolithic phenomenon; the discourse of feedback can vary widely depending on how an AI is programmed or prompted, leading to different pedagogical "personalities". In our case, LLM-A's examiner-like feedback and LLM-B's mentor-like feedback represent two ends of a spectrum. This diversity has significant implications for EFL students' agency in the writing process, the locus of pedagogical authority, and the preparation of teachers to use AI tools effectively.

*Implications for Student Agency*

The manner in which feedback is delivered can either constrain or empower student writers. LLM-A's prescriptive, authoritative style, while clear and actionable, arguably positions students as relatively passive recipients of expert corrections. The danger in such an approach is that students may focus on compliance—fixing each highlighted issue—without fully engaging with the rationale or developing self-regulation skills. If an AI always gives the correction, students might not learn to identify issues independently. In contrast, LLM-B's more suggestive and explanatory style invites students to make decisions (e.g., whether and how to incorporate a suggestion) and to understand the *why* behind needed changes. This has the potential to foster greater ownership of revisions and promote critical thinking about one's writing.

That said, it is important to recognize that not all students may initially know how to respond to facilitative feedback. Lower-proficiency or less confident EFL writers might actually prefer very explicit directives (as often seen with teacher feedback too). Thus, one practical implication is the need to cultivate *AI feedback literacy* among students: training them to interpret and use suggestions critically, and to treat AI feedback not as an infallible authority but as a helpful resource. Teachers could, for example, encourage students to always explain back in their own words why a certain AI-suggested change is needed, or even to occasionally "challenge" the AI by checking if an alternative phrasing might also work. By doing so, students practice a more active role, using the AI feedback as a springboard for learning rather than a mere to-do list.

*Pedagogical Authority and Role of the Teacher*

The contrasting identities of the two LLMs raise questions about *who* (or *what*) holds authority in the feedback process when AI tools are introduced. Traditionally, in EFL writing classrooms, the teacher is the primary authority who evaluates and gives feedback, and through this discourse, the teacher also constructs a supportive or critical persona (which students come to recognize). With AI in the mix, we effectively have a new actor producing teacher-like discourse. If an AI adopts an examiner persona (like LLM-A), it might carry a great deal of weight in the student's eyes—potentially rivalling the teacher's authority. For instance, if the AI gives a grade or uses very confident language ("This essay lacks X, you must do Y"), students might treat it as definitively "correct" feedback. This could undermine the teacher's role, especially if the teacher's own feedback or grading diverges from the AI's comments.

Conversely, if an AI positions itself more as a friendly coach (like LLM-B), it might be seen as a supplement to teacher guidance rather than a replacement for it. In Bearman et al.'s (2023) terms, the "discourse of altering authority" is clearly at play here. Our analysis concretely shows how AI feedback can *alter the locus of authority* by taking on a voice that students typically associate with teacher feedback. This means educators need to proactively manage the integration of AI feedback in their courses. For example, teachers might need to clarify to students that the AI's scores or comments are not official grades, but formative hints. Teachers might even choose which AI tool to use (or how to prompt it) based on the alignment with their desired feedback style: an instructor who values a nurturing, dialogic approach might avoid using a tool that behaves too much like an unforgiving examiner, and vice versa.

Another aspect is the consistency of messaging. In classroom practice, if both teacher and AI provide feedback, discourse alignment becomes important. Students could be confused if, say, the AI praises a certain aspect the teacher does not, or if the AI suggests a correction that the teacher disagrees with. This scenario is likely—AI feedback tools are not infallible and may offer advice that is stylistically or culturally misaligned with the teacher's expectations (for instance, recommending a very flowery style or using phrases the teacher finds inappropriate). The question of authority then arises: whose feedback should the student prioritize? This can affect the student's trust in the AI, in the teacher, or in the process of feedback in general. Pedagogically, one solution is for the teacher to mediate AI feedback. Teachers could discuss AI-generated comments in class, affirming useful points and clarifying or correcting others. In doing so, the teacher reasserts their role as the *human* authority who contextualizes AI input. It also models to students how to critically evaluate feedback sources, which is a valuable skill (tying into AI literacy below).

It's also worth noting that an AI's "persona" can influence classroom power dynamics. If students perceive the AI as an additional teacher (especially one that is always available), they might become less reliant on peer feedback or less inclined to seek the teacher's help. This could reshape classroom interactions. There are potential positives—e.g., quieter students might practice writing revisions with AI feedback and build confidence before interacting with peers or teachers—but also negatives if it leads to reduced human interaction or if the AI's authority goes unchecked. In essence, teachers might need to renegotiate their pedagogical identity when an AI mentor/examiner is introduced into the learning environment. Some teachers might choose to position themselves more as facilitators of students' engagement with AI feedback rather than the sole providers of feedback.

*AI Literacy in Teacher Education*

The findings of this study underscore the importance of incorporating AI literacy into teacher training, specifically focusing on critical understanding of AI-generated feedback. Teachers and future teachers need to be aware that AI tools can embody different discourse styles with various educational implications. As Oravec (2023) argues, educators should be prepared to critically evaluate AI-generated content for deficiencies and inaccuracies—we add that they should also evaluate it for its tone and pedagogical alignment. Teacher education programs can no longer ignore AI; rather, they should include modules where teachers-in-training *use* AI feedback tools, analyze the feedback given, and reflect on how it compares to human feedback best practices.

For instance, a professional development workshop might present participants with examples of LLM-A-like vs. LLM-B-like comments on the same student essay and prompt discussion:

Which is more effective, and why? In doing so, teachers can become more discerning about the AI tools they adopt. They may even learn how to *engineer prompts* to shape the AI's feedback style—for example, instructing the AI to give feedback in a friendly tone and ask questions could make an examiner-type AI more facilitative. Developing such skills would allow educators to better tailor AI tools to their pedagogical needs, rather than having to adapt their pedagogy to the tool.

Another dimension of AI literacy is teaching educators to guide students in *interpreting* AI feedback. This includes setting guidelines or scaffolds for students. For example, teachers might train students in a simple protocol: when you receive AI feedback, categorize each comment (is it a directive? a suggestion? a grammar fix? a content critique?), verify any corrections (since AI can be wrong), and prioritize the changes that impact content and organization before grammar. In essence, students need the meta-cognitive strategies to use AI feedback effectively. Teachers, in turn, need to be literate in how AI behaves to teach those strategies. Our taxonomy can serve as a starting point: teachers can explain to students that "the AI might do several things—it might praise you, point out errors, suggest changes, compare your work to what's expected, etc.—be aware of these, and remember that you remain the writer in charge of your essay." By demystifying the AI's discourse, we prevent students from treating AI feedback as an opaque oracle.

Lastly, AI literacy in teacher education also implies being cognizant of potential biases or cultural mismatches in AI feedback. Although our study did not find overt bias in the limited context of descriptive essays, one can imagine scenarios in which an AI's training data biases might affect feedback (for example, always favoring a certain essay structure that may not align with local rhetorical norms, or failing to appreciate culturally specific content). A critically literate teacher will be on the lookout for such issues. This connects with the social practice level of our CDA: teachers should ask not only "what is the AI telling the student?" but also "whose values or norms is this feedback reflecting?" and "are those appropriate for my context?" If, for example, an AI consistently encourages a very assertive argumentative style that conflicts with the local academic writing culture's preference for nuance, the teacher must intervene.

*Towards Pedagogically-Aligned AI Feedback Systems*

Our findings and discussion suggest a need for AI developers and educators to collaborate in creating AI feedback systems that are pedagogically aligned with educational goals. One size may not fit all—some contexts might benefit from an AI that is more examiner-like (for instance, in high-stakes test preparation where direct correction is valued), whereas others might prefer an AI that acts as a gentle tutor to support learning. Ideally, AI systems could be adjustable on this spectrum. This could involve user settings for "feedback tone" or "feedback strictness". Until such features exist, teachers might achieve some control through prompt engineering as mentioned.

Furthermore, incorporating CDA insights (like the taxonomy we proposed) into AI design can help ensure these tools do not inadvertently communicate in counterproductive ways (e.g., being too harsh or too vague). For example, an AI could be designed to always include an element of affective engagement (to avoid overly critical feedback that demotivates) or to phrase most advice as suggestions rather than commands (to encourage student agency), depending on the pedagogical philosophy.

It is also worth exploring how students themselves perceive the difference. While our analysis is text-centred, an important complementary study would be to gather student feedback on AI feedback: Do students notice the mentor vs. examiner differences? Which do they prefer or find more helpful? Their perceptions will ultimately influence the efficacy of the feedback. It might turn out that a combination is optimal—e.g., students want clear error corrections (examiner) but also encouragement and higher-order suggestions (mentor). This implies that teachers might use AI like LLM-A for certain tasks (grammar checking) and LLM-B for others (content development advice) or even use one AI with a blended approach. The presence of multiple AI feedback tools, each with a distinct discourse style, could itself be leveraged as a learning opportunity: students could compare feedback from both AIs on their essay to get a richer understanding (akin to getting a second opinion). However, managing potentially conflicting advice then becomes an educational task in critical thinking.

It is important to acknowledge the limitations of this study in interpreting the broader implications. Our analysis was based on a specific assignment type (descriptive essays) and two specific LLM models at a certain point in time. Different genres of writing (e.g., argumentative essays, research reports) or different AI systems (including future, more advanced versions) might exhibit different discourse patterns. Additionally, we did not directly measure student outcomes or preferences when using these AI feedback outputs—our discussion of agency and authority implications is inferred from theory and prior research, not from observing actual student behavior in this study. Further research involving student interaction and feedback uptake is needed to empirically validate how these discourse differences play out in practice. Moreover, cultural context plays a role: the notion of a "mentor" vs "examiner" style may be received differently in different educational cultures. In some East Asian contexts, for example, students might be accustomed to and comfortable with authoritative feedback, whereas in Western contexts a mentor style might be more expected (though this is a broad generalization). Educators should consider local expectations when interpreting our findings.

Building on our findings, several avenues for research and practice emerge. First, studies could examine hybrid feedback models where human teachers and AI both contribute: how to balance their voices so that they complement rather than clash? Second, longitudinal studies could track how students' writing and revision behavior evolves when consistently receiving one style of AI feedback vs another—does a mentor-style AI produce more autonomous writers over time? Does an examiner-style AI produce faster improvement in accuracy? Third, from a discourse perspective, analyzing student responses to AI feedback (in terms of revision changes or even replies to the AI if the system allows dialogue) would enrich the CDA, moving into interaction analysis. It would be interesting to see if students ever push back or negotiate with AI feedback (currently uncommon, but if AI chatbots are used, a student could ask the AI for clarification). This would mirror how real mentorship often involves dialogue. Finally, teacher education programs should document and research the impact of explicit AI literacy training. For instance, do teachers who are trained in recognizing AI discourse strategies make different choices about integrating AI in their classrooms than those who are not? Sharing best practices (such as setting norms for students, or fine-tuning prompts for desired feedback style) through case studies will be valuable.

**Conclusion**

This study conducted a critical discourse analysis of AI-generated feedback in an EFL writing context, revealing a six-part taxonomy of discourse moves and illustrating how different LLMs can enact contrasting pedagogical identities through language. The AI feedback we examined was far more than a simple list of corrections—it was a form of educational discourse that can praise, criticize, direct, suggest, explain, and benchmark, much like human feedback. By comparing two LLMs' feedback on the same student essays, we uncovered a striking divergence: one AI consistently behaved like a stringent examiner, while the other functioned more like a supportive mentor. These differences, we argue, have meaningful consequences. They can influence how students perceive and utilize feedback (with impacts on learner agency and engagement), and they effectively shift some of the pedagogical authority from human teachers to AI systems (raising questions about the teacher's evolving role).

Our findings extend CDA into the realm of AI in education, demonstrating that even automated feedback—often perceived as objective or mechanical—contains ideological and interactional subtleties worthy of analysis. The six categories (diagnostic positioning, prescriptive directives, facilitative suggestions, affective engagement, metalinguistic explanation, comparative benchmarking) provide a vocabulary for educators and researchers to discuss what AI feedback actually *does* in communicative terms. We have shown that by using this taxonomy, one can critically evaluate AI feedback tools: for instance, does a given tool lean heavily on prescriptive directives (maybe too heavily?), does it provide metalinguistic explanations to teach the student, does it engage affectively to keep the student motivated? Such questions move the conversation beyond "does AI feedback improve writing scores?" to "does AI feedback communicate in a pedagogically constructive manner?", which is an essential consideration for long-term learning and student well-being.

In practical terms, this research offers several takeaways. Educators looking to incorporate AI feedback should not treat all systems as equivalent—subtle differences in discourse can make one tool more suited to their context than another. If a teacher values student-centred learning, they might favour AI that gives facilitative suggestions and fosters a dialogue; if a context demands rapid error elimination, a more directive AI might be useful, but with caution to ensure students still learn from it. For developers of educational AI, our study underscores the importance of discursive design: the tone and style of feedback should align with sound pedagogical practice. It may even be beneficial to allow customization of that style. Importantly, regardless of the AI used, teachers should prepare students to critically engage with AI feedback. Rather than accepting it as gospel, students (with teacher guidance) should learn to ask: *Why is the AI suggesting this? Is it correct? How does it help my intent?* As Oravec (2023) emphasizes, fostering such critical evaluation is key to AI literacy.

This study focused on a specific scenario (LLM feedback on descriptive essays), and future research will reveal how generalizable the findings are. However, the core insight is likely robust: AI feedback carries an implicit voice, and that voice matters. As we stand at the intersection of TESOL and AI, we have the opportunity—and responsibility—to shape how these new "digital assessors" interact with learners. By applying a critical discourse lens, we can ensure that the integration of AI in writing education is done thoughtfully, upholding principles of good pedagogy and equity. The hope is that AI tools, used wisely, can augment the capacity of teachers and provide learners with more personalized support, without compromising the humanistic elements of education. Achieving that balance will require ongoing dialogue between educators, learners, and technologists, and a commitment to critically evaluating these tools not just for *what* they do, but *how* they do it.

## References

Alnemrat, A., Aldamen, H. A. K., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in Education*, *10*, Article 1614673. https://doi.org/10.3389/feduc.2025.1614673

Baker, T., & Hübner, S. (2024). Algorithmic governance in higher education: A critical discourse analysis of AI policy rhetoric. *Higher Education Quarterly*, *78*(1), 156–175. https://doi.org/10.1111/hequ.12458

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, *86*(2), 369–385. https://doi.org/10.1007/s10734-022-00937-2

Biesta, G. (2010). Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, *29*(5), 491–503. https://doi.org/10.1007/s11217-010-9191-x

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Fairclough, N. (1992). *Discourse and social change*. Polity Press.

Hyland, K., & Hyland, F. (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

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Knox, J. (2020). Artificial intelligence and education in China. *Learning, Media and Technology*, *45*(3), 298–311. https://doi.org/10.1080/17439884.2020.1754236

Luke, A. (2002). Beyond science and ideology critique: Developments in critical discourse analysis. *Annual Review of Applied Linguistics*, *22*, 96–110. https://doi.org/10.1017/S0267190502000053

Oravec, J. A. (2023). Artificial intelligence implications for academic cheating: Expanding the dimensions of responsible human-AI collaboration with ChatGPT and Bard. *Journal of Interactive Learning Research*, *34*(2), 213–237.

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and theories: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, Article 100416. https://doi.org/10.1016/j.edurev.2021.100416

Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, *11*(1), Article 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, *34*(1), 1–20. https://doi.org/10.1080/03054980701476386

Rogers, R., Malancharuvil-Berkes, E., Mosley, M., Hui, D., & O'Garro Joseph, G. (2005). Critical discourse analysis in education: A review of the literature. *Review of Educational Research*, *75*(3), 365–416. https://doi.org/10.3102/00346543075003365

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.

Selwyn, N., & Facer, K. (2022). *The politics of education and technology: Conflicts, controversies, and connections*. Palgrave Macmillan.

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*(2), 187–209. https://doi.org/10.1017/S0958344023000265

Tran, T. T. T. (2025). Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences. *Education Sciences*, *15*(2), Article 232. https://doi.org/10.3390/educsci15020232

van Dijk, T. A. (2014). *Discourse and knowledge: A sociocognitive approach*. Cambridge University Press.

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, *62*, Article 101071. https://doi.org/10.1016/j.jslw.2023.101071

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

Yang, H., Gao, C., & Shen, H. (2024). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies*, *29*, 3837–3858. https://doi.org/10.1007/s10639-023-11991-3

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, *51*, Article 100586. https://doi.org/10.1016/j.asw.2021.100586