

Tone, engagement, and positioning in EFL writing feedback: Human touch versus AI responses

Asma Alshehri

King Saud University (Saudi Arabia)
asmalyahya@ksu.edu.sa

Abstract

Feedback plays an essential role in English as a Foreign Language (EFL) education, particularly in academic writing. As artificial intelligence (AI) language models like ChatGPT are increasingly being explored for potential integration into EFL education, it is important to explore the extent AI-generated feedback can mimic feedback produced by teachers, especially in terms of tone, engagement and learner positioning. The present qualitative study draws on the Appraisal Theory by Martin and White (2005) to analyze and compare the linguistic characteristics of feedback provided by EFL teachers and ChatGPT on student academic essays. More specifically, the research explores authorial attitude (affect, judgment, appreciation), and engagement (dialogic openness and reader positioning). The findings reveal that teacher feedback is richer in digressions, comments and the variety of engagement features, resulting in more interactive and engaging feedback. The study contributes to discussions on the effectiveness of AI-generated feedback, offering suggestions for pedagogical implications for integrating ChatGPT into EFL writing instruction.

Keywords: AI in education, EFL academic writing, appraisal theory, evaluative language, feedback.

Resumen

Tono, compromiso y posicionamiento en la retroalimentación de la escritura en inglés como lengua extranjera: respuestas humanas frente a respuestas generadas por IA

La retroalimentación desempeña un papel esencial en la enseñanza del inglés como lengua extranjera (EFL por sus siglas en inglés), especialmente en la escritura académica. A medida que los modelos lingüísticos de inteligencia

artificial (IA) como ChatGPT se tienen en cuenta para su posible integración en la enseñanza de EFL, resulta importante analizar hasta qué punto la retroalimentación generada por la IA puede imitar la producida por el profesorado, en particular en lo que respecta al tono, el compromiso y el posicionamiento del estudiante. El presente estudio cualitativo se basa en la teoría de la valoración de Martin y White (2005) para analizar y comparar las características lingüísticas de la retroalimentación proporcionada por docentes de EFL y por ChatGPT sobre ensayos académicos de estudiantes. Más concretamente, la investigación examina la actitud tutorial —afecto, juicio y apreciación— y el compromiso—expansión dialógica y posicionamiento del lector—. Los resultados revelan que la retroalimentación del profesorado es más rica en digresiones, comentarios y en la variedad de recursos de compromiso, lo que da lugar a una retroalimentación más interactiva y atractiva. El estudio contribuye al debate sobre la eficacia de la retroalimentación generada por la IA y ofrece propuestas con implicaciones pedagógicas para la integración de ChatGPT en la enseñanza de la escritura en EFL.

Palabras clave: Inteligencia artificial en la educación, escritura académica en EFL, teoría de la valoración, lenguaje evaluativo, retroalimentación.

1. Introduction

The role of feedback has been widely discussed in the literature, especially in the context of EFL academic writing. Effective feedback is essential in EFL writing pedagogy, as it directly impacts students' abilities to improve their writing skills, language proficiency, and self-efficacy (Barrot, 2021; Cui et al., 2021; Hwang, 2025; Hyland, 2006; Johannes & Haase, 2024; Yu & Schunn, 2023). Effective feedback includes information that learners can understand and use, and if done well, can be a very powerful learning tool (Carless & Boud, 2018). According to Hyland and Jiang (2016), effective feedback impacts learner engagement. They proposed three dimensions for learner engagement: (1) cognitive, (2) affective, and (3) behavioral. Relevant to the present study is the affective dimension, which refers to the students' internal states, attitudinal responses, emotional reactions to the feedback they receive. Furthermore, Brookhart (2017) argued that the emotional tone in which feedback is delivered is essential to how learners perceive and use it. Tone refers to the linguistics style and word choices of the feedback message, and how the message will be heard by the receiver. This is especially important in EFL educational contexts, where students may already be anxious about writing assessment.

Recent research highlights the importance of socioemotional factors to the discussion of feedback in writing pedagogy. Carless (2022) noted that with the absence of vocal and facial cues in written feedback, learners can misinterpret it and respond negatively. Learners tend to engage more with positively evaluated feedback as negatively phrased comments may lower self-esteem and perceived self-efficacy. Hwang (2025) investigated the impact of feedback mechanisms on how learners perceive it and on their self-regulation strategies. The findings indicated that caring teacher feedback had positive impact on student engagement. Ruegg (2018) argued that teacher feedback can encourage or discourage positive learner emotions depending on how it is framed. Moreover, Sun (2021) comprehensively reviewed existing literature to examine the effects of teacher interpersonal and linguistics choices on student engagement in EFL classrooms. The research aimed to investigate how teachers care and praise, influence learner behavioral, emotional, and cognitive engagement. Sun (2021) analyzed research on learner engagement, positive psychology, and self-determination theory in EFL contexts. The key findings indicated that feedback which reflects authentic teacher care and constructive praise can predict student engagement. Derham et al. (2022) explored the relationships between grades and the linguistics features of feedback. Their findings revealed that higher grades correlated with positively toned feedback, while lower grades correlated with interrogatives. The study also found the differences in feedback tone explain 44% of the variance in grades. The authors emphasized that teachers should provide balanced feedback, focusing on meaningful guidance and positive framing. In addition, Henderson et al. (2021) encouraged educators to tailor their feedback practices to cater for different student needs and emphasized that feedback should not be one-size-fits-all. However, providing individualized, positively toned feedback to EFL learners remains a universal major task for language teachers (Brookhart, 2017; Mahfoodh, 2017).

Recent developments and integration of Large Language Models (LLMs), such as ChatGPT opened new avenues for language educators and scientific investigation in writing pedagogy. While the adoption of AI tools in EFL contexts remains largely experimental and varies significantly across institutions, emerging research has begun exploring their effectiveness in providing feedback (Dai et al., 2024; Escalante et al., 2023; Guangyuan & Zhaoxia, 2025; Hwang, 2025; Rad et al., 2024; Sari & Han, 2024). Other studies compared AI feedback to traditional teacher feedback (Awidi, 2024;

Çöp, 2025; Cui et al., 2021; Lin & Crosthwaite, 2024; Steiss et al., 2024; Zou et al., 2025). For example, Lin and Crosthwaite (2024) compared written corrective feedback by human teachers to feedback by ChatGPT in a second language context. Their comparison revealed that teachers' feedback addressed both local and global issues with relative consistency. However, ChatGPT showed significant variability and redundancy with a focus on local issues only. Also, ChatGPT failed to provide tailored responses to learners' proficiency levels. The authors recommended that ChatGPT is used as complementary tool to support teacher-led feedback and practices. Another study highlighting the complementary potential of ChatGPT to teacher feedback was conducted by Zou et al. (2025) who found that the majority of the students preferred teacher feedback due to its personalized and clear nature. Unlike technical feedback, teacher empathy, praise and caring express encouragement and motivate learners and deeply impact their engagement and learning behaviors. In addition, Awidi (2024) investigated the effectiveness of ChatGPT in providing personalized feedback on reflective essays in comparison to teachers. The findings indicated that ChatGPT feedback is particularly useful when integrated into language learning education with the oversight of human evaluators. Based on these studies, it can be argued that while ChatGPT shows promise, the strengths of human evaluators in providing personalized learner feedback, remain unique. One possible reason is the absence of a reader built-in model that can inherently understand the audience readers. Jiang and Hyland (2025) argued that ChatGPT lacks the intrinsic awareness of who reads its output, and hence, fails to adjust its writing accordingly. This shortcoming is due to the fact that ChatGPT is trained on huge amounts of texts from various genres and registers with different audiences and purposes. This means that specific readership is not recognized and ChatGPT operates with a generic target reader (Milano, 2023). ChatGPT generates coherent texts based on patterns it learned from its training data, but it does not understand context as humans do (OpenAI, 2023). Consequently, ChatGPT cannot adapt its tone and rhetorical style to suit different types of readers (Jiang & Hyland, 2025; Sardinha, 2024; Steiss et al., 2024; Su et al., 2023). The literature suggests that texts generated by ChatGPT are dialogically closed and lack audience awareness, showing reduced involvement compared to human generated texts.

Although the literature on AI generated texts is rich, the interpersonal elements of feedback generated by AI compared to that generated by their

human counterparts remains scarce and unclear. Teacher feedback does not only evaluate performance but also communicates appreciation and encouragement. Many studies found that supportive, empathetic, and engaging teacher feedback increases positive learner engagement and strengthens teacher-student rapport (Awidi, 2024; Brookhart, 2017; Carless, 2022; Henderson et al., 2021; Sun, 2021; Zou et al., 2025). Thus, understanding the linguistic qualities and emotional framing of AI generated feedback is vital for assessing its suitability as an alternative or supplement to teacher feedback. To this end, the present study draws on the Appraisal Theory Framework by (Martin and White, 2005) to analyze and compare the evaluative language resources employed by both human teachers and ChatGPT when providing feedback to EFL learners. By exploring the appraisal systems of (1) attitude (affect, judgment, appreciation), and (2) engagement (dialogic openness and reader positioning), this research sets to explore the extent to which ChatGPT can generate feedback with the same frequency and function of individualized, supportive, and empathetic tone, characteristics that are typically associated with teacher feedback. The comparison aims to provide insights into the integration of AI tools that can leverage the strengths and at the same time address the limitations of AI in EFL writing feedback practices. Therefore, the current study aims to address this gap by answering the following research questions:

1. How do teachers and ChatGPT use linguistic resources of affect, judgment, and appreciation when constructing written feedback?
2. How do teachers and ChatGPT position and engage with EFL learners in their feedback?

2. Methodology

This study adopts a qualitative comparative analysis approach using Appraisal Theory by Martin and White (2005) to scrutinize the linguistic and evaluative aspects of teacher and ChatGPT generated feedback. Appraisal theory is concerned with how writers voice their opinions, take a stance, and position their readers. Typically, in written feedback, teachers engage with their learners based on their understanding and knowledge of their learners including implied attitudes and responses. Thus, the teacher's knowledge of the audience (i.e., learner) shapes feedback and defines the rhetorical context. Martin and White (2005) argue that appraisal theory goes beyond

exploring the writer’s values and attitudes to highlighting the relations of rapport and authority between the producer of the text (i.e., teachers and ChatGPT) and the receivers (i.e., EFL learners). Appraisal theory has three subsystems as indicated in Figure 1.

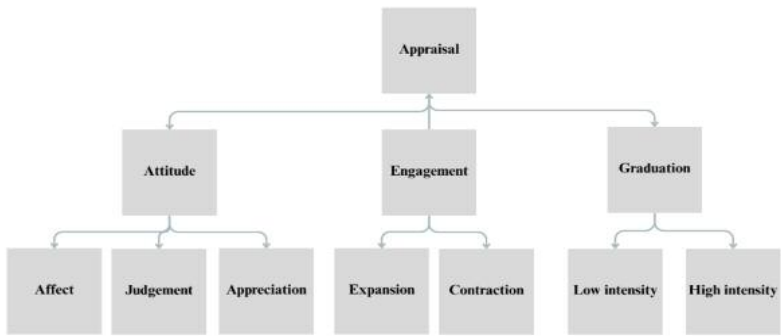


Figure 1. Appraisal Theory Subsystems.

Appraisal’s three subsystems provide writers with the means to be critical, to value, to reject, to accept, and to challenge other arguments. The first subsystem is ‘attitude’ that attends to the resources used by writers/speakers to negatively or positively evaluate phenomena. Attitude is divided into three semantic categories: affect, judgement and appreciation. Affect is how writers express their feelings (e.g., feeling happy, interested, sad, concerned, or bored), judgment deals with moral judgments of human behavior (e.g., admire, criticize, praise, or condemn) and appreciation attends to the assessment of objects but not people behaviors. Second, ‘engagement’ attends to the degree of expansion or contraction of the dialogic space. Engagement refers to the linguistics resources that writers use to take a stance towards the positions and values in the text and with consideration of their readers. Third, ‘graduation; deals with the values by which writers decrease or increase the interpersonal impact of their propositions. In other words, it is concerned with grading (i.e., toning-down or toning-up) feelings, judgements and assessments. In the context of the present study, the two subsystems of appraisal theory will be used to analyze how language used in feedback conveys affect (the evaluator’s emotional tone), judgment (evaluative positioning of the learner), and appreciation (how the text’s quality is evaluated). For example:

- (1) I enjoyed reading your ideas and descriptions. (evaluator's positive tone)
- (2) You need to improve and stop thinking in Arabic. (Judgment of learner's behavior)
- (3) This is a well-structured paragraph. (appreciation of the learner's work)

3. Data Analysis

The data included two sets of feedback on twenty academic texts written by EFL intermediate level first-year university students on the topic of advertisement. For student privacy, all texts were anonymized before the teachers evaluated them. Then, each text was evaluated by thirty experienced teachers and ChatGPT (version used: GPT-4). ChatGPT was prompted to use the sandwich approach in giving feedback (positive, constrictive, positive) and the same rubric the teachers used. Teacher evaluators had three days to provide feedback and it required approximately 25–30 min of total rater time per essay. The wordcount for teacher feedback was 14923 and 5070 for ChatGPT feedback. To normalize the data, raw counts were converted to percentages for both datasets. Appraisal occurrences were calculated separately in each set and frequencies were adjusted based on the totals in each set.

For coding reliability, the data were coded twice. First, all feedback texts were imported into the Appraisal Theory Tagger Tool for coding. The Appraisal Theory Tagger is free and automatically annotates texts based on the three subsystems of appraisal (Alshehri et al., 2025). Second, the researcher recoded a stratified sample of 80% of the clauses (balanced across ChatGPT and teacher feedback and attitude vs. engagement instances) to guarantee consistency. The tagger's labels were visible during the recoding but were treated as provisional. Therefore, whenever the researcher's interpretation differed, Martin and White's (2005) criteria was followed and the original label was overridden. In addition, borderline and multi-functional cases (e.g., utterances that could be read as both predictive and epistemic, or as simultaneously countering and entertaining a proposition) were coded for the dominant interpersonal function in their local text, while secondary possibilities were not counted. This procedure entails interpretive judgement, and it is acknowledged as a limitation of the present analysis.

4. Findings

Drawing on the framework of appraisal theory, language preferences differed with attitude formulations more frequent in EFL teachers’ feedback than in ChatGPT-generated feedback. Below, the results are discussed in more detail.

4.1. Attitude in Teacher and ChatGPT Generated Feedback

The first research objective was to explore the attitudinal expressions employed in teacher feedback compared to ChatGPT-generated feedback. In both sets of feedback, there was a preference to evaluate the writing process rather than judge the learners or express affect, as presented in Table 1.

	Affect		Judgement		Appreciation		Total	
	No	%	No	%	No	%	No	%
Teacher Feedback	51	7%	27	3%	706	90%	784	100
ChatGPT Feedback	0	0%	11	6%	162	94%	173	

Table 1. Attitude in Teacher and ChatGPT Feedback.

First, in relation to the subsystem of affect, teachers expressed their feelings in their feedback, while in the ChatGPT’s feedback, there were no instances of affect. Examples of affect in the teachers’ feedback are demonstrated below:

- (4) I admire [Attitude; affect; positive] your insightful observations.
- (5) I appreciate [Attitude; affect; positive] your choice of vocabulary.
- (6) I really like [Attitude; affect; positive] the ...
- (7) I was impressed [Attitude; affect; positive] with how much your ideas....
- (8) I look forward [Attitude; affect; positive] to your next essay.

Teachers expressed their feelings toward the learners’ ideas, writing and vocabulary positively and there were not instances of negative affect towards the EFL learner or the writing process. Second, the analysis showed a low frequency of judgment which accounts for 3% in the teachers’ feedback and 6% in ChatGPT generated feedback as presented in Table 1. Teachers and ChatGPT made no negative judgements of the EFL learners, but

constructed positive judgments as describing them to be good *readers*, and *professionals* as presented in the examples below:

Teacher feedback:

- (9) You are a good reader [Attitude; judgement; Positive (admire)] and that you pay attention to your use of vocabulary.
- (10) You have started to build ideas like a professional [Attitude; judgement; Positive (admire)].
- (11) You have done well [Attitude; judgement; Positive (admire)]
- (12) You continue to show signs of an experienced writer [Attitude; judgement; Positive (admire)].

ChatGPT feedback:

- (13) The impact of your arguments, demonstrating a good grasp of the topic [Attitude; judgement; Positive (admire)].
- (14) You've successfully outlined [Attitude; judgement; Positive (admire)] several key elements of effective advertising.

Third, appreciation was clearly the predominant attitudinal subsystem in both sets of feedback. It accounted for 90% in the teachers' feedback and for 94% in ChatGPT's feedback, as presented in Table 1. According to Martin and White (2005), appreciation helps the writer objectify his/her involvement in the text. Therefore, when teachers and ChatGPT encoded values of appreciation in their feedback, they were depersonalizing and objectifying their evaluations. Appreciation like affect and judgement has positive and negative evaluations. In the present data, no negative instances of affect or judgment were found in the feedback. However, there were instances of negative appreciation in both sets of feedback, as illustrated in Table 2.

	Positive Attitude		Negative Attitude		Total	
	No	%	No	%	No	%
Teacher Feedback	766	97%	18	3%	784	100
ChatGPT Feedback	162	94%	11	6%	173	

Table 2. Positive and Negative Attitude in Teacher and ChatGPT Feedback.

Positively encoded appreciation values accounted for 97% in the teachers' feedback and for 94% in ChatGPT's feedback, while negatively encoded appreciation accounted for 3% in the teachers' feedback and for 6% in ChatGPT's feedback. The following extracts provide examples of positive and negative appreciation:

Teacher feedback:

- (15) The organization of the passage is logical [Attitude; Appreciation; Positive].
- (16) Many grammatical errors and lack of organization and cohesion made the essay difficult [Attitude; Appreciation; Negative] to read and understand.
- (17) It appears that you have tried to first write your text in Arabic and then translate it into English which accounts for wrong [Attitude; Appreciation; Negative] word choices, organizational and punctuation mistake
- (18) The poor organization [Attitude; Appreciation; Negative] of the task makes it difficult to understand.

ChatGPT feedback:

- (19) The text has a clear [Attitude; Appreciation; Positive] introduction.
- (20) Some sentences are quite long and complex [Attitude; Appreciation; Negative], which may make it difficult for readers to follow.
- (21) Your phrases are awkwardly [Attitude; Appreciation; Negative] constructed.

The high frequency of appreciation resources and the low frequency of judgement formulations in both sets of feedback reflects an objective voice. According to Martin and White (2005) this frequent use of appreciation allowed writers to depersonalize their evaluations and objectify their involvement in the feedback. There were, however, no affect formulations in ChatGPT generated feedback, and this can be due to the design of the AI model, which prioritizes maintaining a formal and objective tone in academic writing feedback (Jiang & Hyland, 2025; OpenAI, 2023; Sardinha, 2024).

4.2. Learner Positioning in Teacher and ChatGPT Generated Feedback

The second research objective was to explore how the authorial voice in both sets of feedback engaged with the learners and how the EFL learners were viewed and positioned in the dialogue. Thus, the framework of engagement was drawn upon to provide a systematic account of how writers linguistically voice their values and how they position their readers. In the present study, engagement refers to the linguistics choices that the language teachers and ChatGPT employed to construct meaningful feedback and to engage with the EFL learners. Martin and White (2005) differentiated between two types of engagement, contraction and expansion. When the dialogue is open and allows for alternative voices and positions, it is considered expansive, but when it restricts or challenges alternative positions, it is dialogically contractive. The dialogic openness in both sets of feedback is presented in Table 3.

	Contractive		Expansive		Total	
	No	%	No	%	No	%
Teacher Feedback	281	58%	204	42%	485	100
ChatGPT Feedback	42	26%	121	74%	163	

Table 3. Contractive vs. Expansive Resources in Teacher and ChatGPT Feedback.

Teachers were more contractive 58% than expansive 42% in their feedback to EFL learners. This means that teachers tended to contract the dialogic space and preferred to restrict the scope of the learners' alternative point of view. However, feedback generated by ChatGPT was dialogically more open (74%) and allowed for alternative positions. Below, the results of contraction, expansion, and learner alignment will be discussed in more details.

4.2.1. Contracting the Dialogical Space

As explained by Martin and White (2005), the dialogical space between the writer and the readers can be contracted (i.e., closed) by employing the resources of disclaim and proclaim. Disclaim refers to meanings that reject dialogic alternatives or consider them as not applying, while proclaim, involves explicit authorial interventions that act to challenge and confront

dialogic alternatives in the ongoing discussion. Teachers and ChatGPT employed more resources of disclaim than proclaim in their feedback, as illustrated in Table 4.

Engagement resources		Teacher	ChatGPT
Contract	Disclaim Deny	8%	0%
	Counter	41%	18%
	Concur	1%	0%
	Proclaim Pronounce	3%	1%
	Endorse	5%	7%
Expand	Entertain	42%	72%
	Attribute Acknowledge	0%	2%
	Distance	0%	0%
Total		100%	100%

Table 4. Engagement Resources in Teacher and ChatGPT Feedback.

This preference suggests that both teachers and ChatGPT avoided explicit authorial interventions in their feedback that limited the scope of alternative propositions. Instead, they negated or replaced alternative positions by ones they perceived as more appropriate. To do so, teachers employed resources of deny (negating a proposition) and counter (countering a proposition), while ChatGPT only employed counter formulations. Counter is conveyed through concessive conjunctions and connectives such as *however, but, despite, although*; through attitudinal expressions such as *unfortunately and fortunately* and through continuatives such as *only, still, even, yet*. Examples of counter and deny from the data follow:

Teacher feedback:

- (22) There is no [Engagement; contract; disclaim; deny] need for the whole number of words.
- (23) It will not [Engagement; contract; disclaim; deny] impede the meaning.
- (24) Writing about only [Engagement; contract; disclaim; counter] 3 appears to show a lack of research in your ideas.
- (25) Well written task despite [Engagement; contract; disclaim; counter] the simple short sentences.

ChatGPT feedback:

- (26) The word “creativity” is used multiple times in the same sentence and paragraph. Instead [Engagement; contract; disclaim; counter], try synonyms or rephrasing to avoid repetition.
- (27) By refining your sentence structure, grammar, and transitions, your writing can communicate even [Engagement; contract; disclaim; counter] more effectively.
- (28) The comma after “...” in the first sentence should be a period or a semicolon instead [Engagement; contract; disclaim; counter].

The low frequency of proclaim resources, such as *concur*, *pronounce*, and *endorse* in both sets of feedback suggests that the writer is not in complete agreement with the learner’s writing nor considers it as valid. For instance, when the authorial voice (i.e., teacher and ChatGPT) endorses the proposition, it is in complete alignment with it. Examples 29 and 30 show how the authorial voice endorses certain aspects of the learner’s writing:

Teacher feedback:

- (29) Your writing demonstrates [Engagement; contract; proclaim; endorse] that you have understood and addressed the expectations of the task.

ChatGPT feedback:

- (30) This demonstrates [Engagement; contract; proclaim; endorse] a good understanding

4.2.2. Expanding the Dialogical Space

As mentioned earlier, feedback generated by ChatGPT was perceived as more expansive 74% than contractive 26%. This is due to the high frequency of entertain resources in the feedback 72%. Entertain is the first major subtype of dialogic expansion and it refers to the propositions presented by the authorial voice as one possibility among others, hence, it entertains other possible alternatives and opens the dialogic space with the learners. Entertain was also the most frequent subtype in the teachers’ feedback which accounted for 42%. Entertain is achieved via modal auxiliaries such as *could*, *can*, *will*, *might*, *may*, *etc.*; via modal attributes such as *it is likely*, *it is possible*, *etc.*; via modal adjuncts such as *probably*, *maybe*, *perhaps*, *etc.*; and mental verb/attribute projections such as *I think*, *I believe*, *I doubt*, *etc.* Consider the following examples from the data:

Teacher feedback:

(31) Your essay will [Engagement; expand; entertain] be fantastic.

ChatGPT feedback:

(32) Some sentences are overly long and could [Engagement; expand; entertain] be split for better clarity.

The remaining subtypes of dialogistic expansions are acknowledge and distance. Acknowledge refers to meanings where some proposition is attributed to an external voice while disassociated from the authorial voice. Distance refers to the authors attempts to distance him/herself from the material attributed. In the present data, formulations of acknowledge accounted for 2% in the ChatGPT's feedback, and there were no instances of distance in both sets of feedback. Therefore, in both sets of feedback the authorial voice did not show detachment or disassociation with the learners.

4.2.3. Alignment with EFL Learners

From the perspective of appraisal, writers do not only express their attitudes and positions but also use signals that indicate how they align or disalign with their addressees or putative readers. Writers come into the text with imagined or envisaged readers that are considered either in agreement, disagreement, or neutral. Therefore, engagement formulations are used to signal both the position of the authorial voice towards the propositions being discussed in the colloquy, and at the same time, how the putative reader is positioned in the text. According to the appraisal theory framework, when writers employ resources of concur, endorse and counter, they assume a relationship of alignment with their readers. However, when the use resources of deny, pronounce and distance, they show a relationship of disalignment and assume their readers do not agree with them. This is evident in the frequent authorial interventions, rejections and distancing resources. Furthermore, when writers entertain other possible alternatives and open the dialogic space by constructing meaning through the resources of entertain, they are assuming that their readers have split points of view and are divided. Finally, when authors acknowledge information and disassociate themselves from it, they are showing no expectations of their readers. The relations of alignment, disalignment and solidarity in the feedback are presented in Table 5.

Resources		Teacher Feedback	ChatGPT Feedback
Aligned	Concur	1%	0%
	Endorse	5%	7%
	Counter	41%	18%
	Total	47%	25%
Not Aligned	Deny	8%	0%
	Pronounce	3%	1%
	Distance	0%	0%
	Total	11%	1%
Divided	Entertain	42%	72%
Neutral	Acknowledge	0%	2%

Table 5. The Authorial voice – Putative Reader Relationship in Teacher and ChatGPT Feedback.

The EFL teachers entered the discourse assuming that the majority of the EFL learners are either in alignment or divided. They positioned 47% of their students as aligned and in agreement with their feedback, while 42% as divided and have mixed points of view; some may agree with the feedback and others may not. The teachers also assumed that 11% of the learners were not aligned nor in agreement with the teacher's feedback. The absence of acknowledge resources indicate that the EFL teachers knew their learners and had clear expectations of solidarity with them. Thus, the learners were not positioned as neutral readers in the teacher feedback. On the other hand, in the feedback generated by ChatGPT, the majority of the learners (72%) were viewed as having alternative points of views about the quality of their own writing, and therefore, the authorial voice employed formulations of entertain to maintain a relationship of solidarity with the learners. Also, 25% of the learners were envisaged as in complete agreement with the generated feedback. In the ChatGPT generated feedback there were rare instances of authorial explicit intervention and rejections, hence the low frequencies of pronounce, deny, distance and acknowledge. Therefore, the authorial voice did not remain aloof from any expectations of solidarity with the learners, instead, it had clear expectations of the learners' alignment.

5. Discussion

Given the importance of feedback in EFL writing instruction and its impact on learners' engagement, self-efficacy, and self-esteem, it is considered a key

component of language development (Barrot, 2021; Brookhart, 2017; Cui et al., 2021; Hwang, 2025; Hyland, 2006; Johannes & Haase, 2024; Yu & Schunn, 2023). At the same time, providing individualized feedback demands substantial time and effort from EFL instructors (Henderson et al., 2021; Mahfoodh, 2017). In this context, the gradual introduction and integration of AI tools such as ChatGPT in higher-education EFL writing courses opens new possibilities for language instructors and scientific investigation. Accordingly, the present study aimed to compare the linguistic characteristics and rhetorical style of feedback provided by EFL teachers and ChatGPT on student academic essays. Drawing on the framework of appraisal theory by Martin and White (2005), the findings revealed differences in the use of attitudinal expressions, engagement preferences, and learner positioning.

With regard to the attitudinal expressions employed in teacher feedback compared to ChatGPT-generated feedback. The findings revealed that in relation to the subsystem of affect, teachers expressed their feelings towards the students writing, while in the ChatGPT-generated feedback, there were no instances of affect or personal emotional expressions. The analysis also showed that in both sets of feedback, there were no negative judgments of the learner. Additionally, in both sets of feedback, the authorial involvement was objectified and depersonalized through the high frequency of appreciation resources in the data. Thus, the high frequency of appreciation resources and the low frequency of judgement and affect formulations in both sets of feedback reflects an objective authorial voice. The findings also indicated that the polarity of the attitudinal expressions in the feedback was significantly positive in both sets of feedback, reflecting a positive tone. The key difference lies in the teacher's willingness to express personal feelings (e.g., "I admire...", "I look forward to your next essay"), contrasted with ChatGPT's lack of explicit affective self-disclosure. This finding aligns with recent research by Jiang and Hyland (2025) who observed that AI-generated language shows reduced involvement compared to human language. Similarly, Sardinha (2024) found that AI tools maintain an objective formal tone and exhibit reduced interpersonal involvement in academic writing.

However, the absence of affect in ChatGPT's feedback should be interpreted with caution. ChatGPT output is highly sensitive to prompt design, and in the present study ChatGPT was instructed to produce feedback on decontextualized essays, without any learner profiles. Under such conditions, it is expected that the feedback would reflect a generic,

polite teacher persona focused on evaluations of text quality rather than on individualized emotional engagement with the learner. Therefore, the findings of the study do not claim that ChatGPT is incapable of producing affective language or learner-sensitive responses. They, rather show that ChatGPT did not spontaneously adopt an affect-rich tone in the absence of explicit prompts or learner history.

Turning to engagement strategies and how learners are positioned within both teacher feedback and ChatGPT-generated feedback. Teacher feedback was slightly more contractive than expansive, while ChatGPT generated feedback employed more expansive engagement markers, particularly entertain formulations. This indicates that the dialogic space was more open in ChatGPT generated feedback compared to the teachers' feedback. Additionally, the engagement resources employed by the teachers positioned the learners as in alignment with the teacher's feedback. On the other hand, in feedback generated by ChatGPT, the majority of the learners were viewed as having divided positions and points of views about the quality of their own writing. The findings align with the those reported in the literature, such as Markey et al. (2024) who found that ChatGPT was unable to model human audience awareness and envisage reader needs. Jiang and Hyland (2025) argued that AI tools rely on statistical patterns and fail to understand the readers so they produce grammatically coherent texts that are not interactive. Sardinha (2024) attributed the limitations of ChatGPT to its underlying algorithms and the language data used to train language models. Similarly, the findings of this study indicate that despite the impressive feedback generated by ChatGPT, it still cannot model human teachers' digressions or comments. Such shortcomings can affect teacher- learner rapport building, and feedback efficacy. This highlights the importance of human input in constructing dialogic, engaging and learner- aware feedback.

It is important to distinguish between two capacities. First, LLMs can mimic an affective tone when they are explicitly prompted or provided with learner history. Second, developing a theory-of-mind-like model of a specific learner's beliefs, emotions and reactions is a more demanding capacity. The present study sheds light on the second capacity. Given the constraints of the prompt and context, ChatGPT did not demonstrate a sustained audience model comparable to the knowledge teachers have of their students.

Finally, teachers expressed more affect than AI and produced more elaborations, digressions and engagement strategies, whereas ChatGPT

tended to compress its feedback into rubric-oriented comments. This explains the length differences between the two sets of feedback and is in line with emerging studies reporting that AI feedback is often shorter and more condensed than human feedback (e.g., Guangyuan & Zhaoxia, 2025). Considered together, these findings suggest that ChatGPT can approximate certain aspects of human feedback, such as positive appreciation and dialogic openness. Nevertheless, it does so through a generic, text-oriented persona that lacks the interpersonal richness and audience awareness characteristic of teacher feedback. This has direct implications for how AI feedback should be positioned in EFL writing pedagogy.

6. Conclusion

The present study provides important qualitative evidence of the interpersonal and engagement differences between human and ChatGPT-generated feedback on EFL academic writing. By applying appraisal theory to both datasets, it shows that while ChatGPT can generate grammatically coherent and dialogically open feedback, teachers' feedback remains richer in affect, digressions and audience-aware engagement strategies. Although such findings are essential to language education, one limitation is the study's focus on interpersonal meaning and authorial voice under a single prompt configuration, which might be anticipated as one of AI's shortcomings and does not capture the full range of possible prompting scenarios.

Pedagogically, the study underscores the importance of authorial tone, and engagement strategies in constructing effective feedback that supports learner motivation, rapport and agency. While the findings showed that ChatGPT shows potential in generating meaningful feedback, the EFL teachers exhibited more ability in employing affective feedback. Teachers expressed empathy, encouragement, and emotional support to motivate learners and promote student agency. In the present study, although ChatGPT did not reproduce the personal and affective dimensions of teacher feedback, it still provided accurate comments on grammar, mechanics, and organization. In EFL higher-education, this suggests a division of labor in which AI tools are primarily used for lower-level, form-focused feedback, while teachers concentrate on higher-order aspects (e.g., content, argumentation, disciplinary conventions) and on individualized

feedback grounded in their knowledge of students (Brookhart, 2017; Carless & Boud, 2018; Hwang, 2025; Sun, 2021).

AI feedback can become both a scaffold and an object of metalinguistic inquiry, rather than a substitute for teacher judgement. Students can be guided on how to use ChatGPT critically and responsibly. For example, students can use AI-generated feedback to identify surface-level issues or explore alternative formulations, and then compare it with teacher feedback to reflect on rhetorical style, stance and engagement strategies. In this way, students can utilize ChatGPT in their writing and become more autonomous learners. In addition, this can potentially allow teachers to focus on evaluating the subjective aspects of writing, such as creativity and originality.

Finally, inquiry in AI language models, such as ChatGPT continue to prove the limitations of AI generated content. Although, one cannot deny the impressive abilities of AI tools, caution, balance and human oversight in the use of AI feedback is necessary. The study shows that, under a standard feedback prompt and without prior learner information, ChatGPT defaults to a generic persona. Future research can explore how individualized and personalized prompting might influence ChatGPT's use of affect, engagement and alignment, and whether these adjustments can bring AI feedback closer to the interpersonal complexity of human feedback.

In interpreting these findings, four limitations should be acknowledged. First, the study focused on interpersonal meaning and authorial voice under a single prompt configuration. ChatGPT was instructed to provide sandwich feedback on decontextualized essays, with no access to learner profiles or prior interaction history. Therefore, the observed absence of affect and generic alignment patterns may reflect the specific prompt and context rather than an inherent inability of LLMs to produce affective or learner-sensitive feedback. Second, it should be noted that while this study explores the potential of AI-generated feedback, the actual integration of such tools in EFL higher education remains tentative and context-dependent. Thus, the findings should be interpreted as contributing to informed decision-making rather than reflecting current widespread practice. Third, the analysis relied on a combination of automatic tagging and manual recoding of 80% of the data, with borderline and multi-functional cases coded for their dominant interpersonal function. While this procedure followed Martin and White's (2005) guidelines, it inevitably involved interpretive judgement, and some utterances (such as predictive vs. epistemic uses of will) could be classified

differently. Fourth, the study examined only one AI model (i.e., GPT 4) and one EFL context (i.e., higher education), which limits the generalizability of the findings to other models and educational settings.

Acknowledgements

I would like to thank the Ongoing Research Funding program (ORF-2026-1829), King Saud University, Riyadh, Saudi Arabia, for funding this research.

Article history:

Received 14 August 2025

Received in revised form 10 February 2026

Accepted 30 March 2026

References

- Alshehri, A., Alshehri, E., & Aljamal, N. (2025). Appraisal Theory Tagger (25-12-33034135). S. A. F. I. Property. <https://appraisal-tagger.netlify.app/>
- Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, 6, 100226. <https://doi.org/10.1016/j.caeai.2024.100226>
- Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning*, 36(4), 584-607. <https://doi.org/10.1080/09588221.2021.1936071>
- Brookhart, S. (2017). *How to give effective feedback to your students*. ASCD.
- Carless, D. (2022). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*, 23(2), 143-153. <https://doi.org/10.1177/1469787420945845>
- 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>
- Çöp, A. (2025). AI vs. Traditional Teaching: Evaluating PaperRater and ChatGPT in Student Engagement and Writing Achievement. *Language Education and Technology*, 5(1), 1-20.
- Cui, Y., Schunn, C. D., Gai, X., Jiang, Y., & Wang, Z. (2021). Effects of trained peer vs. teacher feedback on EFL students' writing performance, self-efficacy, and internalization of motivation. *Frontiers in psychology*, 12, 788474. <https://doi.org/10.3389/fpsyg.2021.788474>
- Dai, W., Tsai, Y.-S., Lin, J., Aldino, A., Jin, H., Li, T., Gašević, D., & Chen, G. (2024). Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education: Artificial Intelligence*, 7, 100299. <https://doi.org/10.1016/j.caeai.2024.100299>
- Derham, C., Balloo, K., & Winstone, N. (2022). The focus, function and framing of feedback information: Linguistic and content analysis of in-text feedback comments. *Assessment & Evaluation in Higher Education*, 47(6), 896-909. <https://doi.org/10.1080/02602938.2021.1969335>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Guangyuan, Y., & Zhaoxia, L. (2025). GPT as book reviewer: A move and syntactic complexity analysis of GPT-generated versus scholar-written academic book reviews. *Journal of English for Academic Purposes*, 76, 101533. <https://doi.org/10.1016/j.jeap.2025.101533>

- Henderson, M., Ryan, T., Boud, D., Dawson, P., Phillips, M., Molloy, E., & Mahoney, P. (2021). The usefulness of feedback. *Active Learning in Higher Education*, 22(3), 229-243. <https://doi.org/10.1177/1469787419872393>
- Hwang, S. (2025). Unpacking the impact of writing feedback perception on self-regulated writing ability: The role of writing self-efficacy and self-regulated learning strategies. *Behavioral Sciences*, 15(2), 1-24. <https://doi.org/10.3390/bs15020100>
- Hyland, K. (2006). Representing readers in writing: student and expert practices. *Linguistics and Education*, 16(4), 363-377. <https://doi.org/10.1016/j.linged.2006.05.002>
- Hyland, K., & Jiang, F. (2016). "We must conclude that...": A diachronic study of academic engagement. *Journal of English for Academic Purposes*, 24, 29-42. <https://doi.org/10.1016/j.jeap.2016.09.003>
- Jiang, F., & Hyland, K. (2025). Rhetorical distinctions: Comparing metadiscourse in essays by ChatGPT and students. *English for Specific Purposes*, 79, 17-29. <https://doi.org/https://doi.org/10.1016/j.esp.2025.03.001>
- Johannes, C., & Haase, A. (2024). The impact of feedback mode on learning gain and self-efficacy: A quasi-experimental study. *Active Learning in Higher Education*, 25(3), 391-407. <https://doi.org/10.1177/14697874221131970>
- Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, 127, 103529. <https://doi.org/10.1016/j.system.2024.103529>
- Mahfoodh, O. H. A. (2017). "I feel disappointed": EFL university students' emotional responses towards teacher written feedback. *Assessing Writing*, 31, 53-72. <https://doi.org/10.1016/j.asw.2016.07.001>
- Markey, B., Brown, D. W., Laudenbach, M., & Kohler, A. (2024). Dense and disconnected: Analyzing the sedimented style of ChatGPT-generated text at scale. *Written Communication*, 41(4), 1-30. <https://doi.org/10.1177/07410883241263528>
- Martin, J. and White, P. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- Milano, E. (2023). *How to ask ChatGPT anything and get the right answers: Learn to prompt AI LLMs effectively*. Self-published.
- OpenAI. (2023). ChatGPT: Optimizing language models for dialogue. <https://openai.com/blog/chatgpt>.
- Rad, H. S., Alipour, R., & Jafarpour, A. (2024). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 32(9), 5020-5040. <https://doi.org/10.1080/10494820.2023.2208170>
- Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *The Language Learning Journal*, 46(2), 87-102. <https://doi.org/10.1080/09571736.2014.958190>
- Sardinha, T. B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1), 100083. <https://doi.org/10.1016/j.acorp.2023.100083>
- Sari, E., & Han, T. (2024). The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. *Journal of Computer Assisted Learning*, 40(5), 2065-2080. <https://doi.org/10.1111/jcal.13004>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Sun, Y. (2021). The effect of teacher caring behavior and teacher praise on students' engagement in EFL classrooms. *Frontiers in psychology*, 12, 746871. <https://doi.org/10.3389/fpsyg.2021.746871>
- Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, 147, 107857. <https://doi.org/10.1016/j.chb.2023.107857>
- Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: a comparison study. *Computer Assisted Language Learning*, 1-30. <https://doi.org/10.1080/09588221.2024.2447279>

Asma Alshehri is an Associate Professor of Applied Linguistics and the Assistant to the Chair of the English Skills Department at King Saud University. Dr. Alshehri's research interests involve academic writing, artificial intelligence in language teaching, and discourse analysis, specifically utilizing Appraisal Theory. She is an expert in the field, having received a patent for the "Appraisal Theory Tagger" from the Saudi Authority for Intellectual Property in 2025.