VNU Journal of Foreign Studies

Journal homepage: https://jfs.ulis.vnu.edu.vn/

COMPARATIVE STUDY OF FEEDBACK ON IELTS SPEAKING PERFORMANCE: PRE-SERVICE TEACHERS VERSUS AUTOMATIC SPEECH RECOGNITION APPLICATIONS

Vu Thi Hong Duyen*, Le Thi Phuong Thao, Nguyen Thu Hien, Tong Van Bac

Hanoi University, Km 9, Nguyen Trai Street, Nam Tu Liem District, Hanoi, Vietnam

Received 13 May 2025
Revised 18 June 2025; Accepted 25 June 2025

Abstract: The integration of AI-powered tools, particularly Automatic Speech Recognition (ASR) apps, in assessing and providing feedback on oral proficiency has gained significant attention lately. This comparative study analyzes Chat-GPT 4o and ELSA Speech Analyzer AI delayed feedback in comparison with feedback provided by three pre-service teachers on a student's IELTS speaking performance. The research employed a quantitative design with a total of 27 sets of feedback from the three sources. The data were analyzed according to the five criteria adapted from the feedback framework of Steiss et al. (2024). All feedback was found to be positive and adhere to IELTS speaking band descriptors. Nevertheless, Chat-GPT's focus was only on grammatical and vocabulary errors, missing the aspects related to pronunciation. The tool did not identify audio input as human speech and, instead, provided corrections based on assumptions only. ELSA gave elaborate details in feedback regarding pronunciation which might be too much information for some learners. Pre-service teachers provided holistic feedback but lacked specific analysis of clear directions for improvement. Besides, the teachers-to-be provided inaccurate pronunciation corrections at times. These insights emphasize the importance of integrating AI and human interaction when learning a language. From an educational standpoint, this method covers the structural and personal technical requirements of each individual student, resulting in a more interactive and positive atmosphere.

Keywords: delayed feedback, Automatic Speech Recognition (ASR) applications, Chat-GPT, ELSA Speech Analyzer, IELTS speaking assessment

* Corresponding author.
Email address: duyenvth@hanu.edu.vn
https://doi.org/10.63023/2525-2445/jfs.ulis.5516

SO SÁNH NHẬN XÉT CỦA GIÁO SINH VỚI PHẢN HỒI CỦA ỨNG DỤNG TRÍ TUỆ NHÂN TẠO ĐỐI VỚI BÀI THI NÓI IELTS

Vũ Thị Hồng Duyên, Lê Thị Phương Thảo, Nguyễn Thu Hiền, Tống Văn Bắc

Trường Đại học Hà Nội, Km9, Nguyễn Trãi, Quận Nam Từ Liêm, Hà Nội, Việt Nam

Nhận bài ngày 13 tháng 5 năm 2025
Chỉnh sửa ngày 18 tháng 6 năm 2025; Chấp nhận đăng ngày 25 tháng 6 năm 2025

Tóm tắt: Việc tích hợp AI, đặc biệt là các ứng dụng nhận diện giọng nói tự động, vào việc đánh giá kỹ năng nói đang nhận được nhiều sự quan tâm. Nghiên cứu này phân tích và so sánh nhận xét từ Chat-GPT 4o, ELSA Speech Analyzer và ba giáo sinh trong phần nói IELTS của một học viên. Nghiên cứu định lượng này thu được 27 nhận xét từ ba nguồn trên, và dữ liệu được phân tích theo khung đánh giá của Steiss et al. (2024). Tất cả nhận xét đều mang tính tích cực và tuân thủ các tiêu chí đánh giá của IELTS. Tuy nhiên, Chat-GPT tập trung vào lỗi ngữ pháp, từ vựng và bỏ qua lỗi về phát âm. Công cụ này không nhận diện được lời nói và chỉ nhận xét dựa trên giả định về các lỗi thường gặp. ELSA phản hồi chi tiết về phát âm, nhưng có thể gây quá tải thông tin. Trong khi đó, nhận xét của các giáo sinh tương đối toàn diện nhưng thiếu phần gợi ý cách luyện tập giúp người học tự cải thiện. Các giáo sinh đôi khi còn mắc lỗi về phát âm. Các phát hiện này khẳng định tầm quan trọng của việc kết hợp AI và yếu tố con người trong quá trình học ngôn ngữ nhằm đáp ứng được yêu cầu về cấu trúc và kỹ thuật cá nhân của học viên, giúp tạo môi trường học tập tích cực hơn.

Từ khóa: nhận xét trì hoãn, ứng dụng nhận diện giọng nói tự động (ASR), Chat-GPT, ELSA Speech Analyzer, đánh giá kỹ năng nói IELTS

1. Introduction

Artificial Intelligence (AI) technology is gaining traction in the teaching and learning of the English language, particularly due to its transformative potential in enhancing spoken communication (Dennis, 2024). AI-driven voice recognition applications have demonstrated considerable promise in supporting students' oral proficiency (Muhonen, 2021), yet they remain uncommon in Vietnamese post-secondary education (Tran & Vu, 2024). A number of case studies in Vietnam reveal that the integration of this technology greatly improved English-as-a-foreign-language (EFL) learners' speaking skills and increased student motivation and engagement (Nguyen, 2021; Nguyen, 2022; Nguyen, 2024; Le & Vo, 2014; Phan, 2021; Vo & Vo, 2020). Notwithstanding, much less is known about AI formative feedback given that feedback is widely recognized as one of the most powerful influences on learning in general (Hattie & Timperley, 2007) and as a crucial motivating factor in the process of teaching and learning speaking (Sallang & Ling, 2019).

The use of AI tools in learning to speak is increasingly commonplace. Learners can practice anytime and anywhere with internet access and receive instant feedback (Kenchakkanavar, 2023). Learning takes place in a non-judgmental and less stressful environment. Comparative studies on AI-generated and human teachers' feedback have thus been conducted to provide insights into how human teachers can exploit AI tools in their feedback giving practice. These studies primarily focus on in-service teachers, which presents a gap in our understanding of how pre-service teachers (PSTs) engage with feedback provided

by emerging AI tools. This demographic group is trained and prepared to navigate in a technology-enhanced teaching context yet has received little empirical attention. It is imperative that PSTs be informed of how to incorporate AI feedback into their assessment practices. The current lack of research-informed guidelines poses great challenges for novice teachers to effectively do so when assessing students' oral performance. This prompted our investigation into how feedback by different AI tools differs from each other and from feedback provided by pre-service teachers guided by the two following research questions:

1. What are the strengths and limitations of AI feedback on IELTS speaking performances?
2. What are the strengths and limitations of pre-service teachers' feedback on IELTS speaking performances?

This study is significant for both pre-service teachers and EFL learners, especially those who are preparing for the IELTS speaking test. By examining the feedback from different sources, the study offers pre-service teachers insights into the strengths and limitations of AI feedback and how it complements their assessments of IELTS speaking. EFL learners also benefit from the systematic analysis of AI-generated feedback when engaged in self-monitored learning. From a broader educational viewpoint, the findings contribute to the development of AI-enhanced assessment frameworks in the EFL context.

2. Literature Review

The review, although limited in scope, informs the direction of the current study. First, a notable portion of prior research was based in Vietnam and other countries considering AI feedback tools like Chat-GPT and ELSA are becoming increasingly commonplace. These studies, however, largely focus on students' perceptions of AI-generated feedback, in comparison with teachers' and peers' feedback, rather than analyzing feedback content and specific linguistic aspects that teachers could leverage for effective feedback strategies.

2.1. Feedback, Delayed Corrective Feedback and Characteristics of Effective Feedback

According to the Oxford Learner's Dictionaries (n.d.), feedback is defined as information on how good/helpful something's or someone's effort is and advice or criticism. Hattie and Timperley (2007) defined feedback as information on one's performance or comprehension that is given by an agent (e.g., teacher, peer, book, parent). Wirantaka (2019) emphasized the role of instructors, positing that feedback is information given by instructors in response to students' performances to monitor the students' learning progress and to ultimately achieve good learning outcomes. As a learning tool, feedback highlights the discrepancies between actual performance and intended performance and motivates behavioral changes (Molloy & Boud, 2014). Feedback is provided verbally through standard classroom instruction or non-verbally in writing as notes and symbols (Wirantaka, 2019). In second language acquisition, feedback is beneficial in several regards including enhancing learners' noticing of linguistic forms, guaranteeing linguistic accuracy and increasing motivation (Ellis, 2009; Hylland & Hyland, 2006; Li, 2010).

Feedback can be immediate or delayed (Ellis, 2009; Shabani & Safari, 2016). Immediate corrective feedback affords instructors the opportunity to seek clarifications from learners and to enhance their understanding of the feedback provided (Li, 2010; Lyster & Saito, 2010). Some scholars advocated for a delayed approach to corrective feedback, positing that errors should be addressed after a certain interval to avoid interrupting learners' speech flow, foster deeper

processing, and complement students' better memorization of their mistakes (Fanselow, 1977; Loewen et al., 2009; Sheen, 2006). The intervals associated with delayed feedback are valuable for memory reinforcement, knowledge retention, and cue utilization (Jones & Bourne, 1964; Renner, 1964; Magilow, 1999; Muhsin, 2016). Delayed feedback provides an analysis of the nature and the type of personal-specific errors and the appropriate corrective measures (Taipale, 2012). In assessment contexts, delayed feedback receivers consistently outperformed those receiving immediate feedback, particularly in terms of long-term retention (Kulik & Kulik, 1988; Mackey et al., 2003; McDonough, 2005; Oliver & Mackey, 2003).

Hattie and Timperley (2007) drew from Hattie's (1999) synthesis of over 500 meta-analyses on feedback and highlighted how effective feedback should provide corrective information, be delivered at appropriate timing, be specific and clear, and promote self-assessment and error detection skills among learners. Building on Hattie and Timperley's work, Wiggins (2012) examined feedback in general education and posited that helpful feedback should be *goal-referenced*, *tangible and transparent*, *actionable*, *user-friendly*, *timely*, *on-going*, and *consistent.* In language learning, Steiss et al. (2024) reconstructed a list of criteria to assess the quality of human and Chat-GPT written feedback in English writing classes and five of them were adopted to construct the grading rubric of this research: (1) *alignment with marking criteria,* (2) *providing clear directions for improvement,* (3) *accurate,* (4) *prioritizing essential features,* and (5) *using a supportive tone*. These five criteria, grounded in existing literature, are highly relevant and applicable for assessing the quality of the feedbacks generated in the written form of the current research setting. The criterion "*timely"* in Steiss' framework was excluded in the grading rubric of this research since Chat-GPT's feedback is set to be delivered immediately after the performance of students and, thus, in a timelier manner than human feedback. In a study whose focus is on feedback content like the current one, timeliness is considered less relevant.

2.2. Automatic Speech Recognition (ASR) Technology

Automatic Speech Recognition (ASR) refers to the utilization of AI to identify and process human speech (Jacko, 2012). In this context, two most current AI tools were employed for the research: ELSA Speech Analyzer and Chat-GPT.

Designed to improve learners' speaking skills, particularly pronunciation, ELSA Speak offers real-time, personalized feedback in various contexts (Jayanti, 2023; Sholekhah & Fakhrurriana, 2023). Its premium feature, ELSA Speech Analyzer, analyzes spontaneous or recorded speech and provides written reports on pronunciation, fluency, grammar, and vocabulary with predicted scores. The report includes transcriptions, overall and breakdown scores aligned with common test scales (e.g., IELTS, TOEFL, CEFR), and highlights issues such as phonetic deviations, intonation errors, fluency patterns, grammatical accuracy, and vocabulary range (Anguera et al., 2023).

Introduced in 2022 by OpenAI, Chat-GPT is a generative AI model for natural language processing (Deng & Lin, 2022). The latest version, Chat-GPT 4o, stands out for its multimodal processing capabilities and faster performance (Celik et al., 2025). Widely used for instant, personalized feedback (Celik et al., 2025; Huang & Li, 2023; Wang, 2025; Yildiz, 2024), Chat-GPT provides written evaluations of IELTS speaking tasks based on the four assessment criteria: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy, and Pronunciation. The feedback includes an estimated overall band score and detailed comments with examples and corrections. In some cases, it has offered more diverse vocabulary input than human teachers (Cao & Zhong, 2023), and its responses have been shown to foster a supportive

learning environment that enhances learners' communication skills (Muniandy & Selvanathan, 2024; Wang, 2025).

2.3. Human Teacher's vs. AI Written Feedback

Seßler et al. (2025) conducted a comparative analysis of feedback from AI and human teachers in the context of scientific inquiry and language education, using expert raters and a structured evaluation protocol. Their findings showed that human feedback was superior in linguistic quality, largely due to teachers' deeper understanding of technical terminology - consistent with observations by Cao and Zhong (2023). Human teachers also provided more personalized responses, drawing on their experience to tailor feedback effectively. In contrast, while AI demonstrated strength in content-related aspects, it struggled to contextualize errors and offer meaningful insights.

In terms of length, AI feedback was concise and well-suited for classroom use, whereas human feedback tended to be more detailed and supportive in tone. Despite these differences, the overall effectiveness of both types of feedback was found to be relatively similar, leading Seßler et al. (2025) to recommend that teachers leverage AI feedback as a supplementary tool without needing to drastically change their teaching methods. Similarly, in EFL writing classrooms, Steiss et al. (2024) found that trained, well-compensated, and time-efficient teachers produced higher-quality feedback than Chat-GPT. However, when training was not a factor, AI feedback closely resembled that of human instructors. Notably, Chat-GPT was especially effective during early writing stages, where its feedback encouraged timely revisions and greater learner engagement.

2.4. Research Gaps

Previous studies, if quantitative, primarily relied on surveys or on grading and analyzing students' work. When it comes to feedback analysis, most research has focused on writing skills. Besides, these studies tend to emphasize students' perceptions of AI-generated feedback rather than examining the content of the feedback itself or the specific linguistic features that teachers could utilize to develop effective feedback strategies. Moreover, there has been little attention paid to the current reality in which teacher feedback is increasingly delivered in written form for speaking assignments - especially those given as homework or extra practice outside the classroom in Vietnam's EFL context. At the same time, students practicing speaking with AI tools are now mostly receiving written feedback from software. This shift has created a growing need for teachers to adopt tools that can help them provide accurate, time-efficient assessments on a large scale. This is particularly important for pre-service teachers, who often lack experience and are still adapting to the demands of the profession.

3. Methodology

3.1. Research Approach

This study adopts a quantitative comparative research approach, designed to evaluate the quality of delayed written feedback provided by different sources. While the data initially takes the form of written text, a quantitized content analysis is employed to systematically code each feedback instance based on a scoring rubric adapted from Steiss et al.'s (2024) (Appendix). Through this process, qualitative content is converted into numerical scores aligned with key assessment criteria, allowing for consistent cross-source comparison.

The primary goal of this approach is to determine the relative accuracy, completeness,

and consistency of feedback from three sources: ChatGPT-4o, ELSA Speech Analyzer, and three pre-service teachers. To control for variability in language performance, all feedback was based on three IELTS mock speaking recordings produced by a single B2-level learner. Each of the three sources evaluated all three speaking samples, resulting in a total of 27 feedback instances.

Although the sample size is modest, it is methodologically appropriate and justifiable for this type of focused comparative analysis. Each feedback instance serves as a distinct, analyzable unit, and the controlled design enhances internal validity by reducing noise from learner-level variation. Given the exploratory nature of the study and its aim to examine feedback quality patterns rather than to generalize to a broader population, this research approach is both valid and effective for addressing the research questions.

3.2. Research Setting and Participants

This study examined corrective feedback that is delayed in timing and written in modality. As elaborated in the Literature Review section, the nature of the IELTS speaking mock test calls for delayed corrective feedback so that the speaker can finish his/her speech without interruptions. Moreover, the feedback by the AI tools employed in this study was written by default, so the teachers were requested to provide written feedback with supplementary materials (e.g., recording for correction of pronunciation, tutorial video for the articulation of certain sounds, definition and usage of certain lexical items, useful learning materials) to ensure consistency and comparability across data sources. By standardizing feedback mode, we aimed to minimize variability caused by differences in feedback delivery and focus more on the content of the feedback itself.

We chose Chat-GPT 4o and ELSA Speech Analyzer considering how widely-used they are for self-study or assessment purposes. We employed a convenience sampling method and selected three pre-service teachers who met the requirements regarding speaking proficiency level and experience in teaching and assessing IELTS speaking performance. The teachers possessed validated IELTS speaking scores of 7.5 to 8.0. At the time of the study, they had had one to two years of experience in teaching and assessing IELTS speaking performance and were in charge of an IELTS speaking course at three different English centers.

3.3. Data Collection and Analysis

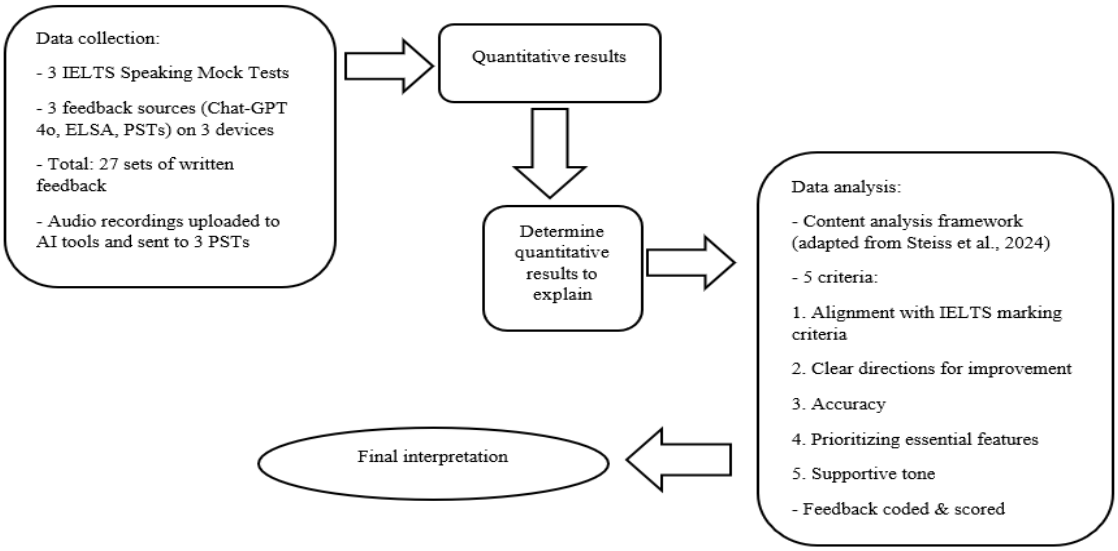
The mock tests were conducted via video-conferencing software Zoom, with one of the co-authors, who is an IELTS speaking instructor and is familiar with the test format, as the mock examiner. The student sat for three mock tests (MT1, MT2, MT3), each lasting 11-14 minutes. The sets of questions used were adopted from the Cambridge IELTS series. The audio transcripts and recordings were uploaded on Chat-GPT 4o using three different accounts (CHAT1, CHAT2, CHAT3) and on ELSA Speech Analyzer using three different accounts (ELSA1, ELSA2, ELSA3) and emailed to the pre-service teachers (PST1, PST2, PST3). The prompt keyed into Chat-GPT 4o was as follows: “Provide feedback on the IELTS speaking test based on the four assessment criteria. I will give you the questions and the corresponding answers.” It should be noted that Chat-GPT’s voice assistant requires the speaker to manually press a button and speak, after which the model transcribes the audio input using its built-in ASR technology and evaluates the transcript. However, for consistency and ease of analysis and comparison across different sources, we chose to upload transcripts of the student’s responses instead of using the voice assistant mode. A total of 27 sets of written feedback (3 mock tests x 3 sources x 3 runs/raters) were collected, analyzed and compared in terms of

content quality. The small sample size allowed for in-depth comparison, although it limits generalizability.

The raters were required to present detailed assessment and evidence for every statement. Cycles of coding, discussion, and refinement of the criteria were repeated until the data analysts exhibited high degrees of interrater agreement. The entire process of coding and assessing the feedback was conducted in consultation with a subject-matter expert in the field of technology integration in foreign language teaching.

Figure 1

Research Design



4. Findings

This section presents a comparative analysis of the feedback content provided by the three sources according to the five criteria adopted from Steiss et al. (2024). Below is a breakdown of the mean scores assigned to 27 sets of feedback from three sources.

Table 2

Total Score Summary

Criterion \ Feedback Source	Chat-GPT 4o	ELSA Speech Analyzer	Pre-service teachers
Alignment with Marking Criteria	3.17	4.5	4.38
Clarity of Directions for Improvement	4.23	5	1.43
Accuracy of Feedback	3.66	4.33	4.8
Prioritization of Essential Features	1.33	3	2.23
Supportive Tone	4	5	2.86
Mean Score	**3.27**	**4.36**	**3.14**

Overall, ELSA was rated the highest (4.36), followed by CHAT (3.27) and PSTs (3.14). In greater detail, ELSA's feedback was found to align most closely with IELTS speaking marking criteria (4.5). One interesting finding is that the scores assigned to all providers remained above 3.0.

Table 3

Examples of Feedback Rated High in Criterion 1

Scoring Category	Example
5: All feedback consistently and explicitly references 4 marking criteria.	(Source: ELSA1 – MT1) *Overall Speaking Score: 47%* *Your English speaking score is Lower Intermediate. Keep it up!* ***- Pronunciation:*** *+ Score: 36% (Beginner)* *+ Your Top Errors and Suggestions for Improvement […]* ***- Intonation:*** *+ Score: 42% (Lower Intermediate)* *+ Pitch Variation: Keep your Pitch Variation within the target range shown in green below […]* *+ Tips for Improvements […]* ***- Fluency:*** *+ Score: 25% (Beginner)* *+ Pace Score (62 wpm – Natural), Pausing Score (50%), Hesitations (Too many) […]* *+ Tips for Improvements […]* ***- Grammar:*** *+ Score: 64% (Upper Intermediate)* *+ Your Grammatical Range […]* *+ Your Top Grammatical Errors […]* ***- Vocabulary:*** *+ Score: 69% (Upper Intermediate)* *+ Areas to improve […]*
5: All feedback is highly specific with examples taken from the student's speech	(Source: ELSA1 - MT1) ***- Lexical Resource:*** *+ You said: "Certainly (adverb)" -> we suggest: "surely, definitely"* *+ You said: "think" -> we suggest:" consider"* ***- Grammar Accuracy:*** *+ Verb tenses (You used the wrong form of the verb "start"): You said: "I start use social media" => Correction: "I started using social media"* *+ Non-finite verbs (You should have used a gerund here): You said: "I start use social media" => Correction: "I started using social media"* *+ Verb tenses (You used the wrong form of the verb "teach"): You said: "my friend teach me" => Correction: "my friend taught me"* *+ [...]** ***- Pronunciation:*** *+ Sound /t/: You forgot to pronounce /t/: absolutely /ˈæbsəluːtli/, It's /ɪts/, just /dʒʌst/, about /əˈbaʊt/** *+ Sound /ə/: You forgot to pronounce /ə/: opinion /əˈpɪnjən/, nowadays /ˈnaʊədeɪz/, to /tə/, can /kən/** *+ Sound /n/: You forgot to pronounce /n/: account /əˈkaʊnt/, in /ɪn/, continue /kənˈtɪnjuː/** *+ […] (A breakdown of intonation, pausing, pace, and hesitation was demonstrated in charts and a color-coded transcript)*

Another striking finding that stands out is that ELSA provided a clear and detailed action plan for the student to avoid making the same mistakes, receiving the highest mean score (5.0), while PSTs provided almost no such suggestion, thus the lowest score (1.43).

Table 4

Examples of Feedback Rated High in Criterion 2

Scoring Category	Example
5: All feedback consistently provides clear directions for improvement	(Source: ELSA1 - MT2) ***- Pronunciation:*** *(with tutorial videos)* *+ For /r/ sound: This is an /r/ sound. Press your tongue against your upper gums behind your front teeth to stop the air from coming out, and then release it.* *+ For /t/ sound: Your mouth should be almost closed for /t/. This will help you get your tongue up high. People also generally round their lips for /t/.* *+ For /i/ sound: The /i/ vowel is similar to /i/ (ee), as in "see," but it's more relaxed. To practice, try saying /i/ and then relax your lips and tongue: /i/.* ***- Intonation:*** *Make the following types of words more prominent by saying them louder and with energy: Nouns, Main verbs, Adjectives, Adverbs (A color-coded transcript with words highlighted to reflect their prosodic prominence)* ***- Fluency:*** *Fluency means being able to control your delivery in a way that feels natural to the conversational setting. To improve:* *+ Pause at the end of complete sentences or after key ideas.* *+ Ensure your speech is constantly moving forward during phrases. Your pace will naturally slow down at the next punctuation mark or logical stop.* *+ Never interrupt the flow! It's common for English learners to stop speaking when searching for a specific word. To avoid interrupting the flow of a conversation, use the language you know to describe a missing concept or idea.* ***- Additional Tips:*** *+ Use punctuation as your guide! Commas (,), colons (:), semicolons (;), and periods (.) show where pauses should take place in written text.* *+ Divide ideas or sentences into "chunks" of about 5-7 words and pause after each chunk. Make sure the words in each chunk belong together grammatically!*

In terms of accuracy, the mean scores of all sources remained relatively high (3.0-5.0), with PSTs rated the highest (4.7-5.0).

Table 5

Examples of Feedback Rated High in Criterion 3

Scoring Category	Example
5: All feedback is accurate	(Source: PST1 - MT3) ***- Fluency and Coherence:*** *+ The first section seems more fluent than the remaining sections* *+ Still prolong the pauses than necessary* ***- Lexical:***

	+ Lack of topic-based vocabulary; synonyms; collocations; and phrasal verbs *+ "Entertain" is repeated too regularly -> "relax"; or change its word formation* *+ "Job" is also repeated -> "occupations"; "positions"* *+ "I think" can be substituted by "as far as I'm concerned"; "I argue that"; ....* ***- Grammar:*** *+ I'm not usually get -> delete "am"* *+ I think is -> I think that* *+ It help them -> It helps them* *+ It's help them -> It helps them* *+ "More" cannot stand before "improve"* *+ They not attract -> they do not attract* *+ They not following -> they don't follow* ***- Pronunciation:*** *+ Lack of ending sounds: Biggest; supermarket** *+ Vegetables: /ˈvedʒ/ instead of /vege/* *+ The sound /θ/ in anything; I think; something is mispronounced* *+ Study: /ʌ/ is mispronounced into /u/* *+ Apps instead of app* *+ Homework cannot be added "s" at the end* *+ Do exam -> do exams* *+ Daily life: "life" lacks the ending sound /f/* ** Audio of the correct pronunciation was attached*

Perhaps the most unexpected finding is one regarding the identification of persistent and crucial mistakes where all sources were rated lower than 3.0, with CHAT being the lowest (1.33).

Table 6

Examples of Feedback Rated High in Criterion 4

Scoring Category	Example
3: Even mix of prioritizing essential and non-essential features	(Source: PST2 - MT2) *- Sometimes, the speaker adds redundant /s/ at the end of some words, such as:* *+ "it" (she pronounced it as "its")* *+ I think (I thinks)* *- Some collocations are used incorrectly like "quality air", "my knowledge of not clear"*

As for the feedback's tone, ELSA's feedback was most supportive (5.0), followed by CHAT's (4.0). One unanticipated result was that feedback of PSTs (2.86) was rated significantly lower in this regard compared to that of their AI counterparts.

Table 7

Examples of Feedback Rated High in Criterion 5

Scoring Category	Example
5: Balance of compliments and criticism; suggestive and respectful tone	(Source: ELSA1 - MT2) *Overall Speaking Score: 43%* *Your English speaking score is Lower Intermediate.* ***Keep it up!*** ***- Pronunciation:*** *+ Your English speaking score is Lower Intermediate.* ***Let's get to 'Intermediate' together!*** *Make a list of "tricky" words or sentences you find hard to pronounce, then say them outloud in Speech Analyzer. You'll get immediate feedback on your pronunciation!* *+ Your Top Errors and Suggestions for Improvement […]* ***- Intonation:*** *+ Your intonation level is Beginner. Let's get to the next level this week!* ***We want everyone to understand you!*** *So remember to always speak loudly and clearly during conversations.* *+ Pitch Variation […]* *+ Tips for Improvements […]* ***- Fluency:*** *+ Your Fluency level is Beginner. Here's a tip to help you improve…* *+ Pace Score, Pausing Score, Hesitations […]* *+ Tips for Improvement […]* ***- Grammar:*** *+ Your grammar level is Intermediate. Now let's make sure you're not a "one-trick pony"* *+ Your Grammatical Range:* ***Good job!*** *You managed to include the following structures […]* *+ Your Top Grammatical Errors: […]* ***- Vocabulary:*** *+ Your level is Intermediate. Time to get active! Your passive vocabulary includes words you know and understand, but that you cannot use comfortably yet. To reach the next level, you'll have to activate newly discovered words by using them in conversation as soon as possible.* *+ Vocabulary Distribution […]* *+ Your Top Performance:* ***[…] Keep up the good work!*** *Using advanced words made your speech more engaging.* *+ Expand Your Active Vocabulary Bank: […]* *+ Be mindful of using informal language: […]*

A closer look at individual sources reveals their distinct strengths and weaknesses. CHAT's feedback was highly regarded in its ability to provide clear directions for improvement (4.23) and its supportive tone (4.0). The chatbot's score for accuracy and alignment with marking criteria was moderate, at 3.66 and 3.17 respectively. Its major weakness lies in the lack of prioritization of important features (1.33). Meanwhile, ELSA demonstrated the most consistent performance across all five criteria. It was superior for its provision of clear suggestions for improvement (5.0) and supportive tone (5.0). ELSA's feedback also closely aligned with the IELTS marking criteria (4.5) and was highly accurate (4.33). Its primary weakness was found to be its prioritization of essential features (3.0). Feedback by PSTs

showed the highest level of accuracy (4.8) and the strongest alignment with marking criteria (4.38). However, PSTs failed to deliver clear, specific and actionable suggestions for improvement (1.43), a hierarchy of important features (2.23), as well as supportive-sounding tone (2.86) in their feedback of all mock tests.

5. Discussion

The mean scores across five criteria reveal disparities in content quality. Overall, ELSA Speech Analyzer delivered quality feedback more consistently than their counterparts. The tool provided users with a detailed action plan for improvement with supplementary resources for demonstration which were not found in feedback by Chat-GPT 4o and pre-service teachers. ELSA's feedback was also more encouraging and supportive in tone, balancing between compliments and criticism. In the following section, we present explanations for and interpretations of our findings in greater depth, drawing on examples from the raw data pool to justify our evaluations.

5.1. Strengths and Limitations of AI Feedback on IELTS Speaking Performances

AI tools, such as Chat-GPT and ELSA Speech Analyzer, offer a range of strengths and limitations in providing feedback on IELTS speaking performances.

Strengths: AI feedback is consistent in its delivery, offering structured feedback across all marking criteria such as grammar, vocabulary, fluency, and pronunciation. ELSA Speech Analyzer, for example, delivers highly specific and detailed feedback, supported by charts and phonetic explanations, which are beneficial for pronunciation improvements. AI feedback can also be instantly available, giving learners the opportunity to review their performance quickly and without delay. This immediacy can promote self-paced learning and allow for repeated practice, which is crucial for improving speaking skills. Additionally, AI feedback can provide a wide range of suggestions for improvement, fostering learner autonomy. Nguyen et al. (2024) found that AI feedback provides structured learning paths that foster learner autonomy. AI tools also maintain a supportive and motivational tone, which encourage learners and promote their learning progress.

Limitations: On the downside, AI feedback lacks the ability to evaluate coherence in speech, especially in terms of the logical organization of ideas and the use of cohesive devices. Chat-GPT's feedback can be too generic, with some vague suggestions, such as "use more varied vocabulary" or "avoid repeating words," without providing specific examples to illustrate common errors. This limitation aligns with findings from Lehman et al. (2020) and Seßler et al. (2024), who noted that AI-generated feedback often suffers from unclear communication of errors, making it difficult for learners to identify and correct specific language problems. Moreover, Chat-GPT cannot process audio input, which limits their accuracy in offering pronunciation feedback. Its feedback on pronunciation was based on its assumptions of common errors among learners of a specific level rather than on the actual mistakes. While ELSA Speech Analyzer offers some phonetic details, its technical terminology may overwhelm low-intermediate learners. Furthermore, AI-generated feedback is sometimes inconsistent across different versions, leading to discrepancies in the suggestions and feedback quality. Dakhil et al. (2025) also noted that AI-mediated speaking assessment tools tend to provide feedback on grammar, vocabulary, intonation, and fluency, but often fall short when it comes to pronunciation.

5.2. Strengths and Limitations of Pre-Service Teachers' Feedback on IELTS Speaking Performances

Strengths: Pre-service teachers (PSTs) generally provide feedback that aligns well with the IELTS speaking band descriptors. Their feedback is often accurate and based on their understanding of the criteria, especially in terms of grammar and pronunciation. The teachers' feedback is typically reliable and rooted in direct observation, making it more personalized and specific compared to AI-generated feedback. Additionally, PSTs can provide structured guidance, helping students improve by offering more targeted advice on how to avoid repeated mistakes. This complements Vu and Nguyen's findings (2021) where human teacher feedback was found to offer structured guidance on addressing errors.

Limitations: PSTs' feedback can be inconsistent, particularly regarding coherence and fluency. While PSTs often address grammar and pronunciation in great detail, comments on fluency and vocabulary may be more general and lack actionable examples. There is also a tendency to overlook persistent errors or patterns in learners' speech, reducing the diagnostic depth of their feedback. This finding contrasts with Lakhdari (2020), who reported that teachers often paid close attention to minor and frequently occurring errors in students' speech. This issue may be due to the limited experience of some PSTs, which affects their ability to detect recurring issues in student performances. Unlike AI tools, PSTs do not typically provide estimated band scores which could help students gauge their progress in a more test-like manner. Additionally, some PSTs may not be as thorough in suggesting specific steps for improvement, such as providing examples of how to expand vocabulary or improve fluency.

In summary, both AI tools and pre-service teachers offer valuable feedback, each with its strengths and weaknesses. AI tools provide immediate and structured feedback, especially beneficial for pronunciation and grammar, but struggle with providing coherent feedback and context-specific examples. Pre-service teachers offer more personalized feedback, though their lack of experience may sometimes lead to vague or incomplete advice. These findings are consistent with the literature, where studies like those from Dakhil et al. (2025) and Lehman et al. (2020) highlighted the potential and challenges of AI-generated feedback, while Vu and Nguyen (2021) and Lakhdari (2020) emphasized the importance of teacher feedback in providing more targeted and structured guidance.

6. Conclusion

The current study highlighted that ELSA Speech Analyzer was the most consistent and comprehensive (strengths, weaknesses, explanations, examples, and suggestions for improvement in texts or visual aids), especially regarding pronunciation aspects. Its feedback was also clear, well-structured, and motivational in tone. The second feedback source, Chat-GPT 4o, was rated high for its proposal of actionable, specific directions for improvement and supportive tone. However, its inability to process pre-recorded audio input limited the accuracy of pronunciation and fluency feedback. Feedback also varied across different versions, raising concerns over the tool's reliability. As for the pre-service teachers, their feedback was generally more accurate and rubric-aligned compared to their AI counterparts, especially in pronunciation and grammar. The feedback would have been more helpful if a clear hierarchy of prominent errors or a statement of persistent errors, and a clear action plan for speaking improvement had been included.

These findings support research-oriented recommendations from previous studies (Cao & Zhong, 2023; Nazeretsky & Kaser, 2024; Seßler et al., 2025; Dakhil et al., 2025) of teachers

exploiting ChatGPT as a complementary feedback source to enhance the quality of traditional teacher feedback. While acknowledging the integration of AI in teachers' feedback practices, Tseng and Yeh (2019) put forth that it is crucial to align AI and teachers' feedback with students' perspectives. The present content analysis offers pre-service teachers with insights into how to strategically exploit AI in their own feedback-giving practices, thus improving the content quality and the efficiency of the process. Tools like ELSA Speech Analyzer and Chat-GPT can assist with tasks at which they excel, such as *identifying grammatical and pronunciation mistakes* and *suggesting action plans for improvement*. This frees up the time and the cognitive resources for teachers to focus on tasks like *tracking recurring errors* and *prioritizing crucial mistakes* and *personalised feedback to suit their students' needs at best*.

7. Limitations and Recommendations

Several limitations threaten our study's validity. First, the small sample limited generalizability as the researchers only collected one English learner and three pre-service teachers as a sample size. Therefore, it is recommended to study more diverse samples leading to comparing the ASR and human feedback in a more efficient manner in speaking performance. Furthermore, the study only examined two AI tools - Chat-GPT 4.0 and ELSA Speech Analyzer - which do not fully represent the variety of ASR-based feedback applications available. Including a wider range of AI tools in future studies would help provide a more complete picture of their effectiveness. Moreover, feedback was given in written form rather than in real-time interactions. Future research can further examine oral (immediate and delayed) feedback in EFL speaking contexts.

References

Anguera, X., Proença, J., Gulordava, K., Tarján, B., Parslow, N., Dobrovolskii, V., & Girard, R. (2023). ELSA Speech Analyzer: English Communication Assessment of Spontaneous Speech. In *Proceedings of 9th Workshop on Speech and Language Technology in Education (SLaTE)* (pp. 95-96).

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Cao, S., & Zhong, L. (2023). Exploring the Effectiveness of ChatGPT-based Feedback Compared with Teacher Feedback and Self-feedback: Evidence from Chinese to English Translation. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2309.01645

Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a Virtual Speaking Tutor to Boost EFL Learners' Speaking Self-efficacy. *Australian Journal of Applied Linguistics*, *8*(1), 102418. https://doi.org/10.29140/ajal.v8n1.102418

Darsih, E., Wihadi, M., & Hanggara, A. (2021). Using ELSA App in Speaking Classes: Students' Voices. In *Proceedings of the 1st Universitas Kuningan International Conference on Social Science, Environment and Technology, UNiSET 2020*. https://doi.org/10.4108/eai.12-12-2020.2304993

Deng, J., & Lin, Y. (2022). The Benefits and Challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81–83. https://doi.org/10.54097/fcis.v2i2.4465

Dennis, N. K. (2024). Using AI-Powered Speech Recognition Technology to Improve English Pronunciation and Speaking Skills. *IAFOR Journal of Education*, *12*(2), 107–126. https://eric.ed.gov/?id=EJ1440171

Ellis, R. (2009). Corrective Feedback and Teacher Development. *L2 Journal*, *1*(1), 3–18. https://doi.org/10.5070/l2.v1i1.9054

Fanselow, J. F. (1977). The Treatment of Error in Oral Work. *Foreign Language Annals*, *10*(5), 583–593. https://doi.org/10.1111/j.1944-9720.1977.tb03035.x

Hattie, J. (1999). *Influences on student learning Influences on student learning*. University of Auckland.

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

Huang, J., & Li, S. (2023). Opportunities and Challenges in the Application of ChatGPT in Foreign Language Teaching. *International Journal of Education and Social Science Research*, *06*(04), 75–89. https://doi.org/10.37500/IJESSR.2023.6406

Hyland, K., & Hyland, F. (2006). Feedback on Second Language Students' Writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/s0261444806003399

Jacko, J. A. (Ed.). (2012). *Human-computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications* (3rd ed.). CRC Press

Jayanti, J. (2023). The Effectiveness of Elsa Speak Application to Improve Students' Pronunciation Ability at Smpn 1 Tandukkalua. *Repository Universitas Sulawesi Barat*.

Jones, R. E., & Bourne, L. E. (1964). Delay of Informative Feedback in Verbal Learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *18*(4), 266.

Kenchakkanavar, A. Y. (2023). Exploring the Artificial Intelligence Tools: Realizing the Advantages in Education and Research. *Journal of Advances in Library and Information Science*, *12*(4), 218-224.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of Feedback and Verbal Learning. *Review of Educational Research*, *58*(1), 79–97. https://doi.org/10.3102/00346543058001079

Ladkhdari, S. (2020). The Role of Teacher's Feedback in Enhancing EFL Learners' Speaking Skill: The Case of EFL Students at Biskra University. *Univ-Biskra.dz*. http://archives.univ-biskra.dz/handle/123456789/16138

Le, X. M., & Vo, K. H. (2014). Factors Affecting Secondary-School English Teachers' Adoption of Technologies in Southwest Vietnam. *Language Education in Asia*, *5*(2), 198–215. https://doi.org/10.5746/leia/14/v5/i2/a03/le_vo

Lehman, B., Gu, L., Zhao, J., Tsuprun, E., Kurzum, C., Schiano, M. A., Liu, Y., & G. Tanner Jackson. (2020). Use of Adaptive Feedback in an App for English Language Spontaneous Speech. *Lecture Notes in Computer Science*, *121*, 309–320. https://doi.org/10.1007/978-3-030-52237-7_25

Li, S. (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language Learning*, *60*(2), 309–365.https://doi.org/10.1111/j.1467-9922.2010.00561.x

Loewen, S., Li S., Fei, F., Thompson, A., Nakatsukasa, K., Ahn, S., & Chen, X. (2009). Second Language Learners' Beliefs About Grammar Instruction and Error Correction. *The Modern Language Journal*, *93*(1), 91–104. https://doi.org/10.1111/j.1540-4781.2009.00830.x

Lyster, R., & Saito, K. (2010). Oral Feedback in Classroom SLA: A Meta-Analysis. *Studies in Second Language Acquisition*, *32*(2), 265–302. https://doi.org/10.1017/s0272263109990520

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional Input and the Incorporation of Feedback: An Exploration of NS–NNS and NNS–NNS Adult and Child Dyads. *Language learning*, *53*(1), 35-66.

Magilow, D. H. (1999). Case Study #2: Error Correction and Classroom Affect. *Die Unterrichtspraxis/Teaching German*, *32*(2), 125. https://doi.org/10.2307/3531752

McDonough, K. (2005). Identifying the Impact of Negative Feedback and Learners' Responses on ESL Question Development. *Studies in Second Language Acquisition*, *27*(1). https://doi.org/10.1017/s0272263105050047

Molloy, E. K., & Boud, D. (2014). Feedback Models for Learning, Teaching and Performance. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 413–424). Springer Science+Business Media. https://doi.org/10.1007/978-1-4614-3185-5

Muhonen, R. (2021). *Using ASR Technology in English Pronunciation Teaching: Finnish Teachers' and Pupils' First Impressions*. https://trepo.tuni.fi/bitstream/handle/10024/130750/MuhonenRiikka.pdf?sequence

Muhsin, A. (2016). The Effectiveness of Positive Feedback in Teaching Speaking Skill. *Lingua Cultura*, *10*(1), 25. https://doi.org/10.21512/lc.v10i1.873

Muniandy, J., & Selvanathan, M. (2024). ChatGPT, a Partnering Tool to Improve ESL Learners' Speaking Skills: Case Study in a Public University, Malaysia. *Teaching Public Administration*. https://doi.org/10.1177/01447394241230152

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024). AI or Human? Evaluating Student Feedback Perceptions in Higher Education. *Lecture Notes in Computer Science*, 284–298. https://doi.org/10.1007/978-3-031-72315-5_20

Nguyen, L. T. H. (2021). Teachers' Perception of ICT Integration in English Language Teaching at Vietnamese Tertiary Level. *European Journal of Contemporary Education*, *10*(3), 697–710. https://eric.ed.gov/?id=EJ1324207

Nguyen, T. M. N. (2022). Effects of Using Computer-Based Activities in Teaching English Speaking at a High School in Ho Chi Minh City, Vietnam. *Social Science Research Network.*

Nguyen, T. T. H. (2024). Examining the Issues of English-speaking Skills That University EFL Learners Face. *International Journal of Advanced Multidisciplinary Research and Studies*, *4*(2), 36–40.

Nguyen, X. H., Le, T. T., Do, Đ. K., Hoang, Q. N., & Nguyen, T. T. H. (2024). Students' Attitudes Toward Utilizing AI-based Technologies to Their Speaking Proficiency: A Case Study. *Journal of Science*, *21*(5). https://doi.org/10.54607/hcmue.js.21.5.4270(2024)

Oliver, R., & Mackey, A. (2003). Interactional Context and Feedback in Child ESL Classrooms. *The Modern Language Journal*, *87*(4), 519–533.

Oxford University Press. (n.d.). Feedback. In *Oxford Learner's Dictionaries.* Retrieved April 20, 2025, from https://www.oxfordlearnersdictionaries.com/definition/english/feedback?q=feedback

Phan, T. K. T. (2021) Vietnamese Undergraduates' Attitudes Towards the Use of Facebook for English Language Teaching and Learning. *17th International Conference of the Asia Association of Computer-Assisted Language Learning (AsiaCALL 2021)*, 181–195. https://doi.org/10.2991/assehr.k.210226.022

Renner, K. E. (1964). Delay of Reinforcement: A Historical Review. *Psychological Bulletin*, *61*(5), 341.

Sallang, H., & Ling, Y. L. (2019). The Importance of Immediate Constructive Feedback on Students' Instrumental Motivation in Speaking in English. *Britain International of Linguistics Arts and Education (BIoLAE) Journal*, *1*(2), 1–7. https://doi.org/10.33258/biolae.v1i2.58

Seßler, K., Bewersdorff, A., Nerdel, C., & Kasneci, E. (2025). Towards Adaptive Feedback with AI: Comparing the Feedback Quality of LLMs and Teachers on Experimentation Protocols. *arXiv*. https://arxiv.org/abs/2502.12842

Shabani, K., & Safari, F. (2016). Immediate vs Delayed Correction Feedback (CF) and Accuracy of Oral Production: The Role of Anxiety. *Theory & Practice in Language Studies (TPLS)*, *6*(11).

Sheen, Y. (2006). Exploring the Relationship Between Characteristics of Recasts and Learner Uptake. *Language Teaching Research*, *10*(4), 361–392. https://doi.org/10.1191/1362168806lr203oa

Sholekhah, M. F. & Fakhrurriana, R. (2023). The Use of ELSA Speak as a Mobile-Assisted Language Learning (MALL) towards EFL Students' Pronunciation. *Journal of Education, Language Innovation, and Applied Linguistics*, *2*(2), 93–100. https://doi.org/10.37058/jelita.v2i2.7596

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Carol Booth Olson. (2024). Comparing the Quality of Human and ChatGPT Feedback of Students' Writing. *Learning and Instruction*, *91*(no), 101894–101894. https://doi.org/10.1016/j.learninstruc.2024.10189

Taipale, P. (2012). *Oral Errors, Corrective Feedback and Learner Uptake in an EFL Setting*. https://jyx.jyu.fi/jyx/Record/jyx_123456789_37544

Tran, D. K., & Vu, T. K. C. (2024). Investigating Learners' Perspectives on ELSA Speak Integration to Enhance Autonomy and Oral Language Proficiency in English Classes. *Proceedings of the AsiaCALL International Conference*, *6*(6), 182–192. https://doi.org/10.54855/paic.24613

Tseng, S. S., & Yeh, H. C. (2019). The Impact of Video and Written Feedback on Student Preferences of English Speaking Practice. *Language Learning & Technology*, *23*(2), 145–158. https://doi.org/10125/44687

Vo, L., & Vo, L. (2020). EFL Teachers' Attitudes Towards the Use of Mobile Devices in Learning English at a University in Vietnam. *SSRN Electronic Journal*, *11*(1). https://doi.org/10.2139/ssrn.3581340

Vu, T. Q., & Nguyen, D. H. (2021). Impacts of Feedback Posted on Google Classroom On Students' Speaking Skill. *TNU Journal of Science and Technology*, *226*(3), 58–63. https://doi.org/10.34238/tnu-jst.4088

Wang, Y. (2025). A Study on the Efficacy of ChatGPT-4 in Enhancing Students' English Communication Skills. *SAGE Open*, *15*(1). https://doi.org/10.1177/21582440241310644

Wirantaka, A. (2019). Investigating Written Feedback on Students' Academic Writing. In *Proceedings of the Third International Conference on Sustainable Innovation 2019 – Humanity, Education and Social Sciences (IcoSIHESS 2019)*. https://doi.org/10.2991/icosihess-19.2019.1

Yildiz, C. (2024). ChatGPT Integration in EFL Education: A Path to Enhanced Speaking Self-Efficacy. *Novitas-ROYAL (Research on Youth and Language)*, *18*(2), 167–182. https://eric.ed.gov/?id=EJ1446766

Appendix

Table 1

Scoring Rubric (Adapted from Steiss et al., 2024)

	Alignment with Marking Criteria	**Clarity of Directions for Improvement**	**Accuracy**	**Prioritization of Essential Features**	**Supportive Tone**
5	1.1. All feedback consistently and explicitly references 4 marking criteria. 1.2. All feedback is highly specific with examples taken from student's speech	All feedback consistently provides clear directions for improvement	All feedback is accurate	All feedback prioritizes essential features or points out persistent errors	Balance of compliments and criticism; suggestive and respectful tone
4	1.1. Most feedback explicitly references 4 marking criteria, but some feedback does not explicitly reference criteria 1.2. Most feedback is clearly described and supported with examples, but a few points are generic and formulaic	Most feedback is usable; few directions are not spelled out	Most feedback is accurate; one piece of feedback is somewhat inaccurate.	Most feedback prioritizes essential features or points out persistent errors	Most feedback has a slight imbalance of compliments and criticisms; suggestive and supportive tone
3	1.1. Half of the feedback explicitly references 4 marking criteria, or All feedback explicitly references 2-3 out of 4 marking criteria 1.2. Half of the feedback is clearly described and supported with examples from student's speech, and half is generic and formulaic	Even mix of specific and vague suggestions	Some feedback is accurate, 1+ pieces are clearly inaccurate.	Even mix of prioritizing essential and non-essential features	Lack of compliments; even mix of supportive and directive tone

2	1.1. Most feedback does not explicitly reference any marking criteria, or All feedback explicitly references 1 out of 4 marking criteria 1.2. Most feedback is generic and formulaic; one piece of feedback is clearly described and supported with examples from student's speech	Lack of actionable next steps	Feedback is mostly inaccurate.	Most feedback focuses on non-essential features; no explicit mention of persistent errors or essential features	Most feedback mainly includes criticisms; directive tone
1	1.1. No feedback references any marking criteria 1.2. No feedback is clearly described and supported with examples from student's speech	No concrete steps for improvement	Feedback is inaccurate or irrelevant to the student's speaking performance.	Feedback does not mention any persistent errors or essential features	No positive comments; disrespectful, condescending or discouraging tone