Huang, M. & Hu, F. (2025). From comparison to integration: A hybrid feedback model based on student perceptions of AI and teacher support in EFL speaking. *Accents Asia, 20*(2), 12-35.

From Comparison to Integration: A Hybrid Feedback Model Based on Student Perceptions of AI and Teacher Support in EFL Speaking

Huang Mengqi[1]
Yamaguchi University
Hu Fei[2]
Fudan University

ABSTRACT

This study explores the development of a hybrid feedback model that combines AI-generated and teacher-provided feedback in EFL speaking tasks. Using a mixed-methods design, the research investigates how Chinese university students perceive the clarity, usefulness, and learning impact of both feedback types. Findings indicate that learners appreciate the immediacy and accessibility of AI tools, while valuing the contextual depth and emotional support offered by teachers. The study proposes a pedagogically meaningful integration of both systems—where AI facilitates frequent autonomous practice, and teachers offer targeted guidance on discourse and pragmatics. These insights highlight the potential of hybrid feedback models to optimize EFL oral instruction and suggest new directions for prompt design and teacher-AI collaboration. The paper also addresses methodological limitations, such as the small participant pool and single-institution context, framing the study as a preliminary exploration of learner perceptions within a rapidly evolving technological landscape.

Keywords: AI-generated feedback, ChatGPT, spoken English, EFL learners, teacher feedback, hybrid model, learner perception

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Professor Ohshima Naoki and Professor Fukuyo Kazuhiro for their invaluable guidance, insightful feedback, and unwavering support throughout the course of this research. Their academic rigor,

[1] Huang Mengqi is a PhD student at Yamaguchi University, enrolled in the Graduate School of Science and Technology for Innovation. Her major is Systems and Design Engineering. Before starting her doctoral studies, she had taught English for 15 years.
Email: d502wcu@yamaguchi-u.ac.jp

[2] Dr. Hu Fei received his PhD in Optics from Fudan University. He is now an entrepreneur and the founder of a well-known STEM education company in China, Shanghai Zhanlu Education Technology Co., Ltd.
Email：hufei@fudan.edu.cn

patience, and encouragement not only shaped the direction of my study but also inspired me to grow as a researcher. Without their expert supervision and kind mentorship, this work would not have been possible. I am sincerely thankful for the opportunity to learn under their instruction.

I would also like to extend my heartfelt thanks to the reviewers and editors of Accents Asia for their constructive comments and careful editorial guidance. Their valuable suggestions have greatly improved the clarity, coherence, and overall quality of this manuscript. I am truly grateful for their time and effort in helping me refine my work.

BACKGROUND

In recent years, artificial intelligence (AI) has rapidly transformed the landscape of language education. Among various AI applications, generative AI chatbots—such as OpenAI's ChatGPT—have shown increasing potential in supporting spoken English development through real-time feedback and interaction. While previous studies have examined AI's role in reading and writing, relatively little research has explored its impact on spoken English acquisition, particularly through combined AI-human feedback mechanisms in English-as-a-foreign-language (EFL) contexts. Furthermore, most existing studies focus on system performance rather than learner perceptions, leaving a gap in understanding how students experience and evaluate AI-mediated feedback in authentic tasks.

Chinese university EFL learners often lack access to immersive English-speaking environments, which limits their opportunities to develop oral fluency and communicative competence. Traditional teacher-led feedback, though valued for its contextual awareness and emotional intelligence, is constrained by time and availability. AI-powered feedback, on the other hand, offers immediacy, consistency, and scalability—raising the question of how the two modes can be integrated effectively.

This study proposes and explores a hybrid feedback model that blends AI-generated and teacher-provided support for oral English learning. By comparing student perceptions of both feedback sources through a mixed-methods approach, the study investigates how learners interpret the clarity, usefulness, and emotional impact of feedback. Findings from this research offer practical insights into how a hybrid model—combining autonomous chatbot practice with guided teacher support—can enhance spoken language instruction. The study further contributes to the growing body of work on teacher-AI collaboration in second language education.

LITERARUTE REVIEW

Teacher and AI Feedback in EFL Speaking Instruction

Teacher feedback is traditionally viewed as essential in language learning. According to Ellis (2009), teacher feedback supports learners' interlanguage development by providing scaffolded, individualized responses. However, in large-class contexts, timely feedback can be inconsistent, and oral feedback may lack permanence unless recorded or transcribed. This challenge is echoed by Carless (2006), who points out that in overcrowded classrooms, teachers often struggle to provide timely, individualized oral feedback. Moreover, as Hyland and Hyland (2006) argue, oral feedback tends to be ephemeral and may not leave lasting impressions unless captured through recordings or follow-up written notes.

Moreover, scholars like Boud and Molloy (2013) redefine feedback as a dialogic and learner-driven process, where students are actively engaged in interpreting and using feedback to regulate their own learning. This aligns with the role of teacher mediation but contrasts with AI systems, which often do not facilitate such reflective or self-regulatory engagement.

Research has also examined the ways in which different types of teacher feedback affect learner outcomes. Hyland and Hyland (2006) differentiate between direct and indirect feedback, with direct feedback providing corrections and indirect feedback prompting learners to identify and correct their own mistakes. Both forms play crucial roles depending on the learners' proficiency level and the instructional context. Wiliam (2011) further emphasizes formative assessment as an integral part of effective teaching, where feedback functions not only as information delivery but as an ongoing dialogue that shapes learning.

In terms of learner responses to feedback, several studies have highlighted the importance of emotional and motivational factors. Zhang and Hyland (2018) noted that students' attitudes toward feedback are significantly shaped by their previous experiences, their confidence level, and the perceived fairness and relevance of the feedback. Learning attitudes play a crucial role in determining the effectiveness of corrective feedback. Lyster and Saito (2010) argue that timely and context-appropriate corrective feedback promotes learner uptake and long-term language development. These findings underscore the complexity of feedback interactions and the need for sensitivity in providing correction.

Additionally, scholars such as Carless and Chan (2017) point out that students may not always understand or trust the feedback they receive, especially when it lacks transparency or actionable guidance. This brings attention to the importance of clarity and the co-construction of meaning in feedback exchanges. To achieve this, feedback must go beyond surface-level correction and invite learners into a dialogic process. Effective teacher feedback not only addresses linguistic accuracy but also engages learners in a reflective process that builds metacognitive awareness and learner agency (Nicol & Macfarlane-Dick, 2006).

Recent advancements in generative AI have introduced new pedagogical tools in language learning. In addition to AI chatbots that simulate interactive dialogue, offer grammar corrections, and generate contextual prompts for learners, other technologies such as automated speech recognition (ASR) systems are increasingly showing promise in speaking instruction. Meta-analytic evidence suggests that ASR interventions have a medium effect size (g = 0.69) in enhancing EFL learners' pronunciation performance. Similarly, practical implementations using ASR combined with peer correction have demonstrated significant improvements in pronunciation and overall speaking skills. Moreover, studies such as Xie et al. (2023) and Lee & Kim (2022) indicate that engaging with AI systems can reduce learner anxiety and foster autonomous speaking practice—thus pointing toward the multifaceted value of AI apps in supporting EFL spoken language development.

While AI feedback offers immediacy and scalability, it tends to focus on surface-level features such as grammar, pronunciation, and vocabulary (Li, Wang, & Liu, 2023; Xie et al., 2023). In contrast, teacher feedback extends beyond correctness and includes affective, strategic, and discourse-level support (Hyland & Hyland, 2006; Wiliam, 2011). Teachers are more likely to respond to non-verbal cues, scaffold student responses, and adjust their tone based on learner needs (Wiliam, 2011). Particularly in oral performance tasks involving opinion, emotion, or cultural nuance, human instructors are better equipped to provide adaptive and empathetic support (Goldstein, 2005). Particularly in oral performance tasks that elicit personal stories, emotional reflection, or culturally nuanced opinions, human instructors are better equipped to provide adaptive and empathetic support. These tasks require not only linguistic accuracy but also affective and pragmatic sensitivity from the interlocutor, dimensions in which teacher feedback remains uniquely valuable (Goldstein, 2005).

Student Perceptions of Feedback and Learner Engagement

Student engagement with feedback is closely linked to its perceived usefulness, emotional connection, and clarity (Zhang & Hyland, 2018). While AI tools like ChatGPT offer immediate and consistent feedback, they often lack the emotional sensitivity and adaptive interaction found in teacher support, which can affect learner motivation and retention (Liu, Zhang, & Wang, 2024).

Recent studies, such as Li et al. (2023) and Zou and Li (2022), suggest that combining AI and teacher feedback can significantly enhance speaking fluency and learner satisfaction over time. This points to the potential effectiveness of hybrid feedback models, which strategically leverage the strengths of both modalities.

Importantly, Carless and Boud (2018) emphasize that student perceptions of feedback are deeply influenced by how well the feedback promotes self-regulation and long-term improvement. Learners value feedback that not only informs but also empowers them to act. As Carless and Boud (2018) argue, sustainable feedback should foster learner agency and enable students to use feedback independently in future learning contexts. As a result, teacher feedback—particularly when dialogic and responsive—is more likely to be internalized and applied, a view also supported

by Nicol and Macfarlane-Dick (2006), who emphasize the importance of feedback that develops students' self-regulation and engagement with learning. AI feedback, while accessible, must evolve to support this kind of sustained learner engagement.

Additionally, learners' willingness to engage with feedback is often mediated by their emotional reactions to it. Yu, Jiang, and Zhou (2020) found that students are more likely to revise their work and reflect on errors when the feedback tone is supportive and non-judgmental. This suggests that emotional resonance plays a critical role in feedback uptake, a dimension where AI still lags significantly behind human instructors.

Sociocultural and Assessment Constraints on Feedback Effectiveness

Another critical limitation of current AI models is their difficulty in recognizing cultural appropriateness and pragmatic usage. As Ishikawa (2021) notes, AI systems often flag informal or colloquial expressions as errors, even when they are contextually appropriate in real-world conversations. Such tendencies may lead to overcorrection, which risks discouraging learners from developing authentic communication styles. Teachers, who possess cultural competence and context-awareness, are able to explain when and why such expressions are acceptable (Ishikawa, 2021).

In high-stakes exam contexts, such as IELTS or TOEFL, Qiu and Zhang (2023) found that AI-generated feedback may not align with nuanced scoring rubrics, especially when it fails to address discourse structure or register. In contrast, human instructors are adept at tailoring advice to individual student weaknesses within exam frameworks (Cheng, 2008; Weir, 2005).

Boud and Molloy (2013) argue that effective feedback must be "fit-for-purpose," embedded in specific learning contexts, and co-constructed with learners. This perspective reinforces the value of teacher mediation in socioculturally and pedagogically complex environments like oral performance assessments in Asia. Furthermore, Lee (2017) points out that feedback is often interpreted through the lens of institutional norms, learner expectations, and the perceived authority of the feedback provider—all of which influence its effectiveness.

In conclusion, the literature underscores the multi-faceted nature of feedback in language education. While AI offers exciting new avenues for increasing access and reducing instructor burden, it cannot yet replicate the relational, adaptive, and pragmatic depth of human feedback. A hybrid approach that draws on the strengths of both AI and teacher input may be best suited to meet the diverse needs of EFL learners in a range of contexts.

However, few empirical studies have examined how such hybrid feedback functions specifically in oral English learning, particularly from the learner's perspective. In light of its potential to combine the immediacy and scalability of AI with the empathy and contextual responsiveness of human instructors, this study explores how Chinese university EFL learners evaluate the clarity, usefulness, and

instructional value of AI-generated and teacher-provided feedback in task-based speaking activities.

Although the integration of AI in feedback provision is growing in prominence, there remains limited understanding of learner preferences and their perceived effectiveness of each feedback type, particularly in the context of spoken English development. This gap in the literature forms the impetus for the current study.

This study is guided by the following research questions:

1. How do Chinese university EFL learners perceive the usefulness of AI-generated feedback, teacher feedback, and their combination in oral English tasks?
2. What types of feedback (AI, teacher, or hybrid) are most likely to influence learner engagement and uptake during task-based speaking activities?
3. How do learners evaluate the strengths and limitations of AI and teacher feedback in terms of immediacy, clarity, emotional resonance, and contextual relevance?
4. What value do learners place on the use of AI chatbots compared with traditional teacher-based learning? What are some commonly shared perceptions learners have regarding the use of AI chatbots?

By addressing these questions, the study seeks to contribute empirical insight into the implementation of hybrid feedback models in EFL contexts and offer pedagogically grounded recommendations for optimizing spoken English instruction in technologically mediated learning environments.

METHODOLOGY

A mixed-methods research design was employed, combining quantitative analysis of Likert-scale survey responses using SPSS, and qualitative analysis of open-ended questions, which were coded thematically to examine how students evaluate the clarity, usefulness, and instructional value of AI versus teacher feedback.

While the primary focus of this study was to explore learners' in-depth perceptions of AI-generated feedback, data from the control group (teacher feedback group) were collected to provide contextual background and serve as a comparative reference during the quantitative phase. However, in the qualitative phase, only open-ended responses from the experimental group (AI feedback group) were thematically coded and analyzed. This design choice was made to better understand learners' experiences and evaluations of AI feedback specifically, which remains an underexplored area in current literature. Although no direct qualitative comparison was made between the two groups, the inclusion of a control group allowed for clearer interpretation of quantitative trends and provided a baseline for understanding learner perceptions.

Participants

Ten undergraduate students (7 female, 3 male), aged 19 to 25, from universities in mainland China participated in the study. All participants were non-English majors with College English Test Band 4 (CET-4) scores ranging between 425 and 500, equivalent to the B1–B2 range on the CEFR scale (roughly comparable to IELTS 5.0–6.0 or TOEFL iBT 45–70).

Participants were recruited through a third-party educational company—Zhanlu (Shanghai) Educational Technology Co., Ltd.—which screened applicants based on specific eligibility criteria provided by the researcher. These included: being a non-English major, willingness to participate in AI-assisted language learning experiments, and agreement to complete pre-tests, training tasks, post-tests, and surveys. The recruitment process and experimental procedures were conducted in accordance with the official cooperation agreement between the researcher and the company, with all sessions taking place at the company's designated research classroom in Shanghai.

Prior to the experiment, none of the students had substantial experience using AI tools (e.g., ChatGPT, Duolingo, etc.) for language learning. Due to access restrictions to ChatGPT in mainland China, none of the participants had prior experience using ChatGPT specifically.

Data Collection

This study adopted a mixed-methods research design, incorporating both quantitative and qualitative elements. The quantitative component involved a Likert-scale survey aimed at gauging learners' perceptions of AI-generated feedback in terms of clarity, usefulness, and instructional value. These responses were analyzed using SPSS to extract descriptive trends.

The qualitative component employed an inductive design (Creswell, 2009; Thomas, 2006), with a focus on participants' written responses to open-ended questions embedded in the same questionnaire. These open-ended prompts invited students to reflect freely on their experiences with AI feedback in task-based speaking activities. All responses were coded inductively using HyperRESEARCH version 4.5.7, and emergent themes were generated following a grounded theory approach (Saldaña, 2009). These themes informed the findings regarding learners' perceptions of AI's impact on their spoken English performance.

Procedure

The experiment was conducted over a two-week period and involved a single group of participants who received AI-generated feedback during oral English tasks (see Appendix 2, ChatGPT Prompt Guide for Oral Performance Assessment). All

participants completed the same set of nine oral communication tasks, which simulated real-life scenarios:

1. ordering food
2. asking for directions
3. booking a hotel
4. shopping
5. returning an item
6. making a doctor's appointment
7. engaging in small talk
8. introducing a city
9. airport check-in

For each task, participants interacted with ChatGPT for at least five turns in each dialogue. Each conversation began with a standardized pre-task prompt that instructed ChatGPT to act as a native speaker and engage the learner in a role-play. To ensure unbiased and individualized interactions, a new session was created for each student, and the cache was cleared between participants so that ChatGPT would not retain any memory of prior sessions. After the interaction, participants used a post-task prompt to request detailed feedback on their grammar, pronunciation, collocation, and natural phrasing. In some cases, they also asked ChatGPT to give a score or rate their speaking performance on a scale of 1–10.

All participants followed the same structured procedure for each speaking task, including identical prompts, timing, and feedback conditions, to ensure consistency and comparability across responses. For each session involving AI-generated feedback, a new ChatGPT interface was opened to prevent carryover effects from previous conversations, as ChatGPT retains conversational history within the same thread. On the post-task questionnaire, they were asked to compare AI-generated feedback with traditional teacher feedback.

RESULTS

This study employed a mixed-methods approach to analyze student perceptions of AI-generated and teacher feedback in EFL speaking tasks. Descriptive statistics, including mean scores and standard deviations, were calculated using SPSS. Table 1 presents the descriptive statistics (mean and standard deviation) for the 19 Likert-scale items from the student perception questionnaire. Among these items, Q20 ("In

the future, I would like to continue using AI as an auxiliary tool for speaking practice") received the highest mean score (M = 4.8, SD = 0.42), indicating strong student interest in integrating AI into future learning. In contrast, Q13 ("Compared with speaking with a person, the conversation with AI feels more natural") had the lowest mean score (M = 2.2, SD = 0.63), reflecting participants' perception that AI still lacks human-like conversational naturalness. These descriptive statistics provide a concise overview of how students perceive various aspects of AI feedback in spoken English learning.

Question No.	Mean	SD
Q4	4.3	0.823272602
Q5	4.1	0.737864787
Q6	4.0	0.816496581
Q8	4.4	0.699205899
Q9	4.0	0.666666667
Q10	4.4	0.843274043
Q11	3.9	1.100504935
Q13	2.2	0.632455532
Q14	4.2	0.632455532
Q15	4.7	0.483045892
Q16	3.4	0.843274043
Q17	2.5	0.971825316
Q20	4.8	0.421637021
Q21	3.1	0.875595036
Q22	4.9	0.316227766
Q24	4.3	0.823272602
Q25	4.2	0.632455532
Q26	3.4	0.843274043
Q27	3.9	0.737864787

Table 1. Descriptive Statistics of Student Perceptions of AI Feedback (N = 10)

To analyze the qualitative data, open coding was conducted using HyperRESEARCH version 4.5.7. A total of 117 responses to the two open-ended questions were coded. This initial coding phase yielded six code categories, which collectively contained 117 unique code instances (see Table 2 for the distribution of codes). At this stage, categories or themes had not yet been established—only individual codes were assigned to text segments.

Here are some examples of how codes were applied using HyperRESEARCH:

- *"AI is fast but confusing"* → Codes: speed, unclear feedback
- *"Teacher knows my level better"* → Codes: personalized feedback, teacher insight

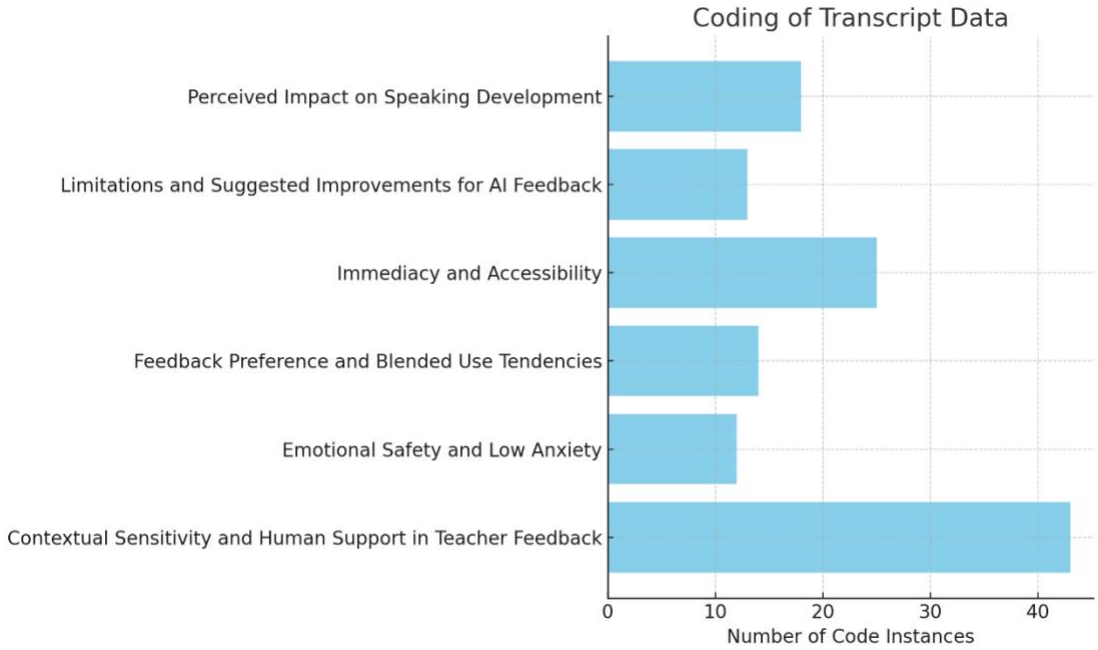
- *"I feel more confident talking to AI"* → Codes: emotional comfort, low anxiety

The three most frequently occurring codes were:

- speed (n = 15)
- personalized feedback (n = 13)
- emotional comfort (n = 12)

These patterns suggest emerging contrasts between student perceptions of AI-generated and teacher-provided feedback. A summary of all initial codes and their frequency is presented in Table 2.

Table 2: Coding of Transcript Data



The next stage involved grouping these individual codes into broader categories and identifying recurring themes across participant responses. Codes such as speed, instant response, and low anxiety were grouped into the category "efficiency and emotional comfort in AI Feedback". Similarly, codes like depth, contextual understanding, and nuance contributed to the theme "depth and pragmatic sensitivity in teacher feedback." Finally, responses that mentioned combining both feedback types or suggested future applications of AI contributed to the third theme, "complementarity and future potential of hybrid models."

Following the initial coding process, similar codes were grouped and refined through further analysis. This process revealed recurring patterns in students' responses, particularly regarding their perceptions of AI-generated feedback in speaking tasks. These patterns were consolidated into three overarching themes that represent the core insights drawn from the data.

1. efficiency and emotional comfort in AI feedback

2. depth and pragmatic sensitivity in teacher feedback

3. complementarity and future potential of hybrid models

These three salient themes reflect students' nuanced perceptions of AI-generated feedback, highlighting both its pedagogical strengths and limitations in supporting spoken English development.

Interpretation and Analysis of Theme 1: Efficiency and Emotional Comfort in AI Feedback

This theme highlights students' positive perceptions of AI feedback in terms of convenience, immediacy, and emotional safety. Participants frequently mentioned that AI provided a nonjudgmental space for oral English practice, which lowered anxiety and encouraged repeated speaking attempts.

S1 described ChatGPT as:

> "A practice partner that is always available...offering low-pressure opportunities for rehearsal and self-correction."

> S8 elaborated further on this emotional comfort:

> "I feel more relaxed practicing with AI ... I don't have to worry about making mistakes in front of a teacher ... It's like I can try again and again without anyone getting annoyed or impatient."

These quotes demonstrate the psychological safety that students associate with AI-based practice environments, contributing to increased confidence and risk-taking in language learning.

> However, despite these advantages, some students pointed out the emotional flatness of AI responses. S6 expressed dissatisfaction with the limited affective range of AI feedback:

> "The AI often just says 'Good job' ... It doesn't feel like real encouragement ... I want something that feels more human, something that shows the teacher really understands how hard I tried."

Such reflections reveal an important nuance: while AI fosters emotional comfort through its neutrality, it may lack emotional resonance and personalization.

This sense of psychological safety mirrors findings by Wang and Zhang (2022), who found that learners often feel more confident using AI tools due to their neutral

and consistent tone. Likewise, Godwin-Jones (2021) noted that AI tools, by offering instant and anonymous feedback, can boost learner autonomy and reduce the social pressure commonly associated with classroom interaction. Ranalli (2018) further critiques AI's inability to deliver context-aware or emotionally responsive feedback, which echoes the concerns raised by students in this study.

Overall, this theme suggests that students appreciate AI's stress-free environment for language learning, but also yearn for more authentic and emotionally intelligent interaction. The tension between emotional safety and emotional depth underscores a key area for future AI development in education.

Interpretation and Analysis of Theme 2: Depth and Pragmatic Sensitivity in Teacher Feedback

Compared to AI, participants viewed teacher feedback as more effective in managing nuanced, contextual communication. Many students emphasized how teachers could respond in real time and adapt their feedback based on subtle cues.

S7 shared:

> "She didn't interrupt me, but waited until I finished ... That kind of feedback feels very human."

S9 also highlighted this responsiveness:

> "I was struggling to finish a sentence, and the teacher waited patiently, then said, 'Maybe you meant…?' That really helped me complete my thoughts without embarrassment."

Students also appreciated that teacher feedback went beyond surface-level correction to include discourse-level instruction on tone, logic, and appropriateness.

S3 observed:

> "My teacher told me that 'actually' sounded rude in my sentence, and explained how tone matters in different situations."

S10 further emphasized the contextual depth of teacher support:

> "It's not just about grammar — my teacher explained how the sentence might be misunderstood in a real conversation. That's something AI can't really do yet."

These examples reveal the ways in which teacher feedback provides both cognitive scaffolding and emotional attunement, allowing students to develop more nuanced communicative competence.

These observations align with Hyland and Hyland (2006) and Goldstein (2005), who stress the interpersonal and pragmatic dimensions of teacher feedback, especially

in L2 settings. Carless (2006) also explored the interpretive gap between teacher intentions and student reception, suggesting that effective feedback requires shared understanding and contextual negotiation — something human teachers are uniquely capable of providing.

This theme reinforces the value of teacher feedback in promoting deeper linguistic awareness, affective support, and pragmatic sensitivity. These features are often lacking in AI systems, pointing to the continuing necessity of human mediation in advanced or emotionally charged learning contexts.

Interpretation and Analysis of Theme 3: Complementarity and Future Potential of Hybrid Models

Rather than viewing AI and teacher feedback as mutually exclusive, many participants described how the two modalities served different yet complementary roles in supporting speaking development. AI was perceived as helpful for independent, low-pressure, and repetitive practice, while teacher feedback was valued for more complex, emotionally sensitive, or high-stakes tasks.

> "When I practice speaking alone, ChatGPT is convenient, but for hard topics like giving opinions in exams, I still want my teacher to help." (S3)

> "I use AI when I want to warm up or test ideas. But when the topic is difficult, I feel safer if my teacher helps me organize my thoughts." (S9)

> "For general speaking, AI is fine. But before the final oral test, I always go to my teacher — it gives me more confidence." (S4)

These statements indicate that students are not choosing between AI and teachers, but rather integrating both based on task complexity and emotional needs.

This finding aligns with earlier studies. For instance, Liu et al. (2024) and Li et al. (2023) found that learners tend to alternate between AI tools and human guidance depending on the task's demands and affective intensity. In this way, learners act as agents who select feedback sources contextually, rather than relying solely on one channel.

Garrison and Vaughan (2008) discuss various types of hybrid models in education that combine technological tools with human facilitation, particularly in blended learning environments. The findings in this study extend this notion by providing empirical, learner-informed insights on how hybrid feedback can function in real-world speaking scenarios. Several participants also proposed ways to improve AI systems to better support hybrid learning environments.

S6 suggested adding casual pre-task interactions to mimic authentic conversation:

> "Maybe ChatGPT can say something like 'How's your day?' at the beginning — just like how teachers talk to us before starting."

This reflects the need for social presence and rapport-building, as discussed in Xie et al. (2023), who argue that conversational warmth significantly influences learner comfort and engagement.

S7 pointed out the need for more adaptive, cumulative feedback:

> "If I keep making the same mistake, AI should realize it and explain more. Not just correct me again and again."

This suggestion resonates with Ishikawa (2021), who emphasized the importance of error tracking and adaptive response in AI-driven feedback to promote long-term improvement.

The findings in this study support a learner-driven hybrid feedback model, one that recognizes the distinct strengths of both AI and teacher feedback in providing support. Rather than replacing human interaction, AI functions as a supplementary tool — particularly valuable for practice and repetition. Teachers, meanwhile, remain essential for personalized, high-stakes, and pragmatic feedback. By highlighting these preferences and concerns voiced by students, this study contributes to the ongoing conversation about how to design and balance roles in providing feedback. While this research does not propose a finalized pedagogical model, it offers grounded insights for future development of more responsive, learner-centered feedback systems.

DISCUSSION

The findings of this study highlight both the potential and limitations of AI-generated feedback on oral tasks in EFL contexts. In terms of efficiency and emotional comfort, students consistently appreciated the convenience, immediacy, and low-pressure nature of AI tools. These features made it easier for them to engage in frequent, independent speaking practice without fear of judgment. Such affordances align with prior studies (e.g., Wang & Zhang, 2022) that emphasize AI's accessibility and its capacity to foster a psychologically safe learning environment. The nonjudgmental tone and availability of AI systems appeared to reduce learners' anxiety and promote confidence, especially during solo practice sessions.

However, when considering depth and adaptability, notable limitations emerged. Participants reported that AI feedback lacked the ability to interpret nuanced discourse and conversational subtleties. It also struggled with providing context-sensitive corrections or emotionally supportive responses. Several students expressed frustration at having to input prompts to clarify ambiguous feedback, indicating that current AI systems often miss the pragmatic and prosodic cues essential for natural communication. These issues mirror previous research findings, such as those of Liu et al. (2024), which highlight the ongoing challenges in achieving human-like emotional intelligence and contextual awareness in AI-driven language tools.

Importantly, learners did not view AI and teacher feedback as interchangeable. Instead, they framed each feedback source as serving distinct yet complementary functions, pointing toward the complementarity and future potential of hybrid models. While in this study, AI was regarded as ideal for repetitive drills, pronunciation correction, and fluency practice, teacher feedback was valued for its empathetic tone, interpretive depth, and ability to tailor responses based on individual learner needs and cultural context. This suggests that a blended approach—leveraging the strengths of both—may offer the most comprehensive support for EFL learners. Rather than choosing between AI or human feedback, participants in this study favored their integration as a dual feedback system that adapts to different learning contexts and emotional needs. This suggests that a blended approach—leveraging the strengths of both—may offer the most comprehensive support for EFL learners.

In real-world applications, such a hybrid model could be implemented in several ways. For instance, language learning platforms could allow students to engage in AI-powered speaking drills outside class hours, receiving instant feedback on grammar and pronunciation. These sessions could then be followed by in-class teacher-led reviews, where educators address higher-order concerns such as discourse strategies, pragmatic appropriateness, and emotional tone. Alternatively, teachers might assign chatbot-based role-play tasks as homework and provide personalized feedback based on the learners' chatbot transcripts in the next lesson. This kind of coordinated approach would take advantage of AI's availability and consistency with the contextual intelligence and empathy of human instructors.

Future research might explore how such hybrid feedback systems can be operationalized effectively—including the design of teacher-AI collaboration workflows, optimal timing for feedback delivery, and learner preferences across different proficiency levels. Moreover, investigating how these systems influence long-term learner outcomes, such as autonomy, retention, and communicative competence, will be crucial for informing evidence-based practice in AI-assisted language education.

Critical to the success of this hybrid model is the role of teachers in designing and crafting effective prompts for AI systems. The success of chatbot-assisted learning heavily depends on the quality and clarity of the prompts fed into the system. To enhance the transparency and replicability of this study, the specific prompts used during the chatbot-assisted speaking tasks have been included in Appendix 2. When teachers carefully align these prompts with their instructional goals, AI-generated feedback can be more pedagogically meaningful and consistent with classroom learning objectives. Therefore, prompt engineering by educators should be considered a key component of future research and professional development in this area.

CONCLUSION

This study explored how Chinese university EFL learners perceive and evaluate AI-generated feedback compared to traditional teacher feedback in oral communication tasks regarding clarity, usefulness, emotional resonance, and instructional value. The

findings suggest that while AI feedback offers accessibility, immediacy, and anxiety-reducing benefits, it lacks emotional sensitivity and adaptive depth. In contrast, teacher feedback provides richer contextual and interpersonal support. In contrast, teacher feedback provides richer contextual and interpersonal support, but it can be limited by factors such as time constraints, varying teaching quality, and inconsistent availability, which may hinder its scalability and immediacy compared to AI.
These insights highlight the pedagogical potential of integrating AI into EFL instruction—not as a replacement but as a way to complement teacher feedback. A hybrid model that combines the efficiency and availability of AI with the nuanced, empathetic guidance of human instructors may optimize both speaking performance and learner motivation.

By combining qualitative and quantitative data in a Chinese university EFL context, this study offers empirical evidence supporting the practical feasibility of AI-assisted speaking instruction. However, due to the small number of participants and the specific context of this study, the findings cannot be generalized broadly. This research should be understood as a preliminary exploration of student perceptions of AI-generated feedback in speaking tasks. As such, the conclusions drawn here serve as a foundation for more extensive, cross-institutional investigations that could reveal broader trends and deeper insights.
In conclusion, this study demonstrates the complementary roles of AI and teacher feedback in oral English learning, pointing toward the potential of hybrid feedback models in future pedagogy.

REFERENCES

Benson, P. (2011). *Teaching and researching autonomy in language learning* (2nd ed.). Routledge.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233.

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

Chen, L., & Zhang, H. (2022). *Exploring the impact of automated feedback on EFL students' reading comprehension and engagement. Computer Assisted Language Learning, 35*(7), 1432–1451.

Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 349–364). Springer.

Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, *1*(1), 3–18.

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. Jossey-Bass.

Godwin-Jones, R. (2021). Emerging technologies: Rebooting language education with learning analytics**.** *Language Learning & Technology*, 25(1), 1–12.

Goldstein, L. M. (2005). *Teacher written commentary in second language writing classrooms*. University of Michigan Press.

Hyland, F., & Hyland, K. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101.

Ishikawa, S. (2021). Pragmatic failure in AI: A corpus-based analysis of appropriateness in chatbot responses. *Journal of Pragmatics*, *178*, 35–49.

Lai, C., & Zheng, Y. (2022). Student agency and autonomous learning with AI chatbots: An exploratory study. *System*, *106*, 102735.

Lee, I. (2017). Classroom writing assessment and feedback in L2 school contexts. *Springer*.

Lee, J., & Kim, H. (2022). Reducing EFL learners' anxiety through chatbot-supported speaking practice. *Language Learning & Technology*, *26*(2), 45–62.

Li, J., Sun, B., & Gao, M. (2023). Hybrid feedback in AI-supported EFL speaking classes: Effects on fluency and motivation. *TESOL Quarterly*, Advance online publication.

Li, J., Wang, X., & Liu, Y. (2023). Enhancing EFL learners' writing performance through AI-powered feedback: A mixed-methods study. *Journal of Second Language Writing, 62*, 101011.

Liu, F., Chen, Y., & Lin, M. (2024). AI is not my teacher: Chinese university students' views on the role of AI in English speaking practice. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 1–20.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis**.** *Studies in Second Language Acquisition*, 32(2), 265–302.

Qiu, Y., & Zhang, H. (2023). Evaluating the effectiveness of AI feedback in IELTS speaking test preparation. *Language Testing in Asia*, *13*(1), 1–16.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674.

Tan, R., Liu, M., & Zhao, Y. (2021). *Artificial intelligence in L2 writing: Effects of automated feedback on accuracy and complexity. ReCALL, 33*(3), 279–296.

Wang, Y., & Zhang, X. (2022). EFL learners' perceptions of AI-based speaking feedback: Benefits and barriers. *ReCALL*, *34*(3), 245–260.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach.* Palgrave Macmillan.

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

Xie, Y., Liu, M., & Wang, T. (2023). Exploring the role of AI chatbots in developing EFL learners' speaking skills: A mixed-methods study. *Computer Assisted Language Learning*, *36*(4), 321–342.

Yu, S., Jiang, L., & Zhou, N. (2020). Chinese university students' emotional responses to teacher written feedback in English writing**.** *Assessing Writing*, 45, 100469.

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102.

Appendix 1

AI Feedback Perception Survey

Basic Information (Fill-in Questions)

1. What is your gender? (Male / Female / Other / Prefer not to disclose)
2. What is your English proficiency level? (Please refer to CET-4 scores)
3. Have you ever used AI (e.g., ChatGPT, Duolingo, etc.) for language learning? (Yes / No)

Please rate the following statements based on your experience (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

- **Clarity of AI Feedback (Likert Scale + Open-ended Question)**

4. The feedback provided by AI is clear and easy to understand, allowing me to recognize my mistakes and corrections effortlessly.

1	2	3	4	5

5. AI feedback is detailed and targeted, avoiding confusion.

1	2	3	4	5

6. AI can distinguish between grammatical, lexical, and pronunciation errors instead of treating them the same way. （AI 在提供反馈时能够区分语法、用词和语音错误，而不会一概而论。）

1	2	3	4	5

Open-ended Question:

7. Have you ever encountered unclear or confusing AI feedback? If yes, please provide examples.

- **Usefulness of AI Feedback (Likert Scale + Open-ended Question)**

8. AI feedback makes it easier for me to identify and correct my mistakes.

1	2	3	4	5

9. The feedback provided by AI is useful for real-life spoken communication, not just for test preparation.

1	2	3	4	5

10. AI feedback makes me more confident in practicing spoken English.

1	2	3	4	5

11. AI provides feedback at an appropriate difficulty level based on my language proficiency, neither too simple nor too complex.

1	2	3	4	5

Open-ended Question:

12. What aspects of AI feedback do you find most helpful? What are the areas you are dissatisfied with?

- **Comparison: AI Feedback vs. Teacher Feedback (Likert Scale + Open-ended Question)**

13. Compared with speaking with a person, the conversation with AI feels more natural.

1	2	3	4	5

14. Compared with teacher feedback, AI feedback is more objective and free from personal bias.

1	2	3	4	5

15. AI feedback is more immediate than teacher feedback, providing instant suggestions.

1	2	3	4	5

16. AI feedback can provide personalized guidance, similar to a teacher, by addressing my specific issues.

1	2	3	4	5

17. AI feedback can replace teacher feedback in terms of motivation and encouragement.

1	2	3	4	5

Open-ended Questions:

18. Do you prefer AI feedback or teacher feedback? Why?

19. In what aspects do you think AI feedback is still inferior to teacher feedback?

- **Willingness to Use AI Feedback in the Future (Likert Scale + Open-ended Question)**

20. In the future, I would like to continue using AI as an auxiliary tool for speaking practice.

1	2	3	4	5

21. I believe AI feedback can partially or completely replace teacher feedback.

1	2	3	4	5

22. If AI feedback becomes more intelligent (e.g., incorporating contextual analysis, improved speech recognition), I would be more willing to use it for language learning.

1	2	3	4	5

Open-ended Question:

23. Under what circumstances would you prefer to use AI for speaking practice? Under what circumstances would you prefer teacher/peer feedback?

- **Impact of AI Feedback on Language Expression, Confidence, and Learning Strategies (Likert Scale + Open-ended Question)**

24. AI feedback makes me more willing to speak English.

1	2	3	4	5

25. AI feedback helps me improve sentence structure and accuracy in expression.

1	2	3	4	5

26. AI feedback encourages me to try using complex sentences and new vocabulary instead of sticking to simple sentences.

1	2	3	4	5

27. AI feedback makes me more aware of self-correction strategies, allowing me to adjust my speech in conversations.

1	2	3	4	5

Open-ended Questions:

28. What do you think is the most significant impact of AI feedback on your spoken English ability?

29. Besides AI feedback, what other methods do you use to improve your speaking skills?

30. Apart from the questions you have already answered, is there anything else you would like to say?

Appendix 2

ChatGPT Prompt Guide for Oral Performance Assessment

Part I

In Part I, participants interacted with ChatGPT in at least five turns in each dialogue. ChatGPT is not required to assess or rate learners' speaking performance.

1. Ordering food at a restaurant
2. Asking for directions in a city
3. Booking a hotel room

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's start with a role-play and practice a restaurant conversation. Imagine I am a customer. Please act as a waiter or waitress. I want to order lunch. Let's begin.

2. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play about asking for directions. Imagine I'm a tourist and you're a local resident. I will ask for help to find a place in the city. Let's begin.
3. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's start a role-play about booking a hotel room. I'll act as a guest and you'll be the hotel receptionist. I want to book a room for two nights. Let's begin.

Part II

In Part II, ChatGPT conducted six authentic real-life dialogues with the participants, and was given prompts to assess and evaluate their speaking performance.

6 Real-Life Tasks (aligned with TBLT):

1. Shopping for clothes
2. Returning an item
3. Making a doctor's appointment
4. Small talk at a party
5. Introducing your city to a foreign friend
6. Airport check-in

Scenario 1：Shopping for clothes

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play about shopping for clothes. I'll be a customer and you'll be the store assistant. I want to ask about sizes and prices. Let's begin.

2. Hi! I'd really appreciate it if you could correct my previous response—especially anything off in grammar, pronunciation, collocations, or natural phrasing. I look forward to your professional feedback. Thanks a lot!

3. How would you rate my speaking?/ Can you score my speaking performance?

Scenario 2：Returning an item

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play about returning a product. I'll act as a customer who wants to return an item I bought yesterday, and you'll be the store clerk. Let's begin.

2. Hi! I'd really appreciate it if you could correct my previous response—especially anything off in grammar, pronunciation, collocations, or natural phrasing. I look forward to your professional feedback. Thanks a lot!
3. How would you rate my speaking?/ Can you score my speaking performance?

Scenario 3：Making a doctor's appointment

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play about making a doctor's appointment. I'm a patient, and you're the receptionist at a clinic. I want to book an appointment for tomorrow. Let's begin.

2. Hi! I'd really appreciate it if you could correct my previous response—especially anything off in grammar, pronunciation, collocations, or natural phrasing. I look forward to your professional feedback. Thanks a lot!

3. How would you rate my speaking?/ Can you score my speaking performance?

Scenario 4：Small talk at a party

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play where we make small talk at a party. You can ask me some friendly questions and I'll try to respond naturally. Let's begin.

2. Hi! I'd really appreciate it if you could correct my previous response—especially anything off in grammar, pronunciation, collocations, or natural phrasing. I look forward to your professional feedback. Thanks a lot!

3. How would you rate my speaking?/ Can you score my speaking performance?

Scenario 5：Introducing your city to a foreign friend

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play where you are a foreign friend visiting my hometown. Please ask me questions about local food, tourist spots, or interesting places. Let's begin.

2. Hi! I'd really appreciate it if you could correct my previous response—especially anything off in grammar, pronunciation, collocations, or natural phrasing. I look forward to your professional feedback. Thanks a lot!

3. How would you rate my speaking?/ Can you score my speaking performance?

Scenario 6：Airport Check-in

1. Hi ChatGPT. You're an expert in spoken English. Please help me practice speaking. Let's do a role-play about airport check-in. I'm a passenger going to London, and you're the airport staff helping me check in. Let's begin.

2. Hi! I'd really appreciate it if you could correct my previous response—especially anything off in grammar, pronunciation, collocations, or natural phrasing. I look forward to your professional feedback. Thanks a lot!

3. How would you rate my speaking?/ Can you score my speaking performance?