Article

RELC Journal
1–17

DOI: 10.1177/00336882251341072
journals.sagepub.com/home/rel

The Role of GenAI in EFL Speaking: Effects on Oral Proficiency, Anxiety and Risk-Taking

Yiwei Chen
Hangzhou Normal University, China

Nuo Ke
The Chinese University of Hong Kong, China

Lan Huang
Hangzhou Normal University, China

Rong Luo
Zhejiang University, China

Abstract

This study investigated the impact of GenAI on students' English oral proficiency, foreign language speaking anxiety (FLSA), and linguistic risk-taking (LRT). Seventy-one Chinese university students learning English as a foreign language (EFL) participated in an 8-week programme for oral practice supported by GenAI. The students were divided into two groups: the high proficiency group (HG, $n = 37$) and the low proficiency group (LG, $n = 34$). Besides an assessment of English-speaking ability, they also completed surveys measuring their perceptions of their oral proficiency, anxiety and risk-taking before and after the programme. Results showed both groups experienced improvements in their English oral proficiency and LRT, accompanied by a noteworthy reduction in FLSA. The HG learners exhibited substantial enhancement in vocabulary among all speaking skills, whereas the LG learners experienced a notable progress in fluency. These findings provide insights into the effectiveness of GenAI in enhancing students' English-speaking learning outcomes and emotional dynamics.

Keywords

GenAI, EFL, oral proficiency, speaking anxiety, linguistic risk-taking

Corresponding author:
Rong Luo, School of International Studies, Zhejiang University, Hangzhou, Zhejiang 310058, China.
Email: luorong@zju.edu.cn

Introduction

Speaking is a key contributor for communication in the target language and shapes learners' overall language development (Kohn and Hoffstaedter, 2017). With insufficient opportunities for interactions in the English as a foreign language (EFL) context, learners often feel anxious and hesitant to speak in front of others (Peng, 2019). Therefore, promoting EFL learners' oral proficiency, reducing their anxiety and encouraging them to take linguistic risks remains a persistent challenge (Fernández-García and Fonseca-Mora, 2022).

To address these issues, a growing body of literature suggested that artificial intelligence (AI) offers innovative opportunities for speaking practice by providing automatic evaluation and feedback (Huang et al., 2021), and by adding naturalness and authenticity in oral practice (Han, 2024). Built on large language models (LLMs), generative artificial intelligence (henceforth GenAI) represented by ChatGPT outperformed early AI in its strong abilities to create human-like and meaningful text on various topics (Dalalah and Dalalah, 2023). Hence, GenAI has begun to be utilised in EFL settings. Despite the growing interest in GenAI for foreign language learning, empirical research on its efficacy in enhancing EFL learners' oral proficiency is still in its infancy. In addition, while there is evidence that early AI has been able to boost positive emotions (Tai and Chen, 2023), the studies on the impact of GenAI on EFL learners' emotions in oral practice remain limited.

To bridge these gaps, this study used ChatGPT 4o with Chinese university students to investigate its effects on oral English proficiency. It also aimed to reveal how GenAI influences Chinese EFL learners' anxiety and risk-taking because both positive and negative emotions are crucial factors affecting speaking achievements and outcomes (Lee and Hsieh, 2019). In addition, some scholars (e.g. Teng and Wang, 2024) have pointed out learners with varying proficiency levels may have different rates of progress and emotional responses in a technology-assisted learning environment. Thus, the current study also explores GenAI's influence on EFL learners of different oral proficiency levels. Our study addresses the following two research questions:

RQ1: What is the impact of ChatGPT on EFL learners' oral proficiency in high proficiency group (HG) and low proficiency group (LG)?

RQ2: How does ChatGPT affect EFL learners' anxiety and risk-taking in high proficiency group (HG) and low proficiency group (LG)?

Literature Review

GenAI and ChatGPT in Speaking Practice

Since its release in 2022, ChatGPT has gained wide attention for its ability to generate human-like text, facilitate authentic interactions and provide personalised feedback. A number of ChatGPT alternatives have been introduced since July 2023, such as Bard and Alpaca. They use deep learning models to create human-like outputs across different formats, including audio, images, text, three-dimensional (3D) objects, and even videos (UNESCO, 2023). Compared with early AI tools, GenAI offers language learners

a 'personalised, interactive, user-friendly and outcome-focused virtual learning environment' (Han, 2024).

The positive impact of GenAI in EFL learners' speaking practice is evident. Dwivedi et al. (2023) stated that GenAI could support improved speaking skills by offering instant feedback and refinement. Zhou et al. (2023) argued that sustained engagement with GenAI contributed to learners' growth in communicative competence. Bin-Hady et al. (2023) contended that English oral speaking practice might be optimised with GenAI's assistance to refine, improve, and enhance learners' responses. However, these advantages are primarily based on observations and lack empirical evidence. More empirical evidence is needed to measure the effect of GenAI thoroughly on speaking practice among language learners.

Speaking Anxiety and Risk-Taking in AI-Supported Language Learning

Emotions, whether positive or negative, significantly correlate with foreign language achievement (Dörnyei, 1998). Foreign language speaking anxiety (FLSA) refers to a psychological state characterised by nervousness and tension experienced during foreign language communication, particularly in front of others (Horwit et al., 1986). FLSA is caused by factors such as limited daily practice opportunities (Richards, 2015), restricted classroom time (Suzuki and Hanzawa, 2022), and the fear of making mistakes (Ebadi and Azizimajd, 2024). AI tools have been integrated into oral practice to ease anxiety and improve language gains. For instance, Du and Daniel (2024) reviewed AI chatbots used in EFL practice and found that half of the studies revealed the potential of AI chatbots in alleviating speaking anxiety. When interacting with AI chatbots, learners felt a 'flow state' focusing on oral activities in a supportive and non-judgemental atmosphere (Zhang et al., 2024). However, other studies argued that technical limitations of early AI technologies, including misinterpretation by natural language processing (NLP) technologies and automatic speech recognition (ASR) systems and inability to maintain coherent conversations may trigger anxiety (El Shazly, 2021; Ji et al., 2023). GenAI chatbots may surpass earlier AI tools in effectively alleviating anxiety among EFL learners. Yet, few empirical studies have touched FLSA in the context of using GenAI for EFL speaking practice.

Linguistic risk-taking refers to learners' readiness to get out of the comfort zone to use the target language in meaningful and authentic settings (Slavkov, 2023). This readiness is influenced by their willingness to communicate, the choices they make, and the educational setting (Sadoughi and Hejazi, 2024). For EFL learners, engaging in controlled risk-taking is vital as it enhances self-confidence and facilitates progress to higher proficiency levels (Brown, 2005). Therefore, fostering an environment that promotes risk-taking is crucial for helping students overcome their fear of communication (Lin and Lin, 2020). Recent studies have suggested that AI tools positively enhanced confidence and engagement by providing a low-pressure environment (Luria, 2024). Similarly, researchers such as Fathi et al. (2024) and Kim and Su (2024) have highlighted the positive role of AI chatbots in improving EFL learners' willingness to communicate and helping learners develop favourable attitudes. While previous studies have explored L2 learners' WTC, motivation, and confidence in chatbot-supported oral communication (Huang and Zou, 2024; Zhang et al., 2024), the crucial role of risk-taking in such processes has received notably less attention.

To address these limitations, this study examines the effects of GenAI on EFL learners' oral proficiency, speaking anxiety and linguistic risk-taking, aiming to enhance our understanding of the role of GenAI in enhancing EFL speaking skills and influencing learners' emotional states.

Methodology

Participants

Our participants are undergraduate students from a southern Chinese university. They were divided into a high proficiency group (HG, $n = 37$) and a low proficiency group (LG, $n = 37$). The HG consists of third and fourth-year English majors, while the LG comprises first-year non-English majors. Participants were selected from eligible volunteers with no prior GenAI-related language learning experience. Participants' English proficiency was first assessed by two experienced raters. Three students in the LG were excluded because their oral English performance was significantly above or below the group average. The final sample comprised two distinct groups: the HG with 37 participants (27 women and 10 men; 19–24 years old) and the LG with 34 participants (27 women and 7 men; 18–20 years old). In terms of English proficiency, HG participants demonstrated IELTS scores ranging from 6.0 to 7.5 (mean = 6.25), while LG participants scored between 3.5 and 5.5 (mean = 4.5) on the IELTS scale.

Research Setting

All the 71 participants volunteered to participate in our 8-week programme utilising ChatGPT 4o for English oral practice. The experimental procedure commenced with participants watching pre-recorded videos introducing the core functionalities of ChatGPT. This was followed by a live video conference, in which researchers presented a series of IELTS oral practice prompts and procedures for using these prompts in an oral conversation with ChatGPT. A question and answer (Q&A) segment was provided to clarify technical or procedural queries from participants. Participants then engaged in daily oral interactions with ChatGPT 4o through both listening and speaking.

Instruments

Oral Proficiency Assessment. Participants' speaking abilities were assessed using the IELTS speaking tests, which were conducted in both pre-tests and post-tests. These tests evaluated four key dimensions of oral proficiency: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. Scores were assigned on a scale from 1 to 9 for each dimension, following the official IELTS speaking criteria. The pre-test questions were randomly selected from the IELTS speaking test question pool for May to August 2024, and the post-test questions were selected from the same source to ensure consistency. To maximise reliability and accuracy in scoring, two experienced IELTS oral examiners independently rated the participants' performance. The inter-rater reliability was confirmed with a Cohen's kappa of 0.92.

Self-Report Oral Proficiency Questionnaire. Participants also completed self-report surveys measuring their perceptions of oral proficiency, anxiety, and risk-taking before and after the programme. The items in the self-report English oral proficiency questionnaire were adapted from Zou et al. (2023), with a Cronbach's alpha of 0.974. These items evaluated learners' comprehensive oral proficiency, covering fluency, pronunciation, reading aloud, grammatical range and accuracy, idea organisation, and presentation skills. Besides, this study incorporated an additional item assessing vocabulary in oral English, aligned with the IELTS speaking criteria. These questions were translated into Chinese to avoid possible misunderstandings. The scale showed high reliability, with a Cronbach's alpha of 0.90.

Foreign Language Speaking Anxiety (FLSA) Scale. A foreign language speaking anxiety (FLSA) scale, modified from Bashori et al. (2021), was applied to evaluate participants' degree of anxiety when communicating with ChatGPT 4o. The scale displayed strong reliability, with Cohen's kappa coefficient of 0.87.

Linguistic Risk-Taking (LRT) Scale. The linguistic risk-taking scale, adapted from Pyun et al. (2014), was employed to evaluate the willingness of EFL learners to engage in challenging linguistic tasks, even at the risk of potential embarrassment. The scale demonstrated high internal consistency with a Cronbach's alpha of 0.71.

Semi-Structured Interview. A semi-structured interview was conducted in Chinese to enrich the quantitative data. A total of 20 participants, ten from the HG and ten from the LG, were interviewed. The interview questions focused on participants' experiences with ChatGPT, exploring their perceptions of English pronunciation, intonation, grammatical competence, fluency and lexical richness. The interview also revealed learners' reflections of speaking anxiety and risk-taking, as well as any technical challenges encountered during the interaction.

Research Procedure

The 8-week programme in the present study was conducted during the summer vacation in 2024. The research procedure consisted of three phases as shown in Figure 1.

In week 1, participants took an IELTS speaking pre-test and were categorised into the HG and LG. Subsequently, both groups completed self-report oral proficiency, FLSA scale and LRT questionnaires, to assess their initial perceptions of English-speaking ability and emotional states. From week 2 to week 7, participants engaged in a conversation with ChatGPT 4o for at least 30 min a day. To ensure task completion, all participants were required to submit timestamped screenshots documenting their daily sessions through the group channel on DingTalk, a widely used instant messaging platform in China. In addition, they needed to follow the RTCF model, which guides them through four steps (role, task, conceptual framework, and feedback) to create effective prompts for ChatGPT (see Figure 2). As illustrated in Figure 2, ChatGPT was first asked to act in a dual role as trainer and interviewer. Subsequently, the participant directed ChatGPT to focus on a topic from the IELTS speaking test. In the third step, the participant instructed ChatGPT to generate a conceptual framework for addressing the

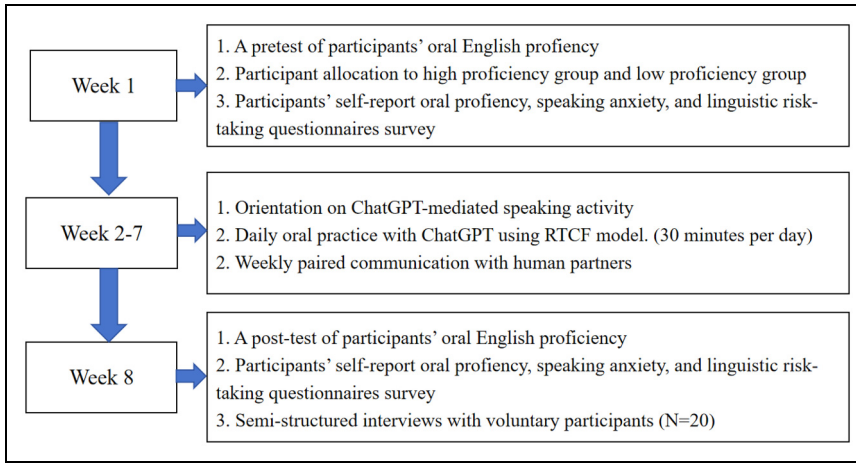

Figure 1. Research procedure.

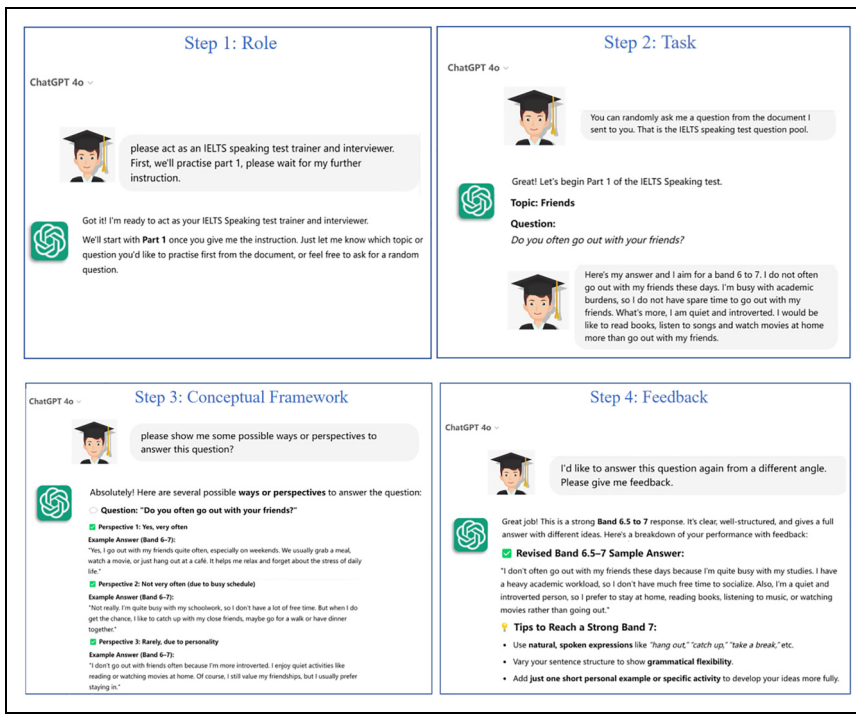

Figure 2. The RTCF model.

designated topic. Finally, following the participant's framework-based response, ChatGPT provided detailed feedback concerning language accuracy, vocabulary and coherence to improve the participant's answer. The topics for daily communication

with GenAI were selected from IELTS speaking test themes, including entertainment, education, personal experiences and future plans. The study concluded in week 8 with a post-test of English oral proficiency, followed by self-report oral proficiency, FLSA and LRT questionnaires survey. Semi-structured interviews were conducted to gather qualitative insights into the impact of ChatGPT on participants' oral performance and emotions.

Data Analysis

In this study, our dataset included pre and post-test results of oral proficiency, self-report questionnaire responses, and interview transcripts from both the HG and LG.

Quantitative Analysis. Quantitative data were analysed using SPSS 27.0. Paired-sample t-tests were conducted to assess within-group differences between pre-test and post-test oral proficiency scores, self-report questionnaire results of oral proficiency and emotional responses.

Qualitative Analysis. Thematic analysis was used to analyse the interview scripts, guided by the framework proposed by Braun and Clarke (2012). The interview data were coded to find key themes and patterns. To ensure accuracy, all interviews were audio-recorded and transcribed into digital format. Participants were also given unique identifiers (e.g. P1, P2…) to preserve their anonymity. The data analysis was conducted by two coders, and the consistency of their coding reached a score of 0.87. Any minor differences in coding were collaboratively discussed to ensure the reliability of our qualitative findings.

Results

Quantitative Results

Pre and Post-Oral Proficiency Test Results. Tables 1 and 2 demonstrate the descriptive data for HG and LG participants' performance across five dimensions of the IELTS speaking criteria. In Table 1, paired-sample t-tests results revealed that HG participants demonstrated significant improvements in vocabulary, grammar, fluency, and overall speaking.

Table 1. Paired-sample t-tests results of oral proficiency for HG.

	Pre-test		Post-test				
Themes	*M*	*SD*	*M*	*SD*	*t*	Sig.	Effect size
Pronunciation	6.662	0.566	6.784	0.641	−2.700	0.010	0.444
Vocabulary	6.243	0.548	7.054	0.587	−13.685	<0.001	2.250
Grammar	6.365	0.402	6.797	0.533	−6.400	<0.001	1.052
Fluency	6.459	0.477	7.068	0.579	−10.385	<0.001	1.707
Overall	6.444	0.427	7.042	0.540	−11.481	<0.001	1.913

N (HG) = 37.

Table 2. Paired-sample t-tests results of oral proficiency for LG.

	Pre-test		Post-test				
Themes	*M*	*SD*	*M*	*SD*	*t*	Sig.	Effect size
Pronunciation	4.471	0.748	5.441	0.860	−7.669	<0.001	1.315
Vocabulary	3.853	0.744	5.353	0.754	−11.097	<0.001	1.903
Grammar	4.000	0.739	5.338	0.725	−10.188	<0.001	1.747
Fluency	4.088	0.753	5.515	0.691	−14.050	<0.001	2.409
Overall	4.088	0.712	5.529	0.685	−12.956	<0.001	2.222

N (LG) = 34.

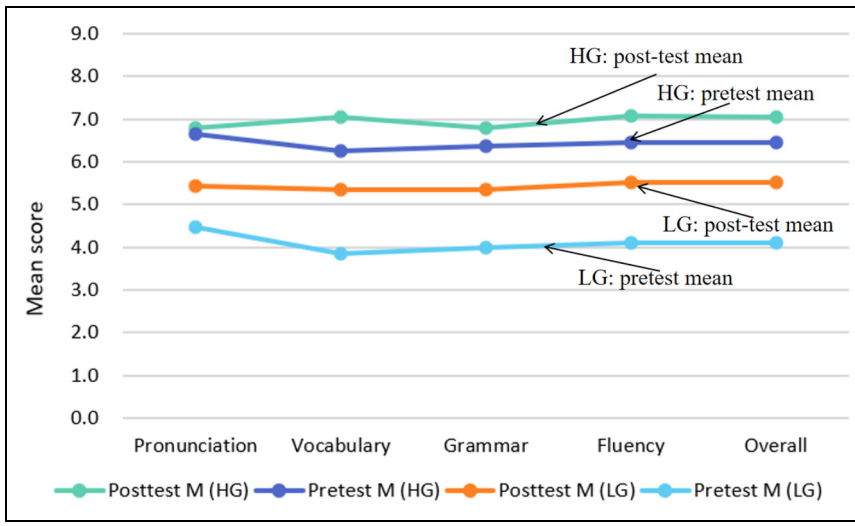

Figure 3. Oral proficiency performance in pre and post-tests for HG and LG.

Table 2 illustrates that LG learners also made significant progress across all dimensions. Figure 3 displays a substantial improvement in all speaking dimensions for both groups. Notably, for HG participants, the increase in vocabulary scores was particularly pronounced, while the improvement in pronunciation was modest. In contrast, LG participants demonstrated substantial gains in fluency. Additionally, the effect size for pronunciation among LG participants (Cohen's d = 1.315) was comparatively larger than that of the HG participants, suggesting that ChatGPT may be more effective in enhancing pronunciation for LG learners.

Comparison of Self-Report Oral Proficiency Results. Table 3 displays a significant self-perceived improvement of oral proficiency among HG learners. On average, their self-report oral abilities exhibited a notable increase. However, HG participants did not report a statistically significant improvement in fluency and intonation, which is different from the paired-sample t-tests results.

Table 3. Paired-sample t-tests results of oral self-evaluation for HG.

	Pre-test		Post-test				
Themes	*M*	*SD*	*M*	*SD*	*t*	Sig.	Effect size
Overall	3.253	0.500	3.787	0.377	−5.860	<0.001	0.554
Fluency	3.568	3.568	3.811	0.569	−1.946	0.059	0.320
Vocabulary	3.000	0.913	3.676	0.580	−3.995	<0.001	0.657
Grammar	2.892	0.875	3.351	0.716	−2.613	0.013	0.430
Pronunciation	3.378	0.953	4.108	0.516	−4.483	<0.001	0.737
Intonation	3.703	0.661	3.730	0.732	−0.255	0.800	0.042
Idea organisation	2.784	0.630	3.622	0.639	−5.167	<0.001	0.849
Reading aloud ability	3.919	0.759	4.243	0.548	−3.402	0.002	0.559
Presentation skills	2.784	0.712	3.757	0.641	−6.005	<0.001	0.963

N (HG) = 37.

Table 4. Paired-sample t-tests results of oral self-evaluation for LG.

	Pre-test		Post-test				
Themes	*M*	*SD*	*M*	*SD*	*t*	Sig.	Effect size
Overall	2.477	0.662	3.355	0.511	−7.993	<0.001	1.354
Fluency	2.765	0.819	3.382	0.604	−4.605	<0.001	0.756
Vocabulary	2.412	0.821	3.059	0.600	−4.293	<0.001	0.705
Grammar	2.353	0.849	3.000	0.696	−3.605	<0.001	0.622
Pronunciation	2.588	1.019	3.529	0.706	−5.026	<0.001	0.835
Intonation	2.559	0.894	3.353	0.646	−5.413	<0.001	0.940
Idea organisation	1.941	0.736	3.294	0.760	−8.971	<0.001	1.594
Reading aloud ability	3.176	1.058	3.824	0.797	−3.823	<0.001	0.660
Presentation skills	2.000	0.853	3.382	0.817	−7.657	<0.001	1.288

N (LG) = 34.

Table 4 illustrates the impact of ChatGPT on LG learners' self-report oral abilities. Overall, they exhibited a remarkable enhancement in their oral proficiency, with their self-evaluations significantly increasing from a pretest mean of 2.493 to a post-test mean of 3.361 with a large effect size (Cohen's $d = 1.351 > 1.2$).

The findings suggest that learners in both HG and LG perceived ChatGPT as beneficial in enhancing their overall oral proficiency and specific dimensions. These results demonstrate the potential effect of ChatGPT in facilitating improvements in EFL learners' English spoken proficiency.

ChatGPT's Impact on EFL Learners' Anxiety and Risk-Taking. The changes in participants' self-perceived FLSA and LRT are represented in Figures 4 and 5. It is evident that both HG and LG experienced a decrease in FLSA and an improvement in LRT. More detailed data are provided in Tables 5 and 6.

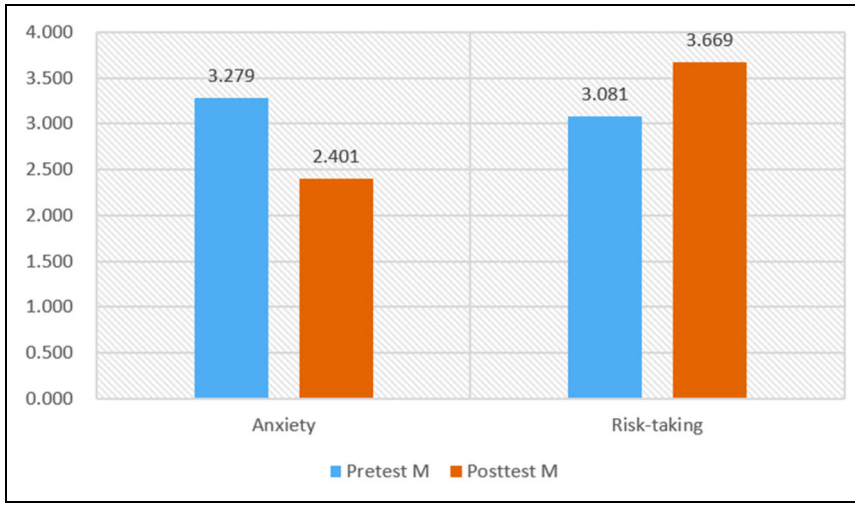

Figure 4. The changes in anxiety and risk-taking for HG.

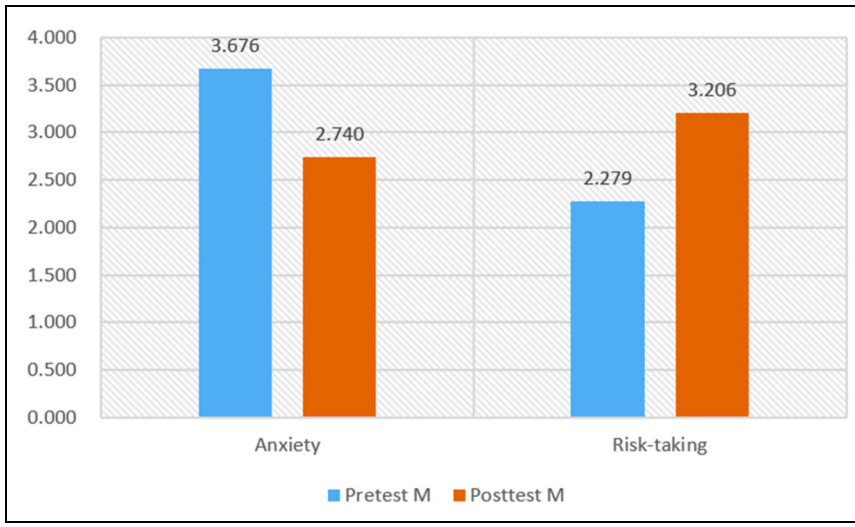

Figure 5. The changes in anxiety and risk-taking for LG.

Table 5 displays the results reflecting emotional states of HG participants. An obvious decrease in the mean anxiety score is manifested. Conversely, risk-taking significantly increased. These findings indicate ChatGPT's positive impact on reducing speaking anxiety and enhancing risk-taking among HG learners. In Table 6, LG participants also experienced a significant reduction in anxiety and a marked increase in risk-taking. Overall, these results suggest that interventions involving ChatGPT effectively reduce anxiety and foster risk-taking for both HG and LG participants.

Table 5. Paired samples t-test results of HG's anxiety and risk-taking.

	Pre-test		Post-test				
Themes	*M*	*SD*	*M*	*SD*	*t*	Sig.	Effect size
Anxiety	3.279	0.865	2.401	0.591	6.639	<0.001	1.092
Risk-taking	3.081	0.731	3.669	0.540	−5.401	<0.001	0.888

N (HG) = 37.

Table 6. Paired samples t-test results of LG's anxiety and risk-taking.

	Pretest		Post-test				
Themes	*M*	*SD*	*M*	*SD*	*t*	Sig.	Effect size
Anxiety	3.676	0.772	2.740	0.650	5.726	<0.001	0.982
Risk-taking	2.279	0.615	3.206	0.716	−6.625	<0.001	1.136

N (LG) = 34.

Qualitative Results

To explore further why EFL learners demonstrated quantitative improvements across nearly all dimensions of oral skills after ChatGPT-mediated interactions, semi-structured interviews were conducted in Chinese. This study used thematic analysis (Braun and Clarke, 2012) to reveal themes reflecting EFL learners' attitudes and perceptions towards the use of ChatGPT. These themes were categorised into perceptions of oral proficiency changes and emotional states after GenAI-mediated oral practice. We numbered the 20 participants from number 1 to 20, with P1 to P10 categorised as the HG interviewees and P11 to P20 as the LG interviewees.

Theme 1: Enhancement of Vocabulary and Idea-Generation Skills among HG Learners. In the HG group, many interviewees emphasised ChatGPT's capacity to enhance vocabulary, foster logical reasoning, and stimulate idea generation. Notably, they highlighted ChatGPT's role in expanding their vocabulary on specific topics and aiding the transition of passive vocabulary into active use. This enhancement is attributed to ChatGPT's role as an interlocutor in introducing topic-related phrases (P4), providing native-like expressions (P2), and promoting concise communication (P9). P6 shared:

> Since I'm a gym enthusiast, I used to say 'go to the gym'. However, thanks to ChatGPT, I've adopted a more casual phrase, like 'hit the gym'. After hearing it a few times, it slipped into my daily conversations without me even realizing it. My vocabulary has significantly improved and my language has become more refined and sophisticated.

Moreover, many HG interviewees emphasised ChatGPT's role as a concept organiser, offering multiple perspectives on a given topic (P4, P7, P9, P10). P9 provided one example:

> When confronted with unfamiliar topics, I often found myself unsure of where to begin. However, ChatGPT provided me with various angles to approach the question. For instance, during a discussion about disliked advertisements, ChatGPT suggested intrusive or disruptive ads could lead to disdain, offering me an insightful perspective I hadn't previously considered.

Theme 2: Enhancement of Fluency and Grammar among LG Learners. LG interviewees highlighted two aspects less frequently mentioned by HG interviewees: fluency and grammar. Regarding fluency, many LG learners reported that they could now articulate more in a single turn while speaking English. They agreed that ChatGPT served as a language role model in enhancing their fluency (P14, P17, P19, P20). For instance, P19 shared:

> I used to speak hastily without organizing my thoughts, which affected my fluency. However, ChatGPT helped me structure my responses by providing a topic sentence and linking details with transitional words. This enabled me to grasp general speaking patterns, avoid fumbling for words, and become more fluent in my speech.

In terms of grammar improvement, LG interviewees noted a conscious shift towards using more varied and complex grammatical structures, facilitated by ChatGPT's feedback. One example comes from P17:

> Before practicing with ChatGPT, I only used simple sentences but now I dare to use complex attributive clauses with 'which' or 'that'.

Theme 3: Divergent Perspectives on Pronunciation Improvement. HG and LG learners' views on pronunciation improvement diverge. HG interviewers attributed this limitation to ChatGPT's robotic voice. P2, P4 and P8 stated ChatGPT's speech output lacked natural pauses, stress patterns and right intonation, which made it less effective in improving pronunciation. LG interviewees, however, believed that ChatGPT played an effective role in improving their pronunciation. They argued that ChatGPT's inability to recognise certain pronunciations made them more aware of their articulation errors, prompting efforts to correct these mistakes and ultimately leading to pronunciation improvement (P18, P20). P11 stated, '*In order to make sure ChatGPT understands me, I have to pronounce the words clearly and correctly, which ChatGPT helps me achieve'.*

Theme 4: Reduced EFL Speaking Anxiety. Several interviewees (P2, P4, P5, P6, P7, P8, P13, P20) reported that ChatGPT provided a low-pressure environment for practising speaking skills. They appreciated the opportunity to interact with it without feeling self-conscious or judged, creating a safe space for language practice. As P8 stated:

> I used to experience intense anxiety whenever I had to answer questions in English, especially in front of classmates, as I feared being judged or even laughed at. However, ChatGPT helps me improve my oral performance without judgment, as it is robotic. I never feel embarrassed or nervous when talking with ChatGPT.

Furthermore, some learners noted that practising with ChatGPT five times a week over 6 weeks offered ample speaking opportunities in a non-threatening environment. This consistent practice allowed them gradually to feel at ease when speaking in an all-English setting, leading to enhanced fluency and, ultimately, a reduction in anxiety.

Theme 5: Enhanced Language Risk-Taking. Both HG and LG interviewers agreed that ChatGPT provided a relaxing environment and thus increased their willingness to take linguistic risks. Some LG learners strongly believed that their fluency could improve because ChatGPT would not misunderstand them, even with their limited English proficiency (P15, P16, P19). HG learners noted that ChatGPT could easily grasp their intended meaning, even when they occasionally misused complex terms. This provided them with opportunities to test their acquired knowledge, thereby encouraging them to speak more freely and bravely (P3, P4, P5, P9). For instance, P3 remarked:

> ChatGPT can always understand whatever I say and provide appropriate feedback, which encourages me to talk freely. I once described 财阀 as 'the bigger joint company that controls the economy', and ChatGPT intelligently identified it as 'chaebol'. However, in a real classroom, I would just use the words I have already mastered and stick to topics I'm familiar with.

In summary, EFL learners who dared to express themselves to GenAI reported improvements in their language proficiency. By practising newly acquired vocabulary or phrases without fear of making mistakes, learners develop the ability to self-correct and continue practising. As a result, they experienced a reduction of speaking anxiety and an enhancement of risk-taking.

Discussion

This study investigated the impact of GenAI on improving EFL learners' oral proficiency as well as its effects on learners' anxiety and risk-taking. For the first research question, the findings indicated that learners in both high and low proficiency groups made notable progress, with HG learners showing significant improvement in vocabulary and LG learners in fluency. HG learners' vocabulary enhancement was driven by ChatGPT's ability to introduce advanced and contextually specific language. This enabled HG learners to incorporate more complex vocabulary in appropriate situations (Yildiz, 2023). LG learners' gains in fluency resulted from an increased exposure to native-like speech, which helped them shift from interlingual to intralingual strategies, and relied more on the target language's structures and vocabulary (Chen, 2006). These findings diverged from previous findings that chatbots were more beneficial for learners with higher oral proficiency (Coniam, 2014), or for lower-proficiency learners because advanced learners might feel reluctant to engage with chatbots (Yin and Satar, 2020). The inconsistencies in findings may stem from the different AI tools used, which differ in algorithms, models, datasets, and the maturity of the technology. In our study, it was observed that learners' oral proficiency was enhanced across multiple dimensions, as GenAI can dynamically adapt to learners' initial proficiency levels and support their progress through personalised interactions. Another significant finding was that LG learners attained substantial improvement in pronunciation, while HG learners reported no statistically significant gains. This contrasts with previous studies that emphasised the effectiveness of chatbots in enhancing English pronunciation for all users (Du and Daniel, 2024). Although HG learners attributed their less significant improvement in pronunciation to ChatGPT's robotic voice, the actual reason might be that most HG learners already had relatively good pronunciation. Their short-term exposure to AI-mediated interactions did not yield statistically significant enhancement in pronunciation skills.

Therefore, we arrive at the conclusion that GenAI positively impacted EFL learners with varying oral proficiency levels. The outcomes are consistent with previous studies by Bin-Hady et al. (2023) and Mohamed (2024), which emphasised the potential of GenAI in enhancing EFL learners' oral proficiency. The observed improvements may result from ChatGPT's human-like interaction abilities, which is beneficial for the development of fluency and the internalisation of linguistic knowledge (Tai and Chen, 2024). Besides, GenAI tools offer real-time feedback ensuring a smooth conversational flow and minimises communication breakdowns (Dal Pian et al., 2024). Moreover, GenAI facilitates a translanguaging space (Turner and Lin, 2024), in which participants switch freely between Chinese and English as needed to achieve their communicative goals.

For the second question, both HG and LG participants experienced a reduction in anxiety and a greater willingness to take linguistic risks in English speaking practice. These emotional changes are related to the enhanced speaking abilities resulting from sustained practice with ChatGPT. These findings are also aligned with previous research showing that communicative activities can affect anxiety (MacIntyre, 2017) and risk-taking (Dehbozorgi, 2012). Firstly, GenAI optimises learners' performance by offering authentic language input, real-time feedback, and personalised guidance, equipping them with the linguistic resources needed for real-life conversations. The more learners engage with ChatGPT, the more comfortable they become in speaking English, ultimately overcoming the fear of using the target language. Secondly, GenAI provides a non-judgmental, stress-free environment that encourages learners to share their ideas and feelings, thereby alleviating anxiety and fostering greater risk-taking in language production. The findings align with previous studies by Dizon (2017) and Tai and Chen (2023), showing that GenAI enhances motivation and willingness to communicate in English, subsequently facilitating oral competence and emotional wellbeing.

Conclusions and Limitations

This study investigated the impact of ChatGPT, a GenAI tool, on the oral proficiency of EFL learners across varying proficiency levels. Both HG and LG participants demonstrated substantial improvements in English oral proficiency in both speaking tests or self-report results. This progress was accompanied by a notable reduction in speaking anxiety and an increase in risk-taking behaviour throughout the 8-week intervention.

Although this study offers valuable insights, several limitations should be acknowledged. First, the absence of a control group limits our ability to compare the observed changes among our participants with those who did not use GenAI, potentially affecting the generalisability of the findings. Second, the experiment lasted 8 weeks, which may not adequately capture the long-term effects of GenAI tools on EFL learners' oral proficiency and psychological dynamics. Another limitation is the study's focus on the context of higher education. Consequently, future research should integrate GenAI into different educational levels and contexts to gain a more comprehensive view. However, it is a promising direction to integrate GenAI into relatively input-poor EFL learning environments in many countries. This study provides pedagogical implications for EFL learning and for language instructors seeking to transform conventional foreign language classrooms. The implementation of models such as the RTCF framework can facilitate conversational practice in class, enabling students to engage in authentic dialogue exercises centered on curriculum-related topics. Moreover, the supportive, empathetic,

and responsive nature of GenAI interactions enables the design of oral class frameworks that may seamlessly integrate pre-class, in-class, and post-class activities.

In conclusion, this study provides a more comprehensive view of the technology's benefits and challenges. It provides empirical evidence that reinforces the role of generative AI as a facilitator, organiser, and interlocutor in linguistic interactions. Notably, the study also highlights its effectiveness in reducing anxiety and encouraging users to take risks.

Authors' Contribution Statement

Yiwei Chen: writing-original draft, review and editing. Nuo Ke: methodology, data analysis. Lan Huang: methodology, data analysis. Rong Luo: project administration, funding acquisition.

Ethical Approval and Informed Consent Statements

The Ethics Review Committee at Hangzhou Normal University approved our interviews (approval: 2408) on 10 May 2024. Respondents gave written consent for review and signature before starting interviews.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Yiwei Chen https://orcid.org/0009-0003-6493-5585
Nuo Ke https://orcid.org/0009-0007-0580-5229
Lan Huang https://orcid.org/0009-0006-0922-7668
Rong Luo https://orcid.org/0000-0001-5069-6200

Supplemental Material

Supplemental material for this article is available online.

References

Bashori M, Van Hout R, Strik H, et al. (2021) Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System* 99: 102496.

Bin-Hady WRA, Al-Kadi A, Hazaea A, et al. (2023) Exploring the dimensions of ChatGPT in English language learning: a global perspective. *Library Hi Tech*. https://doi.org/10.1108/LHT-05-2023-0200

Braun V and Clarke V (2012) Thematic analysis. In: Cooper H, Camic PM, Long DL, et al. (eds) *APA Handbook of Research Methods in Psychology, Vol. 2. Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological*. Washington, DC: American Psychological Association, pp. 57–71. https://doi.org/10.1037/13620-004

Brown SL (2005) Relationships between risk-taking behaviour and subsequent risk perceptions. *British Journal of Psychology* 96(2): 155–164.

Chen S (2006) A study of communication strategies in interlanguage production by Chinese EFL learners. *Language Learning* 40(2): 155–187.

Coniam D (2014) The linguistic accuracy of chatbots: usability from an ESL perspective. *Text and Talk* 34(5): 545–567.

Dalalah D and Dalalah OM (2023) The false positives and false negatives of generative AI detection tools in education and academic research: the case of ChatGPT. *The International Journal of Management Education* 21(2). https://doi.org/10.1016/j.ijme.2023.100822

Dal Pian MC, Dal Pian LF and Dal Pian M (2024) Learning through conversation in the era of generative artificial intelligence. In: *EDULEARN24 Proceedings*. IATED, pp. 6684–6693. https://doi.org/10.21125/edulearn.2024.1586

Dehbozorgi E (2012) Effects of attitude towards language learning and risk-taking on EFL student's proficiency. *International Journal of English Linguistics* 2(2): 41–48.

Dizon G (2017) Using intelligent personal assistants for second language learning: a case study of Alexa. *Tesol Journal* 8(4): 811–830.

Dörnyei Z (1998) Motivation in second and foreign language learning. *Language Teaching* 31(3): 117–135.

Du J and Daniel BK (2024) Transforming language education: a systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education. Artificial Intelligence* 6: 100230.

Dwivedi YK, Kshetri N, Hughes L, et al. (2023) So what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management* 71(2): 155–187.

Ebadi S and Azizimajd H (2024) Exploring the impact of role-playing in clubhouse on the oral fluency and willingness to communicate of EFL learners. *Education and Information Technologies* 29(7): 7927–7968.

El Shazly R (2021) Effects of artificial intelligence on English speaking anxiety and speaking performance: a case study. *Expert Systems* 38(3): e12667.

Fathi J, Rahimi M and Derakhshan A (2024) Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System* 121: 103254.

Fernández-García A and Fonseca-Mora MC (2022) EFL Learners' speaking proficiency and its connection to emotional understanding, willingness to communicate and musical experience. *Language Teaching Research* 26(1): 124–140.

Han Z (2024) Chatgpt in and for second language acquisition: a call for systematic research. *Studies in Second Language Acquisition* 46(2): 301–306.

Horwit EK, Horwitz MB and Cope J (1986) Foreign language classroom anxiety. *The Modern Language Journal* 70(2): 125–132.

Huang F and Zou B (2024) English Speaking with artificial intelligence (AI): the roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior* 159: 108355.

Huang W, Hew KF and Fryer LK (2021) Chatbots for language learning-are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning* 38(1): 237–257.

Ji H, Han I and Ko Y (2023) A systematic review of conversational AI in language education: focusing on the collaboration with human teachers. *Journal of Research on Technology in Education* 55(1): 48–63.

Kim A and Su Y (2024) How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System* 122: 103256.

Kohn K and Hoffstaedter P (2017) Learner agency and non-native speaker identity in pedagogical lingua franca conversations: insights from intercultural telecollaboration in foreign language education. *Computer Assisted Language Learning* 30(5): 351–367.

Lee JS and Hsieh JC (2019) Affective variables and willingness to communicate of EFL learners in in-class, out-of-class, and digital contexts. *System* 82: 63–73.

Lin Y and Lin Y (2020) The enlightenment of affective filter hypothesis and risk-taking on English learning. *Studies in Literature and Language* 20(2): 51–57.

Luria E (2024) Exploring the use of Alexa as a motivational tool in English language learning among EFL students in Israel. In: Papadakis S (eds) *Iot, AI, and ICT for Educational Applications: Technologies to Enable Education for All*. Cham: Springer Nature Switzerland, pp. 77–98. https://doi.org/10.1007/978-3-031-50139-5_4

MacIntyre PD (2017) An overview of language anxiety research and trends in its development. *New Insights into Language Anxiety: Theory, Research and Educational Implications*. 11–30. https://doi.org/10.21832/9781783097722-003

Mohamed AM (2024) Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. *Education and Information Technologies* 29(3): 3195–3217.

Peng JE (2019) The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English. *System* 82: 161–173.

Pyun DO, Kim JS, Cho HY, et al. (2014) Impact of affective variables on Korean as a foreign language learners' oral achievement. *System* 47: 53–63.

Richards JC (2015) The changing face of language learning: learning beyond the classroom. *Relc Journal* 46(1): 5–22.

Sadoughi M and Hejazi SY (2024) Growth language mindset and willingness to communicate: contributions of linguistic risk taking and learning experience. *Acta Psychologica* 248: 104367.

Slavkov N (2023) Linguistic risk-taking: a new pedagogical approach and a research program. *Canadian Journal of Applied Linguistics* 26(2): 32–59.

Suzuki Y and Hanzawa K (2022) Massed task repetition is a double-edged sword for fluency development: an EFL classroom study. *Studies in Second Language Acquisition* 44(2): 536–561.

Tai TY and Chen HHJ (2023) The impact of google assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments* 31(3): 1485–1502.

Tai TY and Chen HHJ (2024) Improving elementary EFL speaking skills with generative AI chatbots: exploring individual and paired interactions. *Computers and Education* 220: 105112.

Teng Y and Wang X (2024) Full caption, English proficiency and their relationships with behavioral, cognitive and emotional student engagement in Chinese EFL college content-based instruction video learning. *Education and Information Technologies* 29(1): 861–880.

Turner M and Lin AM (2024) Translanguaging: process and power in education. *Linguistics and Education* 83: 101340.

UNESCO (2023) Guidance for generative AI in education and research. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000386693

Yildiz M (2023) Exploring the potential of chatbots and ChatGPT in enhancing vocabulary knowledge. In: Kartal G (eds) *Transforming the Language Teaching Experience in the Age of AI*. IGI Global, pp. 160–188.

Yin Q and Satar M (2020) English as a foreign language learner interactions with chatbots: negotiation for meaning. *International Online Journal of Education and Teaching* 7(2): 390–410. http://iojet.org/index.php/IOJET/article/view/707

Zhang C, Meng Y and Ma X (2024) Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System* 121. https://doi.org/10.1016/j.system 2024. 103259

Zhou J, Ke P, Qiu X, et al. (2023) ChatGPT: potential, prospects, and limitations. *Frontiers of Information Technology and Electronic Engineering* 1–6. https://doi.org/10.1631/FITEE. 2300089

Zou B, Guan X, Shao Y, et al. (2023) Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability* 15(4): 2872.