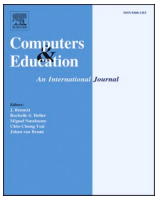
Computers & Education 220 (2024) 105112

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu



Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions

Tzu-Yu Tai [a], Howard Hao-Jan Chen [b,*]

[a] Language Center, Taipei Medical University, Taipei, Taiwan
[b] Department of English, National Taiwan Normal University, Taipei, Taiwan

ARTICLE INFO

Keywords:
Generative artificial intelligence (GAI)
Chatbots
EFL speaking
Cooperative/collaborative learning
Elementary education

ABSTRACT

Generative artificial intelligence (GAI) and automatic speech recognition (ASR) have ushered in promising tools for foreign language learning, notably GAI chatbots. This study investigated the impact of GAI chatbots on elementary school English as a foreign language (EFL) learners' speaking skills, focusing on two interaction configurations—individual and paired. Eighty-five elementary school EFL learners participated in a three-week summer program, engaging in daily 45-min interactions with CoolE Bot. The participants were randomly assigned to three groups: (1) individual interaction with CoolE Bot (I-Bot group), (2) paired interaction with CoolE Bot (P-Bot group), and (3) interaction with teachers and peers in a conventional English classroom (No-Bot group). In each class, participants in the Bot group received worksheets with a topic, prompts, and vocabulary to guide their interactions with CoolE Bot, while those in the No-Bot group also received worksheets for comparable activities. Quantitative (English-speaking tests) and qualitative data (semi-structured interviews) were collected and analyzed. Results revealed that the I-Bot and P-Bot groups' post-test speaking skills were significantly higher than those of the No-Bot group. CoolE Bot significantly improved the speaking skills of EFL learners. Both individual and paired interactions with CoolE Bot demonstrated positive effects, with no significant differences between groups. Interviews highlight CoolE Bot's adeptness in coherent interaction, charismatic conversational style with a human-like voice, diverse topic discussions tailored to learners' interests, and supportive functions. The participants found GAI chatbot-assisted EFL speaking enjoyable, motivating, and engaging appreciating its cartoonish, human-like characters, conversational style, and voice. Additionally, CoolE Bot fostered rapport and a supportive environment enhancing learners' confidence and reducing anxiety regarding EFL speaking. Individual interactions encourage personalized engagement and self-directed learning, whereas paired interactions involve social dynamics, shared learning experiences, and mutual resolution of language challenges.

1. Introduction

Speaking skills significantly impact learners' overall language proficiency (Rahimi & Fathi, 2022). Researchers have advocated the use of artificial intelligence (AI) chatbots for second and foreign language (L2) speaking (Jeon et al., 2023; Tai, 2024; Wang et al., 2023; Yang et al., 2022). AI chatbots are increasingly applied in educational fields to support the learning of various disciplines,

* Corresponding author. Department of English, National Taiwan Normal University, 162 East Hoping Road, Sec. 1, Taipei City, Taiwan.
E-mail addresses: a0937546693@gmail.com (T.-Y. Tai), hjchen@ntnu.edu.tw, hjchenntnu@gmail.com (H.H.-J. Chen).

https://doi.org/10.1016/j.compedu.2024.105112
Received 20 March 2024; Received in revised form 7 July 2024; Accepted 9 July 2024
Available online 9 July 2024

particularly language learning (Wang et al., 2023). Huang et al. (2022) highlight three key features: personalization, user-friendliness, and timeliness. AI chatbots can simulate real-world language contexts, enhancing the authenticity of the learning experience). Furthermore, AI chatbots serve as conversational partners for language learners, providing opportunities for exposure to input and practice with output in a stress-free environment (Jeon et al., 2023; Yang et al., 2022). These chatbots personalize interactions based on learners' interests and proficiency levels (Dizon et al., 2022), facilitate role-playing within a social constructional framework (Sikström et al., 2022), and provide real-time feedback on language use and comprehension (Essel et al., 2022).

Recent advancements in generative artificial intelligence (GAI) and automatic speech recognition (ASR) technologies have significantly affected the domain of AI chatbots (Hwang & Chen, 2023; Wu & Yu, 2024). Powered by sophisticated transformer architectures, such as GPT-4, GAI-based chatbots exhibit significant potential for enhancing the language learning experience by effectively addressing the limitations inherent in conventional AI chatbots (Bishop, 2023). They can predict, comprehend, and generate text in a human-like fashion (Pavlik, 2023), producing coherent and contextually relevant responses over extended interactions (Mohamed, 2023). According to Hwang and Chen (2023), GAI chatbots' adaptability, creativity, and natural language interaction capabilities can meet language learners' developmental needs and learning preferences.

Despite this optimistic outlook, as Liu and Ma (2023) noted, GAI in language education remains an evolving field that requires more empirical studies to understand its effectiveness in L2 learning. Furthermore, according to Divekar et al. (2022), employing a multi-faceted approach that integrates both learner-chatbot and learner-learner interactions is crucial. This approach enables learners to take advantage of chatbots, such as immediate feedback and targeted pronunciation practice, while also fulfilling their preference for collaborative learning with peers (Jeon, 2024). This phenomenon prompts further exploration of the dynamics of individual versus collaborative interactions with GAI chatbots in language learning. Additionally, Wu and Yu's (2024) meta-analysis indicates the role of educational levels in moderating the effects of AI chatbots on learning outcomes. They found that primary and secondary school students did not exhibit superior learning outcomes compared with their counterparts without AI chatbots. The uncertainty regarding whether the efficacy of GAI chatbots, given their advanced nature and the absence of previous AI chatbot limitations, aligns with or differs from that observed in general AI chatbots, constitutes an area warranting further research.

This study aims to address these gaps by examining the effects of GAI-based chatbots on the speaking skills of elementary school EFL learners. The rationale for this exploration is grounded in the limitations of conventional AI chatbots, which may fall short in addressing the dynamic and contextual nature of spoken language learning. Challenges in communication, notably those related to issues with ASR systems, maintaining a coherent dialogue flow, providing relevant responses, and handling complex inquiries, persist (Fryer et al., 2020; Huang et al., 2022; Jeon et al., 2023). Building on this premise, this study developed CoolE Bot, a GAI-based chatbot that incorporates OpenAI's advanced language model, Microsoft's text-to-speech (TTS), and ASR techniques. CoolE Bot's utilization of ChatGPT enhances its communication capabilities, enabling it to understand and generate humanlike responses across a diverse array of topics. Integrating these technologies aims to engage learners in authentic and contextually relevant interactions. In addition, CoolE Bot offers supportive functionality designed to meet the developmental needs and learning preferences of K-12 EFL learners. This study investigates the impact of GAI chatbot-mediated EFL speaking by specifically examining the individual interactions of primary school students with CoolE Bot and collaborative interactions when placed in pairs. Through this exploration, the study aims to provide valuable insights into GAI chatbot-assisted EFL speaking at the elementary school level, offering a foundation for future research and practical implications for enhancing English-speaking experiences for elementary school EFL learners.

2. Literature review

2.1. AI chatbots for language learning

AI chatbots, also referred to as conversational agents, are computer programs designed for dialogues that simulate human-like conversations (Wu & Yu, 2024; Zhang et al., 2023). They employ various AI techniques such as natural language processing, machine learning, neural networks, information retrieval, and deep learning. Their integration into language learning aligns with the interactionist approach (Carhill-Poza & Chen, 2020; Tai & Chen, 2022), which posits that language acquisition advances through comprehensible inputs, meaningful interactions, language outputs, and feedback (Long, 2017). Various AI chatbots are available online, including Apple's Siri, Amazon's Alexa, Google Assistant, and Open AI's ChatGPT.

Noticeable differences exist between traditional and AI chatbots (Wu & Yu, 2023). Conventional chatbots rely on predefined patterns and templates. Their interactions are limited by rules set by their developers, making them less flexible and adaptable to diverse user inputs (Coniam, 2014; Yang et al., 2022). On the contrary, AI chatbots can retain user inputs and learn from previous user inputs, facilitating enhanced engagement and interactions (Nguyen et al., 2022).

Several pedagogical theories support the use of AI chatbots for language learning. Real-time interactivity is a crucial feature of AI chatbots that offers communicative authenticity (Liu et al., 2022). Human-AI chatbot interaction, akin to collaborative dialogue between humans (Swain, 2000), encompasses problem-solving and knowledge-building through interactive and cooperative learning, thus enhancing L2 learning (Wang et al., 2023). AI chatbots are perceived as friendly and patient learning partners (Hwang & Chen, 2023), providing scaffolding and support within learners' zone of proximal development. Learning occurs through observing and interacting with others in collaborative tasks. Beyond acting as interlocutors, they also function as facilitators, promoting peer collaborations and engaging learners in role-playing activities (Cai et al., 2020). These findings align with the principles of social constructivism, which indicate that learners collectively construct knowledge through interactions and collaboration (Vygotsky, 1980). Meaning negotiation occurs when interlocutors adjust their language to achieve agreement and mutual understanding during communication breakdowns, often by asking for clarification, correcting errors, and repeating (Díaz & Nussbaum, 2024; Su & Zou,

2022). Additionally, by replicating real-life language environments, AI chatbots create learning contexts similar to everyday situations, thus enabling learners to apply acquired skills in practical scenarios (Tai & Chen, 2022). This aligns with situated and contextualized learning theories (Zhang et al., 2023), which indicate that learning environments linked to the target language and skills can stimulate content integration into long-term memory and facilitate content retrieval (Brown et al., 2000).

Furthermore, AI chatbots provide individualized learning through tailored exercises, feedback, and content, enabling learners to take ownership of their progress and adapt their learning to their language proficiency, learning style, and objectives (Jeon, 2023), which aligns with self-regulated learning theory. Self-regulated learning theory indicates that optimal learning efficiency occurs when learners can control the content and style of learning and actively engage in learning tasks (Reinders & White, 2016). AI chatbots allow learners to determine their learning pace, engage in language learning ubiquitously, and select learning materials based on individual needs and preferences (Tai, 2024). Lodge et al. (2023, p. 6) describe this kind of interaction as "AI as coach", supporting learners in managing their learning process effectively.

AI chatbots could provide language learners with emotional support, alleviate their anxiety (Hsu et al., 2023), and afford enjoyment (Jeon, 2024). The artificial nature of chatbots provides advantages that help reduce learners' concerns about mistakes and foster ease compared to human interactions (Moussalli & Cardoso, 2020). This reduction in learners' affective filter promotes more effective language learning (Krashen, 1982). Furthermore, engaging in human-AI chatbot interactions can induce a state of flow among EFL learners (Tai & Chen, 2023). According to the flow theory, learners experiencing flow exhibit high learning efficiency and positive affective states (Csikszentmihalyi, 2000). Hence, the specific features of AI chatbots offer considerable potential for L2 learning, supported by theories.

2.2. AI chatbot-assisted L2 speaking

Research on AI chatbots as conversational partners has recently increased Dizon (2020) found that Alexa (Amazon's virtual agent) facilitates EFL learners' speaking because it offers ample opportunities for speaking practice and engages them in various forms of interaction. Tai and Chen (2022) examined the effect of intelligent personal assistants (IPAs) on EFL learners' oral proficiency and compared this effect with that of interactions with L1 and L2 English speakers. They found that IPAs significantly improve the EFL learners' oral proficiency, with a positive effect similar to that of interactions with L1 English speakers. Similarly, Yuan et al. (2023) investigated the efficacy of AI chatbots in improving the English oral proficiency of 74 EFL primary school students. The results revealed that chatbot integration significantly improved oral English proficiency.

Furthermore, AI chatbots have been reported to help increase L2 confidence, enjoyment (Wang et al., 2023), and willingness to communicate (Tai, 2024) while reducing their anxiety (Hsu et al., 2023; Jeon, 2024). Underwood (2017)investigated the effectiveness of including voice-driven AI assistants, such as Amazon's Alexa and Apple's Siri, in an EFL class for elementary school students. Students asked questions, commanded the AI in L2, and constantly reformulated their questions when the AI could not recognize their utterances. The results showed that the students were engaged in and enjoyed interacting with the chatbot. Compared with speaking to humans, they spoke more English when using AI assistants and demonstrated more persistence in speaking to AI assistants in English. Jeon (2024) investigated the affordances of AI chatbots in EFL learning based on interviews with primary school students and log data. Qualitative analysis of the data revealed chatbots' pedagogical affordances (interactional opportunities for students to acquire EFL), technological affordances (technological functionalities supporting EFL learning, e.g., speech recognition techniques, online dictionaries, and search engines), and social affordances (AI's ability to create a stress-free experience). Similarly, Hsu et al. (2023) proposed a chatbot system named TPBOT, which stands for "TOEIC Practice Chatbot," for EFL university students. Furthermore, they found that chatbots helped to eliminate learners' English-speaking anxiety. Tai (2024) explored the effects of Google Assistant-mediated interactions on 92 college EFL learners' willingness to communicate (WTC) in English. Google Assistant, as a supportive and patient learning partner, enhanced the learners' engagement and confidence, and thus their WTC.

However, communication challenges have persisted for AI chatbots in prior studies, primarily owing to deficiencies in the ASR system, and difficulties in sustaining coherent conversations and comprehending user inputs. Some studies found no significant difference between the chatbot and control groups in terms of learning engagement (Liu et al., 2022), confidence (Han et al., 2023), motivation (Kumar, 2021), and performance (Yin et al., 2021). Wu and Yu's (2024) examination of prior empirical research on AI chatbots revealed mixed evidence of their impact on students' learning outcomes. Pérez et al. (2020) noted that learners interacting with chatbots might experience a novelty effect, leading to potential negative emotions that, instead of fostering learning, could have detrimental effects on overall learning outcomes. Furthermore, Huang et al. (2022) found that while chatbots could facilitate students' language learning, they might concurrently increase their extraneous cognitive load. Wu and Yu (2024) found that the effects of AI chatbots differ across educational levels. University students exhibited significantly improved learning outcomes using AI chatbots. However, this positive effect was not observed among primary and secondary school students. According to Jeon (2024), the challenges faced by primary school EFL learners, such as lower language competency, self-directed learning abilities, and digital literacy, may hinder their effective interactions with AI chatbots. They relied more on teachers' guidance, which made it difficult for them to fully engage in learning using AI chatbots (Deveci Topal et al., 2021). According to Divekar et al. (2018), AI chatbots could not assist students as comprehensively as teachers during conversation practice, impacting students' motivation to use them for language learning (Zou et al., 2021). Technical challenges, such as speech recognition issues, triggered anxiety and decreased the motivation of primary school EFL learners to utilize English with chatbots (Jeon, 2024). In Yang et al.'s study (2022) on a voice chatbot named Ellie, Korean EFL learners aged 10–15 expressed a neutral to positive perception of the chatbot as an English conversation partner. They believed that the chatbot helped them improve their speaking skills. However, participants also identified several limitations: the chatbot often failed to understand their utterances, spoke too quickly or at length, and used expressions that were difficult to

comprehend. Despite the promise of AI chatbots for language learning, studies undertaken thus far have provided conflicting evidence concerning the effects of AI chatbots on EFL learners' learning outcomes.

In addition, most prior studies on AI chatbot-based L2 speaking have focused on individual interaction rather than group or paired configurations. As suggested by Zhang et al. (2022), understanding how interaction configurations, individual or collaborative, affect learners' speaking proficiency is essential for tailoring language-learning experiences. Collaborative learning of human learners is a process to enhance deep learning, co-construct knowledge, and solve more complex problems than learning alone (Lodge et al., 2023). Engwall and Lopes (2022) indicate that peer collaboration plays a crucial role in conversation practice with the chatbot to deal with linguistic difficulties or communication problems. Despite these insights, the impact of learners' collaborative interactions with AI chatbots on EFL speaking proficiency remains underexplored (Burkhard et al., 2022). Liu et al. (2024) examined chatbots' influence on elementary EFL learners' interests and pair interactions. Their findings showed chatbots effectively initiated dialogues and facilitated social learning. However, learners' interest decreased significantly, due to the chatbot's inability to decipher communicative intents and adapt to different interaction patterns. In addition, the lack of human identity reduced learners' desire for continued communication, often limiting interactions to one-directional Q&A monologues. Consequently, consistent learning engagement and peer interaction were not guaranteed (Meng et al., 2023). Jeon (2024) investigated chatbot-mediated EFL learning among elementary school students, including individual and group interactions. Most students preferred individual interactions with chatbots, which reduced social anxiety and increased their willingness to communicate in English. These mixed results show the complexity of collaborative learning with AI chatbots, indicating the need for more in-depth comparative studies.

Recent advancements in GAI and ASR technologies have significantly improved the field of AI chatbots (Hwang & Chen, 2023; Nguyen et al., 2022). According to Bishop (2023), GAI, as exemplified by platforms, such as ChatGPT, displays significant potential for enhancing the language-learning experience. The neural network of the language model, which draws from large datasets to form varied strengths of connections, ensures that ChatGPT can produce responses similar to those of human language and generate contextually relevant responses over extended interactions (Mohamed, 2023). Liu and Ma (2023) highlighted the application of GAI-powered language models in various language-learning contexts, potentially revolutionizing language acquisition, instruction, and evaluation. These GAI-based chatbots can predict, comprehend, and generate text in a human-like fashion (Pavlik, 2023). Despite the optimistic outlook for GAI-based chatbots, as Hwang and Chen (2023) noted that GAI in language education is not yet a fully-developed field of research. Liu and Ma (2023) emphasized the need for more rigorous and comprehensive studies to elucidate the effectiveness of GAI-based chatbots.

Therefore, this study addresses the gaps in the literature by investigating three foci: (1) the impact of GAI chatbots on elementary school EFL learners' speaking skills, (2) the effects of individual and paired interactions with GAI chatbots on EFL learners' speaking skills, and (3) EFL learners' perceptions of GAI chatbot-assisted English speaking. Individual interaction involved each student interacting with the GAI chatbot independently, facilitating personalized, one-on-one conversations. In contrast, in paired interaction, two students interacted with the chatbot together, facilitating collaborative learning. This approach might allow students to support each other in overcoming linguistic challenges and enhancing their communicative skills through peer collaboration (Engwall & Lopes, 2022). To achieve these goals, we created CoolE Bot, a GAI-based chatbot that utilizes advanced language models from Microsoft's Azure OpenAI Service along with text-to-speech and automatic speech recognition (ASR) techniques. The bot was designed to engage users in open-ended dialogues and generate coherent and relevant responses. Three research questions were addressed in this study.

(1) Does CoolE Bot significantly promote elementary school EFL learners' speaking skills?
(2) How do individual and paired interactions with CoolE Bot affect the speaking skills of elementary school EFL learners, and what factors contribute to these effects?
(3) How do elementary school EFL learners perceive the effects of CoolE Bot-assisted EFL speaking?

3. Method

3.1. Design

This study investigated the impact of GAI chatbots (CoolE Bot) on elementary school EFL learners' speaking skills. A total of 85 sixth-grade Taiwanese students participated in a three-week summer program, where students engaged in 45-min interactions with CoolE Bot for five days a week. The participants from the three intact classes were randomly divided into three groups to assess the impact of different interaction configurations on English-speaking skills (i.e., fluency, content, pronunciation, grammar, and vocabulary): individual interaction with CoolE Bot (I-Bot group), paired interaction with CoolE Bot and peers (P-Bot group), and interaction with peers and teachers (No-Bot group) in the conventional English class. The speaking topics and tasks assigned to the I-Bot and P-Bot groups were identical. Each speaking task consisted of an interactive session with CoolE Bot and a corresponding written worksheet that offered essential information and prompts for completing the tasks. Participants in the No-Bot group completed similar speaking tasks on the same topics as the two Bot groups, except with the teacher and their peers.

3.2. Participants

The participants included 85 Taiwanese EFL 6-graders recruited from three classes and taught by the same English teacher. They attended 45-min-long English classes every day in a summer program. Students in Taiwan begin to learn English in public schools in the third grade and take lessons once or twice a week each semester. Student observation records offered by their English teacher at the

time of the research provided information about their overall spoken English proficiency. The participants belonged to the beginner level of the Common European Framework of Reference for Languages (CEFR). They were able to perform basic functions of English, such as discussing their interests and greeting friends. They required assistance with vocabulary during these conversations. Furthermore, Taiwan offers a technology-rich learning environment for students (Yu, 2024). According to their teachers, participants were already familiar with digital learning. The demographic data of the participants are summarized in Table 1.

3.3. Instruments

The instruments used in this study were CoolE Bot, English-speaking tests, and semi-structured interviews.

3.3.1. CoolE bot

In this study, we employed CoolE Bot, a GAI-based chatbot accessible through the Cool English website (https://www.coolenglish.edu.tw/) and supported by Taiwan's Ministry of Education. What sets CoolE Bot apart is its innovative integration of advanced language models from Microsoft's Azure OpenAI Service, which incorporates text-to-speech and ASR techniques. Particularly noteworthy is its use of ChatGPT, a sophisticated language model developed by OpenAI that distinguishes it from conventional AI chatbots. Unlike rule-based or simple pattern-matching chatbots, ChatGPT's dynamic language-processing capabilities, are driven by a deep neural network trained on extensive datasets, enabling it to understand and generate human-like responses across a wide range of topics. This capability enhances CoolE Bot's conversational abilities, allowing it to engage users in more natural and contextually relevant interactions. The integration of ChatGPT elevates CoolE Bot to a new level of conversational AI, offering users more authentic and immersive interactions.

CoolE Bot provides a selection of supportive features designed to address the specific needs of EFL learners. These include English-to-Chinese translation support, replay functionality, adjustable speech speed, and a choice between oral and text-based interactions (Fig. 1). The chatbot offers diverse scenarios and characters, allowing learners to explore themes, such as cartoons, famous people, stories, jobs, and games (Fig. 2). Among these themes, cartoons and famous people are designed for learners with basic-level English proficiency. Given that the participants in this study were elementary school students with limited English proficiency, they interacted only with the characters in these two themes. Personalization options extend to the selection of an interactive character, allowing learners to tailor their experiences (Fig. 3). Each character is designed with distinctive responses and varied questioning techniques, contributing to an enriched learning experience and fostering sustained engagement throughout the learning process (Fryer et al., 2020).

3.3.2. English-speaking test

The participants' English-speaking skills were evaluated before and after the experiment using a pretest and post-test, respectively. The speaking test was administered individually, and the audio was recorded. The speaking tests used in this study were adapted from the General English Proficiency Test (GEPT) Kids. GEPT Kids, a test specifically tailored to elementary school students, was developed by the Language Training and Testing Center (LTTC) in Taiwan. The test considers elementary school students' life experiences and local curriculum, equating to an English level of A1 in the Common European Framework of Reference for Languages (CEFR). The quality of the assessment was further verified through evaluation by one assessment specialist. The speaking test comprised three segments: (1) reading simple sentences aloud, (2) using simple sentences to introduce things and people, and (3) responding to questions related to daily life. These segments assessed EFL learners' speaking skills, including fluency, content, pronunciation, grammatical appropriateness, and lexical usage. The test evaluated the learners' ability to narrate, express opinions, and respond appropriately, aligning with the objectives of CoolE Bot-mediated interaction. While the pretest and post-test items remained consistent, their orders differed. The GEPT's elementary speaking rating scale was used by two English teachers who coded the results. The inter-rater reliability was 0.85.

3.3.3. Semi-structured interviews

This study aimed to explore how EFL learners perceived GAI chatbots, specifically CoolE Bot, for English speaking. Individual interviews were administered to the experimental groups to gather in-depth insights into their perceptions of CoolE Bot-mediated English-speaking skills and their interactive experiences with the chatbot (Appendix A). All interviews were conducted in the participants' first language (Chinese). Each interview lasted approximately 10 min and was subsequently transcribed verbatim for thorough examination.

Table 1
Demographic information of the participants.

Group	N	Age	Gender		Years of English Learning
			M	F	
I-Bot	28	12.32	16	12	3.89
P-Bot	28	13.13	15	13	4.24
No-Bot	29	12.01	14	15	3.78

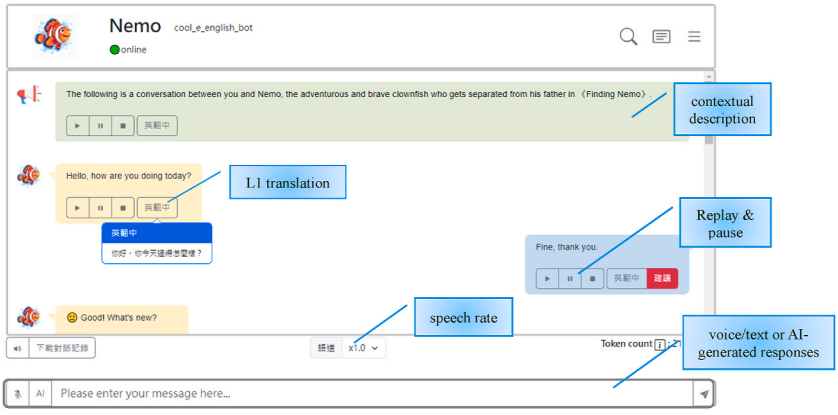


Fig. 1. CoolE Bot functionality.

Fig. 2. Theme selection.

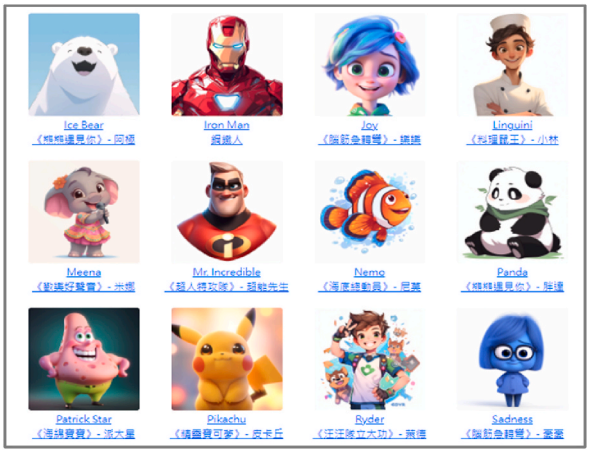


Fig. 3. Character selection.

3.4. Data analysis

This study employed a mixed-methods design (Creswell & Plano Clark, 2018) to address the research questions. To examine the first two research questions, the participants' performances on the speaking pretest and post-test were analyzed using a two-way mixed-design ANOVA (groups: I-Bot, P-Bot, and No-Bot; test: pretest and post-test). A repeated-measures one-way ANOVA was also conducted, followed by Scheffe's post-hoc test to examine within-group differences. To answer the third research question, a

three-phase thematic analysis (Braun & Clarke, 2006) was conducted to interpret the data derived from the audio recordings of the interview sessions. The transcribed interviews were coded to identify core themes indicating the participants' perception of CoolE Bot for EFL speaking skill development. The researchers began by familiarizing themselves with the data through repeated readings of the transcripts, allowing them to gain a holistic understanding of the content. Subsequently, initial codes were generated by systematically identifying meaningful words, phrases, or sentences that captured key concepts or ideas presented in the interviews. These codes were then organized into potential themes, with similar codes grouped based on shared characteristics or meanings. The researchers refined and revised the identified themes through ongoing discussions and comparison, ensuring coherence and relevance to the research questions. Finally, the themes were named and defined, accompanied by illustrative quotations from the interviews to provide concrete examples of each theme. The interview data were used to triangulate and cross-validate the quantitative results obtained from the questionnaires.

3.5. Procedure

The study was implemented in the context of an English class during a summer program. A 45-min English class session was conducted daily. As illustrated in Fig. 4, the researchers first held a meeting with school stakeholders, including the principal, academic affairs director, curriculum chief, and English teachers, and introduced the objectives of the study. Consent forms were obtained from the participants' parents prior to the study. Subsequently, an orientation was conducted to acquaint students with CoolE Bot and the interactive activities. Before the intervention, the participants' English-speaking proficiency was assessed. The researcher also provided orientation on how to interact with CoolE Bot and gave students practice sessions. During the intervention, the teacher initiated each class by introducing interactive topics, goals, and tasks. Topics were chosen based on the participants' interests and familiarity, including hobbies, friends, pets, cartoons, movies, and idols. Bot group participants received worksheets in each class with a designated topic, prompts, and vocabulary to guide their interactions with CoolE Bot. They engaged in tasks such as role-playing conversations with cartoon characters or famous people related to the topic. After interacting with the chatbot, they reflected on their learning and completed their worksheets. One of the two researchers was present in the class to assist students with technology-related issues. The No-Bot group, which performed similar speaking activities, received worksheets with the same designated topics for interactions with their peers. The teacher facilitated and guided learners through these interactive activities to ensure a consistent experience across all participant groups. Following the intervention, all the participants underwent a speaking post-test, and those in the I-Bot and P-Bot groups were interviewed.

4. Results

4.1. Results of speaking tests

A two-way mixed-design ANOVA was performed to evaluate the effect of CoolE Bot on EFL learners' speaking skills. Table 2 presents the descriptive statistics of the scores of the three groups on the speaking skill pretest and post-test, and Table 3 presents the results of the two-way mixed-design ANOVA. As is standard practice, the significant level was set at $p < 0.05$ for all statistical analyses. For the participants' speaking skill tests, there was a significant main effect of groups, $F(2, 82) = 6.06$, $p < 0.001$, $\eta^2 = 0.13$; a significant main effect of time, $F(1, 82) = 275.00$, $p < 0.001$, $\eta^2 = 0.77$; and a significant interaction between group and time, $F(2, 82) = 7.31$, $p < 0.001$, $\eta^2 = 0.15$. Post hoc analyses of the main effect were performed to further investigate the interaction effects.

A one-way ANOVA was conducted to assess the performance of the three groups on the speaking skills pretest and post-test. As shown in Table 4, no significant difference in pretest performance was identified among the three groups ($F(2, 164) = 2.19$, $p = 0.12$, $\eta^2 = 0.05$). However, a significant difference in post-test speaking performance was found among the three groups ($F(2, 164) = 10.24$,

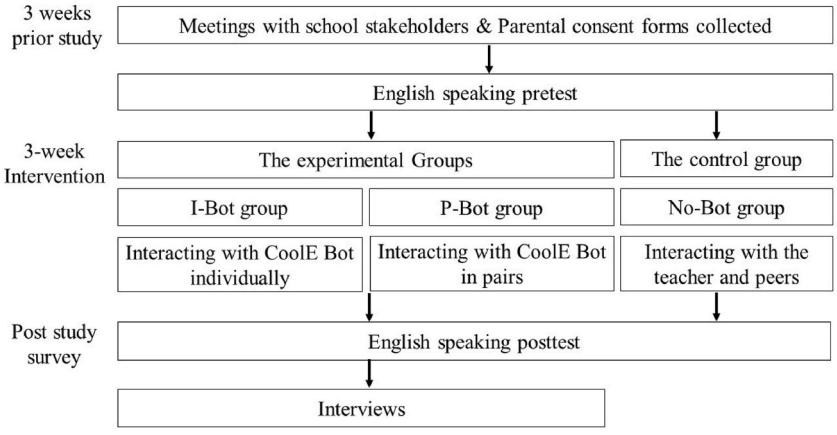


Fig. 4. The experimental procedure.

Table 2
Descriptive statistical results of the speaking pretests and post-tests.

Group	N	Speaking Pretest		Speaking Post-test	
		Mean	SD	Mean	SD
I-Bot	28	31.57	14.62	63.21	16.50
P-Bot	28	24.86	10.08	62.96	10.89
No-Bot	29	28.16	10.90	49.34	11.82

Table 3
Results of the mixed design ANOVA for the speaking pretests and post-tests.

Source	*SS*	*df*	*MS*	*F*	*p*	η^2
Between						
Group	2164.26	2	1082.13	6.06	0.00	0.13
Time	39057.00	1	39057.00	275.00	0.00	0.77
Group x Time	2077.16	2	1038.58	7.31	0.00	0.15
Within						
Subjects	14654.03	82	178.81			
Error	11645.93	82	142.02			
Total	69598.38	169	41498.54			

Note. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

$p < 0.001$, $\eta^2 = 0.20$). Both the I-Bot and P-Bot groups exhibited higher scores than the No-Bot group. Post hoc Scheffé tests indicated a significant difference between the Bot and No-Bot groups, but no significant difference between the I-Bot and P-Bot groups. Both individual and paired human-GAI chatbot positively impacted EFL learners' speaking skills. A repeated-measures ANOVA was performed to examine the differences within the groups between the speaking skills pretest and post-test. The results revealed significant differences between the two tests for the I-Bot ($F(1, 82) = 60.76$, $p < 0.001$, $\eta^2 = 0.69$), P-Bot ($F(1, 82) = 185.55$, $p < 0.001$, $\eta^2 = 0.87$), and No-Bot groups ($F(1, 82) = 74.27$, $p < 0.001$, $\eta^2 = 0.73$). The findings suggest that all three groups demonstrated improvement over time with increased practice. However, it is crucial to note that the study's summer program setting might influence participant engagement and motivation differently than a regular academic year, introducing a potential novelty effect that could impact the generalizability of the findings.

4.2. Interviews with the I-bot and P-bot groups

Table 5 presents the results of interviews with the two Bot groups. Five major themes emerged from the qualitative data analysis of their interview responses.

4.2.1. Theme 1: adeptness in coherent interaction and contextually appropriate responses

Most participants (I-Bot: 89.28%; P-Bot: 82.14%) indicated that CoolE Bot-mediated interaction improved their speaking skills, which can be attributed to several factors: (1) adeptness in coherent interaction and contextually appropriate responses, (2) multi-interactive roles, and (3) immediate feedback. The participants valued CoolE Bot's capacity for authentic communication, creating a realistic and immersive language-learning experience. The chatbot's ability to generate contextually appropriate and personalized responses, coupled with coherent interactions fostered engagement, with the participants expressing a desire to continue conversations.

Over two-thirds of the participants (I-Bot: 64.29%; P-Bot: 75%) reported that learning communication strategies during interaction with CoolE Bot, such as initiating coherent follow-ups, staying on topic, and ensuring clarity in expression, were perceived as applicable in real-world language use scenarios. For example, one participant stated,

Table 4
ANOVA of the pure main effect of group and test.

Source of Variance	*SS*	*df*	*MS*	*F*	*p*	η^2
Group						
Pretest	631.25	2	315.63	2.19	0.12	0.05
Post-test	3610.17	2	1805.09	10.24	0.00	0.20
Error	26299.96	164	320.83			
Test						
I-Bot	14017.79	1	14017.79	60.76	0.00	0.69
P-Bot	20330.16	1	20330.16	185.55	0.00	0.87
No-Bot	6521.12	1	6521.12	74.27	0.00	0.73
Error	11645.93	82	142.02			

Note. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 5
Interview results of I-Bot and P-Bot groups.

	I-Bot		P-Bot	
	N	%	N	%
Adeptness in coherent interaction				
coherent, contextually appropriate interaction	20	71.43	20	71.43
multi-interactive roles	15	53.57	14	50.00
individualized, immediate feedback	18	64.29	19	67.86
Charismatic conversational style, human-like voice				
enhanced engagement and interest in EFL speaking	22	78.57	22	78.57
cartoonish and human-like characters	22	78.57	21	75
Diverse topic discussion tailored to learners' interests				
Enthusiastic, exploring, and learning	17	60.71	19	67.86
expanded vocabulary	14	50.00	13	46.43
learner autonomy in problem-solving	19	67.86	16	57.14
Building rapport and supportive environments				
reciprocal encouragement and emotional support	21	75.00	18	64.29
reduced anxiety	20	71.43	22	78.57
Challenges				
limited English proficiency	7	25.00	5	17.86
non-collaborative counterparts			6	21.43

> *I found out some tips to keep it chatting with me! You should know what it likes, and what it enjoys. Keep actively asking questions to the bots. Stay on topic, or the talking might stop. If the bot doesn't get it, you can ask differently or say it again … again. Be patient!* (P-Bot S3)

The participants self-monitored and evaluated their interactions with the chatbot, adjusting their communicative strategies accordingly.

Furthermore, CoolE Bot's versatile role-playing capabilities allow the participants (I-Bot: 53.57%; P-Bot: 50%) to experience dynamic interactions; one participant stated, "It's super cool! It feels like you are under the sea, chatting with the fish." CoolE Bot created an authentic and immersive English-speaking environment, engaging participants in conversation. In addition, over half of the participants (I-Bot: 64.29%; P-Bot: 67.86%) appreciated the immediate feedback, which created a responsive learning environment that facilitated real-time correction and improvement.

4.2.2. Theme 2: charismatic conversational style coupled with human-like voice

Most participants reported feeling motivated and excited while interacting with CoolE Bot. The participants (I-Bot: 78.57%; P-Bot: 78.57%) appreciated the Bot's charismatic conversational style and human-like voice, creating an enjoyable English-speaking experience. One participant even recognized a familiar voice, stating, "Listen! It's real Elsa! It's her voice in the movie Frozen!" Furthermore, the participants (I-Bot: 78.57%, P-Bot: 75%) expressed appreciation for CoolE Bot's cartoonish and human-like characters. They perceived CoolE Bot as a real conversation partner, with one participant expressing excitement about conversing with Pokémon, stating, "I cannot believe I am talking to Pokémon for real! It felt true." The participants viewed CoolE Bot as having genuine personality characteristics. They described it as cute, kind, generous, engaging, and charming, attributing human-like qualities to its interactions.

4.2.3. Theme 3: diverse topic discussion tailored to learners' interests

More than half the participants (I-Bot: 60.71%; P-Bot: 67.86%) expressed enthusiasm for CoolE Bot's extensive coverage of diverse topics. They appreciated the chatbot's ability to cater to varied interests and promote exploration and learning by engaging in a wide range of topics and conversations. The participants noted that CoolE Bot "shared experiences, offered advice, and exposed them to novel ideas." Consequently, they (I-Bot: 67.86%, P-Bot: 57.14%) expressed a strong interest and willingness to extend their interactions, demonstrating a commitment to overcoming language challenges to sustain conversations; one participant described, "I do my best to make the bot understand what I'm trying to say so that our conversation can continue." In the I-Bot group, some participants employed support functions, such as L1 translation, to facilitate their interactions. Several sought assistance from online resources to ensure accurate pronunciation of words and sentences. For the P-Bot participants, CoolE Bot played the dual roles of interlocutor and moderator, acting as a friend to the participants and promoting discussions. Collaborative efforts with partners have become crucial for overcoming linguistic challenges. For example, one participant stated,

> *CoolE Bot was excellent at chatting. It engaged in conversation with both me and my partner. We discussed and searched online for additional information to better answer its questions.* (P-Bot S11)

Through these interactions, CoolE Bot expanded students' knowledge across diverse topics while fostering self-regulated learning, enhancing their English vocabulary, and refining their speaking skills.

4.2.4. Theme 4: building rapport and supportive environments

The participants (I-Bot: 75%; P-Bot: 64.29%) appreciated the reciprocal encouragement and support provided by CoolE Bot. They

emphasized the chatbot's role as a "friendly and kind" virtual interlocutor, noting that it not only guided them but also offered emotional support. For example, one participant stated,

> *When I said, "I have no friends." Pikachu cheered me up and said, "Don't feel sad. We can be best friends and go on adventures together." It made me happy. Pikachu told me, "We can explore new places and meet other Pokémon!" It sounded like so much fun!* (I-Bot S18)

For the majority (I-Bot: 71.43%, P-Bot: 78.57%), CoolE Bot established a non-judgmental and less anxious atmosphere for EFL speaking. The I-Bot participants considered the chatbot as "a trusted friend who would keep their secrets" and were willing to share personal information with it. They appreciated the individual interaction with CoolE Bot. They emphasized the freedom from worrying about speech mistakes and the judgment of peers or teachers; one participant stated, "It never scolds me."

Several higher-proficiency I-Bot participants stated that they had more time to speak and learn to interact with the chatbot alone. Most P-Bot group members (N = 20, 71.43%) preferred to interact with CoolE Bot with a partner. "My friend and I helped and taught each other. We also asked Google for assistance." They did not feel anxious about peer presence while speaking English; one participant stated, "We can share funny jokes, and interesting stories, and have a good laugh together." In addition to serving as interlocutors, CoolE Bot also functioned as a facilitator, encouraging peer collaborations.

4.2.5. Theme 5: challenges in GAI chatbot-mediated interaction: language complexity and participants' proficiency

Although there were positive aspects to interacting with CoolE Bot, some challenges were encountered. Some participants (I-Bot: 25%; P-Bot: 17.86%) noted that interacting with CoolE Bot posed linguistic and comprehension difficulties, primarily because of advanced vocabulary and overly lengthy sentences. The content sometimes surpassed age-appropriate understanding, hindering a smooth interaction experience. One I-Bot participant expressed frustration.

> *The bot talks too much. Some sentences are too long. It is hard to understand. I wish it could speak Chinese sometimes. Learning is more enjoyable with a friend. I can ask my friend for help.* (I-Bot S17)

Furthermore, despite the potential for collaborative learning in the P-Bot group, there were instances of noncooperative behavior among the paired participants. They reported (N = 6, 21.43%) disagreements over interests and conflicting conversations about characters, reflecting the challenges of maintaining cooperation during group interactions; one participant complained,

> *We have different interests, you know? We both like to share our cool ideas, telling our stories to the bot. But we argue about who gets to talk first and which character to talk to.* (P-Bot S14)

In addition, the frequent use of the participants' first language (Chinese) during peer discussions undermined the intended English-speaking practice facilitated by CoolE Bot. These challenges highlight the areas for improvement and adaptation with the aim of enhancing the user experience and collaborative dynamics within GAI chatbot-mediated interaction. In summary, CoolE Bot positively influenced elementary school EFL learners' speaking skills by creating an engaging and authentic environment. It also promotes collaborative dialogue and fosters mutual support but presents some challenges during interaction.

5. Discussion

The first research question inquired into the impact of CoolE Bot on EFL learners' speaking skills. The statistical results revealed no significant differences in pretest performance among the I-Bot, P-Bot, and No-Bot groups. On the contrary, the I-Bot and P-Bot groups' post-test speaking skills were significantly higher than those of the No-Bot group in the post-test. Furthermore, the results revealed significant interaction effects between group and time, indicating that different interaction configurations led to varied rates of improvement in speaking skills over time. The significant differences between the Bot and No-Bot groups, coupled with the lack of differences between the I-Bot and P-Bot groups, highlight the potential of the GAI chatbot to significantly enhance EFL learners' speaking skills. Learners who interacted with CoolE Bot exhibited more prominent improvements in their EFL speaking than those who interacted with their teachers and peers. The results are consistent with prior research, suggesting that interactions with an AI chatbot yield superior outcomes in enhancing EFL learners' speaking skills compared to human-human interactions (Tai, 2024; Wang et al., 2023). In contrast to Wu and Yu (2024), the elementary school students in this study exhibited significant improvements in their speaking skills. This disparity may be attributed to the GAI chatbot's ability to maintain a coherent dialogue flow and deliver contextually appropriate responses, thereby serving as a language role model. It sustains meaningful conversations and furnishes responses that align with the conversational context.

Furthermore, the findings indicated the critical influence of CoolE Bot's supportive functionalities and engaging character designs in fostering an engaging and enjoyable environment conducive to elementary school students' speaking practice. The participants actively engaged with the GAI chatbot, posing questions and giving commands in English, and demonstrated persistence by reformulating queries when the GAI chatbot faced difficulty in recognizing their utterances. For example, CoolE Bot offers L1 translation, aiding learners in comprehending the chatbot's responses. Participants also valued the audio and visual on-screen transcript responses provided by CoolE Bot, which supported their interaction and alleviated difficulties in interpreting information. This aligns with Mayer's (2017) multimedia learning principle, which indicates that people learn better from words and pictures than from words alone because processing information through different modalities expands the resources of working memory and increases learning effectiveness. Additionally, CoolE Bot can display the transcript of learners' speech output on screen, aiding them in recognizing what

the ASR system accurately captures and pinpointing specific lexical items causing miscommunication. This helps learners modify their erroneous utterances or consult online resources for self-regulated learning, aligning with the findings of Tai and Chen's (2022) study. These features prove instrumental in overcoming the potential barriers highlighted by Jeon (2024), such as lower language proficiency and dependence on teacher guidance among primary school EFL learners, which may hinder their involvement in AI chatbot-based interactions. According to Lodge et al. (2023), such interactions position the GAI chatbot as a coach, prompting learners to self-regulate by self-monitoring, evaluating their responses against the standards, clarifying meaning, and adapting their interaction strategies.

In addition, CoolE Bot provides discussions on various topics tailored to the learners' interests. CoolE Bot fostered participants' explorations and inquiries about the related content and expanded their vocabulary. Motivated by the desire for a deeper understanding, the participants actively sought additional information on various topics, leading to notable enrichment of their speaking skills. Through these interactions, the learners delved into diverse language uses and seamlessly integrated their newfound knowledge into their speaking skills. Consequently, the participants improved their English skills and developed positive feelings towards English. These results align with the findings of Wang et al. (2023), indicating that EFL primary school students' interactions with AI bots resulted in enhanced L2 enjoyment and learning. However, it is important to note that the No-Bot group consisted of students who were the youngest and had the fewest years of learning English compared to the Bot groups. This discrepancy in age and prior English learning experience might have influenced the learning outcomes (Artieda et al., 2020; Lightbown & Spada, 2020; Muñoz, 2019). Consequently, the differences in speaking skill improvements between the groups might be partially attributable to these initial disparities rather than solely the intervention itself.

The second research question explored whether individual and paired interactions with CoolE Bot yielded distinct benefits for EFL learners' speaking skills. Although the P-Bot group exhibited greater improvement in scores, no statistically significant differences were found in the speaking skills between the two groups. Interaction with CoolE Bot, either individually or in pairs, contributed to the EFL learners' ability to construct and confirm meaning through sustained discourse and reflection. By communicating with CoolE Bot, the participants received feedback on their utterances and continued to refine their speech. Learners in the I-Bot group followed a more independent trajectory, driving their inquiries into relevant content and broadening their vocabulary through self-initiated Internet searches. This proactive approach, combined with persistent efforts to sustain conversations, not only enhanced their lexical repertoire but also elevated the quality of their oral communication. For the P-Bot participants, collaborative problem-solving and the exchange of perspectives within pairs contributed to overcoming linguistic challenges, potentially fostering deeper processing of the target language (Divekar et al., 2022; Keshanchi et al., 2023). Furthermore, peer discussions and feedback during interactions with CoolE Bot played crucial roles in solidifying cognitive resolution, addressing linguistic challenges, and promoting shared understanding. These findings align with those of Delgado-Garcia et al. (2021), indicating that paired learners can regulate each other's cognitive activities, co-construct complex language structures, and improve their language learning.

Furthermore, CoolE Bot's multi-interactive roles and supportive functions facilitated participants' speaking practices and fostered continuous improvement. For the I-Bot participants, interacting with CoolE Bot provided personalized responses and promoted collaborative learning, where participants co-constructed messages with the chatbot in English, thereby enhancing their speaking skills (Keshanchi et al., 2023). On the contrary, in the P-Bot group, the collaborative nature of speaking activities with CoolE Bot and peers reflects a different facet. Collaborative tasks demonstrate a teaching presence that emphasizes joint learning experiences and shared goals (Wang et al., 2023). Learners engage in mutual support and share strategies, and navigate language challenges during their interactions, which fosters a supportive and interactive learning environment (Divekar et al., 2022; Kuhail et al., 2023). The facilitation role is not solely borne by CoolE Bot; instead, peers actively contribute to the facilitative function.

The third research question inquired into the Bot group's perception of AI chatbot-assisted EFL speaking. The participants thought that it was enjoyable, motivating, and engaging. They particularly appreciated the cartoonish and human-like characters coupled with a charismatic conversational style and a human-like voice. The chatbot's versatile role-playing capabilities, which allowed them to engage in dynamic interactions resembling those of native speakers, were also well-received. Furthermore, most participants appreciated CoolE Bot's ability to establish rapport and foster supportive environments. The participants perceived CoolE Bot as a supportive and encouraging virtual interlocutor, providing emotional support, guidance, and advice tailored to learners' input, which alleviated anxiety and stress. Consistent with Guo et al. (2023), fostering a social presence, where learners establish a connection with CoolE Bot, is pivotal in cultivating a positive and conducive atmosphere for English-speaking practice. In turn, this enhances learners' self-confidence, motivation, and speaking skills.

The I-Bot participants expressed a positive attitude towards speaking English, attributing their comfort and willingness to communicate to the absence of judgment from their human partners. These findings align with Jeon's (2023) finding that interaction with chatbots helps reduce anxiety related to speaking, thus fostering a more comfortable environment for students to communicate in English. Contrarily, the P-Bot participants emphasized the importance of peer scaffolding and peer sharing in AI chatbot-meditated interactions. Collaboration enhances the L2 learning effect when learners co-construct messages in their target language (Keshanchi et al., 2023). The findings support Divekar et al.'s (2022) multifaceted approach in the human-AI community, incorporating both learner-chatbot and learner-learner interactions. Furthermore, the participants in the P-Bot group did not feel embarrassed by peer presence when interacting with CoolE Bot. The positive attitudes expressed by the P-Bot participants towards peer presence, comments, and corrections in the AI-mediated setting further underscore the valuable role of GAI chatbots in fostering a conducive learning atmosphere. While prior research has indicated that EFL learners experience speaking anxiety during class-based language practice, particularly in the face of peer comments (Shafiee Rad, 2024; Wang et al., 2023), our findings reveal a distinct shift.

Nevertheless, the interviews revealed several challenges encountered by the participants during their interactions with CoolE Bot, highlighting linguistic and comprehension obstacles, such as advanced vocabulary, lengthy sentences, and content exceeding age-

appropriate levels. They conveyed a sense of inadequacy for full engagement with CoolE Bot due to their low English proficiency, which aligns with the findings of Deveci Topal et al. (2021), and Wu and Yu (2024). Despite deliberate efforts during the design phase to provide CoolE Bot with prompts for language simplification, it remains challenging for elementary school EFL learners. To address this concern and enhance the EFL speaking experience of young learners, future designs of GAI chatbots should prioritize the implementation of language simplification techniques, such as breaking down complex sentences and incorporating age-appropriate content. In addition, the inclusion of supportive functionality ensures that learners receive the necessary assistance, enhancing their engagement and comprehension (Jeon, 2024). For example, a chatbot capable of communicating in L1 and L2 could prove to be more effective for beginner-level learners, as indicated by the interview data. These approaches might create a more accessible and engaging experience for young learners, ensuring that the language and topics align with their developmental stage while fostering positive and effective GAI chatbot-mediated interactions.

Notably, several I-Bot participants, particularly those with limited English proficiency, expressed a preference for learning partners. This highlights the diversity in learners' needs and suggests that while the I-Bot configuration is beneficial for many learners, there may be instances where external assistance or learning partners are desired to overcome language obstacles. On the contrary, in the P-Bot group, non-collaborative pairs, influenced by differing interests or proficiency levels, exhibited a lack of engagement in the interaction. As Divekar (2022)noted, placing learners in a collaborative environment does not ensure collaborative learning. The findings indicate thoughtful consideration of the interaction configuration, accounting for learners' proficiency levels, interests, and needs in GAI chatbot-mediated interaction.

6. Conclusion

This study investigated the impact of the GAI-based chatbot, CoolE Bot, particularly regarding individual and paired interaction configurations, on elementary school EFL learners' speaking skills. The results revealed that CoolE Bot significantly improved participants' speaking skills, with a more positive effect than interactions with teachers and peers in a conventional English class. A detailed examination of the interviews revealed that CoolE Bot's adeptness in authentic communication, contextually appropriate and personalized responses, and coherent interactions fostered participants' engagement with a notable desire to continue conversations. Immediate feedback was also highly valued, creating a responsive learning environment conducive to real-time correction, thereby promoting continuous improvement in English speaking. Furthermore, CoolE Bot's versatile role-playing capabilities and diverse topic discussions provided the participants with dynamic interactions. In addition, the charismatic conversational style and human-like voice of CoolE Bot were appreciated, creating an enjoyable English-speaking experience. Beyond linguistic functionality, learners expressed a strong sense of rapport with the chatbot, highlighting its reciprocal encouragement and support. The I-Bot participants individually engaged with CoolE Bot, which fostered personalized cognitive exploration and the resolution of linguistic challenges. However, the P-Bot group experienced collaborative cognitive processes, shared exploration, and mutual resolution of challenges, striking a balance between individual and group dynamics.

From a pedagogical perspective, this study provides evidence supporting the integration of GAI chatbots into elementary school students' EFL speaking. Furthermore, whether students interact with the chatbot individually or in pairs, they achieve similar improvements in their speaking skills, making the chatbot a flexible tool adaptable to various educational contexts. Thus, this study recommends a thoughtful consideration of individual and paired human-GAI chatbot interactions, accounting for learners' proficiency levels, interests, and individual characteristics. In addition, incorporating language simplification techniques and supportive functionality into future GAI chatbot designs is essential for enhancing the GAI-mediated EFL speaking experience at the elementary school level.

However, this study has several limitations that merit attention in future research. The brief duration of the intervention raises concerns about potential novelty effects by introducing a variable that may affect the reliability of the results. Furthermore, the random pairing of the participants in the P-Bot group, without considering factors such as interest, proficiency level, or personality, hindered collaborative learning, which may have influenced the results. Future investigations should incorporate these critical factors into the pairing or group assignment processes. To ensure strong internal validity, interviews with the No-Bot group are crucial. Future research should compare engagement levels and quality between bot and non-bot participants. In addition, the No-Bot group consisted of students with the youngest age and the fewest years of learning English, potentially placing them at a disadvantage. Future research should control for these variables by ensuring more balanced groups in terms of age and prior English learning experience to provide a more accurate assessment of the impact of GAI chatbots on EFL speaking skills.

Moreover, because this study specifically examined the impact of CoolE Bot, caution is warranted when generalizing these findings to other GAI chatbots. Future studies should explore a broader range of technologies. Furthermore, the education culture in Taiwan is characterized by teacher-centered, large-class settings with limited speaking practice time (Tseng, 2024). The use of GAI chatbots, which offer personalized, learner-centered interactions, has significant benefits for students in such an environment. Additionally, Taiwan offers a technology-rich learning environment for students, which might impact their language learning and strategies (Yu, 2024). Participants were already familiar with digital learning, potentially increasing their engagement and interaction with GAI chatbots and consequently enhancing their speaking skills. Therefore, caution is warranted when generalizing the findings to different cultural and educational settings. Finally, the relatively small sample size underscores the need for caution when drawing broad conclusions. To enhance the generalizability of the results across diverse EFL learner populations, future studies should use larger and more diverse samples, including different age groups, educational environments, and cultural backgrounds. Further investigation into implementing these findings in real-world educational settings is essential for informed decision-making by practitioners regarding GAI chatbots in curricula.

CRediT authorship contribution statement

Tzu-Yu Tai: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Resources, Visualization. **Howard Hao-Jan Chen:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

The data that has been used is confidential.

Appendix A

Interview questions

1. Can you describe how you used CoolE Bot?
2. Did you find using CoolE Bot helpful when practicing speaking English? Why?
3. What difficulties did you have when studying with CoolE Bot?
4. How did you overcome these difficulties?
5. Do you want to continue to use CoolE Bot in English class? Why?

References

Artieda, G., Roquet, H., & Nicolás-Conesa, F. (2020). The impact of age and exposure on EFL achievement in two learning contexts: Formal instruction and formal instruction+ content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism, 23*(4), 449–472. https://doi.org/10.1080/13670050.2017.1373059

Bishop, L. (2023). A computer wrote this paper: What ChatGPT means for education, research, and writing. *SSRN*. https://doi.org/10.2139/ssrn.4338981

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, J., Collins, A., & Duguid, P. (2000). Situated cognition and the culture of Learning. *Psychology of Education: Major Themes: Pupils and Learning, 2*, 207–230. https://doi.org/10.3102/0013189X01800103

Burkhard, M., Seufert, S., Cetto, M., & Handschuh, S. (2022). Educational chatbots for collaborative learning: Results of a design experiment in a middle school. In *Proceedings of the 19th International Conference on Cognition and exploratory Learning in digital age (CELDA 2022). Lisbon, Portugal*. https://files.eric.ed.gov/fulltext/ED626892.pdf.

Cai, M. Y., Wang, J. Y., Chen, G. D., Wang, J. H., & Yang, S. H. (2020). A digital reality theater with the mechanisms of real-time spoken language evaluation and interactive switching of scenario & virtual costumes: Effects on motivation and learning performance. In *2020 IEEE 20th international conference on advanced learning technologies (ICALT)* (pp. 295–299). IEEE. https://10.1109/ICALT49669.2020.00096.

Carhill-Poza, A., & Chen, J. (2020). Adolescent English learners' language development in technology-enhanced classrooms. *Language, Learning and Technology, 24*(3), 52–69. http://hdl.handle.net/10125/44738.

Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text & Talk, 34*(5), 545–567. https://doi.org/10.1515/text-2014-0018

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE.

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-bass.

Díaz, B., & Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education, 105071*. https://doi.org/10.1016/j.compedu.2024.105071

Delgado-Garcia, M., Velez, S. C., & Cruz, M. D. T. (2021). Cooperative learning at university: Opinion of students and application of the instrument cooperative learning questionnaire (CLQ). *Innovations in Education & Teaching International, 59*(2), 1–10. https://doi.org/10.1080/14703297.2021.1932557

Deveci Topal, A., Dilek Eren, C., & Kolburan Geçer, A. (2021). Chatbot application in a 5th grade science course. *Education and Information Technologies, 26*(5), 6241–6265. https://doi.org/10.1007/s10639-021-10627-8

Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning, 35*(9), 2332–2360. https://doi.org/10.1080/09588221.2021.1879162

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language, Learning and Technology, 24*(1), 16–26, 10125/44705.

Dizon, G., Tang, D., & Yamamoto, Y. (2022). A case study of using Alexa for out-of-class, self-directed Japanese language learning. *Computers and Education: Artificial Intelligence, 3*, Article 100088. https://doi.org/10.1016/j.caeai.2022.100088

Engwall, O., & Lopes, J. (2022). Interaction and collaboration in robot assisted language learning for adults. *Computer Assisted Language Learning, 35*(5–6), 1273–1309. https://10.1080/09588221.2020.1799821.

Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education, 19*(1), 57. https://doi.org/10.1186/s41239-022-00362-6

Fryer, L. K., Coniam, D., Carpenter, R., & Lăpuşneanu, D. (2020). Bots for language learning now: Current and future directions. *Language, Learning and Technology, 24*(2), 8–22. http://hdl.handle.net/10125/44719.

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education, 203*, Article 104862. https://doi.org/10.1016/j.compedu.2023.104862

Han, S., Hamilton, X., Cai, Y., Shao, P., & Liu, M. (2023). Knowledge-based chatbots: A scale measuring students' learning experiences in massive open online courses. *Educational Technology Research & Development, 71*, 2431–2456. https://doi.org/10.1007/s11423-023-10280-7

Hsu, T. C., Huang, H. L., Hwang, G. J., & Chen, M. S. (2023). Effects of incorporating an expert decision-making mechanism into chatbots on students' achievement, enjoyment, and anxiety. *Educational Technology & Society, 26*(1), 218–231. https://www.jstor.org/stable/48707978.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). 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. https://doi.org/10.1111/jcal.12610

Hwang, G. J., & Chen, N. S. (2023). Editorial position paper: Exploring the potential of generative artificial intelligence in education: Applications, challenges, and future research directions. *Educational Technology & Society, 26*(2). https://doi.org/10.30191/ETS.202304_26(2).0014

Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning, 36*(7), 1338–1364. https://doi.org/10.1080/09588221.2021.1987272
Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1–2), 1–26. https://doi.org/10.1080/09588221.2021.2021241
Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education, 206*, Article 104898. https://doi.org/10.1016/j.compedu.2023.104898
Keshanchi, E., Pourdana, N., & Famil Khalili, G. (2023). Differential Impacts of pair and self-dynamics on written languaging attributes and translation task performance in EFL context. *English Teaching & Learning, 47*(2), 173–194. https://doi.org/10.1007/s42321-022-00113-w
Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York: Pergamon Press.
Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*, 973–1018. https://doi.org/10.1007/s10639-022-11177-3
Kumar, J. A. (2021). Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. *International journal of educational technology in higher education, 18*(1), 1–28. https://doi.org/10.1186/s41239-021-00302-w
Lightbown, P. M., & Spada, N. (2020). Teaching and learning L2 in the classroom: It's about time. *Language Teaching, 53*(4), 422–432. https://doi.org/10.1017/S0261444819000454
Liu, C. C., Chiu, C. W., Chang, C. H., & Lo, F. Y. (2024). Analysis of a chatbot as a dialogic reading facilitator: Its influence on learning interest and learner interactions. *Educational Technology Research & Development*, 1–29. https://doi.org/10.1007/s11423-024-10370-0
Liu, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education, 189*, Article 104576. https://doi.org/10.1016/j.compedu.2022.104576
Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2023.2240316
Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology, 39*(1), 1–8. https://doi.org/10.14742/ajet.8695
Long, M. H. (2017). Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition, 1*, 7–44. https://doi.org/10.1558/isla.33314
Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning, 33*, 403–423. https://doi.org/10.1111/jcal.12197
Meng, L. M., Li, T., & Huang, X. (2023). Double-sided messages improve the acceptance of chatbots. *Annals of Tourism Research, 102*, Article 103644. https://doi.org/10.1016/j.annals.2023.103644
Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a foreing language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11917-z
Muñoz, C. (2019). A new look at "age": Young and old L2 learners. In J. W. Schwieter, & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 430–450). Cambridge University Press. https://doi.org/10.1017/9781108333603.019.
Nguyen, Q. N., Sidorova, A., & Torres, R. (2022). User interactions with chatbot interfaces vs. Menu-based interfaces: An empirical study. *Computers in Human Behavior, 128*, Article 107093. https://doi.org/10.1016/j.chb.2021.107093
Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education, 28*(6), 1549–1565. https://doi.org/10.1002/cae.22326
Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism and Mass Communication Educator, 78*(1), 84–93. https://doi.org/10.1177/10776958221149577
Rahimi, M., & Fathi, J. (2022). Employing e-tandem language learning method to enhance speaking skills and willingness to communicate: The case of EFL learners. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2064512
Reinders, H., & White, C. (2016). 20 years of autonomy and technology: How far have we come and where to next?. http://llt.msu.edu/issues/june2016/reinderswhite.pdf.
Shafiee Rad, H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of speeko application. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2024.2309539
Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education, 188*, Article 104564. https://doi.org/10.1016/j.compedu.2022.104564
Su, F., & Zou, D. (2022). Technology-enhanced collaborative language learning: Theoretical foundations, technologies, and implications. *Computer Assisted Language Learning, 35*(8), 1754–1788. https://doi.org/10.1080/09588221.2020.1831545
Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.
Tai, T. Y. (2024). Comparing the effects of intelligent personal assistant-human and human-human interactions on EFL learners' willingness to communicate beyond the classroom. *Computers & Education, 210*, Article 104965. https://doi.org/10.1016/j.compedu.2023.104965
Tai, T. Y., & Chen, H. J. (2022). The impact of intelligent personal assistants on adolescent EFL learners' speaking proficiency. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2022.2070219
Tai, T. Y., & Chen, H. H. J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments, 31*(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801
Tseng, Y. H. (2024). Critical literacy in an indigenous elementary EFL classroom. *Linguistics and Education, 80*, Article 101234. https://doi.org/10.1016/j.linged.2023.101234
Underwood, J. (2017). Exploring AI language assistants with primary EFL students. In K. Borthwick, L. Bradley, & S. Thouësny (Eds.), *CALL in a climate of change: Adapting to turbulent global conditions* (pp. 317–321). Southampton, United Kingdom.
Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, P. M., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194*, Article 104703. https://doi.org/10.1016/j.compedu.2022.104703
Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology, 55*(1), 10–33. https://doi.org/10.1111/bjet.13334
Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL, 34*(3), 327–343. https://doi.org/10.1017/S0958344022000039
Yin, J., Goh, T. T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research, 59*(1), 154–177. https://doi.org/10.1177/0735633120952067
Yu, Q. (2024). Effects of motion-sensing technology on language learning: Evidence from a meta-analysis. *Interactive Learning Environments*, 1–17. https://doi.org/10.1080/10494820.2024.2324328
Yuan, C. C., Li, C. H., & Peng, C. C. (2023). Development of mobile interactive courses based on an artificial intelligence chatbot on the communication software LINE. *Interactive Learning Environments, 31*(6), 3562–3576. https://doi.org/10.1080/10494820.2021.1937230
Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: Pedagogical approaches, implementations, factor leading to effectiveness, theories, and future directions. *Interactive Learning Environment, 1-29*, 10.1080.10494820.2023.2202704.
Zhang, R., Zou, D., Cheng, G., & Xie, H. (2022). Implementing technology-enhanced collaborative writing in second and foreign language learning: A review of practices, technology, and challenges. *Education and Information Technologies, 27*, 8041–8069. https://doi.org/10.1007/s10639-022-10941-9
Zou, D., Zhang, R., Xie, H., & Wang, F. L. (2021). Digital game-based learning of information literacy: Effects of gameplay modes on university students' learning performance, motivation, self-efficacy and flow experiences. *Australasian Journal of Educational Technology, 37*(2), 152–170. https://doi.org/10.14742/ajet.6682