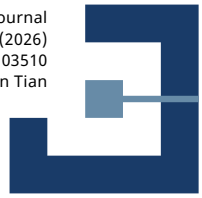
The JALT CALL Journal
ISSN 1832-4215, 22, 103510 (2026)
https://doi.org/10.29140/jaltcall.2026.103510

Castledown



# Investigating the effectiveness of corpus and AI integration in developing Chinese EFL university students' speaking proficiency and attitudes

**Hsueh Chu Chen**
*The Education University of Hong Kong, Hong Kong*
*hsuehchu@eduhk.hk*

**Mei Jung Wang**
*National Kaohsiung University of Hospitality and Tourism, Taiwan*
*sebrina@mail.nkuht.edu.tw*

**Ching Hang Justine Chan**
*The Education University of Hong Kong, Hong Kong*
*s1143997@s.eduhk.hk*

**Jing Xuan Tian**
*The Education University of Hong Kong, Hong Kong*
*s1126315@s.eduhk.hk*

This study aims to construct a Corpus and AI-aided self-directed English-speaking training programme, evaluate the effectiveness of this approach, and assess English learners' attitudes toward the application of this self-directed English-speaking training approach. 31 English as a Foreign Language (EFL) learners participated in a pre-test, a five-session Corpus and AI-aided English-speaking training, a post-test, a self-reflection, a survey, and a follow-up interview. Participants' overall speaking performances and performances on the four subskills (fluency and coherence [FC], lexical resource [LR], grammatical range and accuracy [GA], and pronunciation [PN]) improved after the training with the greatest progress in LR, while the least progress in PN. Participants agreed that the integration of spoken corpus helped them to identify the linguistic features which may influence English oral proficiency and the use of AI tools, including ChatGPT and Murf, in English speaking training was effective. Participants agreed that the integration of spoken corpus helped identify the linguistic features which may influence

English oral proficiency and the use of AI tools was effective. Interactive AI tools also helped create an interactive self-directed speaking learning environment. The majority of participants expressed positive attitudes toward the application of Corpus and AI-aided speaking training approach and showed willingness to interact with the interactive AI tools even though the tools have some limitations.

**Keywords:** English speaking, speaking teaching and learning, English speaking proficiency, spoken corpus, AI in speaking teaching and learning



## Introduction

Speaking is an important component in language teaching and learning. For non-native English speakers, achieving adequate oral proficiency in their second language (L2) is crucial, particularly for fulfilling their academic and career goals (Saito *et al.*, 2016). Speaking in English is an interactive process involving producing, receiving, and processing information, characterized by spontaneity, openness, and constant evolution (Florez, 1999). Foreign language learners often regard speaking as one of the most challenging and anxiety-inducing language skills, as they may feel stressed when speaking the target language in public (Kralova & Tirpakova, 2019). In recent years, speaking instruction has increasingly incorporated computer-assisted language learning (CALL) tools into classroom teaching (e.g., Ratnaningsih *et al.,* 2019). However, despite the use of technology, learners still commonly rely on teachers or peers for feedback during or after class (e.g., Astawa & Ardiasa, 2022; Fan, 2019; Lin & Hwang, 2018). Lai *et al.* (2016) stated that enhancing the self-directed use of technology is crucial for maximizing its potential in language learning (LL). Artificial intelligence (AI) has become commonly used in daily life and intelligent AI models can mimic human intelligence (Collins, 2021). Collaboration with multiple CALL tools is essential to effectively enhance all four dimensions in improving oral proficiency proposed by Chen and Tian (2025). This study aims to echo this study and develop a speaking learning process by integrating AI technology with a spoken corpus.

Corrective feedback (CF) plays an important role in language teaching and learning and is usually provided by teachers or peers, which aims to improve learners' linguistic accuracy (Ellis, 2009). However, sometimes peer feedback is not 100% accurate and may contain some problems since students are unfamiliar with subject knowledge (e.g., Gielen *et al.*, 2010; Yeh *et al.*, 2019). With the current development of AI technology, interactive AI simulates communication between humans. We used interactive AI technology in the current study to replace teachers or peers in providing feedback on English speaking to English learners. Additionally, in this research, we attempted to combine technologies with different functions (spoken corpora & AI tools) and use these two tools to form a self-directed speaking learning approach.

## Literature review

### *Corpus-aided LL*

Using language corpora could aid learners in inputting authentic language, motivate them to analyze the language independently, and encourage them to engage in speculation about language structures, thus developing their ability to identify language patterns through the use of authentic language data (Gut, 2021). Additionally, the use of language corpora could also raise learners' awareness (e.g., Chen & Tian, 2022; Chen & Wang, 2016) by getting access to corpus data. Language corpora have been effectively integrated into L2 classrooms due to several benefits to LL, and the integration of learner corpora is highly encouraged with significant pedagogical value (e.g., McEnery & Wilson, 2011; Sinclair, 2004). Previous studies on corpus-aided LL mainly focused on pragmatics (e.g., Bouzekria *et al.*, 2023) or writing instruction (e.g., Qiu, 2024), while research utilizing spoken corpora remains relatively limited.

A spoken corpus is a collection of speech recordings with written transcriptions and annotations at various levels, serving different purposes (Adolphs & Knight, 2010). Previous studies identified the effectiveness of spoken corpora in pronunciation learning (e.g., Chen & Tian, 2020). However, the lack of feedback from spoken corpus was the main weakness of corpus-assisted pronunciation learning. Feedback on pronunciation was mainly provided by teachers in these studies. Limited studies were identified on investigating the effectiveness of spoken corpora in oral learning. A potential reason is the lack of spoken corpora that include speech data from learners at different proficiency levels and provide detailed linguistic analysis of these samples. In this study, learners followed previous studies by using the spoken corpus, The English Speech Corpus with Different Proficiency Levels (Chen, 2022), to raise awareness of linguistic features which influence speaking performance and perform self-reflection on whether learners themselves have the common features. Another new function for corpus is to double check feedback from AI tools.

### *AI and language teaching*

In recent years, AI is identified as systems that are equipped with cutting-edge technological advancements like machine learning and neural networks have the ability to carry out cognitive tasks, specifically learning and problem-solving (Zawacki-Richter *et al.*, 2019). With the development of AI technology, CALL has evolved into intelligent CALL (ICALL), which improves the quality of interaction between learners and technology (Kannan & Munday, 2018). AI models include analytical, functional, interactive, textual, and visual (Sarker, 2020). Interactive AI and textual AI have become widely integrated into language teaching and learning. Interactive AI (e.g., ChatGPT) plays a crucial role in facilitating efficient and interactive communication automation. Textual AI encompasses various applications such as textual analytics and natural language processing (e.g., text recognition and speech-to-text conversion). Shaikh *et al.* (2023) found ChatGPT useful for formal English LL, highlighting its ability

to generate coherent responses and enhance interactive dialogues. They cautioned against replacing human instruction, suggesting ChatGPT as a valuable supplement when combined with human guidance. Feedback provided by interactive AI tools is effective in LL (e.g., Kim *et al.*, 2021).

Previous research has shown that AI tools are effective in grammar and vocabulary learning (e.g., Hsu *et al.*, 2023; Jia *et al.*, 2022). Fathi *et al.* (2024) explored AI's influence on EFL learners' speaking abilities, including fluency, coherence, lexicon, grammar, pronunciation, and their willingness to communicate. EFL learners engaged in AI-mediated activities. These activities were more effective in enhancing speaking skills and willingness to communicate, with learners responding positively to AI-mediated instruction.

Platforms such as SpeechNotes and Murf are built on AI algorithms, offering speech-to-text (STT) and text-to-speech (TTS) functions. Abdel-Reheem Amin (2024) used a mixed-methods approach to investigate students' perceptions of using AI-powered TTS tools for learning English pronunciation. Participants had positive views of AI TTS apps for pronunciation learning while Abdel-Reheem Amin (2024) recommended integrating AI TTS tools both inside and outside the classroom due to their beneficial pedagogical impact. However, there is still a gap in research on how to use AI tools to acquire speaking linguistic features. This study envisions AI becoming a personalized LL partner for students, leveraging modern technology to address individual speaking needs and provide tailored support.

In 2025, studies attempted to integrate spoken corpora and AI technology in language teaching. Chen *et al.* (2025b) developed an online corpus-based and AI-integrated English-speaking instruction on a learning management system and examined the effectiveness of this training. Chen *et al.* (2025a) examined the integration of corpora and AI tools in online English pronunciation instruction. Spoken corpora in both studies aimed to improve learners' linguistic awareness (e.g., features that could influence learner's speaking performance). Generative AI tools provided learners with feedback, and TTS AI tools were used to generate samples. Spoken corpus data was also used to help check the accuracy of feedback from generative AI tools. However, both studies were conducted in the online speaking and pronunciation context. Whether the combination of corpora and AI technology is equally applicable in general face to face classroom settings needs further support. Moreover, the two aforementioned studies did not report whether they included participants with low AI literacy. The present study aims to fix these gaps since this study was conducted shortly after interactive AI tools were released, and all participants' AI literacy was quite low at that time.

## Research questions

Specifically, two research questions are addressed:

1. To what extent have Chinese EFL learners improved their performances in the four subskills of speaking (FC, LR, GA, and PN)?

2. What are Chinese EFL learners' attitudes toward the integration of corpus and AI technology into English LL?



## Methodology

Participants participated in a 5-session Corpus and AI-aided speaking training and evaluated this training approach.

### *Subjects*

A total of 31 sophomore undergraduate students (21 females and 10 males), all native speakers of Chinese [L1]) and learners of EFL, were invited to participate in this study. All participants majored in Applied English and had an intermediate level of English proficiency, with the International English Language Testing System (IELTS) speaking band scores ranging from 5.0–5.5.

### *Procedure*

There are six stages: a pre-test, a 5-session training (60 minutes per session), a post-test (one week after the training), a self-reflection, a survey, and a follow-up interview. The self-reflection aimed to self-record participants' interactions with the interactive AI tools and reflect on their learning experience after each session. After the training, participants were invited to complete an online survey on a volunteer basis to gain feedback on the integration of corpus and AI technology into speaking learning. 26 out of the 31 participants answered the survey. 11 out of the 26 participants who answered the survey volunteered to attend a group interview to further express their comments.

**Pre- and Post-tests.** For the pre- and post-tests, all participants performed an IELTS speaking part two (Long run) mock exam. All participants were given one minute to prepare a two-minute speech on a given topic, followed by several guided questions. A task example is shown in Figure 1. We selected six topics for the two tests (three for each). These topics were not introduced in the training session. Each participant was randomly assigned to one of the three topics.

**Topic: A holiday you want to go on in the future**

Describe a holiday you want to go on in the future
You should say:
Where you would like to go
When you would like to go
Who you would like to go with
What you would like to do there
And explain why you want to have this holiday

**Figure 1.** An example of the topic used in the tests

*Training details*

**Speech corpus.** The corpus used in this study is The English Speech Corpus with Different Proficiency Levels (hereafter referred to as the English speech corpus), which can be accessed at https://corpus.eduhk.hk/english_speech_corpus/, constructed by Chen (2022). This corpus contains 78 sets of high-quality spontaneous speech data (IELTS speaking topics) and 15 sets of classroom presentation delivered by speakers with IELTS scores ranging from 5.0 to 9.0. The spontaneous speech data is annotated for the four speaking subskills (FC, LR, GA, and PN) based on the primary dimensions of the IELTS speaking test. Key indicators of the four subskills were shown in Appendix A.

**Corpus and AI-aided speaking training.** The Corpus and AI-aided speaking training consisted of one recorded asynchronous session and four synchronous online sessions via Google Meet. Contents of each training session are listed in Appendix B. Introduction of the Corpus and AI-aided speaking learning procedure, the English speech corpus, IELTS speaking criteria, and the features which influence oral proficiency were included in the recorded session. In the four online sessions, participants learned and practiced how to give prompts to the interactive AI tool (ChatGPT 3.5) for story generation on a given topic and feedback on the four subskills, how to use ChatGPT voice extension (Voice Control for ChatGPT) to form a communicative environment, and how to use the TTS tool, Murf (https://murf.ai/), to generate tailor-made native speaker samples. By the end of each training session, participants were given an after-training activity, in which they shared the prompts given and feedback received after interacting with the interactive AI tool on each subskill.

Participants first identified the linguistic features of FC, LR, GA, and PN by browsing or searching corpus data. Then, they were taught how to give prompts to the interactive AI tool to receive feedback and gain access to authentic speech production with the TTS platform. Participants were instructed to

provide prompts related to the features influencing English oral performance, as listed in Appendix A. Sample prompts revised based on (Chen *et al.*, 2025b) were shown to the participants in the training, and they were encouraged to follow the sample prompts. For example, one prompt for LR is "Can you help me to identify the inappropriate collocation, inappropriate expression, and inappropriate word choice for my following speech? Please list the changes you have made. + your speech". One prompt for PN is "Can you help me to find out my pronunciation mistakes on English consonants for my following speech? + your speech". All participants were encouraged to follow the sample prompts and revise the sample prompts based on the key indicators that influence English speaking performance and their personal needs. For example, learners were encouraged to revise the prompt for LR as "Can you help me to add 5 idioms for my following speech? Please list the changes you have made. + your speech". However, it is worth noting that the AI tool provided feedback based on generalized pronunciation patterns stored in its database, rather than offering a personalized analysis of the learners' actual speech input. Learners then should get access to the corpus data again for a double-check of the correctness of the feedback from AI tools. In this study, learners were encouraged to doublecheck all feedback they received from AI tools, as their proficiency level was not yet sufficient to independently assess the accuracy of the feedback using their own linguistic knowledge. For example, the feedback on English consonant /ð/ was received (/ð/ in 'the' and 'this' – ensure that the tongue is lightly touching the back of the top front teeth when pronouncing this sound). Learners needed to check the corpus data if learners in the corpus produced this feature.

Learners then used the TTS platform, Murf, to generate tailor-made native speakers' samples by inserting pauses, adjusting speech speed and pitch, and selecting tones, and mimicked the samples. Additionally, they were instructed to install TTS extensions on the interactive AI tool (e.g., Voice Control on ChatGPT), allowing participants to interact with the AI tool orally. The AI tool would then provide spoken responses using its TTS function. Participants could thus interact with the AI tool through the spoken mean in addition to the normal textual mean. Figure 2 illustrates an example of the Corpus and AI-aided task.

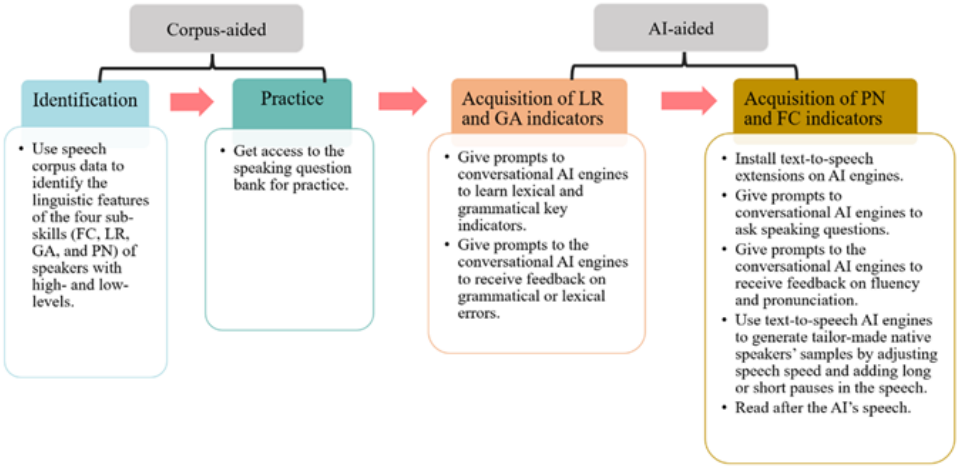



**Figure 2.** An example of corpus and AI-aided English teaching and learning tasks

### *Data analysis*

Since the pre- and post-tests in this study were IELTS speaking mock exam, we followed the IELTS Speaking assessment procedure for the pre- and post-tests speech data analysis. One qualified experienced IELTS speaking teacher, who obtained master's degrees in English Language Teaching and Applied Linguistics, rated the pre- and post-tests based on the IELTS speaking test marking criteria. The rater received IELTS official training and had good understanding of the IELTS speaking test and the 1-9 band level description. Band scores for the four subskills were then used to calculate a mean as the overall score. The rater also described each participant's common features of the four subskills based on the features in Appendix A. 22.58% random sample of the band scores for the four subskills was rated by a second rater. There was a substantial agreement (Landis & Koch, 1977) between the two raters (Cohen's Kappa: 0.66).

Qualitative data (interview & self-reflection) were coded based on 1) use of AI tools and 2) effectiveness of corpus and AI-aided training. When discussing the use of AI tools, the quality and manner of feedback, the stability and reliability of AI tools and the willingness to interact with AI tools were covered. For the effectiveness of corpus and AI-aided training, in addition to the general effectiveness, the effectiveness of the training in enhancing each sub-skills (i.e., FC, LR, GA, PN) were covered.

## Results

Participants' pre- and post-tests results will be reported to address the first research question. Intermediate self-reflection results, questionnaire results and in-depth interview results will then be reported for the second research question.

*Pre- and post-tests results*

To address the first research question, To what extent have students improved in the four aspects of speaking (FC, LR, GA, and PN)?, participants' performances on the four subskills are summarized below.

Pre- and post-tests scores were compared using Wilcoxon signed-rank tests. Results showed statistically significant improvements across all components (Table 1). Participants' overall speaking performances improved significantly from 5.40 in the pre-test to 6.21 in the post-test ($Z = -4.626$, $p < .001$) with a large effect size (r = 0.587). Before the training, the 31 participants received 5.26 for FC, 5.48 for LR, 5.65 for GA and 5.00 for PN. After the training, significant improvements were found for all subskills and the overall score. Participants' FC performance was improved to 5.90 significantly with $Z = -4.743$, $p < .001$ and a large effect size (r = 0.602). Such growth could also be observed in the LR performance (pre-test: 5.48; post-test: 6.40; $Z = -4.444$; $p < .001$; r = 0.564). Participants performed the best in GA among the four subskills in both pre- and post-tests, scoring 5.65 and 6.47 respectively. This improvement was significant, with $Z = -3.939$, $p < .001$ and a large effect size (r = 0.500). The subskill, PN, was scored the lowest among the four subskills in the pre-test. The Wilcoxon signed-rank test revealed that participants' PN performance showed significant improvement after the training with $Z = -4.696$, $p < .001$ and a large effect size (r = 0.596). Despite this significant improvement, PN remained the lowest scoring subskill in the post-test. Notably, participants demonstrated the most significant improvement in LR, with a mean score increase of 0.92 from the pre-test to the post-test. GA followed closely behind with an improvement of 0.82. FC and PN showed improvements of 0.64 and 0.61, respectively.

**Table 1.** Participants' performances on the four subskills in the pre- and post-tests (N=31)

| | Pre-test | | | Post-test | | | Test | Effect |
|---|---|---|---|---|---|---|---|---|
| Subskills | Mean | SD | Median | Mean | SD | Median | statistics | size |
| FC | 5.26 | 0.34 | 5.0 | 5.90 | 0.35 | 6.0 | Z = −4.743*** | r = 0.602 |
| LR | 5.48 | 0.54 | 5.5 | 6.40 | 0.42 | 6.5 | Z = −4.444*** | r = 0.564 |
| PN | 5.00 | 0.37 | 5.0 | 5.61 | 0.38 | 5.5 | Z = −4.696*** | r = 0.596 |
| GA | 5.65 | 0.64 | 5.5 | 6.47 | 0.50 | 6.5 | Z = −3.939*** | r = 0.500 |
| Overall | 5.40 | 0.37 | 5.5 | 6.21 | 0.34 | 6.0 | Z = −4.626*** | r = 0.587 |

Note. FC = Fluency and coherence; LR = Lexical resource; GA = Grammatical range and accuracy; PN = Pronunciation; *** p < .001

The rater also provided detailed descriptions of each participant's performance on the four subskills for the two tests, aiming to identify specific features that influenced the rater's understanding of the participants' speech. The common positive and negative features of the four subskills are listed in Tables 2 and 3. The number of participants who produced these features and the proportion of participants who had these features were included. Difference of the employment of each feature was calculated by deducting the percentage of pre-test by

that of post-test, with the positive values indicating the increased employment of the feature in post-test. Therefore, for positive features, positive values in the difference represented an improvement in employing the feature, whereas for negative features, negative values in the difference represented the same improvement (i.e., employing fewer negative features).

**Table 2.** Employment of positive features among participants' pre- and post-tests (N=31)

| Features | Pre-test | Post-test | Difference (Post-test – Pre-test) |
|---|---|---|---|
| **LR** | | | |
| Use of less common words | 61.29% | 58.06% | −3.23% |
| Use of idiomatic expressions | 16.13% | 29.03% | +12.9% |
| Use of paraphrases | 80.65% | 96.77% | +16.12% |
| Use of appropriate collocations | 77.42% | 96.77% | +19.35% |
| **GA** | | | |
| Use of adverbial clause | 83.87% | 87.10% | +3.23% |
| Use of noun clause | 48.39% | 77.42% | +29.03% |
| Use of relative clause | 64.52% | 61.29% | −3.23% |
| Use of passive voice | 35.48% | 19.35% | −16.13% |

For FC, participants produced inappropriate silent pauses, overused some connectives and discourse markers (e.g., and), and had frequent repetition and self-correction. The number of participants who produced these FC features decreased from the pre-test to the post-test. Participants tended to avoid adding inappropriate silent pauses (pre-test: 90.32%; post-test: 87.10%), using limited connectives and discourse markers (pre-test: 74.19%; post-test: 0.00%), and using frequent repetition and self-reflection (pre-test: 16.13%; post-test: 6.45%).

In the pre-test, 61.29% of the participants used less common words (e.g., cozy); 16.13% of them used idiomatic expressions (e.g., hustle and bustle); 80.65% of them produced paraphrases (e.g., essential → vital → crucial → play a role); and 77.42% of them produced appropriate collocations (e.g., interact with). The proportions of participants who used idiomatic expressions, paraphrases, and appropriate collocations increased from the pre-test to the post-test (29.03%, 96.77%, and 96.77%, respectively). For the negative features, inappropriate collocations (e.g., I can't think ^*of* a better place) and inappropriate word choices (e.g., enhance *responsiveness*: should be 'enhance *sense of responsibility*') were frequently found. After the training, fewer participants produced these two negative features in the post-test (use of inappropriate collocations: decreased from 48.39% to 22.58%; inappropriate word choice: decreased from 77.42% to 48.39%).

**Table 3.** Employment of negative features among participants' pre- and post-tests (N=31)

| Features | Pre-test | Post-test | Difference (Post-test – Pre-test) |
|---|---|---|---|
| **FC** | | | |
| Inappropriate silent pause | 90.32% | 87.10% | −3.22% |
| Overuse some connectives and discourse markers | 74.19% | 0.00% | −74.19% |
| Frequent repetition and self−correction | 16.13% | 6.45% | −9.68% |
| **LR** | | | |
| Use of inappropriate collocations | 48.39% | 22.58% | −25.81% |
| Inappropriate word choices | 77.42% | 48.39% | −29.03% |
| **GA** | | | |
| Noun errors | 32.26% | 25.81% | −6.45% |
| Verb errors | 87.10% | 51.61% | −35.49% |
| Article errors | 22.58% | 22.58% | 0% |
| Preposition errors | 12.90% | 32.26% | +19.36% |
| **PN** | | | |
| Omission of a consonant in a consonant cluster | 35.48% | 19.35% | −16.13% |
| Omission of final consonant | 16.13% | 16.13% | 0% |
| /z/ to (is replaced by) /s/ | 16.13% | 9.68% | −6.45% |
| /m/ to /n/ | 12.90% | 9.68% | −3.22% |
| /eɪ/ to /e/ | 41.94% | 29.03% | −12.91% |
| /ɪ/ to /i/ | 38.71% | 38.71% | 0% |
| /æ/ to /e/ | 22.58% | 19.35% | −3.23% |
| Extra stressed syllables | 74.19% | 74.19% | 0% |
| Inappropriate primary stress | 22.58% | 19.35% | −3.23% |
| Stressed syllables share similar duration with unstressed syllables | 93.55% | 90.32% | −3.23% |

Participants attempted to produce complex structures in the pre-test, including adverbial clause (83.87%), noun clause (48.39%), relative clause (64.52%), and passive voice (35.48%). In the post-test, the proportions of participants who produced adverbial clauses and noun clauses increased to 87.10% and 77.42% respectively. Four types of common grammatical errors were reported, noun errors (32.26%), verb errors (87.10%), article errors (22.58%), and preposition errors (12.90%). After the training, the proportions of participants who made noun errors and verb errors showed decreases to 25.81% and 51.61% respectively.

Consonant-related errors identified in the pre-test included omitting a consonant in a consonant cluster (e.g., pronouncing "complex" /kɒmpleks/ as /ˈkɒmples/), and omitting final consonants. Participants also frequently substituted sounds, replacing /m/ with /n/ (e.g., "dream" /dri:m/ pronounced as /dri:n/), /eɪ/ with /e/ (e.g., "place" /pleɪs/ pronounced as /ples/), /æ/ with /e/ (e.g., "relax" /rɪlæks/ pronounced as /rɪleks/). The substitution of /ɪ/ with /i/ (e.g., "enjoy" /ɪnˈdʒɔɪ/ pronounced as /inˈdʒɔɪ/) was also observed. For word stress, common

errors included adding extra stressed syllables (e.g., pronouncing "MAjor" as "MAJOR") and misplacing primary stress (e.g., pronouncing "caLLIgraphy" as "CAlligraphy"). Many participants struggled with syllable duration, often not differentiating the length of stressed and unstressed syllables. The post-test revealed decreases in omitting a consonant within a consonant cluster (dropped from 35.48% to 19.35%), substituting /s/ for /z/ (from 16.13% to 9.68%), substituting /n/ for /m/ (from 12.90% to 9.68%), substituting /e/ for diphthong /eɪ/ (from 41.94% to 29.03%), and substituting/e/ for /æ/ (from 22.58% to 19.35%). Regarding word stress, the proportion of participants producing errors in primary stress placement decreased slightly from 22.58% to 19.35%. However, there was less noticeable improvement in syllable duration, with 90.32% of participants still not clearly differentiating stressed and unstressed syllables in the post-test, compared to 93.55% in the pre-test.

### *Evaluation of the corpus and AI-aided speaking training approach*

To address the second research question, What are students' feedback on the integration of corpus and AI technology into English LL?, both quantitative and qualitive data are used, including 26 participants' questionnaire data aiming gathering participants' feedback and comments regarding the integration of corpus and AI technology into English speaking training, 31 participants' self-reflection task data aiming sharing their experience using the corpus and AI-aided English speaking learning approach, and 11 participants' follow-up interview data aiming gathering their feedback and comments on the Corpus and AI-aided English speaking training.

**Post-training survey.** A post-training online survey was administered. 26 out of 31 participants completed the survey. Results of the survey are shown in Appendix C.

Regarding the evaluation of the integration of spoken corpus, 96.15% (25 out of 26) of the participants agreed that they knew the different linguistic features of speakers with different proficiency levels as well as the benefits of using learner spoken corpus data in learning English speaking. 92.31% (24 out of 26) of them believed the corpus had provided useful information in learning English speaking and preparing speaking tests, as they became more aware of the differences between their language features and those of speakers who achieved their desired IELTS speaking score. 88.46% of them (23 out of 26) would recommend the spoken corpus to other English learners. 80.77% (21 out of 26) of the participants found that the corpus could help them understand better about the LR, whereas 69.23% (18 out of 26) found that the corpus was helpful in clarifying the GA indicators. 61.54% (16 out of 26) of the participants believed the tool helped improve FC performance. PN was believed to be least improved through the speaking data in the corpus (30.77% found helpful; 8 out of 26 participants).

For the application of AI tools in speaking training, participants generally considered the AI tools introduced were useful for improving English speaking.

88.46% (23 out of 26) of them believed that the interactive AI tool provided sufficient feedback, and 92.31% (24 out 26) of them found the feedback was effective for improving their speaking performance. Specifically, 96.15% (25 out of 26) of the participants found that the feedback was efficient for grammar and vocabulary problems. 80.77% (21 out of 26) of the participants also agreed that the feedback was helpful in pointing out pronunciation problems. 88.46% (23 out of 26) believed that feedback from the interactive AI tool could help them identify problems on pausing. 92.31% of the participants knew how to improve their speaking based on the feedback.

The application of AI tools in speaking learning successfully created an interactive speaking learning environment (92.31%, 24 out of 26), and participants confirmed that they received opportunities to practice English speaking (88.46%). Although 23.08% (6 out of 26) of them found that the feedback from the interactive AI tool in general did not motivate them to practice speaking with the tools, 88.46% (23 out of 26) of them would still recommend the AI tools to their fellow learners.

**Self-reflection and interview.** Based on the qualitative data (self-reflection and follow-up interview), participants' feedback on the corpus and AI technology integration is reported from 4 aspects: 1) how the Corpus and AI-aided approach helped them improve speaking performance; 2) willingness to interact with the interactive AI tool; 3) creation of an interactive speaking learning environment; and 4) challenges and limitations of this approach. Twenty-nine participants (out of 31) expressed positive attitudes towards the application of Corpus and AI-aided speaking training approach. All the 11 interviewees agreed that this approach outperformed the usual self-learning approaches to improve speaking performance, such as referring to printed references and employing online resources (e.g., YouTube videos).

1) Improving speaking performance

a) On GA & LR

Eleven of the 26 participants stressed that the Corpus and AI-aided speaking training approach could help improve grammatical accuracy and lexical richness.

> Participant (P)1: I think this AI tool is very useful because it can help me correct some grammatical errors, and I can also ask him to help me change some inappropriate words or sentences.

The 11 interviewees also confirmed that the interactive AI tool is most helpful on LR and GA because it could paraphrase common expressions with synonyms and less frequently used words, as well as correct grammatical mistakes and suggest complex structures, in which participants could learn from and internalize the advice into their own speaking resources.

> P6: I think for vocabulary, ChatGPT can help me modify some inappropriate

use and replace some advanced vocabulary. After it gave me the answers, I can also learn the expressions that I can use for the same meaning.

b) On FC & PN

Eight of the 26 participants specifically pointed out that using the Corpus and AI-aided speaking training approach is beneficial for improving FC, and the TTS AI tool helped participants produce appropriate pauses.

P25: For Murf, the most impressive part was when I added the pause (using the TTS AI tool). I could realize whether I should pause or not and I could speak the sentences smoothly.

Sixteen out of the 31 participants reported that the application of the Corpus and AI-aided speaking training is effective in helping 1) practice English pronunciation; 2) self-identify pronunciation errors from interactive AI's feedback; 3) correct pronunciation errors; and 4) perceive different English accents (TTS tools).

P1: The most important thing is that I can also use AI to practice my pronunciation. AI can tell you whether your speech is correct or not.

P31: ChatGPT can directly give you suggestions (on the features) such as English consonants, vowels, stress and so on.

P31: Murf is a good website to improve listening and speaking. The website offers lots of countries' special accents, making it easy to understand their accent.

The 11 interviewees emphasized the TTS AI tool's role in PN and FC learning and enabling users to insert pauses and vary accents in their delivery. They could learn by listening to and mimicking speech samples produced by the TTS AI tool.

P5: I think Murf is very helpful in improving fluency. You can practice sentence by sentence slowly and then combine it. You can add some pauses, so it will make your conversation more fluid, that is, when to stop and when to continue speaking. It will not feel rushed, and it will sound more comfortable.

2) Willingness to interact with the interactive AI tool

Thirteen (out of 31) participants showed willingness to interact with the interactive AI tool. They also pointed out that Corpus and AI-aided speaking training approach was an easier and more convenient way for self-learning.

P22: These AI tools are very convenient and useful. We can use these kinds of tools often.

P24: Using ChatGPT allows me to practice English without having to find others to practice. I can practice by myself.

3) Creation of an interactive speaking learning environment

Six interviewees (out of 11) who were relatively high achievers among the 31 participants preferred the Corpus and AI-aided learning approach over self-learning methods like Google searches and online modules for creating interactive speaking environments. They found interactive AI tools beneficial for providing revisions, feedback, and accessibility, contributing to an engaging learning atmosphere.

> P8: The ChatGPT experience taught in class and the corpus method will be more interactive. If we were to answer questions on a website today, it would really be that you had read it, and then it would be over after you selected the options. There would not be any more supplements or anything like that, so it would be different.

Five out of the 11 interviewees, who made imitated improvement after the training, noted that interactions with the interactive AI tool were not as effective as interactions with teachers. They acknowledged the benefits of interactive AI tools for revisions, feedback, and convenience. However, they believed the tools did not foster an interactive learning environment. While the interactive AI tools were not seen as helpful for improving PN and FC, the TTS AI tool was viewed as more beneficial.

> P6: I think interacting with AI means that if you ask it, it will answer exactly what you need. But if you interact with a real person, he will not only give you what you need, but he may also give you some other suggestions and so on. I think that interacting with real people would actually be more interactive.

4) Challenges and limitations

Some participants reported negative attitudes that the tools were hard to use, and the responses received from the interactive AI tool were not 100% correct. Three participants faced challenges when interacting with the interactive AI tool, expressed concerns regarding the stability of the interactive AI tool and level of difficulty operating the AI tools.

> P14: Sometimes it (the AI tool) cannot understand prompts.

> P4: As for Murf, I felt that the operation was a bit complicated in the beginning, because it has quite a lot of functions, so I might not be able to absorb those operation methods at once.

Three participants also acknowledged the limitations of the interactive AI tool, for example, formal response tone, rigid responses, and fancy responses which hardly match with the contexts.

> P20: Some of the words it (the AI tool) substitutes are too formal and are typically used in English tests. In real-life communication, it's more casual, and you may not need overly formal words.

## Discussion

### *Improvements in English speaking performance*

Rater's rating results of the pre- and post-tests identified that participants' general speaking performances improved significantly in all four aspects as well as the overall performance after receiving the Corpus and AI-aided speaking training, which is in line with previous studies (e.g., Gut, 2021; Hsu *et al.,* 2023; Hwang *et al.,* 2022). Even though limited previous studies combined corpora and AI tools in speaking learning, these two tools were applied in LL separately. Corpora were encouraged to be integrated in LL (Gut, 2021). The effectiveness of AI tools in LL has been identified (Hsu *et al.,* 2023). The combination of these two tools in English speaking learning was also effective. However, the improvements made by participants in LR and GA were greater than FC and PN, being consistent with Jia et al. (2022), written feedback from the AI tools being identified effectively in language learners' vocabulary and grammar acquisition. Participants' interview results also supported this finding. They pointed out that the interactive AI tools were more effective in LR and GA learning. Improvements on the FC and PN identified in this study are new findings. Extensions which can enhance learners' oral communication with interactive AI tools and the TTS AI tools are useful for language learners' speaking learning.

Limited previous studies attempted to explore how the integration of corpora and AI tools help learners acquire linguistic features in LL. This study fixed this gap. Participants' performances on FC showed improvements after receiving the training. Generally speaking, participants produced more positive features and made fewer errors in LR and GA after the training except for preposition errors. Prepositions are challenging for language learners (e.g., Lorincz & Gordon, 2012). After identifying the linguistic features which influence a speaker's speaking performance, participants in this study attempted to use complex structures. For pronunciation, the Corpus and AI-aided learning approach was more useful in learning segmental features. For word stress and syllable duration, the effectiveness was limited. Findings of this study supported previous studies. Word stress and rhythm have been considered challenging features for language learners to acquire (e.g., Jekiel, 2022; Roach, 2009). More learning time is needed for these two features. There were only five sessions in this study. For future research, additional sessions could be included to study the effectiveness of the Corpus and AI-aided speaking learning approach on challenging linguistic features in this study.

### *Participants' attitudes towards the integration of corpus and AI technology into English LL*

In general, participants' attitudes toward using the Corpus and AI-aided speaking learning approach were positive. Participants showed willingness to interact with the interactive AI tools, which supported Hwang et al. (2022). They pointed out that communication willingness could be enhanced when learners used AI tools in L2 speaking learning.

Feedback from the interactive AI tools was instant and effective in speaking learning, especially on LR and GA. This function helped fix the limitation of the corpus-aided speaking learning approach. Detailed and instant feedback provided by the interactive AI tools was useful in speaking learning. Since the feedback provided by the interactive AI tools was mainly written feedback, using a single type of AI tool is not sufficient for improving FC and PN. This study attempted to combine interactive AI tools and TTS AI tools and created an interactive speaking learning environment. However, some participants showed negative attitudes toward the feedback from the interactive AI tools and reflected that the feedback was not 100% correct. Checking corpus data to confirm the correctness and reliability of the feedback is a significant component in Corpus and AI-aided speaking learning, which reaffirms the importance of the integration of spoken corpus in AI-aided speaking learning. In addition, being newly exposed to a spoken corpus and several emerging AI tools, some participants appeared overwhelmed by the number and novelty of tools. A more gradual introduction of corpus and AI, together with explicit AI-literacy support, may help reduce this cognitive load in future implementations.

The application of corpus and AI tools created an interactive speaking learning environment, which is in line with Kannan and Munday (2018). Integration of AI technology in LL could enhance the interactions between learners and technology (Kannan & Munday, 2018). Compared with other self-access speaking learning approaches, the Corpus and AI-aided learning approach is more interactive.

The challenges of using the Corpus and AI-aided learning approach were also reported, which are new findings. Previous studies mainly examined the effectiveness of the application of corpora or AI tools in speaking learning. However, few studies identified the limitations. Participants questioned the stability of the interactive AI tool and reported difficulties operating the AI tools. Instructions on how to use the tools should be provided.

## Implications

There are eight steps in the Corpus and AI-aided speaking learning procedure. Steps 1 and 2 belong to the corpus-aided speaking learning part. Learners first use corpus data to identify the linguistic features which influence oral proficiency by comparing speakers with high proficiency levels and speakers with low proficiency levels and then identify their own potential linguistic features by browsing the corpus data of speakers who are from the same proficiency level as participants themselves. In steps 3 and 4, learners start to learn how to use interactive AI tools to build up ideas and produce speech based on these ideas. Learners then give prompts to the interactive AI tools on the linguistic features which influence the oral performances for feedback. Learners make revisions based on the feedback after confirming the correctness of the feedback by checking corpus data. In the last step, learners learn how to get access to the TTS AI tools to generate tailor-made native speaker samples and mimic the samples. If participants' production on some linguistic features do not meet

the requirements of their target proficiency level, they need to go back to step 5, seek for feedback from the interactive AI tools, and make revisions. Figure 3 illustrates the Corpus and AI-aided speaking learning procedure. Language teachers could develop Corpus and AI-aided speaking learning tasks or lessons based on this procedure.

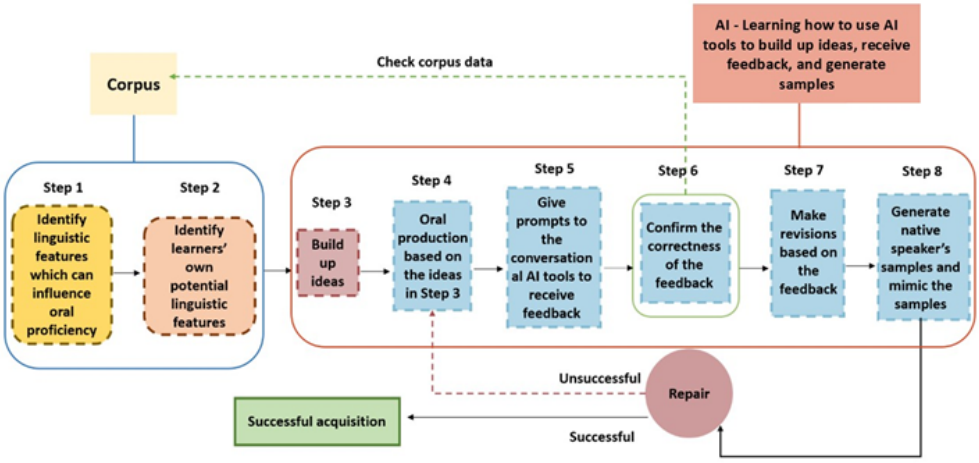



**Figure 3.** Corpus and AI-aided speaking learning procedure

## Conclusion

The study focused on integrating Corpus and AI tools into English speaking learning to enhance oral speaking skills across four subskills (FC, LR, GA, and PN) and foster interactive learning environments. Participants showed overall improvement in speaking, with the most progress in LR and the least in PN. Gains were also observed in Fluency/Coherence and Grammatical Accuracy. In addition, learners found the spoken corpus helpful for identifying linguistic features affecting oral proficiency and agreed that AI tools (e.g., ChatGPT) were effective for speaking practice. Interactive AI tools were noted for creating a self-directed learning environment. Ultimately, most participants viewed the combined corpus-AI approach favorably and expressed willingness to use such tools. We also proposed a structured Corpus and AI-aided speaking learning process for developing English speaking lessons and tasks. Weaknesses of using the Corpus and AI-aided English-speaking approach in English speaking classroom were also reported.

## Limitations and future research

For future studies and English-speaking lesson design, adjustments could be made based on the results of this study. However, there are limitations. This study only recruited one group of participants to investigate the effectiveness of the Corpus and AI-aided speaking learning approach. A control group with

participants who use the traditional speaking learning approach is needed. Another limitation is that a longer training session is needed. The current study was just an attempt to investigate the effectiveness of the Corpus and AI-aided speaking learning approach, and only 5 sessions were included. Longitudinal influences of this approach on learners' speaking performance could be examined for future studies. Additionally, future research could also compare AI feedback and human judgement to further demonstrate the reliability of AI tools in enhancing speaking development.

## References

Abdel-Reheem Amin, E. (2024), EFL Students' Perception of using AI text-to-speech apps in learning pronunciation. *SSRN*. http://dx.doi.org/10.2139/ssrn.4746800

Adolphs, S., & Knight, D. (2010). Building a spoken corpus. In A. O'Keeffe & M. McCarthy (Eds.) *The Routledge handbook of corpus linguistics*. (pp. 38–52). Routledge.

Astawa, I. G., & Ardiasa, I. K. (2022). Enhancing English-speaking skills for food and beverage services for tourism students through a collaborative learning. *Jurnal Kajian dan Terapan Pariwisata, 2*(2), 26–36. https://doi.org/10.53356/diparojs.v2i2.55

Bouzekria, H., Mashaqba, B., Al Khalaf, E., & Huneety, A. (2023). Production of pragmatic routines by Algerian EFL learners: The effect of corpus-based instruction. *Ampersand (Oxford, UK), 10,* Article 100122. https://doi.org/10.1016/j.amper.2023.100122

Chen, H. C. (2022). *The English speech corpus with different proficiency levels*. The Education University of Hong Kong. http://corpus.eduhk.hk/english_speech_corpus/

Chen, H. C. & Tian, J. X. (2020). Using a corpus-aided pronunciation instruction on Mandarin tones training for Cantonese learners. *Taiwan Journal of Chinese as a Second Language, 21,* 65–97. https://doi.org/10.29748/TJCSL.202012_(21).0004

Chen, H. C., & Tian, J. X. (2022). Developing and evaluating a flipped corpus-aided English pronunciation teaching approach for pre-service teachers in Hong Kong. *Interactive Learning Environments, 30*(10), 1918–1931. https://doi.org/10.1080/10494820.2020.1753217

Chen, H. C., Tian, J. X., & Chan, C. H. J. (2025a). Enhancing English pronunciation production and perception through corpus and AI-aided pedagogy. *Computer Assisted Language Learning,* 1–22. https://doi.org/10.1080/09588221.2025.2577364

Chen, H. C., & Wang, Q. (2016). Development and application of a corpus-based online pronunciation learning system for Chinese learners of English. *English Teaching & Learning, 40*(2), 83–114. https://doi.org/10.6330/ETL.2016.40.2.04

Chen, H. C., Zhou, X., & Tian, J. X. (2025b). Examining the effectiveness of integrating corpus-based and AI approaches for English speaking practice. *ReCALL,* 1–18. doi:10.1017/S0958344025100372

Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal: An electronic refereed journal for foreign and second language educators, 1*(1). http://dx.doi.org/10.5070/l2.v1i1.9054

Fan, N. (2019). An investigation of oral corrective feedback in an ESL listening and speaking class. *Journal of Language Teaching and Research, 10*(1), 197. http://dx.doi.org/10.17507/jltr.1001.22

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121,* 103254. https://doi.org/10.1016/j.system.2024.103254

Florez, M. C. (1999). Improving Adult English language learners' speaking skills. *ERIC Digest.* (ERIC Document Reproduction Service No. ED: 435204)

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and instruction, 20*(4), 304–315. https://doi.org/10.1016/j.learninstruc.2009.08.007

Gut, U. (2021). Spoken Corpora. In M. Paquot & S. T. Gries (Eds.) *A practical handbook of corpus linguistics* (pp. 235-256). Springer International Publishing.

Hsu, T. C., Chang, C., & Jen, T. H. (2023). Artificial Intelligence image recognition using self-regulation learning strategies: effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interactive Learning Environments, 1–19.* https://doi.org/10.1080/10494820.2023.2165508

Hwang, W. Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning,* 1–27. https://doi.org/10.1080/09588221.2022.2095406

Jekiel, M. (2022). L2 rhythm production and musical rhythm perception in advanced learners of English. *Poznan Studies in Contemporary Linguistics, 58*(2), 315–340. https://doi.org/10.1515/psicl-2022-0016

Jia, F., Sun, D., Ma, Q., & Looi, C. K. (2022). Developing an AI-based learning system for l2 learners' authentic and ubiquitous learning in English language. *Sustainability, 14*(23), 15527. https://doi.org/10.3390/su142315527

Kannan, J. & Munday, P. (2018). New trends in second language learning and teaching through the lens of ICT, networked learning, and artificial intelligence. *Circle of Linguistics Applied to Communication, 76,* 13–30. http://dx.doi.org/10.5209/CLAC.62495

Kim, H. S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics, 21,* 712–734. https://doi.org/10.15738/kjell.21..202108.712

Kralova, Z., & Tirpakova, A. (2019). Nonnative EFL teachers' speaking anxiety: Post-communist country context. *Sage Open, 9*(2), https://doi.org/10.1177/2158244019846698

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159. https://doi-org.ezproxy.eduhk.hk/10.2307/2529310

Lin, C. J., & Hwang, G. J. (2018). A learning analytics approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *Journal of Educational Technology & Society, 21*(2), 205–219. https://www.jstor.org/stable/26388398

Lorincz, K., & Gordon, R. (2012). Difficulties in learning prepositions and possible solutions. *Linguistic Portfolios, 1*(1), 14. https://repository.stcloudstate.edu/stcloud_ling/vol1/iss1/14

McEnery, A., & Wilson, A. (2011). ICT4LT module 3.4. http://www.ict4lt.org/en/en_mod3-4.htm

Qiu, X. (2024). Exploring the effect of corpus-based writing instruction on learner–corpus interaction in L2 revision: A study of Chinese EFL disciplinary writers. *TESOL Quarterly, 58*(3), 1108–1137. https://doi.org/10.1002/tesq.3308

Ratnaningsih, D., Purba, D., Wiratno, D., & Nofandi, F. (2019). The influence of computer-assisted language learning (CALL) to improve English speaking skills. In S. Madya, W. A. Renandya, M. Oda, D. Sukiyadi, A. Triastuti, Ashadi, E. Andriyanti, & N. Hidayanto (Eds.), *English linguistics, literature, and language teaching in a changing era* (pp. 144–149). Routledge.

Roach, P. (2009). *English phonetics and phonology: A practical course.* Cambridge University Press.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37,* 217–240. https://doi.org/10.1017/S0142716414000502

Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science, 3*(2), 158. https://doi.org/10.1007/s42979-022-01043-x

Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal english language learning. *European Journal of Investigation in Health, Psychology and Education, 13*(9), 1937–1960. https://doi.org/10.3390/ejihpe13090140

Sinclair, J. M. (2004). How to use corpora in language teaching. Benjamins.

Spada, N., & Fröhlich, M. (1995). COLT. *Communicative orientation of language teaching observation scheme: Coding conventions and applications.* National Centre for English Language Teaching and Research.

Yeh, H. C., Tseng, S. S., & Chen, Y. S. (2019). Using online peer feedback through blogs to promote speaking performance. *Journal of Educational Technology & Society, 22*(1), 1–14. https://www.jstor.org/stable/26558824

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0



## Appendix A

*Key indicators annotated in the English speech corpus with different proficiency levels*

| Subskill | Key indicators |
|---|---|
| **FC** | – Repetition<br>– Self-correction<br>– Discourse markers and connectives<br>– Pausing (including the number and duration of filled pause, silent pause, and silent pause at the end of a phrase, a clause or a sentence)<br>– Speech length<br>– Speech rate |
| **LR** | – Collocation (both appropriate and inappropriate)<br>– Idiomatic expressions (both appropriate and inappropriate)<br>– Inappropriate word choice<br>– Less common words<br>– Paraphrase |
| **GA** | – Grammatical errors (including article and possessive errors, noun and pronoun errors, verb errors, preposition errors, conjunction errors, adjective and adverb errors, structural errors, and other errors)<br>– Complex structure (including passive, emphatic 'do', perfect continuous, and inversion)<br>– Sentence form (including coordinate clause, relative clause, noun clause, and adverbial clause) |
| **PN** | – Segmental features (including vowels and consonants substitution, insertion, and elision)<br>– Suprasegmental features (including word stress, intonation, consonant-vowel linking, and rhythm)<br>– Mispronunciation |

Note. FC = Fluency and Coherence; LR = Lexical Resource; GA = Grammatical Range and Accuracy; PN = Pronunciation.

## Appendix B

*Training details of corpus and AI-aided speaking training*

| Online session | Contents |
|---|---|
| 1 (Asynchronous) | a) Introduction of the AI and corpus-aided speaking learning procedure<br>b) Introduction of The English Speech Corpus with different Proficiency Levels and the two functions<br>c) Introduction of the IELTS speaking criteria and the features which may influence oral proficiency<br>d) Introduction of the key indicators annotated in the speech corpus<br>e) Introduction of the two functions (Browse function & Search function) of the English speech corpus<br>f) Demonstration of ChatGPT accounts application and extension installation |
| 2 | a) Introduction of the interactive AI tool (ChatGPT) and extensions that will be used in the training<br>b) Giving support for ChatGPT account application and extension installation<br>c) Giving prompts to the interactive AI tool to build up ideas on IELTS speaking topics<br>d) Generating learners' own answers using the phrases, collocations, and idioms provided by the interactive AI tool<br>e) Getting access to the speech corpus and checking whether the length of the answer meets the learners' target level<br>f) Making revisions to produce the answers with appropriate length |
| 3 | a) Introduction of the GA and LR features produced by speakers with different proficiency levels<br>b) Introduction of how to give prompts to the interactive AI tool and receiving feedback on each grammatical and lexical key indicator<br>c) Hands-on activity to correct the grammatical and lexical errors of IETLS speaking topics pointed out by the interactive AI tool<br>d) Hands-on activity to add positive grammatical and lexical features provided by the interactive AI tool |
| 4 | a) Introduction of PN key indicators annotated in the speech corpus<br>b) Introduction of common PN features by speakers with different proficiency levels<br>c) Introduction of PN features which may influence oral proficiency<br>d) Installation of ChatGPT voice extension<br>e) Hands-on activity to give prompts to the interactive AI tool and receiving feedback on each pronunciation key indicator<br>f) Hands-on activity to correct the pronunciation errors pointed out by the interactive AI tool |
| 5 | a) Introduction of FC key indicators annotated in the speech corpus<br>b) Introduction of common FC features by speakers with different proficiency levels<br>c) Introduction of FC features which may influence oral proficiency<br>d) Hands-on activity to give prompts to the interactive AI tool and receiving feedback on discourse markers and connectives<br>e) Hands-on activity to correct the pronunciation errors pointed out by the interactive AI tool<br>f) Hands-on activity to use the TTS tool (Murf) and adding pauses to generate native-like samples and mimic the samples |

## Appendix C

*Results of the post-training online survey*

| Items | 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| **Participants' perception of the spoken corpus** | | | | | |
| I know the different linguistic features that speakers with different proficiency levels have. | 0.00% | 3.84% | 26.92% | 69.23% | 3.65 |
| The spoken corpus provides me useful information in learning English speaking. | 0.00% | 7.69% | 34.62% | 57.69% | 3.50 |
| In general, the speaking data in the corpus is useful for my speaking preparation. | 0.00% | 7.69% | 42.31% | 50.00% | 3.42 |
| The spoken data in the learner spoken corpus matches with my expectations. | 0.00% | 15.38% | 38.46% | 46.15% | 3.31 |
| I will recommend this spoken corpus to other English learners. | 0.00% | 11.53% | 30.77% | 57.69% | 3.46 |
| **Participants' perception of the AI tools** | | | | | |
| In general, the AI tools are useful for my speaking preparation. | 3.85% | 7.69% | 50.00% | 38.46% | 3.23 |
| In general, the AI-aided speaking learning provides me with sufficient feedback. | 3.85% | 7.69% | 30.77% | 57.69% | 3.42 |
| In general, feedback from the AI tools helps me to identify my speaking problems. | 3.85% | 19.23% | 23.08% | 53.85% | 3.27 |
| Feedback from the AI tools helps me to identify my grammar problems. | 3.85% | 0.00% | 26.92% | 69.23% | 3.62 |
| Feedback from the AI tools helps me to identify my vocabulary problems. | 3.85% | 0.00% | 23.08% | 73.08% | 3.65 |
| Feedback from the AI tools helps me to identify my pronunciation problems. | 3.85% | 15.38% | 34.62% | 46.15% | 3.23 |
| Feedback from the AI tools helps me to identify my word stress problems. | 3.85% | 15.38% | 30.77% | 50.00% | 3.27 |
| Feedback from the AI tools helps me to identify my pause problems. | 3.85% | 7.69% | 34.62% | 53.85% | 3.38 |
| In general, I know how to improve my speaking performance after receiving feedback from the AI tools. | 3.85% | 3.85% | 42.31% | 50.00% | 3.38 |
| In general, the AI-aided speaking learning creates an interactive speaking learning environment. | 3.85% | 3.85% | 38.46% | 53.85% | 3.42 |
| Communicating with the AI tools provides me opportunities to practice English speaking. | 3.85% | 7.69% | 42.31% | 46.15% | 3.31 |
| In general, I'm motivated to practice speaking with the AI tools. | 3.85% | 19.23% | 30.77% | 46.15% | 3.19 |
| I will recommend the AI tools to other language learners. | 3.85% | 7.69% | 38.46% | 50.00% | 3.35 |
| **Participants' perception of the spoken corpus** | | | | | |
| In general, the speaking data in the corpus helps me get a better understanding of the following criteria [you can choose more than one option]: | FC | LR | GA | PN | |
| | 61.54% | 80.77% | 69.23% | 30.77 | |

Strongly disagree = 1; Strongly agree = 4; N=26.
Note. FC = Fluency and Coherence; LR = Lexical Resource; GA = Grammatical Range and Accuracy; PN = Pronunciation.