ACADEMIC JOURNAL PERSPECTIVE:
Education, Language, and Literature
https://jurnal.ugj.ac.id/index.php/Perspective/index
p-ISSN 2354-7340
e-ISSN 2621-6981

AI SPEAKING PRACTICE AND ENGLISH LITERATURE STUDENTS' FLUENCY, INTERACTION, AND ANXIETY: A QUALITATIVE STUDY

Latifah, Nurul Anam*, Widhya Putriana
Universitas Bakti Indonesia, Banyuwangi, East Java, Indonesia
rajakota1992@gmail.com*

Abstract
Speaking is the skill that many learners of English as a foreign language (EFL) find hardest to practice, and voice-based AI speaking applications now offer a private space for doing so. This qualitative descriptive study explored how English Literature students experience such applications with regard to speaking fluency, interactional competence, and speaking anxiety. Ten purposively selected students who had used voice-based AI speaking applications took part: four used a pronunciation-feedback application, two a conversation-practice application, and four a general-purpose application with a voice feature. Data came mainly from semi-structured interviews, supplemented by observation notes and voluntary reflective journals, and were analyzed through reflexive thematic analysis. Three themes were developed around one central idea, private practice with friction: a private space that lowers the cost of speaking; rehearsal of repair strategies, with initiative left largely to the AI; and friction from word-finding demands, recognition errors, heavy feedback, and demanding topics. The findings suggest that participants valued the applications as rehearsal spaces, while their perceived progress remains self-reported. Proposals for teachers and program leaders are offered, and directions for future research are discussed.
Keywords: AI speaking applications, EFL speaking, fluency, interactional competence, speaking anxiety

INTRODUCTION

Speaking is often described as the most demanding productive skill for learners of English as a foreign language (EFL). A speaker has to retrieve words, build grammatical structures, and manage the exchange with another person, all within a few seconds (Skehan, 2018). For many learners, the chances to practice these things are limited to a few minutes in class, and those minutes are often overshadowed by the worry of being judged (Horwitz, Horwitz, & Cope, 1986). This worry is not trivial. A meta-analysis of the literature reports a moderate negative association between second language anxiety and achievement (Teimouri, Goetze, & Plonsky, 2019), which helps to explain why many learners avoid speaking when they could be building fluency. Indonesian studies tend to examine how teaching techniques such as video-assisted project work (Kasdianti, Tusino, & Setiyono, 2024) or podcast-based projects (Pratiwi, Rukminingsih, & Mumtazah, 2024) affect speaking outcomes, and a qualitative study of Indonesian EFL classrooms reports both challenges and solutions in developing speaking proficiency (Halim, Kasim, & Pratiwi, 2025).

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

Three constructs frame the present study. Fluency is usually described through the speed, smoothness, and continuity of speech (De Jong, 2016; Suzuki, Kormos, & Uchihara, 2021), and listeners' impressions of it are closely tied to measurable features such as speech rate and pausing (Suzuki et al., 2021). Interactional competence goes beyond individual production. It concerns how speakers take turns, develop topics, and repair misunderstandings together (Galaczi & Taylor, 2018; Plough, Banerjee, & Iwashita, 2018). Speaking anxiety, in turn, includes communication apprehension and fear of negative evaluation (Horwitz et al., 1986), and it can coexist with enjoyment in the same learner (Dewaele, Chen, Padilla, & Lake, 2019). The three are related in practice. A learner who fears being judged may speak less, hesitate more, and avoid interactional risk-taking, whereas supportive interaction may ease anxiety and make longer turns possible.

Technology has long been proposed as a way to widen opportunities for practice. A systematic review of chatbot-supported language learning describes the technological, pedagogical, and social affordances of chatbots, but also points to challenges that accompany their use (Huang, Hew, & Fryer, 2022; Hwang & Chang, 2023). Chatbots that combine automatic speech recognition with conversational ability have received particular attention for speaking and listening development (Jeon, Lee, & Choe, 2023). The arrival of large language models has changed what these tools can do. Generative AI can now sustain open-ended conversation, adjust its language to the learner, and give feedback on demand (Kasneci et al., 2023; Kohnke, Moorhouse, & Zou, 2023). A recent systematic review of 49 classroom-based studies published between January 2023 and December 2024 reports that generative AI is used as a feedback provider, tutor, interaction facilitator, and conversation partner, that higher education is the dominant setting, and that most studies investigate learner perceptions such as attitude, self-efficacy, and motivation (Lee, Choe, Zou, & Jeon, 2026).

This article uses two related terms. A voice-based AI speaking application is any application in which the learner speaks aloud and the system responds through speech recognition together with AI-generated conversation or feedback. A voice-based generative AI application is the subset of these in which a large language model produces open-ended spoken replies that adapt to what the learner says. AI speaking tools are therefore not a single technology. Some applications are built mainly around pronunciation feedback, while others offer open-ended conversation, and the affordances that learners experience depend on these design choices (Jeon et al., 2023). How learners use the tools, and how far they benefit, is also shaped by teachers' guidance and by learners' own digital skills (Chiu, Moorhouse, Chai, & Ismailov, 2024; Kohnke et al., 2023). For this reason, accounts of what students actually do and feel during practice are needed alongside studies of outcomes.

Studies focusing on speaking are encouraging. Fathi, Rahimi, and Derakhshan (2024) report that AI-mediated interactions improved EFL learners' speaking skills and willingness to communicate, and Wei (2023) finds benefits of AI-supported instruction for English achievement, motivation, and self-regulated learning. Qualitative work describes how learners use chatbots as conversational partners (Belda-Medina & Calvo-Ferrer, 2022), how they experience ChatGPT in language learning tasks (Xiao & Zhi, 2023), and how they adopt AI tools in informal English learning outside the classroom (Liu, Darvin, & Ma, 2024, 2025). An Indonesian qualitative study reports positive perceptions of an AI chatbot in writing, together with challenges such as over-reliance and shallow responses (Janah, Tusino, & Dewi, 2025). Research on affect points in a similar direction. Indonesian vocational high school students report feeling less anxious when speaking to speech-recognition websites than when speaking to other people, although the experimental part of

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

the same study does not find an overall change in anxiety (Bashori, Van Hout, Strik, & Cucchiarini, 2022). Likewise, voice assistants have been examined in relation to adolescent learners' willingness to communicate (Tai & Chen, 2023).

Despite this growing evidence, one gap is particularly relevant here. Fluency, interactional competence, and anxiety tend to be studied separately (for example, speaking performance in Nguyen Huu, 2025; perceived fluency in Mudawy, 2025; and anxiety in Bashori et al., 2022), and mostly through outcome measures or survey ratings, although in a real practice session they occur together: a learner hesitates, repairs, and feels uneasy in the same moment. The gap is sharpest for interactional competence, which covers turn-taking, topic development, and repair and is understood as co-constructed between speakers (Galaczi & Taylor, 2018; Plough et al., 2018). It is unclear how much of it can be practiced with a partner that responds instantly and rarely misunderstands in the way humans do, and Lee et al. (2026) explicitly call for more research on learners' behavioral interactions with generative AI. Learners' own accounts of friction are as informative as their accounts of benefit, as recent work on bias in speech-recognition chatbots (Jeon, Lee, & Coronel-Molina, 2024) and on possible over-reliance on AI dialogue systems (Zhai, Wibowo, & Li, 2024) suggests. A connected, learner-centered account of the three dimensions during AI speaking practice is therefore still needed.

This gap is examined here in a specific setting. Qualitative studies of AI-supported speaking among Indonesian university students are beginning to appear, including a study of first-semester English Literature students who used ChatGPT as a conversational rehearsal partner (Gultom, 2026) and a study of AI-enhanced academic speaking in an Indonesian university context (Panggua et al., 2025). Other work has examined AI chatbots in relation to learners' anxiety and willingness to communicate (AlTwijri & Alghizzi, 2024; Lai & Lee, 2024; Zhang, 2026). An exploratory search of web-indexed scholarly sources, conducted on 19 October 2026 with combinations of terms such as "Indonesian university students," "generative AI," "voice-based AI," "speaking fluency," "interactional competence," and "speaking anxiety," did not identify studies that examine learners' accounts of fluency, interactional competence, and anxiety together across different types of voice-based AI speaking applications. Because the search was exploratory rather than systematic, this gap should be understood in relation to the sources examined and not as evidence that no such studies exist. English Literature students are a useful case because they must use English not only for everyday communication but also for academic discussion of literary texts, where the pressure to "sound knowledgeable" may be especially strong.

This study therefore explores how English Literature students experience voice-based AI speaking applications. It is guided by three research questions: (1) How do students experience the use of these applications in practicing and developing speaking fluency? (2) How do they perceive the role of the applications in supporting interactional competence? (3) How do they experience speaking anxiety when using them? By examining the three dimensions in the same group of learners, the study offers a connected account that can inform how teachers and learners integrate AI speaking tools into EFL programs.

METHOD

Research Design

This study used a qualitative descriptive design. Qualitative description stays close to the data and to participants' own words, which makes it appropriate when the aim is to describe

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

an experience rather than build a new theory (Hall & Liebenberg, 2024; Kim et al., 2017). The study adopted an interpretive stance: learners' accounts were treated as meaningful descriptions of their practice, while the researcher's observations were used to check and enrich those accounts rather than to overrule them. Data were analyzed with reflexive thematic analysis (Braun & Clarke, 2019, 2021a), in which themes are understood as patterns of shared meaning actively developed by the researcher, not as topics that simply "emerge" from the transcripts.

The study did not compare applications; it asked how learners experience voice-based AI speaking practice as a whole. Because the applications differ in design, the type of application used by each participant is recorded in Table 1 and carried into the Findings so that accounts can be read against it.

Participants and Context

Participants were recruited through purposive sampling. To be eligible, participants had to be undergraduate students in the English Literature program at Universitas Bakti Indonesia and have direct experience using the voice or conversation features of an AI-supported application to practice English. Ten students aged 19–21 participated in the study. To protect their confidentiality, participants were assigned identification codes (P01–P10), which are used throughout this article.

The participants were enrolled in the 4th, 6th, and 8th semesters. Their English proficiency levels were recorded as pre-intermediate, intermediate, or upper-intermediate (Table 1). These categories are reported as recorded in the participant-profile data and are not presented as verified CEFR levels.

The applications differed in their primary functions. ELSA Speak was used mainly for pronunciation practice and feedback, while Speak was used for conversational speaking practice. The four participants in the third group used ChatGPT Voice, a generative-AI voice conversation feature that allows users to engage in spoken, open-ended interactions. These applications were identified according to the features participants reported using during the study. Data collection took place in July–August 2026; however, the specific application versions, update dates, and underlying model versions used by individual participants were not systematically recorded. Therefore, this study does not make claims about the precise technical configurations or model versions involved in each participant's practice. ChatGPT Voice is identified as a generative-AI voice conversation feature, whereas the technological architecture of the specific ELSA Speak and Speak features used by participants was not independently verified. The applications and features reported by participants are presented in Table 1 to distinguish pronunciation-focused practice from conversational practice. This distinction is descriptive and is not intended as a comparison of application effectiveness.

Participants reported using their respective applications for between two months and one year, with a median duration of 4.5 months. Their reported practice frequency ranged from one or two sessions per week to almost daily. Four participants used ELSA Speak, two used Speak, and four used ChatGPT Voice. Their main purposes included improving pronunciation and fluency, practicing everyday conversation, discussing academic topics, preparing for presentations, expanding spoken vocabulary, and developing confidence in spontaneous responses.

The applications differed in their primary functions. ELSA Speak was used mainly for pronunciation practice and feedback, whereas Speak was used for conversational speaking practice. ChatGPT Voice, a voice conversation feature of ChatGPT, was used for open-

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

ended spoken interactions with generative AI. The applications and features reported by participants are identified according to their use during the study period (July–August 2026). However, the specific application versions, update dates, and underlying model versions used by individual participants were not systematically documented. Therefore, this study does not make claims about the precise technical configurations of the applications or attribute all reported experiences to large language models. ChatGPT Voice is identified as a generative-AI voice conversation feature, while the underlying technological architecture of the specific ELSA Speak and Speak features used by participants was not independently verified. The application types are presented in Table 1 to distinguish pronunciation-focused practice from conversational practice. This distinction is descriptive and is not intended to compare the effectiveness of the applications. Table 1 presents the participants' profiles and the applications they used.

Table 1: Profile of the participants

ID	Application	App type	Semester	English level	Length of use	Practice frequency	Main purpose
P01	ELSA Speak	Pronunciation-feedback app	6th	Intermediate	6 months	3 times a week	Pronunciation and fluency
P02	Speak	Conversation-practice app	4th	Intermediate	3 months	2 times a week	Everyday conversation
P03	Voice AI feature	General-purpose app with voice AI feature	6th	Upper-intermediate	4 months	4 times a week	Academic topic discussion
P04	ELSA Speak	Pronunciation-feedback app	6th	Upper-intermediate	1 year	Almost daily	Pronunciation and confidence
P05	Voice AI feature	General-purpose app with voice AI feature	4th	Intermediate	2 months	1–2 times a week	Practice before presentations
P06	Speak	Conversation-practice app	6th	Intermediate	5 months	3 times a week	Spontaneous responses
P07	Voice AI feature	General-purpose app with voice AI feature	4th	Pre-intermediate	3 months	2 times a week	Expanding spoken vocabulary
P08	ELSA Speak	Pronunciation-feedback app	8th	Upper-intermediate	8 months	3–4 times a week	Pronunciation and fluency
P09	Voice AI feature	General-purpose app with voice AI feature	6th	Intermediate	6 months	3 times a week	Everyday conversation
P10	ELSA Speak	Pronunciation-feedback app	8th	Upper-intermediate	3 months	Almost daily	Academic topic discussion

(Source: Research data, 2026)

Note. P01–P10 are participant codes used to maintain confidentiality. Participants were aged 19–21 and enrolled in the English Literature program at Universitas Bakti Indonesia. Semester, length of use, practice frequency, and main purpose are based on the participant-profile information. English levels are reported as recorded and are not verified CEFR levels. Application type describes the principal function of the application or feature used and does not by itself establish whether the interaction involved generative AI.

Ten participants were considered adequate for the focused aim of the study and the specificity of the participants' experience. All participants had direct experience of using voice or conversation features in applications to practice English, which made their accounts relevant to the research questions. This rationale is consistent with the concept of information power, which holds that the adequacy of a qualitative sample depends on factors

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

such as sample specificity, study aim, theoretical framework, quality of dialogue, and analytical strategy (Malterud, Siersma, & Guassora, 2016).

Because themes in reflexive thematic analysis are developed through the researcher's active engagement with the data, sample size was not justified by a claim of data saturation. Sample adequacy was instead considered in relation to the scope of the research questions, the relevance of participants' experiences, and the depth of the analysis (Braun & Clarke, 2021b).

Data Sources and Data Collection

Three sources of data were used: semi-structured interviews, observation notes, and reflective journals. Interviews were the primary source. They were guided by fifteen open-ended questions organized into three sections corresponding to the research questions: fluency (Q1–Q5), interactional competence (Q6–Q10), and speaking anxiety (Q11–Q15). The interviewer began with open-ended questions, encouraged participants to provide concrete examples, and used follow-up prompts such as "Can you tell me more?" and "What happened next?" to obtain richer descriptions of their experiences without directing their responses toward the researcher's assumptions. Each interview lasted approximately 30–45 minutes. With participants' permission, the interviews were recorded and subsequently transcribed. Data collection took place in July and August 2026. Interviews were conducted in Indonesian and English, depending on the language used by each participant; where an interview was conducted in Indonesian, the excerpts quoted in the Findings are English translations.

The second data source consisted of observation notes from speaking-practice sessions involving the ten students. The researcher conducted the observations, with each student observed once in an in-person session lasting approximately 30 minutes. With participants' explicit consent, the sessions were recorded. The observation notes documented participants' observable actions and verbal interactions, distinguishing descriptive records from tentative interpretations. As the researcher was also a lecturer in the English Literature program, the possibility that the lecturer–student relationship and the presence of an observer influenced participants' practice was acknowledged. The observations were therefore treated as supplementary contextual data rather than standardized assessments of speaking ability. The recorded sessions were not used to assign proficiency scores or measure changes in speaking performance.

The third data source was a voluntary reflective journal. Participants were invited to describe their speaking-practice activities, identify aspects they found easy or difficult, reflect on their feelings, explain the strategies they used, and consider what they might do differently in later practice. Journal participation was voluntary, and the journal data were treated as supplementary material rather than a requirement for inclusion in the study. One journal entry from P03 is quoted in the Findings. Observation notes and journal entries were read alongside the interview transcripts to contextualize and cross-check participants' accounts. The Findings cite these supplementary sources only where they illustrate or add detail to an interview account, including four brief observation episodes and the journal entry from P03.

Data Analysis

The analysis followed the reflexive thematic approach described by Braun and Clarke (2019, 2021a). It began with familiarization, during which the researcher read the interview transcripts, observation notes, and reflective journals repeatedly to develop an understanding of participants' accounts. Initial codes were then generated for segments relevant to the

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

research questions, with participants' own words retained where they captured meaningful aspects of their experiences. The coding process focused on participants' descriptions of speaking fluency, interaction with AI, and speaking anxiety, while remaining open to other relevant patterns in the data.

The initial codes were examined in relation to one another and developed into broader patterns of meaning. Candidate themes were reviewed against the coded extracts and the full data set to assess their coherence, distinctiveness, and relevance to the research questions. The researcher then refined the themes and organized them around the central interpretive concept of *private practice with friction.* This concept captures the tension between the perceived benefits of practicing English privately with AI and the difficulties that participants encountered, including speech-planning demands, interactional limitations, excessive feedback, and technical disruptions. Each theme was defined to capture a distinct pattern of meaning, and the final analysis was written using anonymized participant excerpts.

Table 2 illustrates how selected excerpts were interpreted through initial codes and categories in the development of the themes. The examples show the analytical process and do not represent the complete coding framework or the entire data set.

Table 2: Illustration of the coding process

Illustrative excerpt (participant ID)	Initial code	Category	Theme
P01: "I think I can speak for longer now."	Longer speech; grammar concern	Perceived fluency change	Theme 1. A private space that lowers the cost of speaking
P03: "I usually choose one topic, such as a novel or a social issue, and talk to the AI for around ten minutes."	Self-directed repetition	Independent practice	Theme 1. A private space that lowers the cost of speaking
P05: "I become more comfortable, especially because nobody in class is listening."	Absence of audience; lower social pressure	Perceived safety	Theme 1. A private space that lowers the cost of speaking
P03: "I ask it to explain the question in another way."	Clarification request	Repair strategy	Theme 2. Rehearsing repair, leaving initiative to the AI
P07: "I do not always remember to ask a question back."	Limited follow-up questions	Interactional initiative	Theme 2. Rehearsing repair, leaving initiative to the AI
P02: "I need time to find the English words."	Lexical retrieval; pausing	Speech planning	Theme 3. Friction at the point of speaking
P08: "If there are too many corrections, I feel overwhelmed and forget what I wanted to say."	Feedback overload	Cognitive and emotional load	Theme 3. Friction at the point of speaking
P06: "I still feel uncomfortable when the AI cannot recognize a sentence that I believe I pronounced clearly."	Recognition frustration	Technical disruption	Theme 3. Friction at the point of speaking

(Source: Research data, 2026)

The excerpts illustrate the analytical process. Initial codes summarize meaningful features of the excerpts, categories group related codes, and themes represent broader patterns of meaning relevant to the research questions.

The researcher was a lecturer in the English Literature program at Universitas Bakti Indonesia and was directly involved in the data collection process, including conducting the semi-structured interviews and observing the participants' speaking-practice sessions. The researcher also had prior experience using AI-supported speaking applications, which

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

contributed to an informed understanding of the technological features and practical experiences associated with AI-assisted English-speaking practice. This experience, together with the researcher's academic background in English language and literature education, informed an interest in technology-assisted language learning. Nevertheless, the researcher approached the study with the intention of exploring students' experiences and perceptions rather than demonstrating the superiority or effectiveness of any particular application. Reflexive memos were used to consider how the researcher's prior experience, expectations, and assumptions about AI-supported speaking practice might influence data interpretation.

The researcher also had a lecturer student relationship with the participants, representing a potential source of power imbalance. This relationship may have influenced students' willingness to participate or their comfort in expressing critical views about the applications. To minimize potential pressure, participation was voluntary, and participants were informed of their right to decline participation, skip questions, or withdraw without penalty or adverse academic consequences. Participants' identities were anonymized using codes (P01–P10), and their responses were treated confidentially. These measures were intended to reduce potential pressure associated with the researcher's institutional role. Nevertheless, the possibility that the lecturer–student relationship influenced participants' responses or behavior during the study cannot be entirely excluded and is acknowledged as a potential limitation.

The researcher personally transcribed the interview recordings and checked the transcripts against the original audio recordings to ensure the accuracy of the transcribed accounts. Interviews conducted in Indonesian were translated into English by the researcher for reporting purposes. The translations were checked against the original Indonesian transcripts to preserve the intended meanings of the participants' statements. The researcher remained attentive to differences in expression and meaning that could arise during translation so that the English extracts represented the participants' accounts as accurately as possible.

The researcher conducted the coding and theme development using reflexive thematic analysis. Reflexive memos and an audit trail were maintained to document analytical decisions and support critical reflection throughout the process. Analytical discussions with the supervisor and a peer provided opportunities to examine emerging interpretations, question assumptions, and consider alternative readings of the data. These discussions were intended to support analytical reflection rather than establish inter-coder agreement or calculate coding reliability.

In keeping with reflexive thematic analysis, coding and theme development were treated as interpretive and iterative activities rather than as procedures intended to produce a single objectively correct coding structure. The researcher's engagement with the data, theoretical orientation, prior experience, and assumptions about AI-supported speaking practice were therefore recognized as relevant to the analytical process. Reflexive memos were used to consider how these assumptions might shape interpretations. Because the interview guide was organized around the three constructs of speaking fluency, interactional competence, and speaking anxiety, its structure may have shaped how participants' accounts were initially grouped into themes. Accordingly, the links between these dimensions reported in the Discussion are interpretations of patterns across participants' accounts rather than objectively established causal relationships.

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

Trustworthiness and Ethical Considerations

Several strategies were employed to support the trustworthiness of the study (Nowell, Norris, White, & Moules, 2017). First, participants' interview accounts were compared with observation notes and reflective journal entries to identify areas of convergence and divergence across the data sources. Second, coding decisions, analytical reflections, and theme development were documented through reflexive memos to maintain an audit trail. Third, contrasting accounts and negative cases were actively examined to ensure that the analysis represented the diversity of participants' experiences rather than focusing exclusively on recurring patterns. Participants' statements are presented using anonymized identifiers, and identifying information has been removed to protect their privacy.

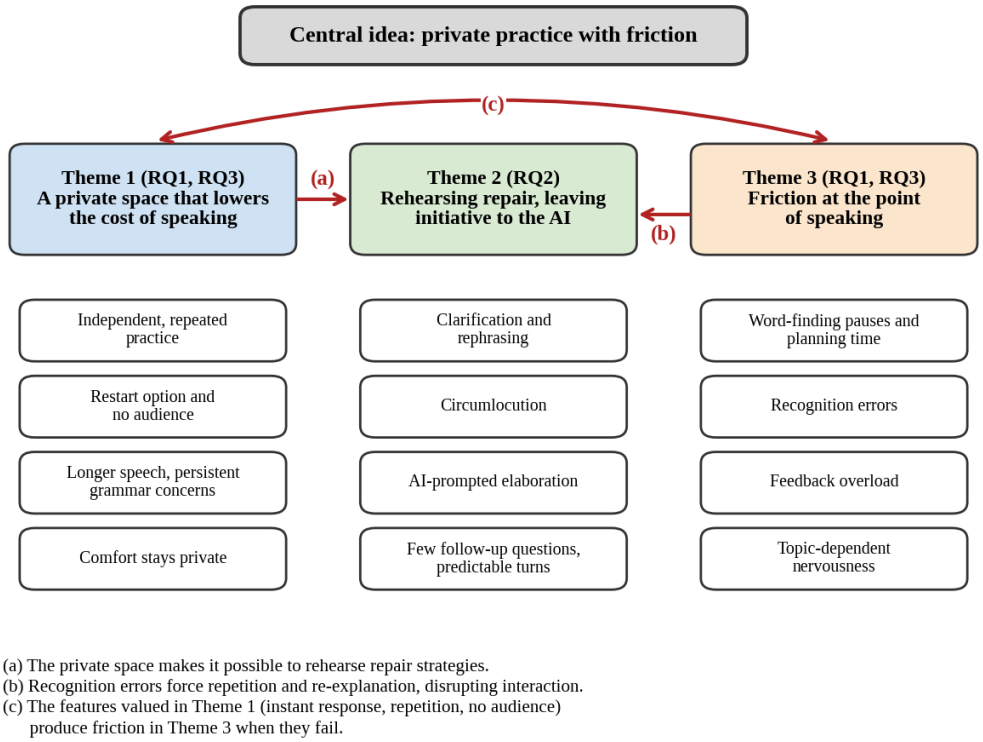
Ethical considerations were addressed throughout the research process. The study was conducted in the English Literature undergraduate program at Universitas Bakti Indonesia. Ethical approval for this research was obtained from the Research Ethics Committee of the Faculty of Language and Literature, Universitas Bakti Indonesia (Approval No. 0023/RS-FBS/UBI/V/2026, dated 15 July 2026). Data collection commenced after ethical approval had been granted on 15 July 2026. Participation was voluntary, and participants were informed about the purpose of the study, the nature of their involvement, and their right to decline to answer any question or withdraw from the study without penalty. Informed consent was obtained from all participants before data collection, and audio recordings were made only with participants' permission. All interview transcripts and research records were anonymized, stored securely, and accessed only for research purposes, and all research procedures were conducted in accordance with the applicable institutional ethical guidelines.

RESULTS AND DISCUSSIONS

Findings

The analysis produced three themes built around one central idea: private practice with friction. Participants valued the applications as a private place to keep trying, and the same features that made practice attractive sometimes produced friction when they did not work as expected. Figure 1 shows the three themes and the links between them, and Table 3 summarizes the themes, the research questions each theme informs, and the participants whose accounts supported them. Letters after participant codes in Table 3 show application type: E for ELSA Speak, S for Speak, and G for the general-purpose application with a voice feature (Table 1). Excerpts are presented as spoken or, where the interview was conducted in Indonesian, in English translation, and square brackets mark the researcher's clarifications. Interview data are the basis of all three themes. The four brief observation episodes and the one journal entry (P03) cited below were consistent with the corresponding interview accounts and serve only to illustrate or add detail to them.

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002



(a) The private space makes it possible to rehearse repair strategies.
(b) Recognition errors force repetition and re-explanation, disrupting interaction.
(c) The features valued in Theme 1 (instant response, repetition, no audience) produce friction in Theme 3 when they fail.

Figure 1. Thematic map of students' experiences with voice-based AI speaking practice. Arrows (a)–(c) mark the links between themes described in the Discussion.

Table 3: Summary of themes, subthemes, research questions, and supporting participants

Theme	Subthemes	RQ	Participants (application type)
Theme 1 (RQ1, RQ3): A private space that lowers the cost of speaking	Independent, repeated practice; the option to restart and the absence of an audience; longer stretches of speech with persistent grammar concerns; comfort that stays private	RQ1, RQ3	P01 (E), P03 (G), P04 (E), P05 (G), P09 (G), P10 (E)
Theme 2 (RQ2): Rehearsing repair, leaving initiative to the AI	Predictable turn-taking; clarification requests and rephrasing; circumlocution; elaboration prompted by the AI; limited follow-up questions; less spontaneity than human talk	RQ2	P02 (S), P03 (G), P04 (E), P06 (S), P07 (G), P08 (E)
Theme 3 (RQ1, RQ3): Friction at the point of speaking	Word-finding pauses and planning time; recognition errors that interrupt flow and shake confidence; feedback overload; topic-dependent nervousness	RQ1, RQ3	P01 (E), P02 (S), P03 (G), P05 (G), P06 (S), P07 (G), P08 (E), P10 (E)

(Source: Research data, 2026)

Theme 1: A Private Space That Lowers the Cost of Speaking

Participants described the applications first of all as a convenient place for repeated oral practice that they organized themselves. Practice was frequent for some (P04 and P10 reported using their application almost every day). P03, for example, chose a topic and stayed with it: "I usually choose one topic, such as a novel or a social issue, and talk to the AI for around ten minutes. If my answer sounds unclear, I try again."

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

The possibility of trying again was central to how learners described their progress. Asked about changes in fluency, P01 gave a measured answer that recurred, in different words, across the data set:

> Researcher: What changes, if any, have you noticed in your speaking fluency since using the application?
>
> P01: I think I can speak for longer now. I still make grammar mistakes, but I do not stop after every sentence as often as before.

Participants reported perceiving longer stretches of speech, but none described their progress as complete. Grammar concerns persisted alongside the longer stretches.

Several participants, including P04, P05, and P09, associated this type of practice with a lower level of social pressure. They attributed this sense of comfort to the privacy of individual practice and the freedom to repeat or restart their responses without feeling embarrassed in front of others. One participant explained: "Practicing with AI feels less embarrassing because I can restart. In class, I worry that people will remember my mistakes." P05 described a change in feelings over the course of a single practice session. The participant reported worrying about pronunciation before the session but also stated that during practice: "I become more comfortable, especially because nobody in class is listening." P04 expressed a similar experience, explaining that responses could be repeated without the embarrassment associated with speaking in front of classmates.

However, this sense of comfort did not necessarily extend to classroom speaking situations. P09 explained: "I am less afraid of making mistakes with AI, but I still feel nervous when speaking in front of my lecturer." This account suggests that the reduced social pressure experienced during private AI-supported practice did not automatically transfer to situations involving face-to-face interaction with a lecturer. In this respect, the applications provided a relatively private space for speaking practice, but participants could still experience anxiety when speaking in a more socially exposed academic setting.

Theme 2: Rehearsing Repair, Leaving Initiative to the AI

Participants did not describe the AI as a human-like conversational partner, but they did describe it as a place where interactional moves could be tried out. Turn-taking was one such move. P06 said: "I answer the question and then wait for the next one. It feels like taking turns, although I know the AI usually respond immediately." The exchange was recognizable as a conversation, but its rhythm was more predictable than talk with another person.

Repair strategies were more clearly rehearsed. When the AI's question was hard to follow, participants asked for rephrasing, shortened their answers, or requested examples:

> Researcher: What do you do when you do not understand the AI's response or when the conversation becomes difficult?
>
> P03: I ask it to explain the question in another way. If I still do not understand, I use a shorter answer and ask for an example.

One observed participant used almost the same formulation, asking, "Could you explain the question in another way?" The strategy was therefore not only reported but also visible in practice. A second strategy, circumlocution, was used when vocabulary failed. P02 said: "When I did not know a word, I described it using easier English. The AI continued the conversation, so I learned that I could explain an idea in another way." P08 described a

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

comparable move, explaining a difficult word in simpler English when the AI did not understand.

The AI also pushed some learners to say more. P07 reported that it sometimes asks why they hold an opinion: "That makes me add reasons, although I do not always remember to ask a question back." P06 raised the same limitation: "I can answer the AI's questions, but I do not always know what follow-up question I should ask." In both accounts the AI took the lead in developing the topic, and the learner's role was mainly to respond. P04 valued the freedom to repeat a conversation without bothering anyone, but added that the exchange "does not feel as unpredictable as talking to my friends."

Accounts of repair and initiative came mainly from users of the conversation-oriented applications (P02, P03, P06, and P07). The two ELSA Speak users who contributed to this theme spoke about circumlocution (P08) and the predictability of the exchange (P04).

Theme 3: Friction at the Point of Speaking

Alongside the benefits, participants described moments when practice became harder. Four kinds of friction recurred: word-finding, recognition errors, heavy feedback, and demanding topics. P02 explained what happened when the right word did not come:

> Researcher: What usually happens when you cannot find the right word or organize a sentence during practice?
>
> P02: Sometimes I know what I want to say in Indonesian, but I need time to find the English words. I pause or use a simpler sentence.

The observation notes point in the same direction. In one session, a participant stopped mid-utterance, said, "What is the word...", and then replaced the word with a simpler term. The researcher treated this as a probable sign of lexical retrieval difficulty and simplification, which P02's account supports, but did not assume that the pause reflected anxiety. Topic familiarity also mattered. P05 noted that practice went better with a familiar subject and that an unfamiliar question called for extra planning time: "When the AI asks a question I have never thought about, I spend more time planning my answer." A similar pattern appeared in the notes, where a participant stayed silent for about four seconds before answering the first question. The researcher recorded the silence as a descriptive observation and left its meaning open until it could be discussed with the participant.

The most visible interruption was technical. When the application did not recognize a sentence, participants repeated it, and the repetition could interrupt their line of thought. P08 said: "When the app misunderstands my pronunciation, I repeat the sentence. It can be useful, but sometimes I lose the idea I was explaining." P07 described the same disruption in stronger terms, saying that a misrecognized pronunciation left them frustrated and sometimes unable to recover their train of thought. The notes recorded a participant repeating a sentence twice after the application displayed a different recognition result. P03's journal adds a similar episode: when the AI misunderstood one sentence, "I became distracted and had to explain my point again." In the same entry, P03 wrote that asking the AI for a follow-up question helped to continue after a pause. P06 said that recognition failures could shake confidence: "I still feel uncomfortable when the AI cannot recognize a sentence that I believe I pronounced clearly. I start doubting my pronunciation." Recognition problems were reported by users of all three application types (P08, ELSA Speak; P06, Speak; and P07 and P03, the general-purpose application).

Feedback could also become excessive:

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

> Researcher: How do you react when the AI corrects, misunderstands, or fails to recognize your speech?
>
> P08: If it corrects me, I try to repeat the sentence. If there are too many corrections, I feel overwhelmed and forget what I wanted to say.

P10 voiced the same ambivalence, calling the corrections useful but sometimes confusing when there were too many. Both participants who described this overload (P08 and P10) used ELSA Speak, the pronunciation-feedback application. Finally, nervousness depended on the topic. P01 and P10 felt at ease with daily topics, but academic topics made them nervous "because I want to use accurate vocabulary and sound knowledgeable."

Discussion

The three themes describe private practice that participants valued and that also carried friction. The first research question is addressed mainly by Themes 1 and 3, the second by Theme 2, and the third by Themes 1 and 3. The sections below interpret these findings in relation to each research question and to earlier studies, and mark where they confirm, qualify, or add to what is known.

Perceived Fluency: Practice Helps, but Perception Is Not Performance

The first research question asked how students experience AI speaking applications in developing fluency. Participants reported speaking longer and stopping less often, while grammar concerns and word-finding pauses remained. This is consistent with the favorable picture in reviews and studies of AI-supported speaking, which associate chatbot practice with more willing and more extended talk (Du & Daniel, 2024; Zhang, 2026), and it resembles the finding that AI speaking tools enhance performance-oriented aspects of speaking more than knowledge-based ones (Nguyen Huu, 2025): participants described longer speech but continuing grammar concerns. A difference should nevertheless be noted. Those studies measure or report effects, whereas the present data describe how learners interpret their own progress, as in other perception-based work on AI conversational tools and fluency (Mudawy, 2025). Listeners' judgments of fluency are closely tied to measurable features such as speech rate and pausing (Suzuki et al., 2021), and no temporal measures were taken here. The participants' impression of fewer stops is therefore plausible but cannot be treated as evidence of improved performance. In this respect the findings sit close to a qualitative study of English Literature students in which AI rehearsal affected readiness and willingness to participate more than immediate linguistic improvement (Gultom, 2026).

The difficulties participants described fit attentional accounts of second language performance, in which accuracy, complexity, and fluency compete for limited attention (Skehan, 2018). P05's extra planning time on unfamiliar topics and P02's word-finding pauses illustrate this: repeated practice may speed access to what is already known, but it cannot supply what has not yet been learned. The technical disruption reported by four participants (P03, P06, P07, and P08) across all three application types adds a further point. When recognition failed, learners lost time, sometimes lost the thread of what they were saying (P07 and P08), and in one case began to doubt their own pronunciation (P06). No single participant described all of these consequences, so they are best read as a pattern across accounts and not as a sequence that learners experienced. This pattern was not described in the studies reviewed for this article and is offered as a tentative observation. The present data cannot show whether misrecognition falls more heavily on some accents, a concern raised in work on bias in speech-recognition chatbots (Jeon et al., 2024). They do

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

show that one learner (P06) questioned their own pronunciation when the tool failed to recognize it.

Interactional Competence: Rehearsing Repair, Leaving Initiative to the AI

The second research question concerned interactional competence. Clarification requests, paraphrasing, and circumlocution are all repair practices, and repair, together with turn-taking and topic management, is a central component of interactional competence (Galaczi & Taylor, 2018; Plough et al., 2018). The participants' accounts, supported by one observed episode, indicate that the applications gave learners a low-risk place to try these practices. This agrees with the finding that rehearsal with ChatGPT helps English Literature students with predictable exchanges but not with improvisation, and that they enter discussions more readily but hesitate when exchanges become unpredictable (Gultom, 2026).

At the same time, interactional competence is co-constructed: speakers adjust to each other and share responsibility for the direction of talk (Galaczi & Taylor, 2018; Plough et al., 2018). Participants described a partner that led the topic, turns that were predictable, and few follow-up questions from the learner. AI practice therefore seems to strengthen responding more than initiating. Social presence and perceived human-likeness influence learners' motivation with chatbots (Ebadi & Amini, 2024), which may be one reason why participants valued the AI as a practice partner without regarding it as equivalent to a human one. The point about initiative is tentative, however, because it rests on the accounts of two learners (P06 and P07) and needs to be tested with transcripts of learner–AI talk.

Speaking Anxiety: Lower Pressure, Conditional Comfort

The third research question concerned speaking anxiety. Participants' sense of being less embarrassed with AI than in class matches the finding that Indonesian students feel less anxious when speaking to speech-recognition websites than to people (Bashori et al., 2022), and syntheses that link AI-supported language learning with lower anxiety (AlTwijri & Alghizzi, 2024; Du & Daniel, 2024). The reason participants gave, the worry that classmates will remember their mistakes, corresponds to fear of negative evaluation, which is associated with foreign language speaking anxiety (Okyar, 2023). An AI partner removes the audience, and with it much of the perceived risk. The same study by Bashori et al., however, finds no overall change in anxiety after practice (Bashori et al., 2022), and the present data support that caution: relief with the AI did not carry over to speaking in front of a lecturer (P09), so comfort during practice should not be read as reduced anxiety over time.

The finding that comfort was conditional adds to this picture. Topic-dependent nervousness (P01 and P10) suggests that for English Literature students the pressure may lie less in speaking in general than in speaking academically, with accurate vocabulary and an appearance of knowledge. Feedback overload was described by two participants (P08 and P10), both of whom used the pronunciation-feedback application, which suggests that it may partly reflect application design rather than AI practice in general; this possibility needs a larger and more balanced sample. Studies that stress lower anxiety and higher willingness to communicate with AI chatbots (Lai & Lee, 2024; Zhang, 2026) are qualified by these accounts, which show three conditions under which comfort gave way: the topic, the amount of correction, and recognition failure. The data were not designed to detect over-reliance on AI tools, which Indonesian students also raise in relation to other tools (Khonirin & Roslaini, 2024).

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

Synthesis: Private Practice with Friction

Viewed together, the themes show that the features participants valued most (immediate response, repetition, and the absence of an audience) were also the features that produced friction when they failed. Immediate response became interruption when the application misheard, repetition became tiring when corrections piled up, and the absence of an audience meant that confidence gained in private had not yet been tested in public. Recognition errors illustrate how the three dimensions may interact: across participants' accounts they appeared as a fluency problem (lost train of thought), an interactional problem (having to repeat and re-explain), and an affective problem (doubting one's pronunciation). Studies that measure only one outcome may therefore miss how difficulties in one dimension spill into the others. This pattern is consistent with the view that conversational AI works best as a preparatory space that complements, rather than replaces, interaction with others (Gultom, 2026).

CONCLUSION

This study explored how English Literature students experience voice-based AI speaking applications in relation to fluency, interactional competence, and speaking anxiety. Across ten participants, the accounts indicate private practice with friction: the applications offer a low-pressure space for repetition and for rehearsing repair strategies, yet participants also describe word-finding pauses, recognition errors, heavy feedback, demanding academic topics, and comfort that does not carry over to speaking before a lecturer, while perceived gains remain self-reported. Tentatively, the accounts suggest that fluency, interactional competence, and anxiety may be better read together, because recognition errors appeared in accounts touching all three, and that AI practice may support responding more than initiating. Both points rest on a small number of accounts and need further testing.

Implications

The following implications are proposals for practice and were not tested in this study. For teachers, we propose that learners plan topics and key vocabulary before a session, that tasks require learners to ask questions and change topics, that feedback is limited to a few priority points where the application allows it, and that AI practice is followed by low-stakes speaking with peers or the lecturer. For students, one participant (P06) doubted their pronunciation after recognition failures, so learners may wish to be cautious before concluding that a recognition error signals a pronunciation problem. For program leaders, the accounts suggest treating applications as different tools, since pronunciation-feedback and conversation-practice applications may serve different needs, and offering guidance on how to choose and use them.

Limitations and Future Research

This study has several limitations. The sample is small and drawn from one program, and the data are mainly self-reported and may invite socially desirable answers. Participants used different applications, including one that gives mainly pronunciation feedback, and this difference is not analyzed; whether the applications used generative AI has not been verified. English proficiency is not examined as a factor, and the proficiency categories are reported as recorded and are not verified CEFR levels. Interviews were conducted in Indonesian and English, and translation of Indonesian extracts may affect how participants' meanings are represented. The interview guide was organized by the three constructs, which may have shaped how accounts were grouped into themes. Each participant was observed only once,

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002

in a brief in-person session, and the observation and journal data are limited in amount and were used as supplementary evidence only. In addition, the researcher was a lecturer in the English Literature program at Universitas Bakti Indonesia and taught the participants. This lecturer–student relationship may have created a potential power imbalance that influenced participants' willingness to participate or their comfort in expressing critical views. Although participation was voluntary and participants were informed of their right to decline, skip questions, or withdraw without academic consequences, the possibility that the researcher's institutional role influenced their responses cannot be entirely excluded.

Future studies could combine interviews with temporal measures of fluency, analyze learner AI transcripts through conversation analysis, compare applications and human partners on the same task, follow learners over a longer period, and examine whether recognition errors affect learners with different accents unequally.

REFERENCES

AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon*, *10*(10), e31053. https://doi.org/10.1016/j.heliyon.2024.e31053

Bashori, M., Van Hout, R., Strik, H., & Cucchiarini, C. (2022). Web-based language learning and speaking anxiety. *Computer Assisted Language Learning*, *35*(5–6), 1058–1089. https://doi.org/10.1080/09588221.2020.1770293

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427. https://doi.org/10.3390/app12178427

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2021a). *Thematic analysis: A practical guide*. SAGE.

Braun, V., & Clarke, V. (2021b). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, *13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with artificial intelligence (AI) based chatbot. *Interactive Learning Environments*, *32*(7), 3240–3256. https://doi.org/10.1080/10494820.2023.2172044

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2). https://doi.org/10.1515/iral-2016-9993

Dewaele, J.-M., Chen, X., Padilla, A. M., & Lake, J. (2019). The flowering of positive psychology in foreign language teaching and acquisition research. *Frontiers in Psychology*, *10*, 2128. https://doi.org/10.3389/fpsyg.2019.02128

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers*

and Education: Artificial Intelligence, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Ebadi, S., & Amini, A. (2024). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive Learning Environments*, *32*(2), 655–673. https://doi.org/10.1080/10494820.2022.2096638

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

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, *15*(3), 219–236. https://doi.org/10.1080/15434303.2018.1453816

Gultom, E. S. (2026). From classroom to chatbot: The use of AI in developing speaking proficiency. *Huele: Journal of Applied Linguistics, Literature and Culture*, *6*(1), 01–12. https://doi.org/10.30598/huele.v6.i1.p01-12

Halim, N., Kasim, N. A., & Pratiwi, D. F. (2025). Developing speaking proficiency in Indonesian EFL classrooms: A qualitative study on challenges and solutions. *The Academic: English Language Learning Journal*, *10*(1), 9–18. https://doi.org/10.52208/aellj.v10i1.1384

Hall, S., & Liebenberg, L. (2024). Qualitative description as an introductory method to qualitative research for master's-level students and research trainees. *International Journal of Qualitative Methods*, *23*, 16094069241242264. https://doi.org/10.1177/16094069241242264

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

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., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, *31*(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615

Janah, R., Tusino, T., & Dewi, P. (2025). Artificial intelligence (AI) chatbot in EFL writing: Exploring students' perspectives. *Scripta: English Department Journal*, *12*(2), 379–385. https://doi.org/10.37729/scripta.v12i2.6811

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*, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Jeon, J., Lee, S., & Coronel-Molina, S. M. (2024). Rethinking AI: Bias in speech-recognition chatbots for ELT. *ELT Journal*, *78*(4), 435–445. https://doi.org/10.1093/elt/ccae035

Kasdianti, E., Tusino, T., & Setiyono, J. (2024). Improving learners' speaking skill through video-assisted project based learning. *Scripta: English Department Journal*, *11*(2), 133–141. https://doi.org/10.37729/scripta.v11i2.5260

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khonirin, F. A., & Roslaini, R. (2024). EFL students' perceptions towards the use of Grammarly application in learning writing skills. *Scripta: English Department Journal*, *11*(2), 187–194. https://doi.org/10.37729/scripta.v11i2.5370

Kim, H., Sefcik, J. S., & Bradway, C. (2017). Characteristics of qualitative descriptive studies: A systematic review. *Research in Nursing & Health*, *40*(1), 23–42. https://doi.org/10.1002/nur.21768

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lai, W. Y. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence*, *7*, 100291. https://doi.org/10.1016/j.caeai.2024.100291

Lee, S., Choe, H., Zou, D., & Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, *34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

Liu, G. L., Darvin, R., & Ma, C. (2024). Unpacking the role of motivation and enjoyment in AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation in the Chinese context. *Computers in Human Behavior*, *160*, 108362. https://doi.org/10.1016/j.chb.2024.108362

Liu, G. L., Darvin, R., & Ma, C. (2025). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, *38*(7), 1632–1660. https://doi.org/10.1080/09588221.2024.2310288

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Mudawy, A. M. A. (2025). Exploring EFL learners' perceptions on the use of AI-powered conversational tools to improve speaking fluency: A case study at Majmaah University. *Forum for Linguistic Studies*, *7*(1), 589–598. https://doi.org/10.30564/fls.v7i1.7774

Nguyen Huu, H. (2025). AI for English speaking practice: A study of effectiveness and engagement among Vietnamese university learners. *Australasian Journal of Educational Technology*, *41*(3), 45–61. https://doi.org/10.14742/ajet.10050

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, *16*(1), 1609406917733847. https://doi.org/10.1177/1609406917733847

Okyar, H. (2023). Foreign language speaking anxiety and its link to speaking self-efficacy, fear of negative evaluation, self-perceived proficiency and gender. *Science Insights Education Frontiers*, *17*(2), 2715–2731. https://doi.org/10.15354/sief.23.or388

Panggua, S., Pongsapan, N. P., Ismail, H., Patandung, Y., Rachel, & Tangirerung, J. R. (2025). AI-enhanced academic speaking skills: A qualitative investigation of digital tool integration in Indonesian EFL university context. *Forum for Linguistic Studies*, *7*(8). https://doi.org/10.30564/fls.v7i8.10763

Plough, I., Banerjee, J., & Iwashita, N. (2018). Interactional competence: Genie out of the bottle. *Language Testing*, *35*(3), 427–445. https://doi.org/10.1177/0265532218772325

Pratiwi, S. A. A., Rukminingsih, R., & Mumtazah, L. D. (2024). The impact of project based learning using podcast on student speaking skill. *Scripta: English Department Journal*, *11*(2), 177–186. https://doi.org/10.37729/scripta.v11i2.5442

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge. https://doi.org/10.4324/9781315629766

Suzuki, S., Kormos, J., & Uchihara, T. (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, *105*(2), 435–463. https://doi.org/10.1111/modl.12706

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

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), 363–387. https://doi.org/10.1017/S0272263118000311

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, *8*(3), 212. https://doi.org/10.3390/languages8030212

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1), 28. https://doi.org/10.1186/s40561-024-00316-7

Zhang, W. (2026). The impact of AI chatbots on EFL learners' oral proficiency and willingness to communicate. *System*, *136*, 103919. https://doi.org/10.1016/j.system.2025.103919

DOI: http://dx.doi.org/10.33603/perspective.v14i2.13002