2026/1 (21. vol.) ERUDITIO – EDUCATIO
45-55 Tanulmányok / Studies

Voice-Based AI Chatbots in English Language Learning: A PRISMA-Guided Systematic Review on Recent Specialized Literature

Lili Sotkovszky – Jenő Bárdos

Abstract
This study, guided by the PRISMA protocol, systematically reviews research on voice-based AI chatbots in English language learning published from 2022 to the present. It addresses two questions: (1) What research questions and what kind of trends characterize recent publications on their use? and (2) What future research recommendations emerge regarding AI chatbots in voice mode for developing English-speaking skills? Findings show a predominant focus on speaking skills development and learners' perceptions, with studies highlighting improvements in fluency, pronunciation, and coherence. Affective factors such as anxiety reduction and willingness to communicate received moderate attention, while contextual and demographic dimensions remain underexplored. Current evidence emphasizes performance outcomes but lacks methodological robustness, with frequent calls for larger, more diverse samples and comparative designs. This review clarifies trends, gaps, and future directions, offering guidance for educators and researchers seeking effective integration of voice-based chatbots in language education.
Keywords: voice-based AI chatbots; English language learning; speaking skills development; systematic review
Subject-Affiliation in New CEEOL: Language and Literature – Applied Linguistics – Computational Linguistics
DOI: 10.36007/eruedu.2026.1.045-055

Introduction

Advancements in artificial intelligence (AI), particularly in speech technologies such as automatic speech recognition, have enhanced chatbots' ability to interact with language learners in ways that increasingly approximate human communication (Kohnke et al. 2023). At the same time, growing educational research has highlighted the use of chatbots in language teaching and applied linguistics. Recent studies, emphasizing different aspects of chatbots, have enriched specialized literature by demonstrating their applications across diverse language learning purposes, including reducing anxiety, fostering willingness to communicate, providing conversational practice, and supporting vocabulary development, among others (Jeon et al. 2023).

Although the use of voice-based AI chatbots in English language learning has expanded in recent years, more comprehensive evidence is needed, as new tools

are emerging rapidly, often without prior research or classroom piloting. To keep pace with this dynamic field, educators require ongoing input, support, and guidance – needs that systematic research can help address. Exploring the use of voice-based AI chatbots in various institutions can offer valuable insights, while English language teachers also need clearer guidance on effective task and activity design for classroom integration.

It is also important to note that researchers employ different types of chatbots with distinct features. Researchers have employed various types of chatbots, ranging from embodied systems with spoken and written modes to voice-only tools, with some developing tailored versions for specific learner groups and others relying on commercially available platforms such as ChatGPT (Jeon et al. 2023).

Some recent reviews have explored voice-based AI chatbots' affordances in language education. For example, Du and Daniel (2024) conducted a systematic review (for the years 2017-2023) showing that, despite ongoing technological and speech recognition challenges, chatbots uniquely enhance interaction and motivation in English language learning. Another review by Jeon et al. (2023) analyzed 37 empirical works on speech-recognition chatbots and proposed a conceptual framework comprising three components – goal-orientation, embodiment ("if the chatbot has a virtual body of some sort" [Jeon – Lee – Choe 2023]), and multimodality – leading to the identification of eight chatbot types. The framework offers an initial model for analyzing chatbots across disciplines and may support researchers and educators in selecting appropriate tools for specific educational objectives.

The study was structured around the following research questions:

RQ1: What research questions and what kind of trends characterize recent publications on the use of voice-based AI chatbots in language learning?

RQ2: What future research recommendations emerge from studies on the use of AI chatbots in voice mode for developing English-speaking skills?

Method

A systematic literature review was carried out following the analytical framework of the PRISMA guidelines (Moher et al. 2009). The literature review revealed that although the development of voice-based AI chatbots dates back to the 1980s, research into their role in students' English language learning only began to gain momentum after 2016 (Koc and Savas 2025), with a notable surge occurring after 2022, most likely in connection with the release of ChatGPT. As the features of voice-based chatbots continue to evolve and improve, it becomes increasingly important for this field of research to focus on the most recent studies. Accordingly, this study limits its scope to publications from 2022 to 2025.

The review focused on two major academic databases: Web of Science, and Scopus. On August 4, 2025, the authors searched for studies on the use of voice-based chatbots in English language learning using a mixture of Boolean operators "AND" and "OR" and the following keywords (TITLE-ABS-KEY): "chatbot*", "conversational agent*", "conversation*", "language learning*", "speaking*",

"EFL*", "foreign language*", "ESL*", and "second language*". The database search retrieved 404 records. After removing duplicates, 377 unique records were screened, and 275 titles and abstracts were identified as unrelated to the study's focus.

From the remaining papers, we selected those that met the following inclusion criteria: (1) employed an empirical study in language education (excluding reviews and conference papers); (2) used voice-based chatbots; (3) focused on developing foreign language speaking skills (excluding studies that targeted only vocabulary or writing skills with chatbots); (4) involved exclusively AI-based chatbots; and (5) examined English language learning only (excluding other languages). In total, 21 articles were included in this review.

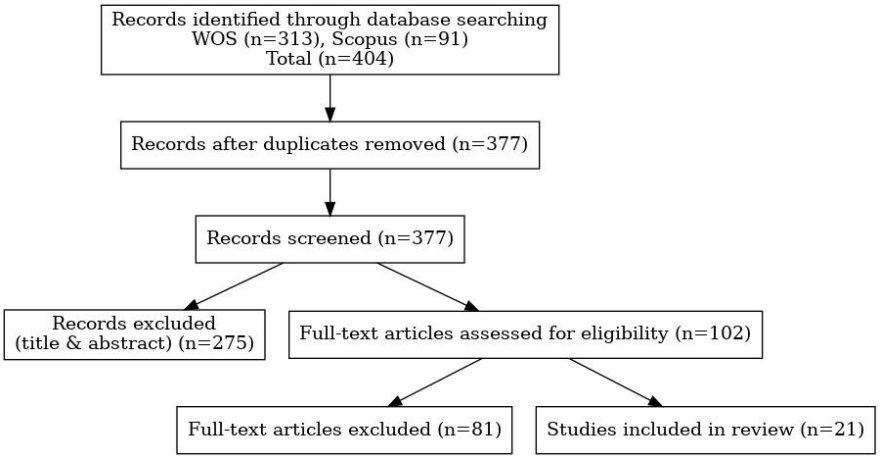

Fig. 1 Prisma flow chart (Moher et al. 2009)

Data extraction, analysis and coding

After selecting the articles according to the inclusion and exclusion criteria, a thematic analysis was conducted to examine and synthesize their findings. To address RQ1, which focuses on research questions and trends, we conducted a content analysis of the studies. Themes were identified and categorized inductively (Braun and Clarke 2006), without relying on a predefined coding scheme, allowing codes to emerge directly from the data, with themes identified at the semantic level (Braun and Clarke 2013, 1-400). The purpose of coding was to support a deeper understanding of the literature and extract relevant insights that address the research questions. Each article was carefully reviewed to record details related to the study's guiding questions, with the aim of characterizing current knowledge on the use of voice-based AI chatbots in language teaching and learning. Initially, the findings were coded independently, then merged and reviewed to identify themes and sub-themes. Based on the coding of the 21 reviewed articles, 8 categories

were identified in response to RQ1, reflecting research questions and trends that characterize recent publications. Regarding RQ2, which addresses the challenges and limitations associated with using AI chatbots in voice mode for developing English-speaking skills, 5 themes and 15 sub-themes were identified.

Categories derived from the analysis of the reviewed articles' research questions are shown in figure 2.

Code	Categories	Article(s) containing the research question
QC01	**Anxiety and confidence in speaking** – speaking anxiety, foreign language anxiety.	JA17, JA19, JA20, JA21
QC02	**Speaking skills development** – fluency, accuracy, vocabulary, pronunciation, grammar, intonation, and communicative competence.	JA01, JA02, JA03, JA04, JA06, JA07, JA08, JA10, JA13, JA14, JA15, JA16, JA19
QC03	**Learners' perceptions and attitudes** – how students view AI chatbots (usefulness, motivation, effectiveness, satisfaction).	JA02, JA04, JA06, JA07, JA09, JA10, JA11, JA13, JA14, JA16, JA18
QC04	**Willingness to communicate (WTC)** – whether interaction with chatbots increases or decreases WTC.	JA02, JA05, JA09, JA12
QC05	**Comparison of learning contexts** – AI-mediated vs. face-to-face interaction, AI-based vs. traditional speaking tasks, chatbot vs. intelligent tutoring system.	JA02, JA05, JA09, JA17, JA20
QC06	**Challenges –** difficulties learners encounter (pronunciation comprehension, system limitations, interaction barriers).	JA03, JA11
QC07	**Emotional and psychological impact** – emotional changes, motivation, self-regulation and self-efficacy.	JA08, JA09, JA15, JA21
QC08	**Learner variables and demographics** – proficiency levels (elementary, intermediate, advanced), age groups (children, undergraduates), gender.	JA21

Fig. 2

Categories derived from the analysis of future research recommendations in the reviewed articles are demonstrated in figure 3.

Categories	Code	Sub-themes	Article(s) containing the research recommendation(s)
Sample and population diversity	RR01	Larger sample size	JA01, JA02, JA03, JA07, JA10, JA12, JA13, JA15, JA17, JA20
	RR02	More diverse participants (cultural, ethnic, educational levels)	JA01, JA02, JA03, JA07, JA08, JA09, JA10, JA12, JA13, JA15
	RR03	Different proficiency levels, ages, gender	JA04, JA07, JA10, JA13, JA15, JA19, JA20, JA21
Study design and methodology	RR04	Longer study duration / longitudinal studies	JA02, JA06, JA07, JA09, JA13, JA15, JA17, JA19, JA20
	RR05	Reduce reliance on self-reports	JA06, JA17, JA21
	RR06	Stronger experimental designs (control groups, randomization)	JA08, JA12, JA13, JA15, JA20
	RR07	Incorporate qualitative insights (interviews, journals, focus groups, observation)	JA07, JA12, JA17
	RR08	Regular follow-up	JA01
Technological development and tool comparison	RR09	Compare different chatbots/ tools (Duolingo, LINE, CoolE Bot, etc.)	JA02, JA03, JA06, JA08, JA13, JA19
Skills' integration	RR10	Speaking skills development	JA02, JA15
	RR11	Broaden skills focus (listening, reading, writing)	JA10
Psychological and affective dimensions	RR12	Study WTC, self-regulation	JA02, JA09, JA15
	RR13	Measure speaking anxiety	JA17, JA20, JA21
	RR14	Study role of self-efficacy in outcomes	JA21
	RR15	Motivation, sustained engagement, gamification	JA20, JA21

Fig. 3

Results

Research questions and trends characterizing recent publications on the use of voice-based AI chatbots in language learning are shown in figure 4.

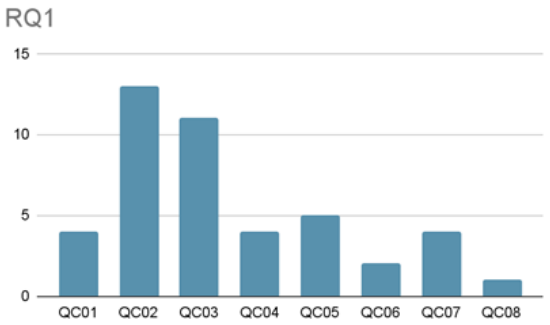

Fig. 4

Figure 4 shows that most studies focused on speaking skills development (n = 13) and learners' perceptions and attitudes (n = 11), whereas substantially fewer articles addressed anxiety and confidence in speaking (n = 4), willingness to communicate (n = 4), or comparisons of learning contexts (n = 5). Challenges and frustrations (n = 2) and learner variables and demographics (n = 1) were only marginally represented.

Future research recommendations emerge from studies on the use of AI chatbots in voice mode for developing English-speaking skills are demonstrated in figure 5.

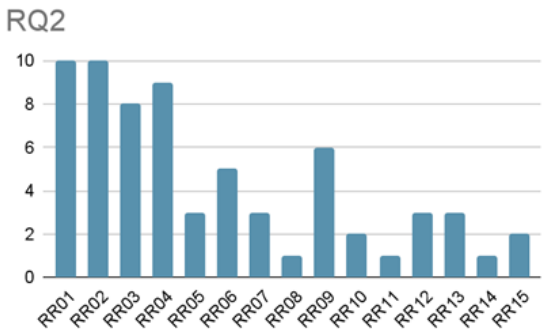

Fig. 5

Figure 5 shows that the most frequent recommendations were to employ larger sample sizes (n = 10) and include more diverse participant groups (n = 10). Nearly as common were calls for longer study durations (n = 9) and for the consideration of different proficiency levels, ages, and genders (n = 8). Comparisons of various

chatbots or tools were suggested by six studies, while stronger experimental designs were recommended in five. The rest were less frequently noted.

Discussion

The analysis of research foci across the included studies (Figure 4) indicates a clear concentration (55%) on speaking skills development and on learners' perceptions and attitudes, suggesting that current scholarship prioritizes both the direct improvement of oral proficiency and learners' subjective experiences with AI chatbots. For instance, Fathi et al. (2024) explored the impact of an AI platform on EFL learners' speaking skills – fluency, coherence, vocabulary, grammar, and pronunciation. Thirty-three learners joined an AI group using the Andy English Chatbot, while thirty-two learners formed a control group with face-to-face peer interaction. Data were collected through IELTS speaking tests and semi-structured interviews. Another study (Zhang 2025) measured speaking proficiency through pre- and post-tests with t-tests analyzing fluency, pronunciation, intonation, and stress, while also using surveys to capture students' perceptions in a mixed-methods design.

A smaller but still notable proportion (27%) addressed affective dimensions, such as anxiety and confidence in speaking, willingness to communicate, emotional changes, motivation, self-regulation and self-efficacy. For example, Khalik (2025) explored whether chatbot-assisted preparation reduces speaking anxiety among non-English majors. In this study, forty students were randomly assigned to chatbot-assisted and control groups, with anxiety measured using a translated version of the Foreign Language Classroom Anxiety Scale (FLCAS). While these studies underscore the importance of psychological factors, their representation remains relatively limited despite their recognized role in language learning (Ballıdağ and Aydın 2025).

11% of the studies examined contextual dimensions, such as comparisons of learning contexts. Deng et al. (2024) studied fourteen students who used an AI chatbot for human-like dialogue practice before engaging in face-to-face interactions with a native English teacher outside the traditional classroom, whereas Zhang et al. (2024) contrasted chatbot use with an intelligent tutoring system. The least examined areas were learner demographics and challenges, suggesting that issues of diversity and barriers to chatbot use remain overlooked. Overall, current evidence is centred on performance- and perception-related outcomes, with less attention to contextual and demographic factors that warrant deeper exploration.

The distribution of research recommendations (Figure 5) highlights a strong methodological emphasis within the current evidence base. The repeated calls for larger and more diverse samples, longer intervention periods, and closer attention to learner characteristics such as proficiency, age, and gender align with the findings related to the research questions discussed above, indicating that existing studies remain limited in both scope and generalizability. The relatively frequent recommendations to compare different chatbots and to strengthen experimental

designs further underscore the need to move beyond small-scale, short-term investigations and pursue more robust, systematic, and comparative research.

At the same time, the future recommendations' relatively limited attention to speaking skills development and affective considerations – such as willingness to communicate, self-regulation, speaking anxiety, and self-efficacy – conforms to the substantial number of studies focusing on the psychological mechanisms of chatbot-mediated language learning and their role in improving oral proficiency.

Conclusion

AI chatbots for English learning are still in the early stages, but research across diverse contexts shows they are gaining significant attention and hold promise for enhancing classroom language learning. In recent years, the number of studies in this field has grown exponentially, with findings being continuously updated as the technology develops at a rapid pace. For this reason, our review concentrated on research published within the last three years in order to capture current and relevant evidence. This study presents an up-to-date systematic review by analyzing the latest research questions and future recommendations on chatbots in language learning, thereby providing a current overview of the field. In doing so, it seeks to clarify the directions in which research is evolving and identify areas where future empirical investigations would be most needed.

The analysis of research questions and recommendations reveals the same pattern: learner demographics and challenges remain underexplored, indicating that issues of diversity and barriers to chatbot use are often overlooked. Therefore, we emphasize the need for larger and more diverse samples, longer intervention periods, and closer consideration of learner characteristics. Calls for stronger experimental designs, qualitative research and comparative studies further highlight the necessity of moving beyond small-scale, short-term investigations toward more rigorous and generalizable research.

We, the authors, would like to extend the discussion by pointing to several additional areas of inquiry that current studies mention only sporadically. First, the extent to which students receive technical assistance deserves closer attention, as learning outcomes may partly depend on the level of support provided. Second, the use of chatbots by absolute beginners and lower-level learners raises important questions about whether chatbots can adapt by employing simplified language. Third, prior technological knowledge may play a role in effective use: such familiarity could assist learners in formulating appropriate prompts and in setting realistic expectations of the tool. Finally, the effectiveness of voice-based chatbot use may depend on learners' prior learning experiences, which warrants systematic investigation.

We also advocate the development of frameworks that can guide research on chatbots, both when evaluating general-purpose large language model chatbots in foreign language speaking skill development and when testing chatbots specifically designed for language learning. The practical value of such frameworks would be

that language teachers could make evidence-informed decisions about selecting and integrating chatbot-based tasks into their lessons, thereby maximizing their impact on student learning outcomes.

Ultimately, such research will be decisive in addressing the fundamental question of to what extent learners can benefit from chatbots in addition to human interaction, a question the answer to which may redefine the future of language learning and teaching.

References

*References marked with an asterisk and accompanied by a code represent the studies included in the review.

*JA18 Alsalem, R. (2024:. EFL Students' Perception and Attitude towards the Use of ChatGPT to Promote English Speaking Skills in the Saudi Context.

*JA20 Ballıdağ, M., and Aydın, S. (2025): A comparison of the effects of AI based chatbots and peer interactions on speaking anxiety among EFL learners. *Future in Educational Research, 3*(2), 224-238.

Braun, V., and Clarke, V. (2006): Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Braun, V., and Clarke, V. (2013): Successful qualitative research: A practical guide for beginners.

*JA08 Chien, Y. C., Wu, T. T., Lai, C. H., and Huang, Y. M. (2022): Investigation of the influence of artificial intelligence markup language-based LINE ChatBot in contextual English learning. *Frontiers in Psychology, 13*, 785752.

*JA01 Cooray, S., Hettiarachchi, C., Nanayakkara, V., Matthies, D., Samaradivakara, Y., and Nanayakkara, S. (2024, April): Kavy: Fostering Language Speaking Skills and Self-Confidence Through Conversational AI. In *Proceedings of the Augmented Humans International Conference 2024* (pp. 226-236).

*JA09 Deng, Y., Wen, K., Dusza, D. G., and Huang, H. W. (2024, March): AI-supported Authentic Communication with Native Speakers: Exploring EFL Learners' Willingness to Communicate and Emotional Changes. In *Proceedings of the 2024 International Conference on Innovation in Artificial Intelligence* (pp. 59-64).

Du, J., and 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.

*JA10 Duong, T. V. T., and Suppasetseree, S. (2024): The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning, Teaching and Educational Research, 23*(3).

*JA02 Fathi, J., Rahimi, M., and Derakhshan, A. (2024): Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254.

*JA11 Hu, W. C., and Škultéty, R. (2024): Unlocking the Learning Potential: ChatGPT as a Virtual Platform for Cross-Interaction in English Language Learning. *Engineering Proceedings, 74*(1), 59.

Jeon, J., Lee, S., and Choe, H. (2023): Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education, 206*, 104898.

Jeon, J., and Lee, S. (2024): Can learners benefit from chatbots instead of humans? A systematic review of human-chatbot comparison research in language education. *Education and Information Technologies, 29*(17), 23329-23360.

*JA17 Khalik, M. F. R. (2025): Reducing English Speaking Anxiety in Classroom Presentations: The Role of Chatbot-Assisted Preparation. *ELP (Journal of English Language Pedagogy), 10*(2), 106-119.

Koç, F. Ş., and Savaş, P. (2025): The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL, 37*(1), 4-21.

Kohnke, L., Moorhouse, B. L., and Zou, D. (2023): ChatGPT for language teaching and learning. *Relc Journal, 54*(2), 537-550.

Lee, S., Choe, H., Zou, D., and Jeon, J. (2025): Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 1-25.

*JA12 Liang, H., and Tse, A. W. C. (2024, September): The Influence of Interacting with Generative AI Chatbots in Informal English Learning Environments on Undergraduate Students' Willingness to Communicate in Mainland China: a Case Study. In *Proceedings of the 2024 the 16th International Conference on Education Technology and Computers* (pp. 26-32).

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., and Jong, M. S. Y. (2024): Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11*(1), 50.

Ma, H., Ismail, L., and Han, W. (2024): A bibliometric analysis of artificial intelligence in language teaching and learning (1990–2023): evolution, trends and future directions. *Education and Information Technologies, 29*(18), 25211-25235.

*JA16 Mahmoud, R. H. (2022): Implementing AI-based conversational chatbots in EFL speaking classes: an evolutionary perspective.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Prisma Group. (2010): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International journal of surgery, 8*(5), 336-341.

*JA03 Muniandy, J., & Selvanathan, M. (2025): ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a Public University, Malaysia. *Teaching Public Administration, 43*(1), 4-20.

*JA19 Naseer, F., Khalid, U., Qammar, M. Z., and Kashif, H. (2024): Chatbots as Conversational Partners: Their Effectiveness in Facilitating Language Acquisition and Reducing Foreign Language Anxiety. *Journal of Applied Linguistics and TESOL (JALT), 7*(4), 238-255.

*JA15 Qiao, H., and Zhao, A. (2023): Artificial intelligence-based language learning: illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in psychology, 14*, 1255594.

Şahin Kızıl, A., Klimova, B., Pikhart, M., and Parmaxi, A. (2025): A systematic review of the recent research on the usefulness of chatbots for language education. *Journal of Computer Assisted Learning, 41*(2), e70001.

*JA05 Zhang, D., Wu, J. G., and Fu, Z. (2024): From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System, 127*, 103501.

*JA06 Zhang, J. (2025): Integrating chatbot technology into English language learning to enhance student engagement and interactive communication skills. *Journal of Computational Methods in Sciences and Engineering, 25*(3), 2288-2299.

Zhang, S., Shan, C., Lee, J. S. Y., Che, S., and Kim, J. H. (2023): Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies, 28*(11), 15223-15243.

*JA07 Zhou, Q., Hashim, H., and Sulaiman, N. A. (2025): Supporting english speaking practice in higher education: the impact of AI chatbot-integrated mobile-assisted blended learning framework. *Education and Information Technologies*, 1-32.

*JA13 Tai, T. Y., and Chen, H. H. J. (2024): Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education, 220*, 105112.

*JA04 Yang, H., Kim, H., Lee, J. H., and Shin, D. (2022): Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL, 34*(3), 327-343.

*JA14 Ye, Y., Deng, J., Liang, Q., and Liu, X. (2022): Using a smartphone-based chatbot in EFL learners' oral tasks. *International Journal of Mobile and Blended Learning (IJMBL), 14*(1), 1-17.

*JA21 Yuan, Z., and Lyu, T. (2024, June): Association Between AI Chatbot Self-efficacy and EFL Student Class-related Anxiety: A Control-Value Theory Perspective. In *Proceedings of the 2024 9th International Conference on Distance Education and Learning* (pp. 371-375).

Wang, F., Cheung, A. C., Neitzel, A. J., and Chai, C. S. (2025): Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research, 95*(4), 623-660.

Wu, X., and Li, R. (2024): Effects of robot-assisted language learning on English-as-a-foreign-language skill development. *Journal of Educational Computing Research, 62*(4), 790-814.