



## A Decade of Korean Research on AI-Mediated English Speaking: A Text-Mining Analysis

Yena Lee (CHA University) · Hye Rang Om (Korea University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: November 3, 2025  
Revised: November 29, 2025  
Accepted: January 10, 2026

Lee, Yena (First author)  
Professor, Department of English  
Linguistics and Language  
Technology  
CHA University  
120 Haeryong-ro, Pocheon-si,  
Gyeonggi-do, 11160  
Seoul, Korea  
Email: [rena01@cha.ac.kr](mailto:rena01@cha.ac.kr)

Om, Hye Rang (Corresponding  
author)  
Adjunct Professor, Department of  
English Language Education  
Korea University  
145 Anam-ro, Seongbuk-gu,  
Seoul, 02841, Korea  
Tel: +82-2-3290-2350  
Email: [ohr323@korea.ac.kr](mailto:ohr323@korea.ac.kr)

### ABSTRACT

Lee, Yena and Hye Rang Om. 2026. A decade of Korean research on AI-mediated English speaking: A text-mining analysis. *Korean Journal of English Language and Linguistics* 26, 84-105.

This study examines abstract-level research trends in AI-mediated English-speaking education through a systematic text-mining analysis of English-language abstracts drawn from 130 Korean journal articles and theses published between 2015 and 2025. The dataset was constructed by retrieving English-language abstracts from RISS, a major Korean academic database and was analyzed using a combination of word-frequency analysis, TF-IDF weighting, N-gram extraction, and co-occurrence network analysis. These quantitative procedures were supplemented by topic modeling and CONCOR clustering to identify dominant thematic structures across the corpus. The results indicate that chatbot-based technologies—most notably PengTalk and ChatGPT—have emerged as the primary platforms for AI-mediated speaking practice, particularly within classroom-oriented instructional designs emphasizing proficiency development, level-based assessment, and feedback provision. Four major thematic domains were identified: classroom-based outcome validation, AI/chatbot-mediated support and proficiency gains, learner-centered task design and interaction, and speaker or user experience. Across these domains, AI is increasingly positioned not merely as an auxiliary tool but as an interactive learning partner that reshapes pedagogical practices and assessment approaches. The findings further suggest a growing emphasis on engagement, interactional authenticity, and formative feedback in AI-mediated speaking research. This study concludes by discussing implications for future classroom-embedded longitudinal research and the development of multimodal (speech–text) evaluation frameworks that foreground fairness, authenticity, and personalization in English-speaking assessment.

### KEYWORDS

artificial intelligence, English speaking, ChatGPT, chatbot, text mining, research trends

## 1. Introduction

The integration of artificial intelligence (AI) into English language education has fundamentally reshaped how speaking is taught, practiced, and assessed. Early applications primarily relied on automated speech recognition (ASR) to diagnose segmental and suprasegmental pronunciation errors and deliver immediate, individualized feedback (Eskenazi 2009). Although initially narrow in scope, these computer-assisted pronunciation training (CAPT) systems inaugurated a broader shift toward data-informed, learner-centered pedagogy. Subsequent advances in natural language processing (NLP) and machine learning extended pedagogical possibilities beyond discrete pronunciation drills, enabling learners to engage in interactive, adaptive, and increasingly naturalistic dialogues with machines (Fryer and Carpenter 2006). More recently, generative AI—exemplified by large language model (LLM)–based systems such as ChatGPT—has marked a further milestone by affording contextually rich interaction, real-time scaffolding, and scalable formative feedback across diverse speaking scenarios (Wang et al. 2024). As this body of research has expanded rapidly, synthesizing how these developments are summarized and foregrounded in the literature has become increasingly important, particularly in relation to how AI-mediated speaking is represented across published research outputs. These developments have intensified long-standing questions in applied linguistics about how communicative competence is conceptualized in human–machine interaction, how validity can be assured in assessment, and how technology might democratize access to effective speaking practice.

In Korea, empirical studies have increasingly examined the educational applications of AI-powered tools in EFL speaking contexts. Studies involving commercial AI speakers and retrieval- or rule-based Chatbots—such as EBS’s AI PengTalk—reported improvements in fluency and accuracy alongside reductions in speaking anxiety, suggesting that always-available, low-stakes interaction can mitigate affective barriers to oral production (Hwang 2021, Kong 2020, Yoon 2021). In parallel, an expanding body of work has examined generative AI platforms in tertiary contexts, frequently foregrounding positive learner perceptions, measurable gains in speaking performance, and task- and proficiency-contingent benefits that highlight the importance of affordance–learner fit (Huh 2024, Lim 2024, Rha 2024). At the same time, Korean scholarship has raised critical concerns regarding intercultural authenticity (e.g., whether LLM-mediated dialogues align with sociolinguistic norms relevant to local curricula), assessment reliability (e.g., the stability and transparency of automated scoring), and the sustainability of short-term interventions that lack classroom-embedded, longitudinal designs (Kim and Lee 2023). Taken together, the abstracts suggest a research landscape that balances optimism about AI-mediated speaking practice with ongoing methodological and pedagogical caution.

Alongside pedagogical research, methodological innovations in text mining have created new opportunities for analyzing large corpora of educational data. Text mining techniques such as clustering and topic modeling have been widely employed to uncover latent themes and map research trajectories in applied linguistics and education (Ahadi et al. 2022, Ferreira-Mello et al. 2019). For example, Charitopoulos et al. (2025) demonstrated how machine learning integrated with text mining can streamline the evaluation of open-ended student responses, while Wang (2024) showed that vocabulary learning systems grounded in text mining significantly enhanced retention and engagement. Compared with simple frequency counts, text-mining approaches can highlight terms that distinguish subtopics (e.g., TF-IDF), capture recurring multi-word expressions (e.g., N-grams), and reveal networks of concepts that tend to co-appear in the literature (e.g., co-occurrence analysis). Yet, despite these methodological advances, text mining has rarely been applied to chart the research landscape of AI-supported English speaking, and even less so in relation to Korea-based scholarship published across journals and graduate theses.

The present study addresses this gap by applying topic modeling and complementary text-mining techniques to a corpus of English-language abstracts drawn from Korean academic journal articles and graduate theses on AI-mediated English speaking. Focusing explicitly on abstract-level discourse, the study examines how research emphases are summarized, framed, and foregrounded in published abstracts, rather than analyzing full texts. In doing so, it aims to identify the dominant thematic clusters and keyword patterns that characterize Korean scholarship on AI-mediated English speaking as a whole, based on publications released between 2015 and 2025. Guided by this abstract-level analytical scope, the study addresses the following research questions:

- 1) What thematic clusters are identifiable in English-language abstracts of Korean research on AI-mediated English speaking through topic modeling?
- 2) What salient keywords and association patterns emerge from word frequency, TF-IDF, N-gram, and co-occurrence analyses of these abstracts?
- 3) How do the identified thematic and lexical patterns characterize Korean AI-mediated English speaking research between 2015 and 2025?

## **2. Literature Review**

### **2.1 AI and English Speaking**

Research on AI-mediated speaking has evolved from pronunciation-focused tools to open-ended conversational systems capable of scaffolding interaction and generating feedback at scale. The earliest strand centered on computer-assisted pronunciation training (CAPT) driven by ASR. Meta-analytic and systematic reviews generally report positive effects on pronunciation, with stronger evidence for segmental targets and adult learners, while also noting a historical bias toward listen-and-repeat drills and limited attention to broader outcomes such as intelligibility and comprehensibility (Amrate and Tsai 2024, Mahdi and Al-Khateeb 2019). These tools demonstrated the feasibility of individualized, data-rich oral practice and laid the foundation for more dialogic applications.

A second line of work has focused on spoken dialogue systems (SDS) and intelligent personal assistants that extend practice beyond discrete phonology to short exchanges and routine tasks. Early studies on ChatBots in language learning highlighted their potential to expand practice opportunities outside class and reduce anxiety in low-stakes interaction (Fryer and Carpenter 2006). Subsequent classroom and field studies with SDS have reported sustained use and generally positive learner experiences in both lower-secondary and adult contexts, although outcomes are moderated by task design and implementation fidelity (Ericsson and Johansson 2023). Reviews of conversational AI in English language teaching (ELT) between 2013 and 2023 similarly document a rapid increase in published studies and broadly positive affective and cognitive effects. At the same time, these reviews observe a concentration of research in Asian EFL settings and a predominance of general-purpose assistants rather than pedagogy-specific conversational agents (Lai and Lee 2024).

The most recent development concerns generative AI (GenAI) and large LLMs as speech-capable interlocutors and coaches. LLM-based systems can sustain multi-turn, context-sensitive dialogue; adapt prompts and roles; and deliver varied feedback (e.g., recasts, metalinguistic explanations, and exemplars) during or after tasks. Synthesis pieces in applied linguistics argue that GenAI has the potential to extend and transform speaking instruction and assessment, while also introducing risks related to hallucination, controllability, and alignment with sociolinguistic

norms. Emerging empirical studies with LLM Chatbots report short-term benefits for lower-proficiency learners and learner preferences for interactive practice over receptive alternatives, but emphasize the need for larger-scale longitudinal studies with stronger learning measures.

A major area of growth is AI-supported assessment. Automated scoring has progressed from research prototypes to operational use in large-scale testing, such as contributory scoring via Speech Rater in TOEFL iBT. Technical reports document concurrent validity and reliability evidence, yet language assessment scholarship continues to caution against premature adoption. Key issues include construct coverage (e.g., interactional competence, pragmatics), cross-accent fairness, score stability across prompts, and the interpretability of system outputs. Current best practice in educational contexts often combines human-referenced rubrics with machine-generated indicators (e.g., speech rate, pause ratio, lexical diversity), accompanied by uncertainty estimates and actionable explanations.

Measurement and reporting, however, remain inconsistent. Recent guidance in applied linguistics and systematic reviews recommend a checklist for comparability: specify model/version and ASR/TTS specifications; detail interaction setup (synchronous vs. asynchronous, single- vs. multi-turn, speech-only vs. multimodal); document task genre and time-on-task; enumerate feedback types and timing; and report both outcome measures (beyond CAF indices to include intelligibility, comprehensibility, and interactional moves) and process traces (e.g., prompts, turns, repair sequences, uptake). Evaluation frameworks from AI research—such as user satisfaction, task success, and error recovery—offer additional benchmarks but require adaptation to L2 constructs.

Despite clear progress, the field faces several open challenges. Many studies remain short-term with small samples or bespoke systems, limiting generalizability and fueling calls for longitudinal, classroom-embedded trials with stronger baselines and fidelity checks. Generative AI also introduces new concerns regarding content accuracy and safety in content-rich speaking tasks, which may be mitigated through constrained scenarios, verification prompts, and teacher-facing dashboards. Issues of equity are equally salient, as factors such as bandwidth, device quality, and acoustic environments shape access to speech-native AI tools. Finally, emerging capabilities in interactional analytics and prosody-aware feedback—now enabled by streaming models—could enhance the measurement of interactional competence and intelligibility if validated against human judgments. Overall, recent synthesis research underscores both the promise of AI-mediated speaking and the necessity of principled design, accountable measurement, and equitable deployment.

## **2.2 Research on AI and English Speaking**

Research on AI-mediated English speaking has unfolded in a gradual yet multi-faceted trajectory. Early inquiries focused heavily on pronunciation, with ASR technologies, computer-assisted pronunciation training (CAPT), and AI-speaker-supported practice providing the first demonstrations of scalable and individualized support. These studies, often targeting accuracy, intelligibility, and short scripted turns, confirmed that learners could benefit from feedback-rich environments even in relatively constrained tasks (Han et al. 2019, Hwang 2021, Jeong and Choi 2021, Jo et al. 2022, Kim 2020, Park 2022, Seo 2017, Sim 2022). While their contributions validated AI as a legitimate pedagogical partner, the emphasis on segmental features and discrete drills also highlighted the limits of narrowly framed practice.

As research interests matured, attention shifted toward more dialogic modes of interaction. Chatbot-mediated speaking and spoken dialogue systems broadened the pedagogical lens to include role-plays, interviews, and everyday exchanges. In these contexts, AI was no longer treated merely as a pronunciation coach but as a low-stakes interlocutor that extended speaking opportunities and scaffolded turn-taking, repair, and questioning

strategies. Studies across elementary to tertiary settings reported that when tasks were clearly framed and scaffolded, learners tended to experience gains in fluency and accuracy alongside positive perceptions of the technology (Kim 2016, 2017, Lee 2019, Lee and Park 2019, Lee et al. 2019, Sung 2019, Yang et al. 2019). Particularly notable was the domestic exploration of EBS AI PengTalk in elementary schools, where researchers documented not only learner outcomes but also usage patterns and classroom design considerations (Choi 2021, Hong 2021, Oh and Back 2022, Oh et al. 2022, Seong and Lee 2021).

The arrival of generative AI and LLMs has further reframed English speaking practice as an open-ended and adaptive exchange. Unlike earlier tools, LLMs can reformulate, question, and elaborate in ways that simulate authentic dialogue. Studies in higher education contexts report improvements in speaking performance and favorable learner perceptions, with benefits often contingent on the careful design of prompts and feedback strategies (Baek 2024, Cheon 2023, Ko 2024, Lim 2024, Noh 2024, Rha 2024). In addition to general conversation practice, targeted applications such as interview training and OPIc-style tasks have gained traction, producing encouraging findings through both developmental and quasi-experimental approaches (Baek 2024, Ko 2024).

Across these developments, two enduring concerns stand out. The first involves assessment. Initial uses of AI in speaking emphasized formative feedback, but more recent scholarship has explored process- and product-oriented evaluation frameworks aligned with rubrics and semi-automated scoring pipelines. Despite evidence of efficiency, questions about stability, interpretability, and fairness remain, leading scholars to advocate triangulation with human ratings and explicit links to curricular constructs (Choi 2020, Han 2020, Kong 2020, Seong and Shin 2020, Shim et al. 2020). The second involves affective and interactional outcomes. Research consistently suggests that AI-mediated practice can reduce anxiety and encourage participation, particularly when learners rehearse in private or preparatory stages. Purpose-driven tasks and opportunities for revision appear especially effective in sustaining motivation and supporting uptake of feedback (Choi 2020, Han 2020, Ki 2022, Kim 2020, Lee 2021, Seo 2017, Yeom 2022).

Taken together, these findings portray a field that balances promise with caution. AI has been shown to expand access to interlocutors, amplify feedback, and diversify speaking practice, yet its effectiveness depends on principled task framing, transparent assessment design, and the presence of teacher mediation. Persistent limitations—short intervention windows, small or narrow samples, and inconsistent reporting of AI configurations and outcome measures—underscore the need for systematic mapping of the research landscape. Against this backdrop, the present study employs topic modeling and complementary text-mining techniques to map abstract-level thematic clusters in Korean scholarship on AI-mediated English speaking within the 2015–2025 publication window.

### **2.3 Application of Text Mining in English Education**

The growing availability of learner-generated digital data has encouraged the use of text mining as a means of analyzing large corpora of essays, open-ended responses, and online interactions in English education. This methodological approach has been increasingly recognized as a way to generate insights into learner performance, instructional design, and pedagogical innovation. A systematic review by Ahadi et al. (2022) mapped the educational text-mining landscape from 2000 to 2022, identifying writing analytics, the analysis of learner forums and MOOCs, and the automated evaluation of assignments as dominant application areas. Their findings established that text mining has become a central technique in educational data science, offering both descriptive and predictive insights.

Several empirical studies illustrate how text mining has been operationalized in English education. Charitopoulos et al. (2025), for example, demonstrated how machine learning and text-mining algorithms could be applied to

evaluate free-text responses such as essays, thereby reducing instructors' grading workload while improving the consistency of feedback. Similarly, Wang (2024) investigated a vocabulary learning system that leveraged corpora of authentic English texts and text-mining-based personalization. Their results indicated that learners using the system significantly outperformed peers engaged in traditional vocabulary study, both in retention and in sustained engagement. Together, such studies highlight how text mining can enable adaptive and scalable feedback that supports learner autonomy.

Beyond classroom applications, text mining has also been mobilized to track research trends and thematic shifts within English language education. Ferreira-Mello et al. (2019) provided a comprehensive overview of educational text-mining methods, emphasizing clustering, topic modeling, and natural language processing as commonly used techniques. These approaches have been employed not only to analyze learning processes but also to visualize the evolution of research topics and pedagogical concerns over time, thereby contributing to evidence-based agenda-setting in applied linguistics.

Despite these advances, important limitations remain. Much of the existing work is concentrated on short-term outcomes such as vocabulary growth or writing accuracy, with relatively few studies addressing longer-term communicative competence or speaking proficiency. Furthermore, the majority of text-mining research relies on written corpora, leaving oral language data comparatively underexplored. Methodological variation—across corpora, preprocessing techniques, assessment metrics, and learner populations—further complicates direct comparison across studies and raises questions about generalizability.

In summary, text mining has proven to be a powerful methodological toolkit for automating assessment, tailoring individualized feedback, and uncovering patterns in learner data. However, its potential for analyzing English speaking—particularly in relation to interactional competence and discourse features—remains underdeveloped. Addressing this gap requires future studies that extend text-mining applications beyond writing and vocabulary learning toward spoken data, longitudinal outcomes, and cross-modal integration, thereby enriching our understanding of language development in digital learning environments.

### **3. Research Method**

#### **3.1 Data Collection**

This study analyzed scholarly materials to identify research trends in AI-mediated English-speaking instruction. Using RISS (Research Information Sharing Service), we retrieved Korean journal articles and MA/PhD theses published between 2015 and 2025 that included English-language abstracts and addressed AI-related tools in English-speaking education. The search was conducted on September 17, 2025, using the following keyword set: (“AI” OR “artificial intelligence”) AND (“English speaking” OR speaking) AND (chatbot OR “AI speaker” OR “voice recognition” OR “ChatGPT” OR “PengTalk”). Records were screened by title and abstract for topical relevance. Studies were included if they (a) were published between 2015 and 2025, (b) were Korean journal articles or graduate theses, and (c) provided an English-language abstract. Records were excluded if the abstract was missing, the focus was not English-speaking instruction or assessment, or AI was only tangentially mentioned. Duplicate records were removed based on identical titles, authors, and publication information. The final corpus consisted of 130 English-language abstracts, which served as the analytical dataset for the subsequent text-mining analyses. Although the corpus size is modest, it reflects the full set of RISS-indexed publications meeting the inclusion criteria within the target period and provides a sufficiently robust basis for identifying structural research trends in this domain.

Textom, a text-mining software supporting morphological analysis, stopword filtering, and co-occurrence matrix generation in a web-based GUI environment, was employed as the main analytical tool. Its automated cleaning features and Korean noun-oriented analytical functions ensured consistency and reproducibility throughout the preprocessing process. Using Textom, this study systematically generated word-frequency analyses, TF-IDF values, N-gram patterns, and network analyses.

The collected abstracts underwent several preprocessing steps for text-mining analysis. Non-essential elements such as tables, figures, references, and special characters were removed, followed by the elimination of stopwords based on Textom's default English stopword list supplemented with a researcher-defined list of high-frequency but semantically uninformative terms. Stemming and lemmatization were then conducted to consolidate lexical variations (e.g., "learn," "learning," "learner"), and POS tagging was applied to classify semantically meaningful units.

To ensure consistent representation of key concepts, frequent lexical variants were inspected and, where appropriate, normalized. For example, "AI" and "artificial intelligence" were treated as a single concept, as were trivial orthographic variants such as "Pengtalk/PengTalk." Where distinctions were theoretically meaningful (e.g., "student" vs. "learner" as role labels), the forms were retained as separate tokens but interpreted comparatively during analysis. Spelling errors in the English text were corrected when the intended form was unambiguous, and duplicate abstracts were removed to finalize the analytical corpus.

The resulting corpus served as the dataset for various text-mining procedures—including word-frequency analysis, TF-IDF analysis, N-gram analysis, co-occurrence matrix analysis, topic modeling, and CONCOR clustering. Through these procedures, this study quantitatively and structurally examined research trends concerning AI in English-speaking education.

### **3.2 Methods of Analysis**

The text-mining procedures employed in this study were designed not as a simple listing of techniques but as a sequential and interconnected analytical workflow. After completing data preprocessing and constructing the analytical units, word-frequency and TF-IDF analyses were first conducted to identify patterns in the emergence of core concepts. Subsequently, N-gram and co-occurrence matrix analyses were used to examine structural relationships among concepts, and finally, topic modeling and CONCOR clustering were performed to extract latent themes and semantic networks. This stepwise analytical design strengthened methodological transparency and ensured reproducibility throughout the study.

To investigate research trends related to the application of artificial intelligence (AI) in English-speaking instruction, this study adopted a text-mining-based analytical procedure. The corpus was constructed by collecting domestic journal articles and theses that included keywords related to English speaking and AI technologies. During preprocessing, frequent lexical variants were manually inspected and normalized to ensure consistent representation of core concepts (e.g., "AI" and "artificial intelligence," "ChatBot" and "chatbot"). In addition, spelling errors in the abstracts were corrected when the intended forms were unambiguous, and a customized stopword list was applied to remove high-frequency but semantically non-informative terms. These procedures were implemented to enhance analytical clarity and replicability at the abstract level. During preprocessing, stopwords were removed, lemmatization was applied, and morphological analysis was conducted to refine the analytical units. The final corpus included academic nouns and foreign-language terms commonly used in scholarly discourse. This selection reflected the assumption that such terms—particularly those originating from international literature—play a central role in interpreting academic concepts within this domain. In contrast, verbs, adjectives, postpositions, prepositions, pronouns, and other parts of speech with low semantic discriminability were excluded from analysis.

Several analytical parameters were determined with reference to both prior text-mining studies and the characteristics of the dataset. POS filtering was applied to retain semantically salient nouns and technical terms, as these units most reliably capture conceptual content in academic abstracts. In co-occurrence and CONCOR analyses, trimming cutoffs and minimum frequency thresholds were introduced to reduce noise and prevent unstable clustering, particularly given the relatively modest corpus size ( $N = 130$ ).

Data analysis was conducted using a multilayered set of text-mining techniques. First, word-frequency analysis identified the most frequently appearing key concepts, enabling the extraction of central research keywords. Second, TF-IDF analysis captured the terms that uniquely characterized the research domain beyond simple frequency. Third, N-gram analysis was conducted to capture associative patterns beyond unigram frequency. While bigrams ( $N = 2$ ) were automatically generated through Textom, higher-order N-grams (trigrams and tetragrams) were also examined using the same tokenized corpus. However, due to the relatively small size of the dataset (130 abstracts),  $N \geq 3$  expressions demonstrated substantial sparsity, with only a small number of meaningful multi-word expressions (e.g., “automated speaking assessment,” “learner-centered task design,” “real-time scaffolding feedback”) appearing at interpretable frequencies. For this reason, the quantitative reporting focuses on bigrams, and the limited coverage of longer expressions is acknowledged as a methodological constraint. Although higher-order N-grams ( $N \geq 3$ ) were initially explored, their frequencies were highly sparse due to the limited size of the abstract corpus. Consequently, the quantitative analysis focuses primarily on bigrams, which provided a more stable and interpretable representation of recurring conceptual associations across studies.

Fourth, a  $25 \times 25$  co-occurrence matrix was constructed using 25 keywords extracted from the top 30 N-gram combinations. Each cell  $M_{ij}$  indicated the number of abstracts in which the two keywords co-occurred. Correlations below .20 were trimmed to avoid spurious ties. The resulting matrix was imported into UCINET and analyzed using the CONCOR algorithm following standard iterative partitioning procedures. Words with fewer than ten occurrences were excluded to prevent unstable clustering. This process yielded four stable clusters that summarize the major conceptual domains of AI-mediated English-speaking research.

Fifth, topic modeling was then performed to extract the latent thematic structures embedded within the literature. LDA models were estimated with  $K = 5$  to  $K = 15$ , and coherence ( $c_v$ ) and perplexity curves were jointly inspected. Perplexity decreased sharply between  $5 \leq K \leq 10$  and plateaued thereafter, while coherence reached a local maximum near  $K = 10$ . Thus,  $K = 10$  was selected as the optimal balance between model fit and interpretability. Topic modeling complemented the structural analyses by revealing latent research themes that extended beyond surface-level keyword co-occurrence.

These analytical procedures were designed to complement one another. Word-frequency and TF-IDF analyses provided quantitative evidence of keyword salience, while N-gram and matrix analyses revealed structural relationships among the terms. Topic modeling identified the latent themes within the literature through the Latent Dirichlet Allocation (LDA) algorithm. The optimal number of topics ( $K$ ) was tested within the range of 5 to 15, and based on comparisons of Coherence Scores and Perplexity values,  $K = 10$  was selected as the most stable model.

The CONCOR clustering further clarified the semantic network and boundaries among themes. To prevent distortion due to excessive node expansion, words with fewer than ten occurrences were excluded. A correlation cut-off value of .20 was applied to ensure stable cluster separation. Through this comprehensive analytical framework, the study offers a multidimensional understanding of research trends in AI-based English-speaking instruction.

TF-IDF analysis was employed to identify terms that distinguish specific research themes beyond simple frequency counts by weighting words that are frequent in particular abstracts but less common across the corpus. N-gram analysis captured recurrent multi-word expressions that reflect patterned conceptual framing, while co-occurrence analysis examined networks of terms that tend to appear together within abstracts, thereby revealing relational structures among key concepts rather than isolated keywords.

## 4. Results

### 4.1 Results of the Frequency Analysis

A word-frequency analysis was conducted on the collected research literature to examine how artificial intelligence (AI) has been discussed within the field of English-speaking instruction. As shown in Table 1, analysis identified the top 30 most frequently occurring terms, which collectively illustrate the current research trends at the intersection of English-speaking education and AI technologies.

**Table 1. Top 30 Words by Frequency (N = 130 Abstracts)**

Rank	Word	Frequency
1	English	764
2	AI	458
3	ChatBot	419
4	student	398
5	study	378
6	learning	356
7	learner	332
8	class	275
9	group	222
10	level	195
11	school	190
12	teacher	189
13	education	170
14	research	169
15	result	168
16	language	146
17	Pengtalk	141
18	intelligence	139
19	ChatGPT	129
20	activity	123
21	feedback	122
22	model	118
23	tool	113
24	proficiency	110
25	application	110
26	skill	105
27	technology	105
28	task	103
29	test	103
30	analysis	102

As illustrated in Figure 1, the word cloud provides an intuitive overview of the dominant terms in the corpus, complementing the numerical frequencies reported in Table 1.



## 4.2 TF-IDF Analysis

To probe research trends more deeply, we conducted TF-IDF analysis, which assesses each term's relative importance across the corpus and highlights topic-differentiating keywords beyond raw frequency. As shown in Table 2, "ChatBot" (TF-IDF = 346.38) ranked highest, underscoring its central role in AI-mediated speaking instruction. "PengTalk" (271.47) and "ChatGPT" (236.63) also ranked highly, reflecting the active uptake of a specific educational chatbot and a state-of-the-art generative model in classroom research. This pattern suggests that AI has become a primary vehicle for improving speaking proficiency rather than a mere auxiliary tool.

**Table 2. Top 30 TF-IDF Terms**

Rank	Word	tf	idf	TF-IDF
1	ChatBot	63	0.8267	346.38
2	PengTalk	21	1.9253	271.47
3	group	46	1.1412	253.34
4	ChatGPT	23	1.8343	236.63
5	class	62	0.8427	231.74
6	speaker	17	2.1366	192.29
7	model	30	1.5686	185.10
8	teacher	55	0.9625	181.91
9	AI	98	0.3848	176.26
10	learner	85	0.5272	175.02
11	strategy	21	1.9253	163.65
12	user	15	2.2618	158.32
13	learning	93	0.4372	155.65
14	task	32	1.5041	154.92
15	education	58	0.9094	154.59
16	feedback	41	1.2562	153.26
17	activity	42	1.2321	151.55
18	variable	5	3.3604	151.22
19	school	65	0.7954	151.13
20	proficiency	37	1.3589	149.48
21	level	68	0.7503	146.31
22	program	26	1.7117	138.65
23	application	41	1.2562	138.19
24	conversation	38	1.3322	134.55
25	intelligence	55	0.9625	133.78
26	convergence	2	4.2767	132.58
27	knowledge	10	2.6672	128.03
28	interaction	33	1.4733	126.70
29	perception	34	1.4435	125.58
30	design	26	1.7117	124.96



(38), and “school–student” (41) point to authentic classroom contexts centered on teacher–learner interaction. Performance- and assessment-related pairs—“English–proficiency” (38), “proficiency–level” (33), “post–test” (23)—suggest frequent use of post-testing and level assessments to validate gains in speaking ability. Pairs such as “learning–tool” (35), “teaching–learning” (28), and “study–effect” (23) reflect ongoing interest in evaluating AI-based tools and instructional strategies. Of particular note, “voice–recognition” (28) and “intelligence–speaker” (26) indicate attention to speech recognition and speaker characteristics, implying growing interest in automating assessments of accuracy and fluency.

Overall, N-gram analysis highlights four streams: (1) AI-Chatbot-based learning support, (2) classroom teaching–learning contexts, (3) achievement and level assessment, and (4) speech-recognition-based speaking evaluation. AI thus appears as a key driver reconfiguring classroom interaction and assessment mechanisms rather than a peripheral aid.

**Table 3. Top 30 Bigrams (N-grams)**

Rank	Term 1	Term 2	Frequency
1	AI	ChatBot	121
2	AI	PengTalk	84
3	English	education	82
4	English	class	78
5	AI	English	48
6	English	teacher	45
7	ChatBot	English	45
8	language	learning	43
9	English	learning	43
10	control	group	41
11	school	student	41
12	teacher	student	38
13	school	English	38
14	English	proficiency	38
15	learning	tool	35
16	student	English	34
17	proficiency	level	33
18	purpose	study	31
19	voice	recognition	28
20	teaching	learning	28
21	ChatGPT	English	28
22	learner	English	27
23	convergence	education	27
24	learner	level	26
25	intelligence	speaker	26
26	English	ChatBot	25
27	result	study	25
28	English	learner	25
29	study	effect	23
30	post	test	23

While Table 3 reports the most frequent bigrams, exploratory inspection of higher-order N-grams (trigrams and

tetragrams) provided additional nuance. Due to the limited size of the corpus, only a small set of multi-word expressions met minimal frequency thresholds, but these included domain-specific phrases such as “automated speaking assessment,” “learner-centered task design,” and “real-time scaffolding feedback.” Although not tabulated here, these higher-order N-grams reinforced the interpretation that current research is moving toward integrated designs that combine automated speaking evaluation, learner-centered task structuring, and interaction-sensitive feedback.

#### 4.4 Co-Occurrence Matrix Analysis

To examine how key concepts are interconnected, we conducted co-occurrence matrix analysis. As shown in Appendix, strong pairs included “English–study” (121), “English–student” (107), “English–result” (99), and “English–AI” (94), confirming English as the pivotal keyword closely tied to learners, research outcomes, and AI. The high association of “English–study” and “English–result” underscores the close coupling of instruction with outcome validation. AI-related couplings—“AI–study” (91), “AI–student” (77), “AI–learning” (65), “AI–research” (58), and “AI–result” (70)—suggest that AI functions as a central mediator in learner-focused research and effectiveness evaluation. Platform-specific links—“AI–Chatbot” (44), “AI–PengTalk” (19), “AI–ChatGPT” (14)—show the prominence of ChatBot and platform-based applications. Classroom context is also salient: “student–study” (100), “student–result” (80), “student–learning” (70), “teacher–student” (38), and “school–student” (41) indicate prevalent pre-/post-testing within school settings. “Class–student” (50) and “group–student” (37) further suggest the normalization of class- and group-level experimental designs. Language learning outcomes are reflected in “language–result” (49), “language–learning” (50), and “language–education” (29), while “proficiency–level” (33) corroborates the centrality of proficiency and speaking level as evaluation indices. Although less frequent, technology-focused couplings such as “voice–recognition” and “intelligence–speaker” recur consistently, pointing to likely expansion in speech-recognition-based speaking assessment and AI speech modeling.

In sum, the co-occurrence structure organizes around four axes: (1) English learning and outcome validation, (2) AI technology and effectiveness research, (3) class/group experimental designs, and (4) language acquisition and proficiency assessment—signaling a field that pursues both technological application and pedagogical verification.

#### 4.5 Topic Modeling

We conducted topic modeling with ten topics to delineate the thematic structures of AI applications in English speaking instruction. As noted in the Method section, models with  $K = 5$  to  $K = 15$  topics were estimated, and  $K = 10$  was selected based on a joint inspection of perplexity and coherence scores, which indicated that ten topics provided the best balance between model fit and interpretability. Based on top-weighted terms within each topic, the overall discourse converged on four axes—“English–student (learner)–AI (chatbot)–learning”—with simultaneous salience of classroom experimental design and the adoption of generative AI.

As shown in Table 4, Topic 4 (English–study–ChatBot–AI–learning) and Topic 10 (English–student–chatbot–AI–class) show high weights for “English” (0.068, 0.062) in conjunction with chatbot and class variables, indicating that ChatBot-mediated tasks and interaction design in speaking classes are core research agendas. Topic 6 (English–class–student–AI–study) highlights class–student contexts, suggesting active classroom implementation and evaluation. Topic 9 (English–study–student–AI–group) includes “group” among its top terms, reflecting the common use of control/experimental group comparisons. Topic 7 (student–English–ChatBot–AI)

assigns a notable weight to “ChatGPT” (0.035), pointing to rapid expansion of work on learner–chatbot interaction since the introduction of generative AI. Topic 3 (learning–English–intelligence–AI–speaker) includes “speaker,” suggesting integration of speaking tasks with speech-based interaction and assessment (e.g., ASR).

Overall, topic modeling indicates: (1) ChatBot-centered instructional scenario design; (2) real-world class/school implementation and between-group experiments; (3) shifts in interaction patterns following the introduction of generative AI and (4) the rise of multimodal evaluation that includes speech variables. Future work should deepen the intersection implied by Topics 7 and 3—“generative AI × speech-based feedback”—and test causal models linking classroom learning experiences (e.g., social presence, fairness, personalization) to achievement indices (fluency/accuracy/complexity) emphasized in Topics 4, 6, and 10.

**Table 4. Topic Modeling: Top Terms and Weights (K = 10)**

Topic	Term 1	Term 2	Term 3	Term 4	Term 5
1	English (.056)	learner (.037)	ChatBot (.029)	AI (.028)	student (.020)
2	ChatBot (.039)	English (.032)	learner (.025)	student (.024)	AI (.023)
3	learning (.037)	English (.030)	intelligence (.024)	AI (.023)	speaker (.021)
4	English (.068)	study (.035)	ChatBot (.034)	AI (.031)	learning (.028)
5	English (.046)	student (.028)	AI (.027)	learning (.021)	study (.021)
6	English (.046)	class (.038)	student (.034)	AI (.026)	study (.022)
7	student (.041)	English (.037)	ChatGPT (.035)	chatbot (.021)	AI (.020)
8	AI (.052)	chatbot (.037)	English (.032)	learning (.027)	learner (.021)
9	English (.037)	study (.033)	student (.030)	AI (.029)	group (.024)
10	English (.062)	student (.032)	ChatBot (.027)	AI (.023)	class (.022)

#### 4.6 CONCOR Clustering

To clarify the conceptual structure of research on AI and English speaking, we conducted CONCOR analysis of keyword co-occurrence to reveal semantic communities among clustered terms. As summarized in Table 5, four principal clusters emerged:

**Table 5. CONCOR Clusters, Member Terms, and Labels**

Cluster	Member Terms (excerpt)	Label
1	English, class, group, level, research, language, PengTalk, tool, skill, test, participant, use, perception, ability, effect, survey	Classroom-Based Learning Outcomes
2	AI, ChatBot, study, learning, learner, school, teacher, education, result, intelligence, ChatGPT, activity, feedback, proficiency, application, analysis, conversation, program, datum, finding, method, improvement, content	AI/Chatbot-Mediated Support & Effectiveness
3	student, model, technology, task, teaching, interaction, strategy, design, type	Learner-Centered Strategies & Task Design
4	speaker, user	Speaker/User Experience

Cluster 1: “English, class, group, level, research, language, PengTalk, tool, skill, test, participant, use, perception, ability, effect, survey.” This cluster centers on classroom context and research design, focusing on learner abilities, perceptions, tools, and effect validation. The prominence of “class, group, level, test” indicates class-level experiments and achievement measurement as a major axis, while “perception, ability, effect” points to learner perceptions, ability change, and outcome verification.



Taken together, the CONCOR results indicate four organizing domains: (1) classroom-based experimentation and outcome validation; (2) AI/chatbot-mediated learning support and proficiency gains; (3) learner-centered strategies, task design, and interaction; and (4) speaker/user experience. The field is evolving toward an integrated perspective that couples technological application with learner experience.

The ten LDA-derived topics and the four CONCOR clusters were not intended to produce identical classifications but to provide complementary perspectives on the conceptual landscape of AI-mediated English-speaking research. The classroom-based outcome validation cluster (Cluster 1) broadly aligned with Topics 4, 6, and 10, which emphasized classroom settings, learner groups, testing procedures, and proficiency levels. This alignment is directly reflected in the co-occurrence matrix: for example, “English–study” (121), “English–student” (107), and “English–result” (99) form some of the strongest pairwise associations in the corpus, mirroring the high-weight classroom and achievement terms in these topics. The AI/chatbot-mediated support cluster (Cluster 2) mapped onto Topics 1, 2, 7, and 8, where high-probability terms such as “AI,” “ChatBot,” “ChatGPT,” and “feedback” indicated strong technological mediation. This relationship is also supported by the co-occurrence structure: “AI–study” (91), “AI–student” (77), and “AI–learning” (65), along with “AI–ChatBot” (44), constitute a dense technological subnetwork consistent with the thematic focus of these topics. The learner-centered task design cluster (Cluster 3) corresponded to Topics 3 and 5, which featured terms related to instructional design, strategy, and interactional processes. Although the top 30 co-occurrence pairs do not fully reflect all task-related terms, strong relational pairs such as “student–class” (50) and “student–group” (37) reinforce the interactional foregrounding that characterizes these topics. Finally, the speaker/user experience cluster (Cluster 4) converged with Topic 3, where “speaker” appeared alongside “intelligence” and “AI,” signaling growing attention to speech-based interfaces. The co-occurrence matrix confirms this pattern: “intelligence–speaker” (26) and “AI–speaker” (not in top 30 but recurrent across abstracts) show consistent co-presence in documents that discuss speech-based AI interaction. Thus, while the LDA model produced ten granular thematic groupings, the CONCOR results synthesized these into four higher-order conceptual domains. The structural relationships in the co-occurrence network therefore corroborate the probabilistic themes identified through topic modeling, enhancing the robustness and interpretive validity of the study’s findings.

## 5. Conclusion & Implications

Using a suite of text-mining methods (word frequency, TF-IDF, N-gram, co-occurrence, topic modeling, and CONCOR), this study systematically mapped research trends in AI-mediated English speaking education across Korean academic journals and theses from 2015 to 2025. The findings demonstrate that AI, particularly ChatBot-based systems such as PengTalk and ChatGPT, has been primarily employed as an interactive learning partner to support speaking proficiency and feedback-driven instruction. The analysis identified four dominant research domains—classroom-based outcome validation, AI/ChatBot-mediated support, learner-centered strategies and task design, and speaker/user experience—indicating a field transitioning from technological adoption toward pedagogical integration. Furthermore, the study highlights how Korean scholarship reflects global tendencies toward generative AI adoption while maintaining a distinct emphasis on empirical classroom validation and learner affect.

The findings of this study align with global research trends that emphasize AI as both a pedagogical agent and an assessment partner (Amrate and Tsai 2024, Lai and Lee 2024). Whereas earlier Korean studies concentrated on the technological feasibility of ASR and CAPT tools, more recent works highlight how generative AI and LLMs enable richer forms of dialogic interaction and personalized feedback. This evolution parallels the global trajectory

identified in studies such as Fryer and Carpenter (2006) and Wang et al. (2024), suggesting that Korea's research community is now contributing to the broader shift from form-focused accuracy training to meaning-oriented communicative learning.

The co-occurrence and topic-modeling results further reveal that Korean scholarship tends to foreground classroom-embedded validation and learner affective responses, contrasting with international research that often prioritizes technical modeling or automated scoring. This difference underscores Korea's strong pedagogical orientation and its potential to inform AI-mediated speaking research that values fairness, authenticity, and contextual appropriateness. Moreover, the growing prominence of "ChatGPT," "PengTalk," and "feedback" in the corpus indicates that the research focus has moved beyond tool adoption toward design-based evaluation and human-AI interaction quality.

The current mapping also provides insight into gaps and tensions within the field. While studies demonstrate short-term gains in speaking proficiency and motivation, longitudinal evidence remains scarce. Thematic clustering shows limited exploration of multimodal feedback (e.g., speech-text-gesture) and ethical dimensions such as transparency or bias in automated assessment. Thus, there is a continuing need to connect AI-mediated speaking research with broader frameworks of communicative competence and emerging standards for AI literacy and responsible design.

From a pedagogical standpoint, the study reinforces that AI and chatbots can serve as authentic interlocutors providing instantaneous, adaptive feedback and expanding speaking opportunities beyond the classroom. Teachers and curriculum designers should therefore view AI not as a peripheral aid but as a catalyst for learner-centered innovation, integrating chatbot dialogues, reflection prompts, and performance analytics into everyday practice.

From a theoretical perspective, the results highlight the need for a convergent research framework that links applied linguistics, learning analytics, and human-AI interaction. Such integration can illuminate how AI shapes linguistic, cognitive, and affective dimensions of L2 speaking development. The shift from CAPT to generative AI also calls for re-conceptualizing construct validity in automated assessment, ensuring that new tools measure not only fluency and accuracy but also interactional competence, discourse management, and pragmatic appropriateness.

From a user-experience (UX) perspective, future AI-mediated learning environments must enhance social presence, personalization, and transparency. Learners' trust and motivation will depend on how clearly systems communicate feedback, manage errors, and adapt to proficiency levels. Cross-disciplinary research combining education, HCI, and ethics can ensure sustainable and equitable adoption in diverse contexts.

This study is subject to several limitations related to corpus scope, preprocessing methods, and the high frequency of platform-specific keywords (e.g., PengTalk). Future research should address these limitations by:

- (1) expanding the dataset to include non-Korean publications for comparative cross-cultural trend analysis;
- (2) linking speech-recognition and learning analytics data to model causal pathways between AI-mediated feedback and proficiency gains;
- (3) conducting longitudinal, classroom-based experiments to validate sustained learning outcomes and affective effects; and
- (4) developing multimodal (speech-text-gesture) AI tutoring architectures that promote authentic interaction and adaptive scaffolding.

In addition, mixed-method research combining quantitative topic modeling with qualitative discourse analysis would provide richer insights into how AI-mediated speaking evolves across contexts and proficiency levels. A

key limitation of the present study is that, although it aims to discuss changes over time, the analyses were conducted on a pooled corpus of abstracts without implementing explicit year-based or longitudinal analyses. As a result, the findings primarily reflect overall thematic distributions rather than systematic temporal shifts in research emphases. Future research should therefore incorporate year-based analytical frameworks—such as keyword trajectory analysis, topic prevalence by publication year, or dynamic topic modeling—to more rigorously capture how abstract-level representations of AI-mediated English speaking research evolve over time.

## References

- Ahadi, A., A. Singh, M. Bower and M. Garrett. 2022. Text mining in education—A bibliometrics-based systematic review. *Education Sciences* 12(3), 210.
- Amrate, M. and P.-H. Tsai. 2024. Computer-assisted pronunciation training: A systematic review. *ReCALL* 37(1), 22-42.
- Charitopoulos, A., M. Rangoussi, D. Metafas and D. Koulouriotis. 2025. Text mining technologies applied to free-text answers of students in e-assessment. *Discover Computing* 28(1), 5.
- Cheon, S. 2023. Educational application of ChatGPT: Its impact on Korean university students' English speaking skills. *The Journal of Linguistic Science* 107, 469-496.
- Choi, W. 2020. Process-oriented speaking assessment of primary English using AI chatbots: Possibilities and limitations. *Primary English Education* 26(1), 131-152.
- Choi, W. 2021. A survey of elementary school students' recognition on the dialogue-based chatbot - EBSe AI Pengtalk. *Journal of the Korea English Education Society* 20(4), 211-231.
- Ericsson, E. and S. Johansson. 2023. English speaking practice with conversational AI: Lower secondary students' educational experiences over time. *Computers & Education: Artificial Intelligence* 5, 100164.
- Eskenazi, M. 2009. An overview of spoken language technology for education. *Speech Communication* 51(10), 832-844.
- Ferreira-Mello, R., M. André, A. Pinheiro, E. Costa and C. Romero. 2019. Text mining in education. *WIREs Data Mining and Knowledge Discovery* 9(6), e1332.
- Fryer, L. K. and R. Carpenter. 2006. Bots as language learning tools. *Language Learning & Technology* 10(3), 8-14.
- Han, C., Y. Yang and K. Kim. 2019. Research on developing a serviceable speaking program based on artificial intelligence Voice Recognition. *Primary English Education* 25(2), 109-125.
- Han, D. E. 2020. The effects of voice-based AI chatbots on Korean EFL middle school students' speaking competence and affective domains. *Asia-Pacific Journal of Convergent Research Interchange* 6(7), 71-80.
- Huh, S., J. Baek and K. Rha. 2024. How did EFL college instructors incorporate into their speaking classes: Focused on using prompts and obtaining feedback? *The Jungang Journal of English Language and Literature* 66(4), 381-401.
- Hong, S. H., S. Lee, J. Jeon and J. Oh. 2021. The perception and satisfaction analysis of AI Pengtalk. *Modern English Education* 22(4), 70-78.
- Hwang, Y. 2021. A study on the educational effectiveness and potential of using an AI speaker for English education. *Multimedia Assisted Language Learning* 47(1), 225-251.
- Jeong, Y. H. and I. Choi. 2021. Exploring the efficacy of an AI chatbot-based automatic speech recognition based on a comparative error analysis. *Multimedia Assisted Language Learning* 24(4), 261-289.
- Jo, K., J. H. Park, B. Nam, C. Sim and H. Choi. 2022. Responses of the AI chatbots to erroneous utterances: AI

- Pengtalk and speaking class. *The Journal of Education* 42(4), 179-195.
- Ki, Y. 2022. *A Study on the Perception and Attitudes of Elementary School Students on Speaking English Education Using AI Technology*. Master's thesis, Hankuk University of Foreign Studies, Seoul, South Korea.
- Kim, H. and S. S. Lee. 2023. A study on task design of AI-based voice chatbot for elementary English speaking. *Multimedia Assisted Language Learning* 26(1), 31-58.
- Kim, M. J. 2020. *A study on Korean Middle School Students' English Speaking Assessment Using Artificial Intelligence Speaker: Focusing on Affective Effect and Objectivity in Assessment*. Master's thesis, Chung-Ang University, Seoul, South Korea.
- Kim, N. Y. 2016. Effects of voice Chat on EFL learners' speaking ability according to proficiency levels. *Multimedia Assisted Language Learning* 19(4), 63-88.
- Kim, N. Y. 2017. Effects of different types of Chatbots on EFL learners' speaking competence and learner perception. *Cross-cultural Studies* 48(1), 222-252.
- Ko, M. 2024. Development of an AI-based OPIc teaching model. *Studies on Humanities and Social Studies* 6(5), 781-794.
- Kong, Y. J. 2020. *Development of Instructional Strategies for English Speaking in Class Using AI Speakers*. Master's thesis, Seoul National University.
- Lai, W. Y. W. and J. S. Lee. 2024. A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers & Education: Artificial Intelligence* 7, 100291.
- Lee, D. H. and S. Park. 2019. A developmental plan for an English conversation learning chatbot through the application of elementary school English textbooks. *Primary English Education* 25(4), 79-100.
- Lee, E. B. 2021. *User Experience Research on AI Speakers as an English Conversation Learning Tool*. Master's thesis, Hongik National University, Seoul, South Korea.
- Lee, J. H., H. Kim, D. Shin and H. J. Yang. 2019. A study on the conversational interaction with chatbot for foreign language learning. *Multimedia Assisted Language Learning* 22(1), 132-153.
- Lee, S. 2019. The effects of gamification-based artificial intelligence chatbot activities on elementary English learners' speaking performance and affective domains. *Primary English Education* 25(3), 75-98.
- Lim, Y. 2024. Learning effects of complementary learning methods in speaking skills using AI-supported devices. *Foreign Studies* 68(6), 91-112.
- Mahdi, H. S. and A. A. Al-Khateeb. 2019. The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education* 7(3), 733-753.
- Noh Y. 2024. The impact of ChatGPT-Assisted learning on English speaking proficiency of Korean university students. *The New Studies of English & Literature* 89, 89-114.
- Oh, J., S. Hong, S. Lee and J. Jeon. 2022. The role and application of AI PengTalk in elementary English education. *Korean Journal of Elementary Education* 33(2), 35-50.
- Oh, Y. and J. Back. 2022. Exploring the use and effectiveness of an artificial intelligence Chatbot in Korean elementary school EFL classrooms: 'AI PengTalk' as a speaking learning tool. *The Journal of Modern British & American Language & Literature* 40(1), 207-235.
- Park, H. J. 2022. The effects of using mobile assisted speech recognition applications on listening proficiency and affective attitude. *English Language Teaching* 34(4), 117-134.
- Rha, K. 2024. Exploring exploitability of ChatGPT as a learning platform for EFL College students' speaking. *Studies in Linguistics* 72, 199-223.
- Seo, J. E. 2017. *Case Study of Early Childhood English Learning Based on Artificial Intelligence Voice Recognition*

- System*. Master's thesis, Chung-Ang University, Seoul, South Korea.
- Seong, S. and S. Lee. 2021. Analyzing learners' and teachers' perceptions of AI PengTalk for English learning and the suggestions for its use. *The Journal of Learner-Centered Curriculum and Instruction* 21(21), 915-935.
- Shim, K. N., E. J. Song, S. Y. Chu, H. K. Kwon and D. G. Min. 2020. A study of the development of AI mediated assessment tasks and their application into practice in primary English classrooms. *English Language Assessment* 15(1), 41-61.
- Sim, S. Y. 2022. *(An) Analysis and Evaluation of Ai-based English Speaking Applications for Young Learners*. Master's thesis, Chung-Ang University, Seoul, South Korea.
- Song, E. J. and K. N. Shim. 2020. A study on the development of AI-based assessment model for English speaking skills in primary school. *English Language Assessment* 15(2), 53-76.
- Sung, M. 2019. Development of a flowchart-based English-speaking chatbot for negotiation of meaning. *Primary English Education* 25(4), 101-122.
- Wang, C., B. Zou, Y. Du and Z. Wang. 2024. The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System* 127, 103533.
- Wang, J. 2024. Utilizing text mining technology to enhance English learners' vocabulary. *International Journal of Electronics and Communication Engineering* 11(9), 86-98.
- Yang, H. J., H. Kim, D. Shin and J. Lee. 2019. A study on adopting AI-based chatbot in elementary English-Speaking classes. *Multimedia Assisted Language Learning* 22(4), 184-205.
- Yeom, J. 2022. *Development of Design Principles for Synchronous Online English Speaking Classes Using Artificial Intelligence Chatbot*. Master's thesis, Seoul National University, Seoul, South Korea.
- Yoon, Y. 2021. Prospects of using AI chatbots in teaching speaking in primary English: With special reference to Dialogflow. *The Journal of Korea Elementary Education* 32(2), 15-28.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary

**Appendix. Co-Occurrence Matrix (Top Terms × Top Terms) – Counts**

	English	ai	chatbot	student	study	learning	learner	class	group	level	school	teacher	education	research	result	language	pengtalk	intelligence	ChatGPT
English	0	94	60	107	121	88	82	61	45	66	65	55	58	80	99	63	21	53	23
AI	94	0	44	77	91	65	61	42	38	48	45	41	40	58	70	46	19	39	14
chatbot	60	44	0	49	56	40	39	32	19	36	33	28	29	36	42	32	7	23	5
student	107	77	49	0	100	70	63	50	37	56	58	50	42	61	80	50	20	40	20
study	121	91	56	100	0	83	76	56	45	63	57	50	50	69	94	61	21	49	19
learning	88	65	40	70	83	0	66	41	38	51	43	38	41	57	69	50	11	36	16
learner	82	61	39	63	76	66	0	41	33	48	40	32	40	54	64	44	11	32	12
class	61	42	32	50	56	41	41	0	26	33	36	30	26	38	44	24	13	25	10
group	45	38	19	37	45	38	33	26	0	29	21	19	19	25	40	27	7	26	11
level	66	48	36	56	63	51	48	33	29	0	37	35	31	37	51	36	9	30	11
school	65	45	33	58	57	43	40	36	21	37	0	37	34	42	49	28	17	31	9
teacher	55	41	28	50	50	38	32	30	19	35	37	0	29	37	43	26	11	24	9
education	58	40	29	42	50	41	40	26	19	31	34	29	0	37	43	29	9	30	11
research	80	58	36	61	69	57	54	38	25	37	42	37	37	0	58	43	8	29	16
result	99	70	42	80	94	69	64	44	40	51	49	43	43	58	0	49	16	41	14
language	63	46	32	50	61	50	44	24	27	36	28	26	29	43	49	0	5	33	14
pengtalk	21	19	7	20	21	11	11	13	7	9	17	11	9	8	16	5	0	9	0
intelligence	53	39	23	40	49	36	32	25	26	30	31	24	30	29	41	33	9	0	7
ChatGPT	23	14	5	20	19	16	12	10	11	11	9	9	11	16	14	14	0	7	0