Contemporary Educational Technology, 2025, **17**(4), ep596
ISSN: 1309-517X (Online)

OPEN ACCESS

Research Article

The effectiveness of artificial intelligence in English instruction for speaking and listening skills: A meta-analysis

Thada Jantakoon [1]
0000-0001-5935-5875

Thiti Jantakun [2*]
0009-0007-4250-7589

Kitsadaporn Jantakun [2]
0009-0004-0856-6524

Weerapa Pongpanich [3]
0009-0005-4540-5719

Rungfa Pasmala [4]
0009-0009-7710-9593

Panita Wannapiroon [4]
0000-0001-8633-5781

Prachyanun Nilsook [4]
0000-0003-3019-3635

[1] Rajabhat Maha Sarakham University, Mahasarakham, THAILAND
[2] Roi Et Rajabhat University, Roi Et, THAILAND
[3] Rangsit University, Pathum Thani, THAILAND
[4] King Mongkut's University of Technology North Bangkok, Bangkok, THAILAND
[*] Corresponding author: thiti100@reru.ac.th

Citation: Jantakoon, T., Jantakun, T., Jantakun, K., Pongpanich, W., Pasmala, R., Wannapiroon, P., & Nilsook, P. (2025). The effectiveness of artificial intelligence in English instruction for speaking and listening skills: A meta-analysis. *Contemporary Educational Technology, 17*(4), ep596. https://doi.org/10.30935/cedtech/17310

ARTICLE INFO

Received: 24 Sep 2024

Accepted: 26 Jun 2025

ABSTRACT

The increasing integration of artificial intelligence (AI) in education has raised significant questions about its pedagogical value, especially in language learning. This meta-analysis examines the extent to which AI contributes to the development of English-speaking and listening skills. A systematic review of the literature was conducted by the preferred reporting items for systematic reviews and meta-analyses guidelines, utilizing peer-reviewed studies indexed in Scopus, ERIC, and EBSCOhost from 2017 to 2024. Nineteen studies met the inclusion criteria, all of which utilized experimental or quasi-experimental designs with measurable learning outcomes. The analysis reveals a substantial overall effect of AI-enhanced instruction (standardized mean difference [SMD] = 0.981, 95% confidence interval [0.571, 1.391], $p < .001$), with particularly notable improvements in speaking proficiency (SMD = 1.033). Although listening outcomes showed a positive trend (SMD = 0.714), the effect did not attain statistical significance. Considerable heterogeneity was noted across the studies, reflecting variations in learner populations, instructional contexts, and AI applications. Quality appraisal using the risk of bias in non-randomized studies of interventions framework indicated a predominantly low to moderate risk of bias. Publication bias analysis, including funnel plot symmetry and fail-safe N, further confirmed the reliability of the results. These findings highlight the advantages of AI in enhancing speaking skills within English instruction and underscore the need for further empirical studies to investigate its impact on listening comprehension. Collectively, the results provide timely, evidence-based guidance for educators and policymakers aiming to integrate AI effectively into language education. Highlight the advantages of AI in enhancing speaking skills within English instruction and underscore the need.

Keywords: artificial intelligence, English instruction, speaking skill, listening skill, meta-analysis

INTRODUCTION

In recent years, the integration of artificial intelligence (AI) into educational contexts has initiated a paradigm shift in instructional methodologies, particularly in the domain of English language education (Ling, 2023). As English continues to serve as the dominant global lingua franca, the demand for innovative and effective pedagogical approaches has intensified, prompting educators and researchers to explore the potential of AI in fostering language acquisition outcomes (Duisenova & Zhorabekova, 2023). Advancements in AI have enabled the development of adaptive learning platforms, real-time feedback systems, and interactive environments, all of which offer unprecedented opportunities to support learners in developing core communicative competencies, most notably, speaking and listening skills (Qiao & Zhao, 2023).

While traditional instruction has long played a foundational role in language education, AI represents a pivotal evolution in pedagogy by enabling instruction that is both responsive and personalized (Chang, 2023). AI-powered tools such as virtual dialogue agents, speech recognition systems, and gamified applications can immerse learners in dynamic, context-rich environments that simulate authentic communication scenarios. These technologies provide immediate, data-driven feedback on elements such as pronunciation, intonation, and fluency components critical to spoken language development (Li, 2024).

Speaking and listening, though fundamental to communicative competence, are often the most challenging skills for learners to master, particularly in contexts where exposure to authentic language use is limited (Choe et al., 2020). Classroom-based instruction may not always provide sufficient opportunities for extended oral practice or feedback, thereby necessitating supplementary tools to bridge the gap (John & Yunus, 2021). AI-based platforms address this need by providing learners with repeated exposure to varied linguistic inputs, including diverse accents and speech patterns, while promoting autonomous language use (Daweli & Mahoub, 2024).

Despite growing interest in AI-enhanced language learning, the literature remains fragmented, and the field lacks a comprehensive synthesis of empirical findings, especially regarding AI's effectiveness in promoting speaking and listening proficiency (Rahmati et al., 2021). Although many individual studies report promising outcomes, their methodological diversity and varying scopes make it difficult to draw generalizable conclusions. A rigorous meta-analytic review is thus warranted to assess the magnitude and consistency of AI's instructional impact (Alsowat, 2020).

This study addresses this gap by conducting a meta-analysis that evaluates the effectiveness of AI-enhanced instruction on the speaking and listening skills of English learners. Synthesizes evidence from empirical studies that utilize AI tools, including speech recognition software, conversational agents, adaptive learning platforms, and virtual reality environments (Madhavi et al., 2023). By aggregating data from diverse educational levels and geographical contexts, the analysis aims to offer a statistically robust and pedagogically meaningful assessment of AI's instructional utility.

The significance of this research lies not only in its contribution to the academic discourse on educational technology but also in its implications for instructional policy, teacher training, and curriculum innovation (Amin, 2023; Konyrova, 2024). As AI becomes increasingly embedded in educational ecosystems, a nuanced understanding of its pedagogical strengths and limitations is essential for informed decision-making and sustainable integration (Yang, 2024).

To guide this investigation, the following research questions are posed:

1. To what extent does AI-enhanced instruction improve English learners' speaking and listening skills compared to traditional instructional methods?
2. Are there differential effects of AI interventions across various types of speaking and listening tasks or learner proficiency levels?

3. How does the instructional effectiveness of AI compare between the domains of speaking and listening?

Through these questions, this study seeks to illuminate both the pedagogical possibilities and practical constraints of AI-supported English language instruction. Also aims to inform the design of future research and the strategic implementation of AI technologies in language learning environments (Wu, 2024; Xatamova, 2024). Given the rapid evolution of AI applications, this meta-analysis provides a timely overview of current research trends and offers a foundation for shaping future innovations in technology-enhanced language education (Bao & Sile, 2024; Yaseen, 2023).

LITERATURE REVIEW

Evolution and Applications of AI in Language Learning

Researchers have been paying more and more attention to the use of AI in English language education over the past ten years, especially for its ability to improve students' speaking and listening skills (Hao et al., 2021). The goal of this section is to bring together what we know so far about how well AI works for language instruction, focusing on important empirical findings, methodological approaches, and gaps in the literature that keep coming up.

AI's growth in language learning is part of a larger trend away from traditional computer-assisted language learning systems and toward more advanced, AI-powered platforms that can analyze speech in real time, give dynamic feedback, and tailor lessons to each student (Sharifi et al., 2018). Early studies in technology-enhanced language learning helped us come up with ideas for how to judge new digital teaching methods. Studies that look at a lot of different studies, like Zhao's (2013), have shown that technology has a generally positive effect on learning a language. This opens the door for more focused research into AI-specific tools.

As AI technologies have advanced, more attention has been paid by researchers to how they might be used to improve oral language development. By providing real-time, automated feedback in a safe, learner-controlled setting, speech recognition systems, for example, have proven effective in enhancing learners' pronunciation, fluency, and prosodic characteristics (Agnes & Srinivasan, 2024; Choe et al., 2020). In parallel, conversational agents driven by AI have been used to mimic real-world conversations, giving students adaptive linguistic support and the opportunity to practice speaking in situationally relevant settings (Qiao & Zhao, 2023).

AI has made it easier to create adaptive listening exercises that instantly adapt to students' skill levels (Lee et al., 2020). These systems improve learners' auditory discrimination and comprehension abilities by exposing them to a variety of speech patterns, regional accents, and contextual listening scenarios (Lin & Liu, 2024). Furthermore, learner error patterns can be recognized by AI algorithms, which can then produce tailored feedback to support more strategic and differentiated listening instruction (Ni et al., 2022).

These new technologies are a big step forward for teaching, but a lot of the research that is already out there is limited in scope, often only looking at short-term interventions or not having strict experimental controls. Also, many empirical designs don't consider how speaking and listening skills are connected, which shows that we need more studies that look at these skills in a variety of contexts over time. To fully understand the teaching value and practical limits of AI-enhanced language learning, it is important to deal with these problems.

Effectiveness and Challenges of AI in Language Instruction

Many studies have shown that using AI to improve speaking and listening skills can have good results. Ling (2023) discovered that students who used AI-enhanced platforms made significant progress in their speaking skills compared to those who learned in a more traditional way. Duisenova and Zhorabekova (2023) also found that adding gamified elements to AI-based apps made students more interested and helped them understand what they were listening to better.

AI works better for teaching languages in some situations and for some types of learners than others. Age, skill level, and cultural background are just a few of the things that can affect how well students respond to

AI-based learning (Peng & Xie, 2021). For instance, Lee et al. (2023) said that how well AI-supported English lessons worked depended on how different the learners' native language was from English.

Recent studies have looked at how to use AI in the classroom. Researchers have looked into how AI could help teachers give more personalized lessons and grade students' work more quickly (Feng et al., 2023). Blended learning models, which mix AI-assisted self-study with traditional teaching, have become more popular as a way to balance the benefits of technology with the need for human interaction in the classroom (Chu & Szlagor, 2023).

Some researchers are worried about relying too much on AI in language learning, even though these are good things. There have been doubts about the accuracy of AI-generated language input and the fact that current systems can't handle the subtlety of human communication (Ho, 2024). Also, fairness and access are still very important when looking at how widely AI tools are used in different types of schools (Murdoch & Lin, 2023).

Research has also examined AI's impact on specific aspects of speaking and listening. In speaking, studies have focused on pronunciation, fluency, vocabulary usage, and grammatical complexity (Madhavi et al., 2023). In listening, AI has been shown to improve learners' ability to distinguish similar sounds, comprehend connected speech, and understand a range of accents and dialects (Yang, 2024).

Current Research Approaches and Future Directions

Researchers in the field of AI-assisted language instruction have used a variety of methods, such as experimental studies, quasi-experimental designs, and longitudinal inquiries (Rahmati et al., 2021). Even though many of these studies say that AI has a positive effect on speaking and listening skills, it's hard to say for sure what the overall effect is because the research designs, sample characteristics, and measurement criteria are all different (Alsowat, 2020).

Recent advances in machine learning and natural language processing have opened up new possibilities for AI in language learning. Smart conversational agents and virtual reality-based learning environments are two examples of new technologies that are expected to make learning more immersive, personalized, and interactive (Qian & Kong, 2024). Researchers face problems because technology is changing so quickly, especially when it comes to making sure that study results are still useful and relevant over time (Yaseen, 2023).

The field needs big, thorough studies that can give us generalizable information about how well AI technologies work for teaching languages (Wang et al., 2024a). Long-term studies that look at the lasting effects of AI-enhanced teaching are especially important for figuring out how valuable it is in the long run (Jerusha & Rajakumari, 2024).

There is a lot of research that shows that AI can help people improve their English speaking and listening skills, but the results are all over the place. A comprehensive meta-analytic synthesis is needed to bring them all together. Such an analysis would help make things clearer about best practices, help plan future research and give evidence-based advice on how to use AI in English language curricula and policy making.

METHODOLOGY

We conducted a meta-analysis to investigate the impact of AI on English classes in terms of speaking and listening skills. To determine the impact of AI on English lessons for speaking and listening skills, we employed a five-step method suggested by previous studies: search strategy, inclusion criteria, data extraction, publication bias assessment, and data analysis.

Search Strategy

This meta-analysis followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines as outlined by Moher et al. (2010) to maintain transparency, accuracy, and methodological rigor throughout the review process. Literature was retrieved from three major academic databases: Scopus, ERIC, and EBSCO. These databases were selected for their strong coverage of high-quality publications in the field of educational technology.

The search strategy used the keywords “AI” AND “English learning”, and the search was limited to publications released between January 2017 and December 2024. The database search was conducted electronically and focused on identifying empirical studies relevant to the application of AI in English language learning contexts.

Inclusion Criteria

Research using a randomized or quasi-experimental design was included for this review. Publications based on the population-intervention-comparison-outcome (PICO) paradigm were also considered. The PICO aspects in this research were as follows:

(1) population: pre-/post-registration students,

(2) intervention: all kinds of AI in English Instruction methods,

(3) comparison: traditional education methods, and

(4) outcomes: speaking and listening skills.

Data Extraction

Authors, publication dates, countries, sample sizes, participant types, intervention methods, and outcomes were among the criteria that two reviewers (KJ and WP) separately gathered.

Publication Bias Test

Ferrer (1998), Tang and Liu (2000), and Song et al. (2002) exemplify studies that utilize the funnel plot to investigate bias. Publication bias, stemming from the tendency for research with favorable outcomes to receive publication approval, is a significant contributor to prejudice. This occurs because studies yielding positive results are more frequently submitted for publication. The bias in the English language is marked by the seldom publication of negative findings in English-language journals compared to positive studies; yet this trend is not always applicable. Citation bias is defined as the tendency for research yielding positive results to be cited more often, hence facilitating their identification and inclusion in databases. An instance of purposeful bias is a pharmaceutical company intentionally concealing negative product information (Eyding et al., 2010). This indicates that prejudice could be deliberate. A symmetric distribution serves as the foundation for evaluating bias. The distribution should exhibit symmetry if all studies provide random evaluations of the identical unbiased mean value. Assume that the studies exhibit bias, including an insufficient quantity of small studies yielding favorable outcomes and substantial effect sizes. The funnel plot displays asymmetry, characterized by a significant gap in the lower half (**Figure 1**).

Fail-Safe N

Fail-safe N is a statistical concept used in meta-analysis to assess the robustness of the findings by estimating the number of non-significant or null studies needed to nullify the observed effect. It indicates the potential impact of publication bias on the overall results.

Calculation of fail-safe N

1. Determine the observed effect size: Identify the effect size obtained from the meta-analysis.
2. Set a criterion for statistical significance: Establish a threshold for statistical significance (commonly $p < 0.05$).
3. Calculate the z-score: Use the observed effect size and standard error to calculate the z-score.
4. Estimate the fail-safe N: Apply Rosenthal’s (1979) formula to estimate the number of non-significant studies needed to bring the overall effect to a non-significant level.

Interpretation

A larger fail-safe N suggests that the meta-analytic result is more robust, as many non-significant studies would be required to overturn the observed effect. Conversely, a smaller fail-safe N raises concerns about the stability of the findings.

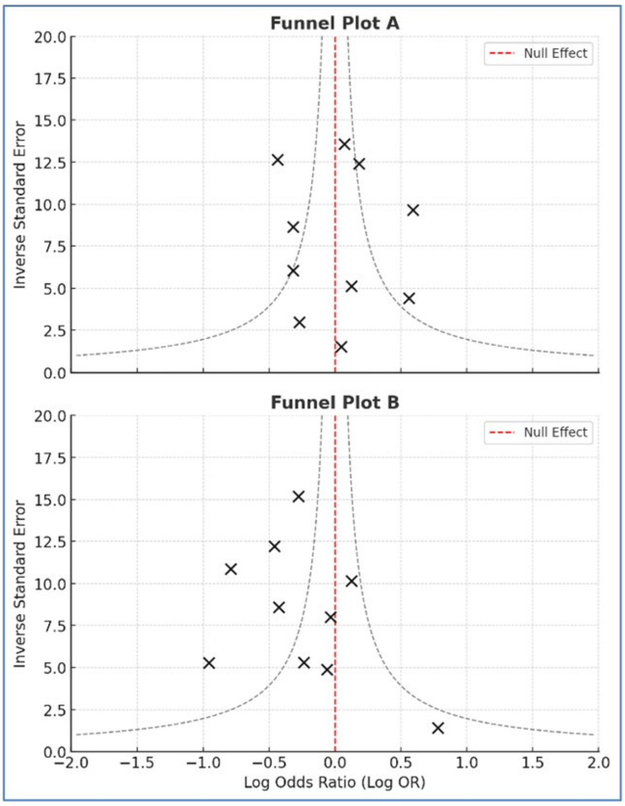

Figure 1. Top panel: Significant point disadvantage with a limited sample size & Lower panel: Restoration of the symmetrical funnel (including open circles) (Source: The authors' own work)

Example

Suppose a meta-analysis yields an observed effect size with a p-value of 0.02 and a fail-safe N of 100. In that case, it suggests that 100 additional non-significant studies with a similar effect size would be needed to render the overall result non-significant. This provides a measure of reassurance regarding the robustness of the observed effect.

Data Analysis

Jamovi desktop software was used during the meta-analysis (Caldwell, 2022). We reported the standardized mean difference (SMD) for continuous data, along with 95% confidence intervals (CIs). In each analysis, I^2 was used to evaluate the statistical heterogeneity between the studies. The selection of the fixed-effect model ($p > .1$, $I^2 < 50\%$) or the random-effects model ($p < .1$, $I^2 \geq 50\%$) was determined by the values of p and I^2 (Higgins & Thompson, 2002). The fixed-effect model was selected based on the values of p and I^2.

FINDINGS AND DISCUSSION

Results of the Literature Search

To ensure a comprehensive and replicable synthesis of relevant research, a systematic search strategy was implemented in accordance with the PRISMA guidelines (Moher et al., 2010). The procedural flow of identification, screening, eligibility evaluation, and final inclusion is illustrated in **Figure 2**.

The literature search was conducted across three major academic databases: Scopus (n = 150), ERIC (n = 44), and EBSCOhost (n = 7615), yielding a total of 7,809 initial records. The search strategy employed a Boolean combination of keywords: "artificial intelligence" OR "AI" AND ("English instruction" OR "English language learning") AND ("speaking skills" OR "listening skills"). Results were filtered to include only peer-reviewed publications in English published between 2017 and 2024. Eligible sources included journal articles, conference proceedings, and academic monographs.

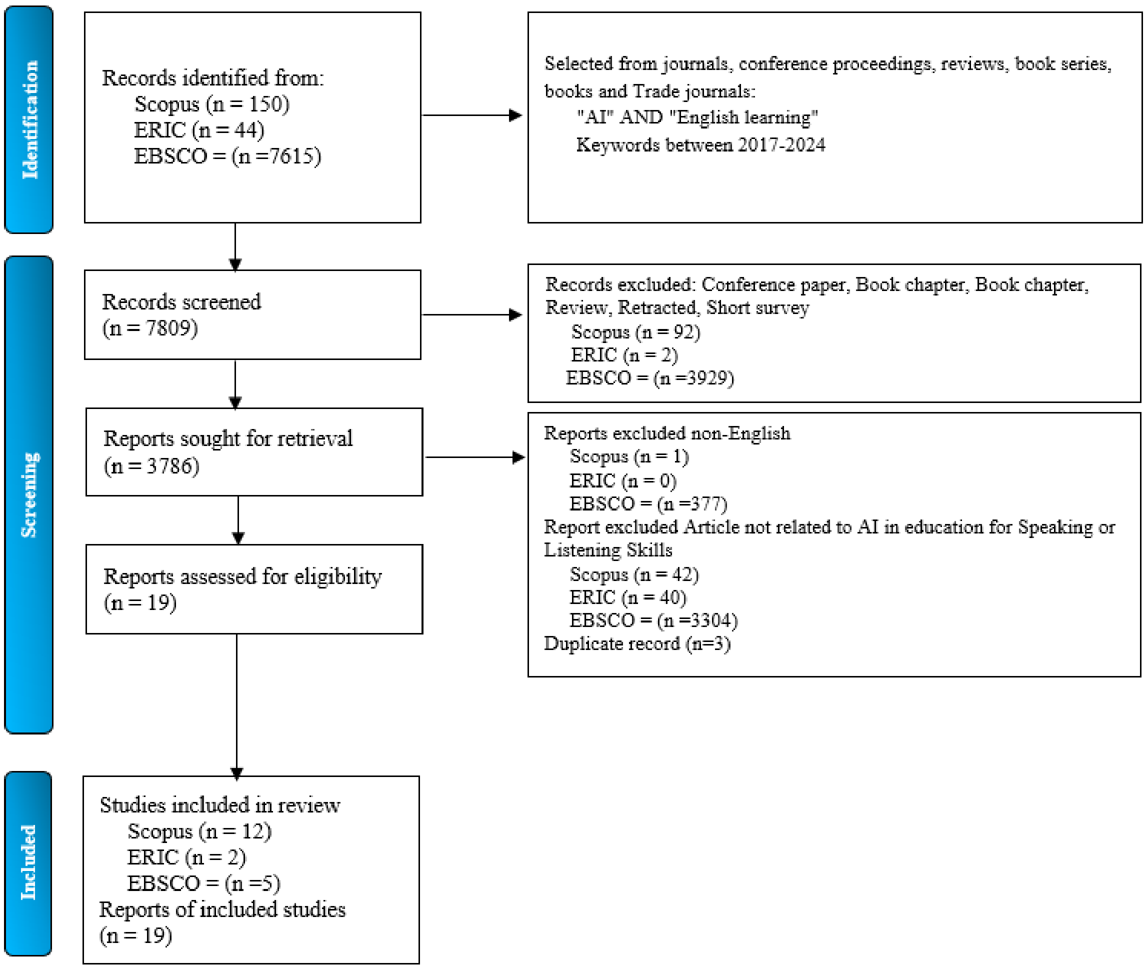

Figure 2. PRISMA article screening flowchart (Source: The authors' own work)

During the initial screening phase, 4,023 records were excluded based on title and abstract due to irrelevance or inappropriate publication type (e.g., literature reviews, retracted papers, short communications, or book chapters). The remaining 3,786 records were reviewed in full text. Further exclusions were made for studies not published in English, those lacking empirical data on AI-supported instruction for speaking or listening, and duplicate entries across databases. Specifically, 42 articles from Scopus, 40 from ERIC, and 3,304 from EBSCOhost were excluded, along with three duplicates.

After rigorous eligibility evaluation, 19 empirical studies met all inclusion criteria and were retained for meta-analysis. These consisted of 12 studies from Scopus, two from ERIC, and five from EBSCOhost. All included studies utilized either randomized controlled trials or quasi-experimental designs, featured both AI-enhanced instructional groups and traditional control groups, and reported quantitative outcomes in either speaking or listening domains.

As shown in **Figure 2**, this systematic approach ensured methodological transparency and minimized selection bias, thereby providing a robust evidentiary base for the subsequent meta-analytic synthesis.

Study Characteristics

This meta-analysis comprises 19 studies that represent a diverse and internationally distributed body of research focused on the effectiveness of AI in English language instruction, particularly in terms of speaking and listening skills. **Table 1** presents a comprehensive summary of the essential characteristics of these studies.

Table 1. Attributes of the 19 included studies

Author (year)	Type of participant	Number of participants: Total (number of groups)	Experimental (AI learning)	Control group (traditional education)	Outcome
Um et al. (2023)	Elementary school students	616 (2)	341	275	Listening skills, Speaking skills
Chen et al. (2022)	Elementary school students	70 (2)	35	35	Speaking skills
Ahmed Ali (2020)	Primary school student	40 (2)	20	20	Listening skills, Speaking skills
Adipat (2023)	Undergraduate students	80 (2)	40	40	Speaking skills
Lui and Chen (2023)	Elementary schools	36 (2)	18	18	Speaking skills
Lin and Mubarok (2021)	Undergraduate students	50 (2)	28	22	Speaking skills
Kim (2022)	Undergraduate students	486 (3)	164	132	Listening skills
Wang and Zheng (2022)	Undergraduate students	60 (2)	30	30	Listening skills, Speaking skills
Qiao and Zhao (2023)	Undergraduate students	93 (2)	47	46	Speaking skills
Zou et al. (2023)	University student	70 (2)	35	35	Speaking skills
Kemelbekova et al. (2024)	University student	51 (2)	26	25	Speaking skills
Yi (2024)	Secondary school students	105 (2)	52	53	Speaking skills
Zheng et al. (2024)	University student	52 (2)	26	26	Speaking skills
Hwang et al. (2024)	Senior high school students	93 (2)	46	47	Speaking skills
Fathi et al. (2024)	University student	65 (2)	33	32	Speaking skills
Zhang et al. (2024)	Undergraduate student	131 (2)	65	66	Speaking skills
Wang et al. (2024b)	Undergraduate student	99 (3)	33	33	Speaking skills
Sonsaat-Hegelheimer and Kurt (2024)	Undergraduate student	15 (3)	5	5	Speaking skills
Ibrahim et al. (2024)	Undergraduate student	81 (2)	40	41	Speaking skills

The studies examined cover a range of geographical regions, including East Asia (China, Korea, and Taiwan), Southeast Asia (Thailand), the Middle East (Egypt), Central Asia (Kazakhstan), and Western academic contexts. The geographic diversity improves the applicability of the findings across various educational systems and sociocultural contexts.

The research encompassed various learner demographics, including elementary school children, secondary school students, and university-level learners. Among the 19 studies, the majority (n = 11) concentrated on undergraduate students, while four studies involved elementary-level learners, two focused on high school students, and two were conducted with primary school pupils. This variety facilitates comparative analysis across educational levels and developmental stages.

The aggregate sample size across all studies was 2,172 participants, with individual study samples varying from 15 to 616. All studies utilized randomized or quasi-experimental designs, with participants assigned to an experimental group (receiving AI-enhanced instruction) and a control group (receiving traditional instruction). Most studies employed a two-group comparison framework, whereas a minority utilized three-group designs to assess various intervention formats.

Quality Assessment

We used the risk of bias in non-randomized studies of interventions (ROBINS-I) tools to check the methodological rigor and internal validity of the studies we included. This tool looks at seven main areas where bias could happen:

(1) bias caused by confounding,

(2) bias in choosing participants,

(3) bias in classifying interventions,

(4) bias caused by not following through on planned interventions,

(5) bias caused by missing data,

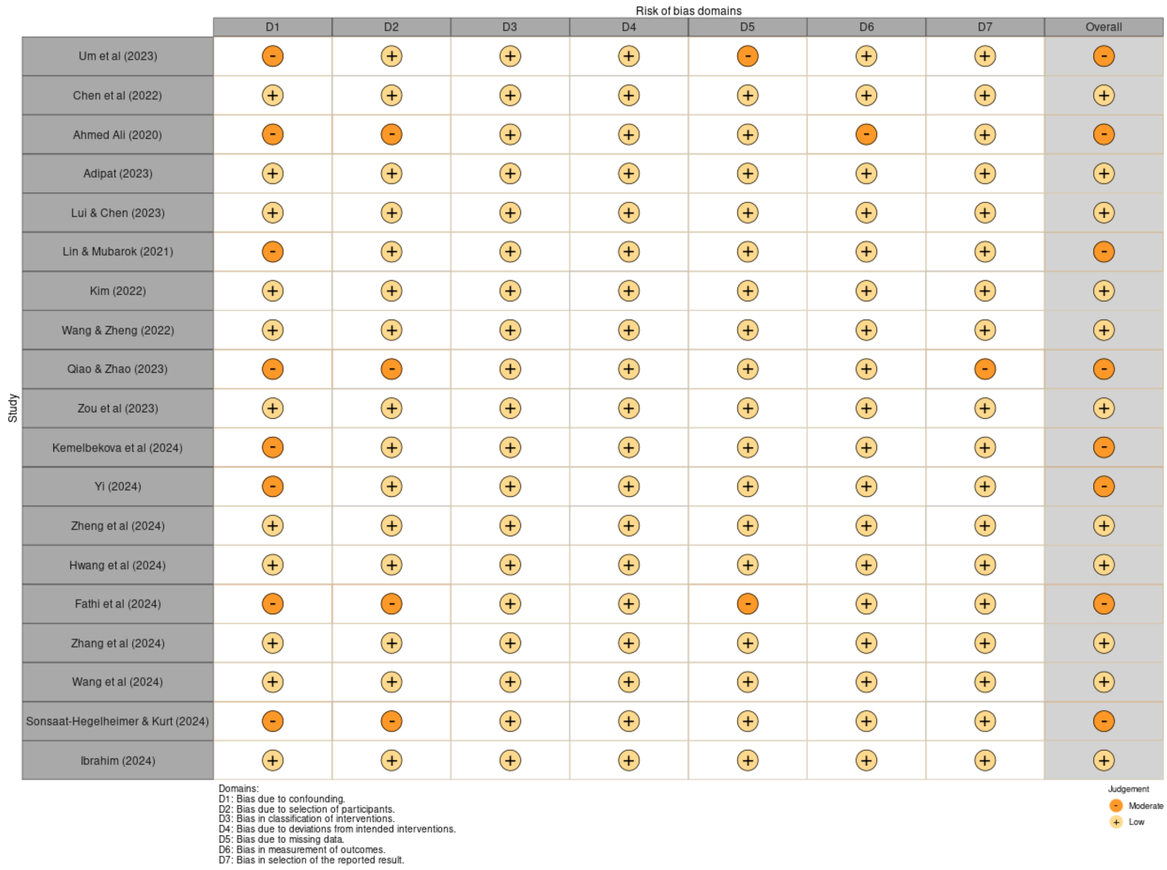

Figure 3. Traffic-light plot (ROBINS-I) (Source: The authors' own work)

(6) bias in measuring outcomes, and

(7) bias in choosing the reported result.

The studies looked at each domain and gave it a low or moderate risk rating.

Figure 3 shows the traffic-light plot, which gives a visual summary of the risk evaluations for each study at the domain level. Most of the studies showed a low risk of bias in all areas. Still, there were some modest risks, especially in D1 (confounding), D2 (participant selection), and D6 (outcome measurement). These were mostly due to not fully adjusting for differences at the start, not using random research designs, or not being able to trust the outcome measures employed in quasi-experimental settings.

Figure 4 shows the overall overview of risk assessments from all the studies in the pool. More than 75% of the domain evaluations were rated as low risk, which means that the research included had a generally high level of methodological soundness. Even so, the fact that there is moderate risk in some areas means that the results need to be carefully looked at, especially when it comes to generalizability and the strength of causal assertions.

It was important that none of the research were found to have a major or critical risk in any area. This makes the meta-analytic results more believable and confirms that the synthesized effect sizes are correct. The fact that all of the research used the same methods makes it possible to draw reasonable conclusions regarding how well AI works for teaching English.

Results of the Meta-analysis

The results of this meta-analysis provide strong evidence that AI is helpful in teaching English, especially when it comes to helping students improve their speaking and listening skills. The analysis combines data from 19 studies and shows statistically significant results that show how AI could change language instruction today.

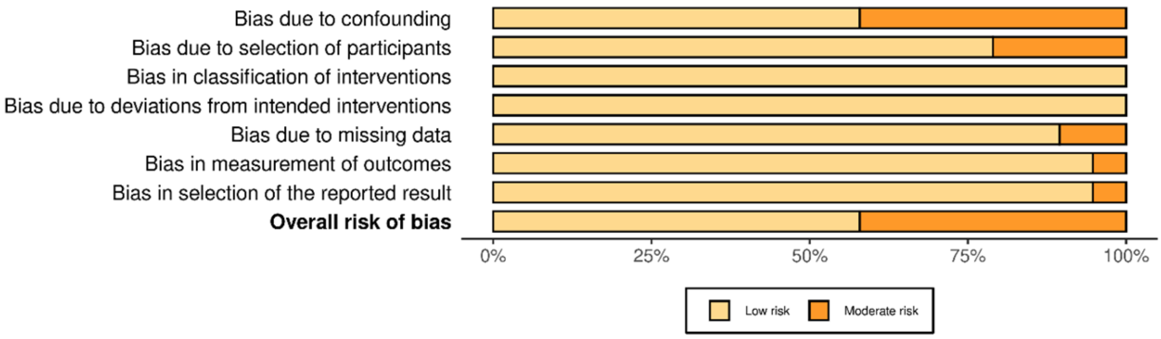

Figure 4. The summary plots (ROBINS-I) (Source: The authors' own work)

Table 2. Random-effects model (k = 19)

	Estimate	SE	Z	p	95% CI	
					Lower bound	Upper bound
Intercept	0.981	0.195	5.03	< .001	0.571	1.391

Note. Tau² estimator: REML. Knapp and Hartung (2003) adjustment used.

Table 3. Heterogeneity statistics

Tau	Tau²	I²	H²	R²	df	Q	p
0.804	0.6459 (SE = 0.2414)	93.42%	15.192	-	18.000	243.204	< .001

Table 4. Model fit statistics and information criteria

	Log-likelihood	Deviance	AIC	BIC	AICc
ML	–23.427	64.172	50.855	52.744	51.605
REML	–22.668	45.336	49.336	51.117	50.136

Overall Effectiveness of AI in English Instruction

The random-effects meta-analysis (**Table 2**) yielded a substantial overall SMD (SMD = 0.981; 95% CI [0.571, 1.391]; $p < .001$), indicating a statistically significant and practically meaningful positive effect of AI-based instruction on learners' speaking and listening skills. This effect size, approaching one standard deviation, signals a substantial instructional benefit. It lends empirical weight to prior claims regarding the pedagogical value of AI in creating interactive, personalized, and immersive language learning environments.

However, the observed heterogeneity was high (I^2 = 93.42%, **Table 3**), suggesting considerable variability across studies in terms of participant characteristics, instructional settings, AI tools used, and measurement approaches. This degree of heterogeneity supports the appropriateness of employing a random-effects model and highlights the importance of further subgroup and moderator analyses

As shown in **Table 4**, the model fit indices support the use of the restricted maximum likelihood (REML) estimation method over the traditional maximum likelihood (ML) approach. Specifically, the REML method yielded a higher log-likelihood (–22.668 vs. –23.427), lower deviance (45.336 vs. 64.172), and reduced values for AIC, BIC, and AICc, indicating a better overall model fit. These findings suggest that REML offers a more accurate estimation of variance components, particularly in the presence of between-study heterogeneity, thereby justifying its selection for the random-effects model in the current meta-analysis.

The p-curve analysis (**Figure 5**) provides further validation of the meta-analytic findings by demonstrating a right-skewed distribution of statistically significant results, thereby mitigating concerns regarding publication bias. This pattern affirms the evidential value of the synthesized outcomes and strengthens confidence in the robustness and credibility of the observed effects.

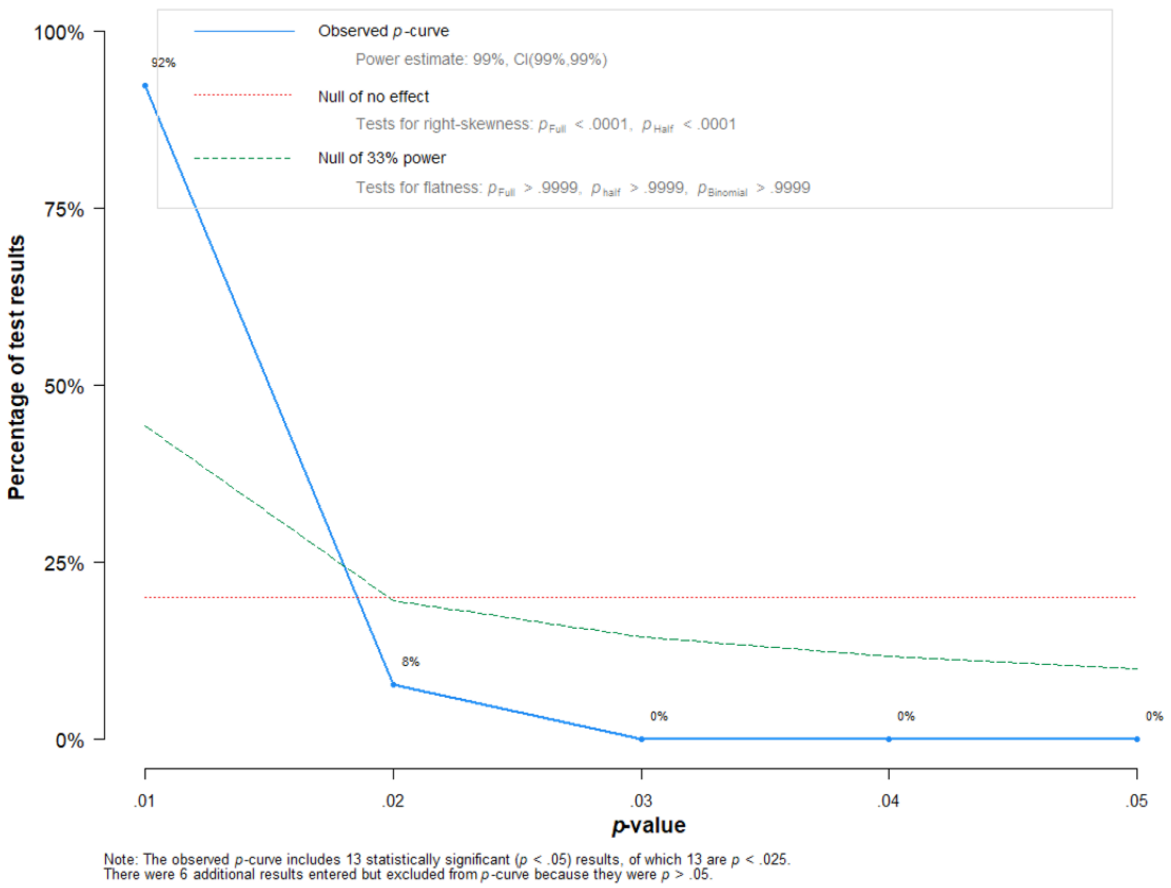

Figure 5. p-curve plot (Source: The authors' own work)

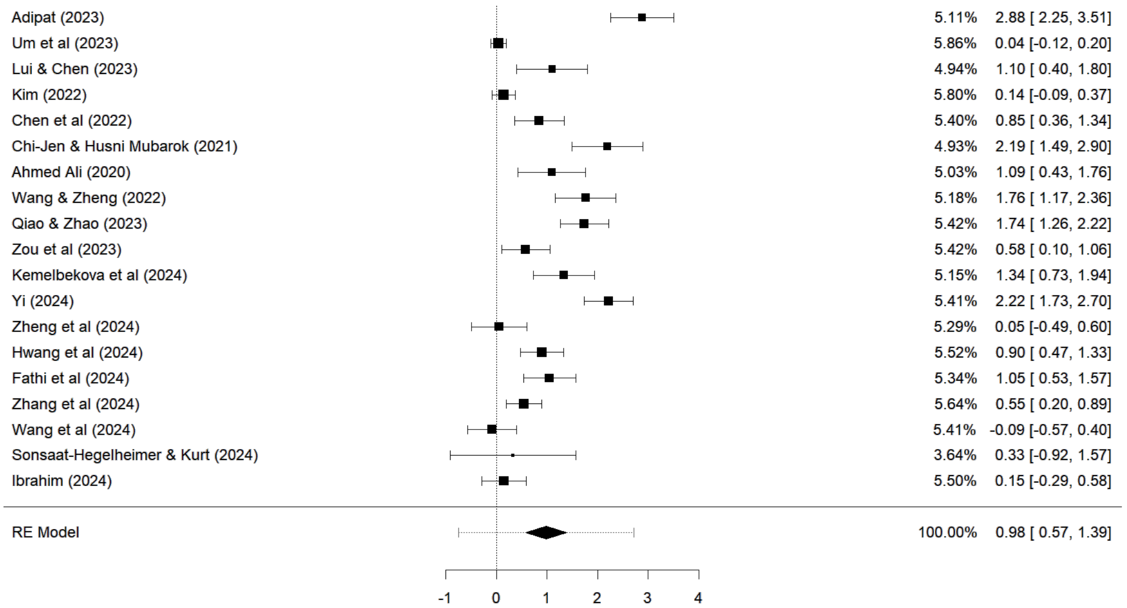

Figure 6. The overall effect of the forest map (Source: The authors' own work)

Figure 6 displays a forest plot in which the vast majority of effect sizes are positioned to the right of the line of no effect, indicating a consistent advantage for AI-enhanced instruction across the included studies. This visual pattern reinforces the quantitative findings, further substantiating the overall efficacy of AI interventions in supporting speaking and listening skill development.

To assess the presence of publication bias, multiple statistical approaches were employed, as summarized in **Table 5**.

Table 5. Publication bias assessment

Test name	Value	p
Fail-safe N	1,653.000	< .001
Begg and Mazumdar's rank correlation	0.333	0.049
Egger's regression	1.369	0.189
Trim and fill number of studies	0.000	-

Note. Fail-safe N calculation using the Rosenthal approach

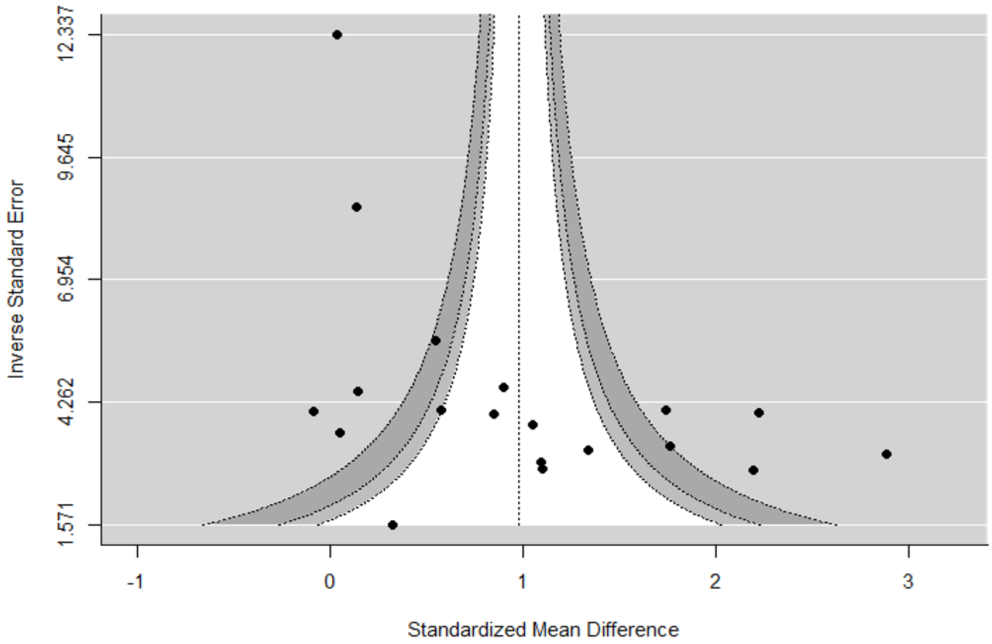

Figure 7. Funnel plot of AI in English instruction for speaking and listening skills overall (Source: The authors' own work)

Table 6. Random-effects model (k = 18)

	Estimate	SE	Z	p	95% CI	
					Lower bound	Upper bound
Intercept	1.030	0.200	5.16	< .001	0.611	1.455

Note. Tau² estimator: REML. Knapp and Hartung (2003) adjustment used.

The fail-safe N was estimated at 1,653, implying that more than 1,600 null-result studies would be required to nullify the significant overall effect–an indication of substantial robustness in the findings. Although Begg and Mazumdar's rank correlation test yielded a marginally significant result ($p = .049$), suggesting slight asymmetry in the funnel plot, Egger's regression intercept test did not support the presence of systematic bias ($p = .189$). Additionally, the trim-and-fill method identified no missing studies, lending further support to the stability and reliability of the meta-analytic estimates. Taken together, these indicators suggest that the synthesized effect sizes are unlikely to be substantially influenced by publication bias.

Figure 7 displays the funnel plot of AI in English instruction for speaking and listening skills overall.

Speaking Skills Subgroup Analysis

Eighteen of the included studies reported outcomes related explicitly to speaking skills. As detailed in **Table 6**, the pooled SMD was 1.033 (standard error [SE] = 0.200, 95% CI [0.611, 1.455], $p < .001$), indicating a large and statistically significant effect favoring AI-based instruction. This suggests that AI interventions substantially outperformed traditional pedagogical methods in enhancing learners' speaking proficiency, exceeding the magnitude of one standard deviation.

Table 7 reports high heterogeneity among these studies ($I^2 = 92.41\%$), supporting the application of a random-effects model. The forest plot presented in **Figure 8** illustrates the consistency of effect sizes across

Table 7. Heterogeneity statistics

Tau	Tau²	I²	H²	R²	df	Q	p
0.800	0.6407 (SE = 0.2481)	93.41%	13.178	-	17.000	228.795	< .001

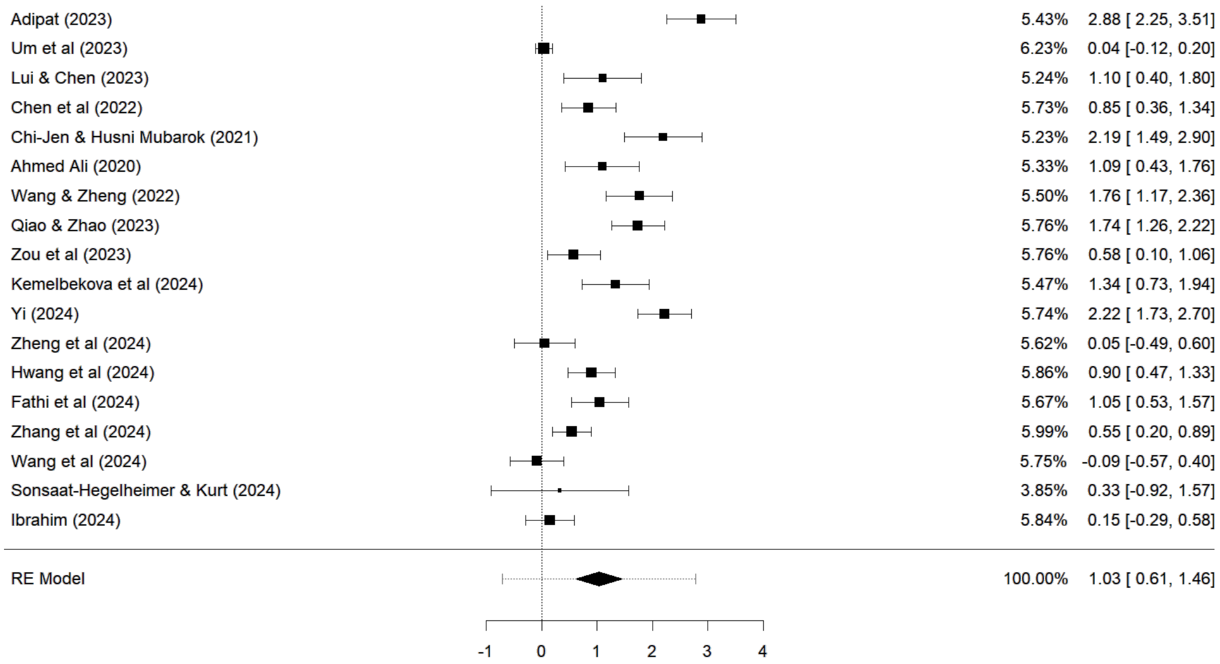

Figure 8. Forest plot of speaking skills (Source: The authors' own work)

Table 8. Model fit statistics and information criteria

	Log-likelihood	Deviance	AIC	BIC	AICc
ML	–22.144	59.154	48.289	50.070	49.089
REML	–21.386	42.772	46.772	48.438	47.629

Table 9. Publication bias assessment

Test name	Value	p
Fail-safe N	1,595.000	< .001
Begg and Mazumdar's rank correlation	0.281	0.112
Egger's regression	1.009	0.328
Trim and fill number of studies	0.000	-

Note. Fail-safe N calculation using the Rosenthal approach

studies, with most estimates falling well to the right of the line of no effect, indicating a clear and uniform advantage for AI-supported instruction.

Table 8 shows the model fit statistics and information criteria.

Regarding publication bias, **Table 9** reveals a high fail-safe N of 1,595, suggesting a substantial degree of result stability. Both Begg's rank correlation test ($p = .112$) and Egger's regression test ($p = .328$) indicate no statistically significant evidence of publication bias. The corresponding funnel plot (**Figure 9**) appears symmetrical, and the absence of imputed studies via the trim-and-fill procedure further supports the robustness of these findings.

This review included 18 studies that focused on how AI tools might help improve speaking skills. The effects reported were quite varied; some studies showed minor impacts, while others showed much stronger ones. Most (about 94%) pointed in a positive direction. After pooling the results, the average effect size was 1.033 (CI [0.6106, 1.4553], $p < .0001$), which is a substantial result favoring AI-based instruction.

Still, not all studies agreed. The heterogeneity was high ($I^2 = 92.41\%$), and the Q-statistic also showed significant variation across studies. The prediction interval went from –0.7079 to 2.7738, while AI usually helped, it did not work equally well everywhere. That is something to consider.

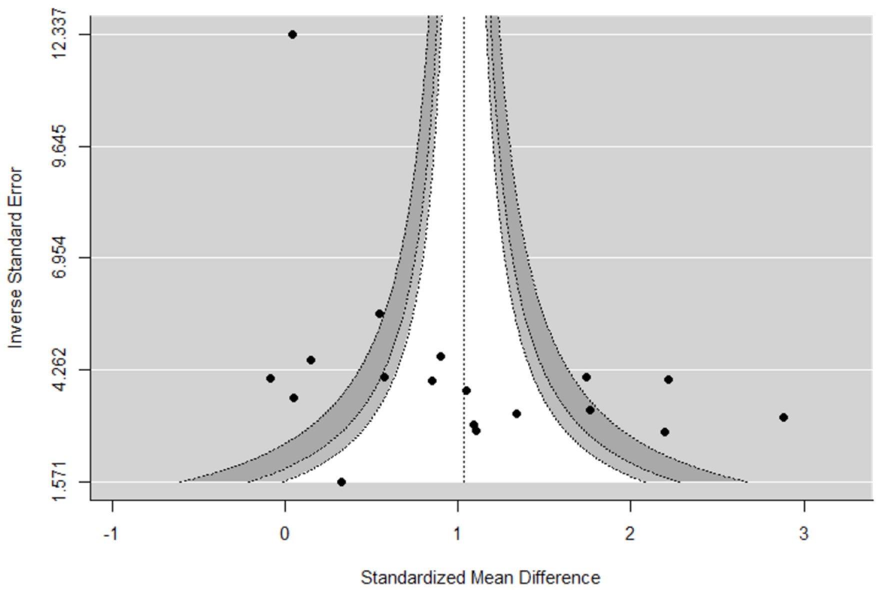

Figure 9. Funnel plot of AI in English instruction for speaking and listening skills (Source: The authors' own work)

Table 10. Random-effects model (k = 4)

	Estimate	SE	Z	p	95% CI	
					Lower bound	Upper bound
Intercept	0.714	0.409	1.75	0.179	–0.587	2.016

Note. Tau² estimator: REML. Knapp and Hartung (2003) adjustment used.

Table 11. Heterogeneity statistics

Tau	Tau²	I²	H²	R²	df	Q	p
0.783	0.6132 (SE = 0.5444)	96.30%	27.048	-	3.000	37.3689	< .001

On a positive note, the analysis did not find any extreme outliers–no studentized residuals beyond ±2.99, and Cook's distance values all looked fine. As for publication bias, the funnel plot looked pretty balanced, and both Begg's (p = 0.1124) and Egger's test (p = 0.3280) did not raise any concerns. No missing studies were added through the trim-and-fill procedure either.

When narrowing the scope to just the 18 studies that addressed speaking skills, the estimated effect size increased slightly (SMD = 1.033, 95% CI [0.611, 1.455], p < .001; see **Table 6**). This adds weight to the idea that AI-driven learning methods are particularly helpful for developing oral language skills. Although the data showed a high level of heterogeneity (I^2 = 92.41%; **Table 7**), no individual studies appeared to skew the results or behave as outliers. Additionally, no signs of publication bias were detected (**Table 9**, **Figure 8**, and **Figure 9**).

Taken together, these results point to the value of AI tools–especially those using conversational agents, real-time pronunciation feedback, or gamified speaking tasks–in helping learners improve their fluency, pronunciation, and overall confidence when speaking. This is in line with findings from recent research (Fathi et al., 2024; Zhang et al., 2024).

Listening Skills Subgroup Analysis

Table 10 summarizes the results of the random-effects model for the four studies examining listening skills. The overall effect size was estimated at 0.714 (SE = 0.409), with a 95% confidence interval ranging from –0.587 to 2.016. Although the estimate suggests a potential benefit of AI-based instruction, the effect was not statistically significant (Z = 1.75, p = .179), possibly due to limited statistical power.

Considerable heterogeneity was detected across studies, as shown in **Table 11**. The I^2 value reached 96.3%, with a tau² of 0.6132 (SE = 0.5444), indicating that much of the variability in effect sizes is likely due to true differences rather than random error. The Q statistics (Q = 37.369, df = 3, p < .001) further supports this conclusion.

Table 12. Model fit statistics and information criteria

	Log-likelihood	Deviance	AIC	BIC	AICc
ML	−4.374	15.256	12.748	11.521	24.748
REML	−3.658	7.316	11.316	9.513	23.316

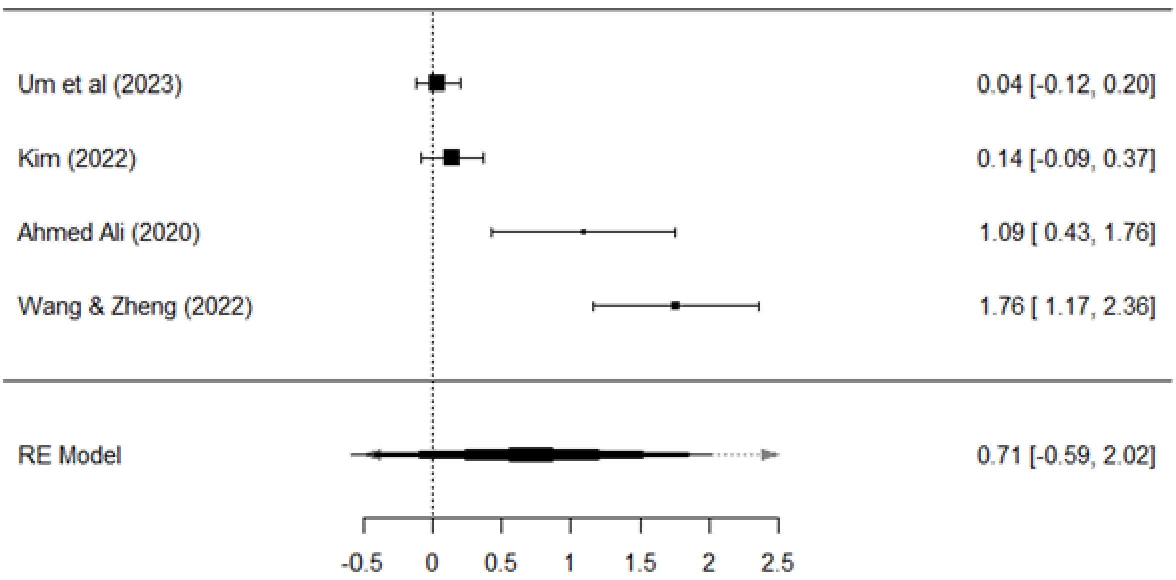

Figure 10. Forest plot of AI in English instruction for listening skills (Source: The authors' own work)

Table 13. Publication bias assessment

Test name	Value	p
Fail-safe N	39.000	< .001
Begg and Mazumdar's rank correlation	0.667	0.333
Egger's regression	3.909	0.060
Trim and fill number of studies	0.000	-

Note. Fail-safe N calculation using the Rosenthal approach

Model fit indices, presented in **Table 12**, favored the REML approach over the full ML estimation. REML yielded a lower AIC (11.316) and BIC (9.513), as well as better log-likelihood and deviance values. These results indicate that REML offers a better-fitting model for estimating the overall effect in the presence of high between-study variance.

Figure 10 shows the forest plot of AI in English instruction for listening skills.

In contrast, the subset of studies focusing on listening skills (k = 4) yielded a moderate but non-significant effect size (SMD = 0.714, 95% CI [−0.587, 2.016], p = .179; **Table 10**). Given the small number of studies and the wide confidence interval, this finding should be interpreted with caution. Notably, the level of heterogeneity was very high (I^2 = 96.3%; **Table 11**), indicating that the included studies varied substantially in terms of design, context, or intervention characteristics.

The limited evidence base restricts the ability to draw firm conclusions regarding the effectiveness of AI tools for listening comprehension. Still, the positive direction of the overall effect is worth noting. A fail-safe N of 39 (**Table 13**) suggests moderate robustness, although the possibility of small-study effects cannot be ruled out entirely.

Figure 11 displays the funnel plot of AI in English instruction for listening skills.

AI applications in listening instruction, such as adaptive comprehension modules or speech-recognition-enhanced tasks, may not yet deliver consistent outcomes across diverse learner groups. This may reflect a need for better integration into instructional design or more refined technological solutions, as pointed out in prior studies (Ahmed Ali, 2020; Kim, 2022).

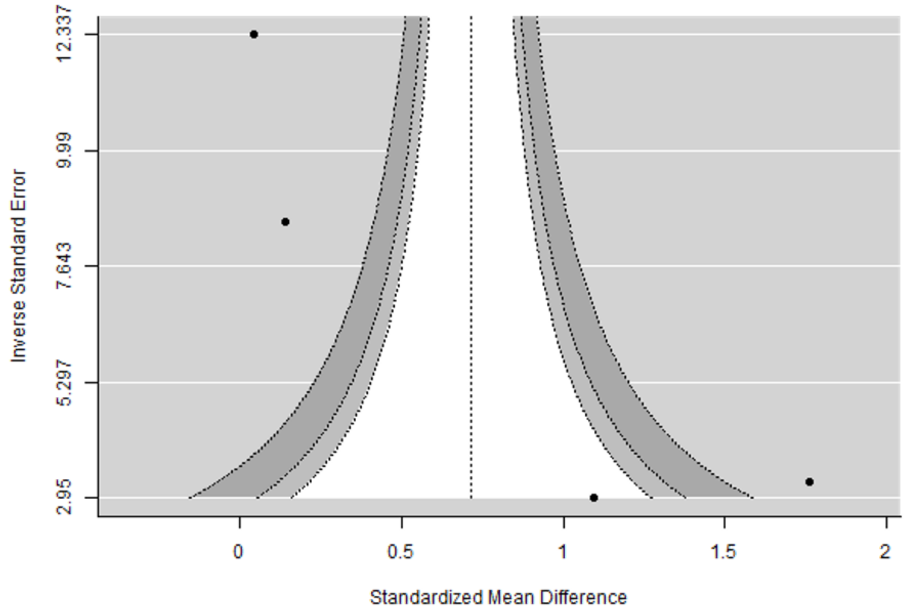

Figure 11. Funnel plot of AI in English instruction for listening skills (Source: The authors' own work)

DISCUSSION

This meta-analysis contributes to the growing body of evidence suggesting that AI can play a significant role in enhancing English language instruction, particularly in the domains of speaking and listening. The overall effect size for speaking proficiency was not only statistically significant but also quite large mirroring previous findings that emphasize the benefits of AI-supported learning environments that allow for individualized instruction, interactive activities, and feedback in real time (Ling, 2023; Qiao & Zhao, 2023).

A closer examination of the subgroup data reveals that AI has had a notably stronger impact on speaking than on listening. This difference may reflect how many AI applications designed for speaking such as chatbots, pronunciation tools, and gamified speaking platforms–actively engage learners and give them immediate, useful feedback. These tools often promote repeated practice and greater learner independence, both of which are important in developing fluency. In contrast, the evidence on listening remains limited, and the effects less consistent, likely because of variation in how listening activities are designed and a smaller number of studies focusing on this area.

The high level of heterogeneity observed across the included studies ($I^2 > 90\%$) implies that outcomes are likely shaped by contextual factors, such as learners' education level, geographic region, type and intensity of AI use, or linguistic background. These findings echo the perspectives of researchers who argue that AI's educational impact is highly context-sensitive (Lee et al., 2023; Peng & Xie, 2021). For instance, younger learners may respond more positively to game-based tools, while older students might benefit more from advanced simulation technologies. Clarifying the role of such moderators will be important for refining future instructional design and increasing transferability.

Although the direction of the effect for listening was encouraging, the lack of statistical significance and small sample base mean that we should interpret the results with caution. It is also possible that limitations in current AI technology such as difficulty understanding accents or processing real-time spoken input contributed to weaker outcomes in this domain. Methodological limitations, such as small sample sizes and a lack of experimental control, may also have played a part.

In practical terms, these results suggest that AI tools are best viewed as supplements to, rather than replacements for teacher-led instruction. Stronger outcomes are likely when AI components are part of a blended learning model that pairs self-directed use of technology with structured guidance from educators. For this to succeed, teachers will need support and training to make informed choices about when and how to use AI effectively, ensuring alignment with both learning goals and student needs.

CONCLUSION

This meta-analysis provides solid evidence that AI can support English language learning, especially in improving speaking abilities. Learners in AI-assisted environments generally performed better in speaking tasks than those in traditional settings, pointing to the potential of these tools in building oral language skills. Although the results for listening were less definitive, the overall positive direction of findings indicates that AI may still offer value in this area and deserves further exploration.

For teachers and educational leaders, these findings suggest that AI technologies–when chosen carefully and used purposefully–can serve as valuable additions to language learning programs. Rather than relying on these tools in isolation, they should be part of broader instructional plans that align with well-established teaching practices. Continued research, especially studies that follow learners over longer periods and in a range of contexts, will help determine how best to use AI tools across different learning environments.

As AI continues to evolve and becomes more widely available, thoughtful integration into English language instruction can help support more equitable access, increase student motivation, and improve the overall quality of teaching and learning worldwide.

Author contributions: TJ: conceptualization, methodology, supervision, writing – review & editing; **TJ:** project administration, validation, writing – review & editing; **KJ:** investigation, data curation, formal analysis, writing – original draft; **WP:** investigation, visualization, resources; **RP:** software, data curation, visualization; **PW & PN:** supervision, resources, formal analysis. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: The authors declared that no ethics committee approval was required since the study is a review of existing literature.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Adipat, S. (2023). An artificial intelligence-enhanced phenomenon-based learning approach for interdisciplinary understanding and speaking skills. *International Journal of Instruction, 16*(3), 531–550. https://doi.org/10.29333/iji.2023.16329a

Agnes, D., & Srinivasan, R. (2024). Fostering vocabulary memorization: Exploring the impact of AI-generated mnemonic keywords on vocabulary learning through Anki flashcards. *World Journal of English Language, 14*(2), 434–445. https://doi.org/10.5430/wjel.v14n2p434

Ahmed Ali, S. G. (2020). Using an artificial intelligence application for developing primary school pupils' oral language skills. *Journal of Education Sohag UNV, 75*(75), 67–110. https://doi.org/10.21608/edusohag.2020.97643

Alsowat, H. H. (2020). Evidence-based practices of English language teaching: A meta-analysis of meta-analyses. *English Language Teaching, 13*(11), 75–94. https://doi.org/10.5539/elt.v13n11p75

Amin, M. Y. M. (2023). AI and Chat GPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies, 4*(4), 554–567. https://doi.org/10.33422/ijhep.v4i4.554

Bao, R., & Sile, L. (2024). Empowering English vocabulary instruction through generative artificial intelligence: The AEFC method. *Journal of Educational Research Progress, 4*(4). https://doi.org/10.55375/joerp.2024.4.4

Caldwell, A. R. (2022). SimplyAgree: An R package and jamovi module for simplifying agreement and reliability analyses. *Journal of Open Source Software, 7*(71), Article 4148. https://doi.org/10.21105/joss.04148

Chang, H. (2023). The effect of AI chatbot-based tourism English instruction on intercultural communicative competence. *STEM Journal, 24*(2), 15–32. https://doi.org/10.16875/stem.2023.24.2.15

Chen, C. H., Koong, C. S., & Liao, C. (2022). Influences of integrating dynamic assessment into a speech recognition learning design to support students' English speaking skills, learning anxiety, and cognitive load. *Educational Technology & Society, 25*(1), 1–14.

Choe, S., Lee, K. A., & So, Y. (2020). The effects of phonemic awareness instructions on L2 listening comprehension: A meta-analysis. *The Journal of AsiaTEFL, 17*(4), 1294–1316. https://doi.org/10.18823/asiatefl.2020.17.4.9.1294

Chu, J., & Szlagor, M. (2023). Multi-intelligence English teaching model based on distance and open education. *International Journal of Web-Based Learning and Teaching Technologies, 18*(1), 1–15. https://doi.org/10.4018/ijwltt.325617

Daweli, T. W., & Mahoub, R. A. M. (2024). Exploring EFL learners' perspectives on using AI tools and their impacts in reading instruction: An exploratory study. *Arab World English Journal, 10*(11). https://doi.org/10.24093/awej/call10.11

Duisenova, M. M., & Zhorabekova, A. N. (2023). The effectiveness of gamification and artificial intelligence in increasing the motivation and effectiveness of students in learning English in elementary school. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(10), Article em2330. https://doi.org/10.29333/ejmste/13670

Eyding, D., Lelgemann, M., Grouven, U., Härter, M., Kromp, M., Kaiser, T., Kerekes, M. F., Gerken, M., & Wieseler, B. (2010). Reboxetine for acute treatment of major depression: Systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ, 341*, Article c4737. https://doi.org/10.1136/bmj.c4737

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

Feng, T., Deng, J., Liu, Y., Gao, Y., Peng, L., & Fu, J. (2023). A study on application of artificial intelligence in college English teaching management and evaluation. In *Proceedings of the 2023 International Conference on Emerging Information Technology for Smart Applications* (pp. 128–132). ACM. https://doi.org/10.1145/3641343.3641447

Ferrer, R. L. (1998). Graphical methods for detecting bias in meta-analysis. *Family Medicine, 30*, 579–583.

Hao, T., Wang, Z., & Ardasheva, Y. (2021). Technology-assisted vocabulary learning for EFL learners: A meta-analysis. *Journal of Research on Educational Effectiveness, 14*(4), 834–862. https://doi.org/10.1080/19345747.2021.1917028

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Ho, P. X. P. (2024). Using ChatGPT in English language learning: A study on I.T. students' attitudes, habits, and perceptions. *International Journal of TESOL & Education, 4*(1). https://doi.org/10.54855/ijte.24414

Hwang, W. Y., Luthfi, M. I., & Liu, Y. F. (2024). AI-enhanced video drama-making for improving writing and speaking skills of students learning English as a foreign language. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2024.2437655

Ibrahim, K. A. A. A., Kassem, M. A. M., & Lami, D. (2024). Intelligent computer-assisted language assessment (ICALA) in philosophy-based language instruction: Unraveling the effects on critical thinking, self-evaluation, academic resilience, and speaking development. *Language Testing in Asia, 14*(1), Article 46. https://doi.org/10.1186/s40468-024-00320-1

Jerusha, A., & Rajakumari, R. (2024). Longitudinal analysis of student progress in AI-integrated English language courses. In *Proceedings of the 2024 International Conference on Knowledge Engineering and Communication Systems* (pp. 1–6). IEEE. https://doi.org/10.1109/ickecs61492.2024.10616489

John, E., & Yunus, M. M. (2021). A systematic review of social media integration to teach speaking. *Sustainability, 13*(16), Article 9047. https://doi.org/10.3390/su13169047

Kemelbekova, Z., Degtyareva, X., Yessenaman, S., Ismailova, D., & Seidaliyeva, G. (2024). AI in teaching English as a foreign language: Effectiveness and prospects in Kazakh higher education. *XLinguae, 17*(1), 69–83. https://doi.org/10.18355/XL.2024.17.01.05

Kim, N. Y. (2022). AI-integrated mobile-assisted language learning: Is it an effective way of preparing for the TOEIC test in classroom environments? *English Teaching, 77*(3), 79–102. https://doi.org/10.15858/engtea.77.3.202209.79

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Konyrova, L. (2024). The evolution of language learning: Exploring AI's impact on teaching English as a second language. *Eurasian Science Review: An International Peer-Reviewed Multidisciplinary Journal, 2*(2), 133–138. https://doi.org/10.63034/esr-42

Lee, J. H., Lee, H., & Lo, Y. Y. (2023). Effects of EMI-CLIL on secondary-level students' English learning: A multilevel meta-analysis. *Studies in Second Language Learning and Teaching, 13*(3), 439–467. https://doi.org/10.14746/ssllt.38277

Lee, S., Kuo, L., Xu, Z., & Hu, X. (2020). The effects of technology-integrated classroom instruction on K-12 English language learners' literacy development: A meta-analysis. *Computer Assisted Language Learning, 35*(1–2), 1–29. https://doi.org/10.1080/09588221.2020.1774612

Li, Y. (2024). A deep learning multimodal feature fusion approach for assessing English study. In *Proceedings of the 2024 4th Asia-Pacific Conference on Communications Technology and Computer Science* (pp. 0047–0052). IEEE. https://doi.org/10.1109/acctcs61748.2024.00047

Lin, C. J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society, 24*(4), 16–35.

Lin, H., & Liu, X. (2024). InVideo AI: Creating instructional audiovisual input for English learning. *RELC Journal*. https://doi.org/10.1177/00336882241267773

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

Liu, P.-L., & Chen, C.-J. (2023). Using an AI-based object detection translation application for English vocabulary learning. *Educational Technology & Society, 26*(3), 5–20. https://doi.org/10.30191/ETS.202307_26(3).0002

Madhavi, E., Sivapurapu, L., Koppula, V., Sreehari, V., & Rani, P. (2023). Developing learners' English-speaking skills using ICT and AI tools. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 32*(2), 142–153. https://doi.org/10.37934/araset.32.2.142153

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007

Murdoch, Y. D., & Lin, Y. (2023). Factors mediating the link between engagement and satisfaction among online English-mediated instruction learners during COVID-19. *Electronic Journal of e-Learning, 21*(3). https://doi.org/10.34190/ejel.21.3.3031

Ni, A., Cheung, A., & Shi, J. (2022). Effects of educational technology on reading achievement for Chinese K-12 English second language learners: A meta-analysis. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1025761

Peng, J., & Xie, X. (2021). English-medium instruction as a pedagogical strategy for the sustainable development of EFL learners in the Chinese context: A meta-analysis of its effectiveness. *Sustainability, 13*(10), Article 5637. https://doi.org/10.3390/su13105637

Qian, M., & Kong, C. (2024). Exploring the gap: The challenge of achieving human-like generalization for concept-based translation instruction using large language models. In *Proceedings of the AAAI Spring Symposia*. AAAI. https://doi.org/10.1609/aaaiss.v3i1.31283

Qiao, H., & 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*. https://doi.org/10.3389/fpsyg.2023.1255594

Rahmati, J., Izadpanah, S., & Shahnavaz, A. (2021). A meta-analysis on educational technology in English language teaching. *Language Testing in Asia, 11*(1), Article 37. https://doi.org/10.1186/s40468-021-00121-w

Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Sharifi, M., Seyedagha Sharifi, M. B., Rostami AbuSaeedi, A., Jafarigohar, M., & Zandi, B. (2018). Retrospect and prospect of computer assisted English language learning: A meta-analysis of the empirical literature. *Computer Assisted Language Learning, 31*(4), 413–436. https://doi.org/10.1080/09588221.2017.1412325

Song, F., Khan, K. S., Dinnes, J., & Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International journal of epidemiology, 31*(1), 88–95. https://doi.org/10.1093/ije/31.1.88

Sonsaat-Hegelheimer, S., & Kurt, Ş. (2024). The impact of generative AI-powered chatbots on L2 comprehensibility. *Journal of Second Language Pronunciation, 10*(3), 339–374. https://doi.org/10.1075/jslp.24053.son

Tang, J. L., & Liu, J. L. (2000). Misleading funnel plot for bias detection in meta-analysis. *Journal of Clinical Epidemiology, 53*(5), 477–484. http://doi.org/10.1016/s0895-4356(99)00204-8

Um, H., Kim, H., Choi, D., & Oh, H. (2023). An AI-based English education platform during the COVID-19 pandemic. *Universal Access in the Information Society, 23*, 1233–1248. https://doi.org/10.1007/s10209-023-01046-2

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024a). 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*, Article 103533. https://doi.org/10.1016/j.system.2024.103533

Wang, X., Zhang, Q., Chen, H., Neitzel, A. J., & Davis, M. H. (2024b). The effects of language learning strategy instruction on college students' English achievement and learner autonomy in mainland China: A meta-analysis. *International Journal of Educational Research, 124*, Article 102442. https://doi.org/10.1016/j.ijer.2024.102442

Wang, Y., & Zheng, L. (2022). A novel deep framework for English communication based on educational psychology perspective. *Frontiers in Public Health, 10*. https://doi.org/10.3389/fpubh.2022.916101

Wu, Y. (2024). Study on the impact of utilizing ChatGPT and other AI tools for feedback in EAP writing classrooms on the discursive writing performance of English major students. *Transactions on Social Science, Education and Humanities Research, 4*, 143–150. https://doi.org/10.62051/4se95x52

Xatamova, N. (2024). The future of legal English learning: Integrating AI into ESP education. *SPAST Reports, 1*(7). https://doi.org/10.52773/tsuull.conf.teach.foreign.lang.2024.8.5/pnxs6025

Yang, T. (2024). Impact of artificial intelligence software on English learning motivation and achievement. *SHS Web of Conferences, 193*, Article 02011. https://doi.org/10.1051/shsconf/202419302011

Yaseen, M. G. (2023). Exploring the evolution of AI integration in English as a foreign language education: A Scopus-based bibliometric analysis (1997-2023). *Mesopotamian Journal of Computer Science, 2*(2), 19–31. https://doi.org/10.58496/mjcsc/2023/019

Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System, 121*, Article 103259. https://doi.org/10.1016/j.system.2024.103259

Zhao, Y. (2013). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal, 21*(1), 7–27. https://doi.org/10.1558/cj.v21i1.7-27

Zheng, C., Chen, X., Zhang, H., & Chai, C.S. (2024). Automated versus peer assessment: Effects on learners' English public speaking. *Language Learning & Technology, 28*(2), 210–228.

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability, 15*(4), Article 2872. https://doi.org/10.3390/su15042872