ISSN 2616-5783 Vol.9, Issue 1: 36-43, DOI: 10.25236/AJHSS.2026.090106

Does Integrating Artificial Intelligence into EFL Instruction Improve Learners' Listening Proficiency? Evidence from a Meta-analysis of Quasi-experimental Studies

Feng Yue[1,a,*]

[1]*School of Foreign Languages and Cultures, Geely University of China, Chengdu, China*
[a]*fengy_guc@163.com*
[*]*Corresponding author*

Abstract: *In recent years, with the vigorous development of artificial intelligence technology, an increasing number of researchers and frontline teachers have applied it to foreign language education scenarios. At present, a small number of studies have explored the impact of integrating AI into EFL teaching on learners' listening ability through quasi-experimental designs. However, these studies only reported whether there were significant differences in listening scores between the experimental class and the control class, without reporting the effect size. Thus, the extent to which the integration of AI technology affects listening proficiency remains unknown. In view of this, this study conducted a meta-analysis of 15 quasi-experimental studies. The fixed-effects model showed that the adjusted combined effect size was 0.42, which reached a statistically significant level, indicating that the integration of AI has a moderate positive impact on EFL listening teaching. This article also discusses the implications of the research results for teaching.*

Keywords: *Artificial Intelligence; EFL Learners; Listening Proficiency; Quasi-Experimental Studies*

1. Introduction

With the deep integration of Artificial Intelligence (abbreviated as AI thereafter) technology into the educational sphere, tools such as intelligent teaching systems and adaptive learning platforms are progressively reshaping the landscape of second language (L2) acquisition[1]. Within English as a Foreign Language (EFL) instruction, listening comprehension, as a core receptive skill, has long been constrained by limitations inherent in traditional classrooms[2]. These limitations include standardized materials, delayed feedback mechanisms, and insufficient adaptation to individual learner differences. In recent years, AI technology, leveraging capabilities such as speech recognition, Natural Language Processing (NLP), and big data analytics, has introduced innovative solutions for listening instruction[3]. These include generating personalized learning pathways, providing real-time error diagnosis, and enhancing input through multimodal resources. For instance, intelligent speech-to-text tools can help learners pinpoint challenging words or phrases, while adaptive algorithms can dynamically adjust material difficulty to match learner proficiency[4]. These functionalities are hypothesized to enhance the precision and efficiency of listening training.

Despite the increasing number of practical implementations of AI-assisted EFL listening instruction, existing research findings exhibit heterogeneity regarding the effect of AI integration and the magnitude of the effect sizes. Most empirical studies demonstrate that AI tools can improve listening performance[5][6], while others report no statistically significant difference compared to traditional methods[7]. This divergence may stem from variations in research design (e.g., sample size, intervention duration), the types of AI tools employed (e.g., speech recognition vs. intelligent recommendation systems), and learner characteristics (e.g., initial proficiency level, learning motivation). Furthermore, most existing studies have employed small-scale quasi-experimental designs, lacking systematic integration regarding the magnitude of effects. Consequently, employing meta-analysis to quantify the overall effect size of AI integration on EFL listening proficiency can not only provide empirical evidence to resolve current controversies but also offer crucial guidance for optimizing the application of AI technology in language pedagogy.

Based on this rationale, this study focuses on quasi-experimental research designs. Through a systematic review of the literature published between 2020 and 2025, it employs meta-analysis to address the following questions: Does the integration of AI technology into EFL listening instruction significantly enhance learners' listening proficiency? Is the resulting effect size substantial? The findings aim to provide data-driven support for educational practitioners.

2. Literature Review

The application of AI technology in EFL listening instruction can be summarized into three core models: adaptive learning systems, intelligent feedback tools, and multimodal input platforms[4]. Adaptive learning systems group learners through initial proficiency tests and dynamically deliver listening materials matched to their difficulty levels (e.g., transitioning from short dialogues to academic lectures) based on real-time response data[8]. Their core mechanism utilizes Item Response Theory (IRT) or deep learning models to achieve personalized learning paths tailored to individual needs[9]. Such systems have been proven to effectively reduce learners' cognitive load, particularly exerting a positive impact on the motivation of low- to intermediate-level learners[10]. Intelligent feedback tools focus on providing precise support during the listening process, with typical functions including: speech-to-text transcription with sentence-level translation, automatic error type annotation, and personalized re-practice of incorrect items[11]. Compared to traditional teacher feedback, AI tools offer advantages in immediacy and traceability but also carry the risk of over-reliance on technology, potentially hindering the development of metacognitive skills[12].

The improvement of listening proficiency results from the combined effects of cognitive abilities (e.g., phonological decoding, contextual inference), affective factors (e.g., anxiety levels, learning interest), and instructional environments (e.g., material authenticity, interaction frequency)[13]. AI technology exhibits differentiated impacts across these dimensions: at the cognitive level, speech recognition technology can enhance phonological perception (e.g., distinguishing between /θ/ and /ð/) through "listen-speak comparison" training[14]; at the affective level, the instant reward mechanisms of adaptive systems (e.g., virtual badges, progress visualization) can reduce listening anxiety, particularly showing significant effects for beginners[15]; at the instructional environment level, AI-generated authentic materials (e.g., adapted BBC news, edited film clips) can improve situational authenticity, addressing the latency issues of traditional textbooks[16].

Empirically, quasi-experimental designs, capable of examining intervention effects in natural instructional settings (without random grouping), have become a mainstream method in educational technology application research. In the field of AI and EFL listening instruction, common quasi-experimental designs include the Nonequivalent Control Group Design and the Interrupted Time Series Design. The former controls selection bias by matching the initial proficiency levels of experimental and control groups (e.g., college entrance English scores, pretest results)[17], while the latter captures the dynamic effects of interventions through multiple pretest-posttest data points[18]. However, inherent limitations of quasi-experimental research may affect the reliability of conclusions: first, selection bias is difficult to completely eliminate, as learners who voluntarily use AI tools may inherently possess higher learning motivation; second, most studies have short intervention durations, making it challenging to assess the long-term effects of AI technology; third, inconsistent reporting of effect sizes, with few quasi-experimental studies explicitly providing Cohen's d or Hedges's g values, hinders cross-study comparisons. These issues present methodological challenges for the meta-analysis in this study and highlight the necessity of systematically integrating existing evidence.

3. Methodology

Meta-analysis is a statistical method for systematically and quantitatively synthesizing previous research results. For the same research topic, inconsistent conclusions often arise due to factors such as research subjects, funding, various environmental influences, and researchers themselves. However, traditional descriptive literature reviews mostly describe without critique and cannot conduct quantitative comprehensive analysis of these research conclusions. The meta-analysis method makes up for this deficiency by quantitatively synthesizing multiple studies with the same research topic. Its basic process is as follows: formulating research questions, comprehensively searching relevant research literature, establishing strict inclusion and exclusion criteria, describing basic information, and conducting quantitative statistical analysis. This study employs meta-analysis to quantitatively synthesize quasi-experimental studies on the impact of integrating AI technology into EFL listening instruction on learners' listening proficiency.

3.1. Literature Retrieval

The literature retrieval covered the period from January 1, 2020, to November 30, 2025. The studies were sourced from databases such as Web of Science, ProQuest, EBSCO, LLBA, and ERIC, with search keywords including "AI/artificial intelligence" and "English/EFL listening". In addition, this study manually screened the references of the latest empirical studies and related review articles to further search for relevant literature.

3.2. Literature Inclusion Criteria

The literature inclusion criteria were as follows: 1) The literature was written in English; 2) It was a quasi-experimental study on the integration of AI into English listening teaching, with research design adopting repeated measures or between-group comparison; 3) The dependent variable included at least English listening proficiency; 4) The effect size could be calculated. Based on the above criteria, a total of 15 studies were finally included in the meta-analysis statistics, among which 14 were journal articles and 1 was a doctoral dissertation. The information on the studies is listed in Table 1.

Table 1: Information on the studies.

#	Studies	M1	M2	S1	S2	n1	n2	Subgroup
1	Abdelsalam & Mahmoud[5]	46.63	70.3	4.64	5.49	40	40	Elementary
2	Ali et al.[6]	14.5	18	2.4	1.7	25	25	Secondary
3	Alrasheedi[19]	21.78	35.14	2.04	2.07	50	50	Tertiary
4	Koç[7]	43.33	40.99	19.65	18.54	19	20	NA
5	Ghareeb[20]	10.53	30.48	1.36	3.78	20	20	Elementary
6	Kim[21]	175.91	194.84	59.92	64.69	132	190	Tertiary
7	Sahito et al.[22]	59.1	78.45	8.3	26.15	30	30	Secondary
8	Lv & Zhang[23]	67.3	77.64	9.45	11.27	30	29	Tertiary
9	Abdellatif et al.[24]	13.83	17.38	3.07	1.6	27	30	Tertiary
10	Zhou et al.[25]	18.82	21.37	1.49	1.26	32	35	Tertiary
11	Wu & Wang[26]	11.4	14.3	4.797	3.12	31	31	Tertiary
12	Jesinth[27]	9.1	15.6	1.11	2.6	30	30	Tertiary
13	Dorgham[28]	15	33.69	3.31	2.69	34	33	Tertiary
14	Jiang et al.[29]	67.73	85.73	9.21	8.932	50	50	Tertiary
15	Loebis[30]	65.7	78.5	5.2	4.8	30	30	Tertiary

Notes: 1, control class; 2, experimental class; M, mean; S, standard deviation; n, sample size.

3.3. Meta-analysis Process

3.3.1. Effect Size Calculation

Meta-Essentials was used for data analysis. Since the included literature in this study contained small-sample studies ($n < 60$), Hedges's g was selected as the effect size indicator. All studies only reported the significance of independent samples t-test results, but not the effect size. Therefore, this study first calculated the effect size of each study based on the mean and standard deviation of the post-test scores of the experimental group and the control group provided in these studies. Referring to Cohen's effect size measurement standards: ES < 0.2 was considered a small effect size; 0.2 < ES < 0.8 was a medium effect size; and ES > 0.8 was a large effect size[31].

3.3.2. Model Selection

In the results of the heterogeneity test, if $I^2 \geq 50\%$ and the Q value was statistically significant ($p < 0.05$) for the studies on the impact of AI integration into listening teaching on listening performance, indicating heterogeneity among studies, a random-effects model was adopted; otherwise, a fixed-effects model was used[32].

ISSN 2616-5783 Vol.9, Issue 1: 36-43, DOI: 10.25236/AJHSS.2026.090106

4. Results and Discussion

4.1. Effect of Integrating AI into EFL Listening Instruction on Learners' Listening Proficiency

According to the statistical principles of meta-analysis, only data with good homogeneity can be combined. Therefore, it is necessary to conduct a heterogeneity test on the results of multiple studies to select an appropriate effect model based on the heterogeneity analysis results. The results of the sample heterogeneity test in this study showed that Q=435.05, p<0.001, and I^2=96.78, indicating significant heterogeneity among the samples. Therefore, a random-effects model should be used for analysis.

Table 2 and Figure 1 show the effect size (Hedges's g), confidence intervals (represented by black lines), and prediction intervals (represented by green lines) of each study. It can be seen that the effect sizes of each study vary to different degrees. In addition, the results of Koç's study showed that the confidence interval included 0, indicating no significant difference in scores between the experimental group and the control group[7]; all other studies showed positive effect values with confidence intervals that did not include 0, indicating that the integration of AI into EFL teaching had a significant positive impact on listening proficiency. As can be seen from the random-effects model in Table 3, the combined effect size of all studies was 2.58, reaching a statistically significant level (p<0.001), which indicates that the integration of AI into teaching has a large and positive impact on students' listening proficiency.

Table 2: Effect size of each study.

#	Studies	Hedges's *g*	CI Lower Limit	CI Upper Limit	Weight
1	Abdelsalam & Mahmoud	4.65	3.84	5.55	6.59%
2	Ali et al.	1.68	1.05	2.36	6.79%
3	Alrasheedi	6.48	5.54	7.52	6.43%
4	Koç	0.12	-0.51	0.76	6.81%
5	Ghareeb	7.07	5.49	8.93	5.47%
6	Kim	0.30	0.08	0.52	7.05%
7	Sahito et al.	0.99	0.46	1.54	6.88%
8	Lv & Zhang	1.00	0.47	1.56	6.87%
9	Abdellatif et al.	1.47	0.90	2.08	6.84%
10	Zhou et al.	1.85	1.29	2.45	6.85%
11	Wu & Wang	0.72	0.21	1.25	6.89%
12	Jesinth	3.26	2.52	4.09	6.66%
13	Dorgham	6.21	5.11	7.46	6.21%
14	Zhai et al.	1.98	1.51	2.48	6.92%
15	Loebis	2.57	1.91	3.30	6.75%

Notes: CI, confidence interval.

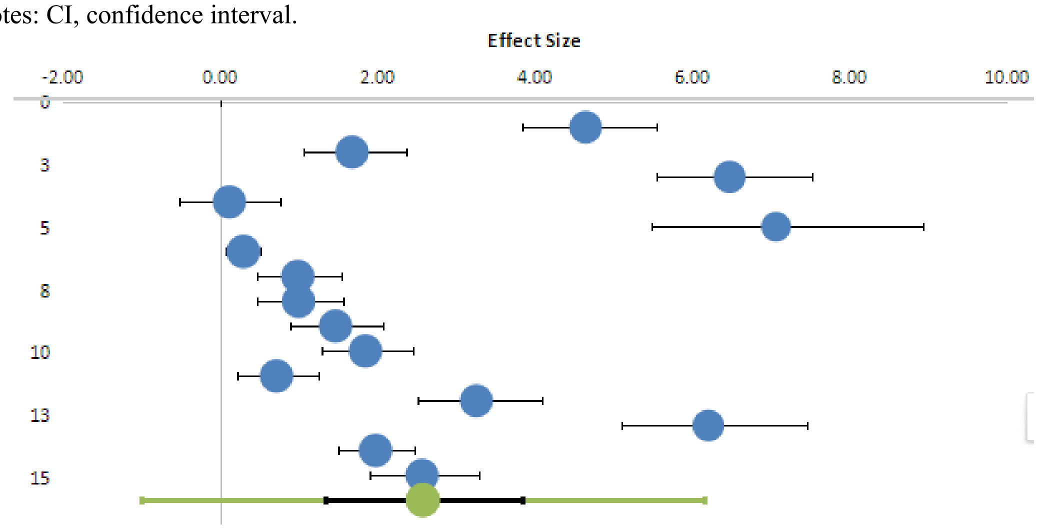

Figure 1: The forest plot regarding the effect size of each study.

ISSN 2616-5783 Vol.9, Issue 1: 36-43, DOI: 10.25236/AJHSS.2026.090106

Table 3: Combined effect size of the studies.

Models	Hedges's *g*	SE	CI LL	CI UL	Z-value	P-value
Fixed Effect Model	1.31	0.07	1.16	1.46	18.87	<0.001
Random Effects Model	2.58	0.58	1.33	3.83		<0.001

Notes: SE, standard error; LL, lower limit; UL, upper Limit.

However, to ensure the accuracy of the results, it is necessary to analyze whether there is publication bias in the research findings. Bias, also known as systematic error, refers to the deviation between the research results or inferred values and the true values. In the field of social science research, reporting bias is widespread. Only by correctly evaluating the degree of reporting bias can its impact on meta-analysis results be minimized, so evaluating reporting bias is indispensable. Due to the small sample size of this study, qualitative funnel plots and quantitative Begg's test were used to detect publication bias. The funnel plot is characterized by its allowing researchers to visually judge whether there is bias in the research results, but visual judgment alone may lead to differences. The Begg rank correlation method (referred to as Begg's test) is a quantitative method to identify bias using rank correlation test, which is also suitable for small-sample studies. If Z > 1.96 and $p < 0.05$, bias exists; if Z < 1.96 and $p > 0.05$, no bias exists.

As can be seen from Figure 2, the points on the funnel plot are not symmetrically scattered around the combined effect size of 2.58, initially indicating the possible existence of publication bias. The results of Begg's test showed Z = 3.41 > 1.96 and $p = 0.001 < 0.05$, indicating the presence of publication bias. This study used the Trim-and-fill method to examine whether publication bias affects the research results. After adjustment, the combined effect size of AI integration into teaching on listening proficiency was 0.42 (a small effect size), with a confidence interval of [-0.49, 1.79]. The adjusted effect size confidence interval crosses zero, indicating that there is no publication bias in the studies on the impact of AI integration into teaching on listening proficiency.

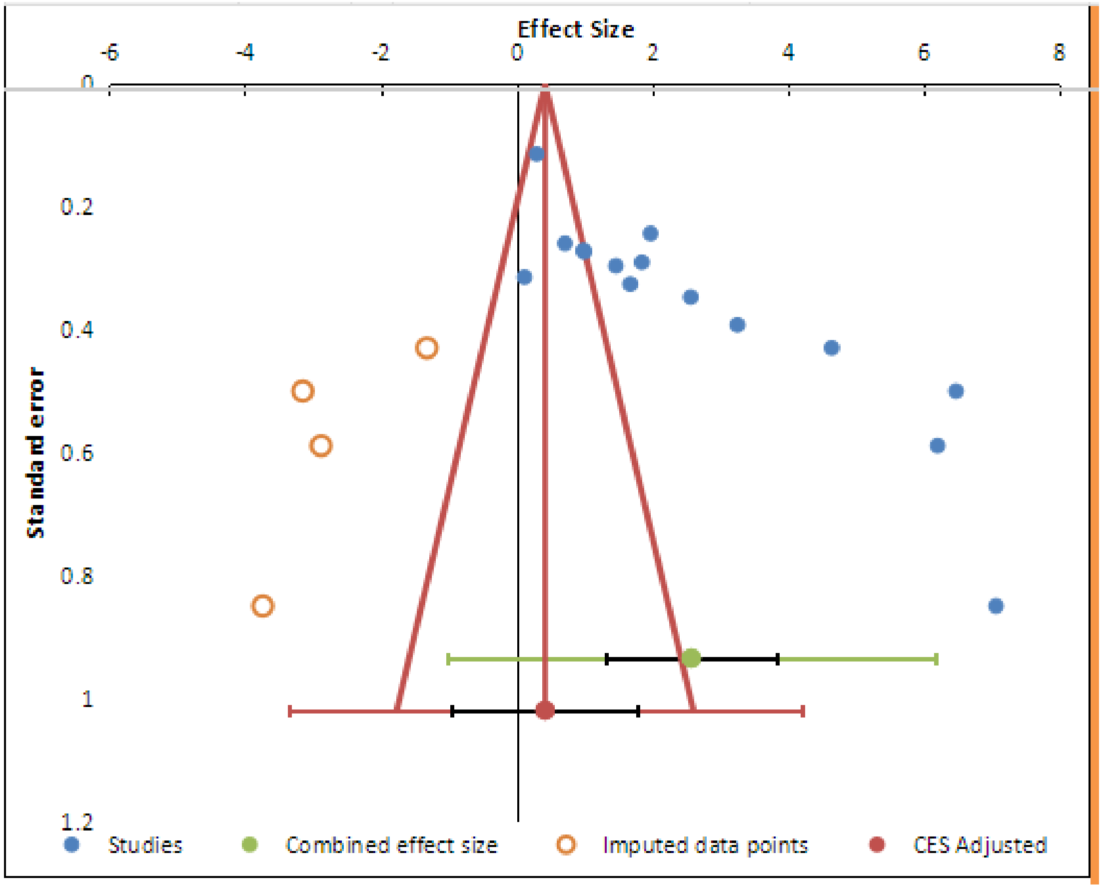

Figure 2: The funnel plot (after adjustment).

4.2. Pedagogical Implications Drawn from the Results

Meta-analysis results show that integrating artificial intelligence into EFL listening instruction exerts a moderately positive effect on learners' listening proficiency—findings that offer multi-faceted guidance for instructional practice.

First, the adaptive affordances of AI markedly boost teaching efficiency[33]. By mining learner data, intelligent algorithms identify individual weaknesses—phoneme discrimination, speech-rate adaptation or lexical blind-spots—and dynamically calibrate both difficulty and content of listening tasks. Students who struggle with connected speech receive targeted shadowing drills; those with limited vocabularies are served level-graded clips auto-captioned by the system. Second, real-time feedback mechanisms fill gaps inherent in conventional classrooms[34]. Speech-to-text engines flag mispronunciations within milliseconds; NLP-driven interactive quizzes probe literal and inferential comprehension, allowing teachers to pinpoint class-wide bottlenecks at once. Third, immersive environments created by state-of-the-art technologies like virtual reality scenarios or voice-enabled smart agents heighten input authenticity[35], bridging the notorious "classroom English"–real-world divide—an especially critical gain for learners who lack naturalistic L2 settings outside school. When learners immerse themselves in authentic language contexts, their listening proficiency is highly likely to get improved.

Nevertheless, technical integration must remain balanced. Analytic dashboards should inform rather than replace teacher judgment; instructors still orchestrate lesson design, using machine-generated profiles to fine-tune group tasks or discussion prompts. Future research should examine how best to blend AI with pedagogical approaches such as task-based language teaching and investigate whether affect-sensitive computing can mitigate listening anxiety during autonomous practice sessions.

5. Conclusions

This study conducts a meta-analysis of fifteen quasi-experimental studies examining the impact of AI integration on EFL listening proficiency—the first systematic synthesis of an under-researched domain. While its findings provide preliminary guidance for classroom practice, they remain inherently provisional: given that only a handful of scholars have addressed this specific topic (with even fewer publishing in high-impact journals), the external validity of pooled effects must await validation from larger primary datasets. Consequently, subgroup comparisons and moderator modeling—analyses requiring more primary studies than currently available—were omitted from the present investigation. Future meta-analyses are encouraged to expand the evidence base, enabling rigorous examination of potential mediating variables such as learner aptitude features (e.g., working memory capacity) and contextual factors (e.g., task type).

Acknowledgements

The author gratefully acknowledges Geely University of China for funding the 2025 key project *Research on the Application of AI-Driven Human-Machine Collaborative Teaching Model in English Listening Instruction* (Grant No.: 2025SGXYJG170) under the "Three Campuses" Program.

References

[1] Wang Z, Chen Z, Li L, et al. Between support and substitution: the impact of artificial intelligence use on anxiety and learning performance among Spanish majors in China[J]. Frontiers in Psychology, 2025, 16: 1710445.

[2] Khan M O, Khan S. Influence of online versus traditional learning on EFL listening skills: A blended mode classroom perspective[J]. Heliyon, 2024, 10(7): e28510.

[3] Ma Y, Tang X J, Huang X. AI-powered adaptive English language learning systems: Leveraging deep learning algorithms and natural language processing for personalized teaching approaches[J]. IEEE Access, 2025, 13: 153189-153198.

[4] Zhang Z. Enhancing English listening comprehension via AI-based adaptive learning platforms incorporating speech-to-text and predictive analytics[J]. Systems and Soft Computing, 2025, 7: 200418.

[5] Abdelsalam H E, Mahmoud N M. Using an artificial intelligence based program to enhance primary stage pupils' EFL listening skills[J]. Sohag Universit Journal of Education, 2021, 83: 1-32.

[6] Ali Y A, Naeem S, Alqarni S, et al. Evaluating the impact of artificial intelligence-based tools on listening comprehension among ESL learners[J]. Contemporary Journal of Social Science Review, 2025, 3(3): 514-522.

[7] Koç F Ş. The development of listening and speaking skills in EFL via an artificially intelligent chatbot

ISSN 2616-5783 Vol.9, Issue 1: 36-43, DOI: 10.25236/AJHSS.2026.090106

application: A quasi-experimental design study[D]. Middle East Technical University (Turkey), 2024.
[8] Izadi M, Izadi M, Heidari F. The potential of an adaptive computerized dynamic assessment tutor in diagnosing and assessing learners' listening comprehension[J]. Education and Information Technologies, 2024, 29(3): 3637-3661.
[9] Baylari A, Montazer G A. Design a personalized e-learning system based on item response theory and artificial neural network approach[J]. Expert Systems with Applications, 2009, 36(4): 8013-8021.
[10] Meng N, Mat Deli M, Abdul Rauf U A. Educational Technology and AI: Bridging Cognitive Load and Learner Engagement for Effective Learning[J]. Sage Open, 2025, 15(4): 21582440251395930.
[11] Huang M, Mahmud M M, Shuang G C. Impact of real time feedback features on listening accuracy of Chinese university English learners via iSmart teaching platforms[J]. Journal of Advanced Research in Social and Behavioural Sciences, 2025, 40(1): 163-177.
[12] Mohebbi A. Enabling learner independence and self-regulation in language education using AI tools: a systematic review[J]. Cogent Education, 2025, 12(1): 2433814.
[13] Yeldham M, Gao Y J. Examining whether learning outcomes are enhanced when L2 learners' cognitive styles match listening instruction methods[J]. System, 2021, 97: 102435.
[14] Sun W. The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation[J]. Frontiers in psychology, 2023, 14: 1210187.
[15] Xiao Y. The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: a randomized controlled trial[J]. Humanities and Social Sciences Communications, 2025, 12(1): 1-14.
[16] Bui H T, Filimonau V, Sezerel H. AI-thenticity: Exploring the effect of perceived authenticity of AI-generated visual content on tourist patronage intentions[J]. Journal of Destination Marketing & Management, 2024, 34: 100956.
[17] Kenny D A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design[J]. Psychological Bulletin, 1975, 82(3): 345-362.
[18] Zuckerman I H, Lee E, Wutoh A K, et al. Application of regression-discontinuity analysis in pharmaceutical health services research[J]. Health Services Research, 2006, 41(2): 550-563.
[19] Alrasheedi S. The Effect of Using AI Applications to Develop EFL Listening Comprehension Skills among University Students[J]. Conhecimento & Diversidade, 2024, 16(44): 601-637.
[20] Ghareeb Ahmed Ali S. Using an artificial intelligence application for developing primary school pupils' oral language skills[J]. Sohag University Journal of Education ,2020, 75(75): 67-110.
[21] Kim N Y. AI-integrated mobile-assisted language learning: Is it an effective way of preparing for the TOEIC test in classroom environments?[J]. English Teaching, 2022, 77(3): 79-102.
[22] Sahito J K M, Panwar A H, Ramzan I. Exploreing the impact of AI on the listening skills of ESL learners[J]. Journal of Applied Linguistics and TESOL (JALT), 2025, 8(1): 1059-1067.
[23] Lv Z, Zhang L. Investigating the effectiveness of artificial intelligence in English listening teaching: A case study of the New Concept English AI[J]. International Journal of Web-Based Learning and Teaching Technologies, 2025, 20(1): 1-19.
[24] Abdellatif M S, Alshehri M A, Alshehri H A, et al. I am all ears: listening exams with AI and its traces on foreign language learners' mindsets, self-competence, resilience, and listening improvement [J]. Language Testing in Asia, 2024, 14(1): 54.
[25] Zhou Q, Hashim H, Sulaiman N A. Integrating AI chatbots in informal digital English learning: impacts on listening competencies in Chinese higher education[J]. Education and Information Technologies, 2025: 1-29.
[26] Wu S, Wang F. Artificial intelligence-based simulation research on the flipped classroom mode of listening and speaking teaching for English majors[J]. Mobile Information Systems, 2021, 2021(1): 4344244.
[27] Jesinth F J. Effects on developing listening skill for ELF student's with artificial intelligence[J]. Journal of Information Systems Engineering & Management, 2025, 10(48s):1221-1229.
[28] Dorgham R. Using artificial intelligence tools for developing EFL students' listening skills and reducing their listening anxiety[J]. Educational and Psychological Studies, 2025, 40: 513-577.
[29] Jiang L, Wei M, Zhai Y. An empirical study of AIGC enabled high school English listening class in China[J]. Journal of Education, Humanity and Arts Research, 2025, 1(1): 1-5.
[30] Loebis I A. Effectiveness of AI-based English language learning apps in improving listening skills[J]. Lingeduca: Journal of Language and Education Studies, 2025, 4(1): 9-16.
[31] Cohen J. Statistical power analysis for the behavioral sciences[M]. New York: Academic Press, 2015.
[32] Borenstein M, Hedges L V, Higgins J P T, et al. Introduction to meta-analysis[M]. John Wiley & Sons, 2021.

ISSN 2616-5783 Vol.9, Issue 1: 36-43, DOI: 10.25236/AJHSS.2026.090106

[33] Lo A W T. The educational affordances and challenges of generative AI in Global Englishes-oriented materials development and implementation: A critical ecological perspective[J]. System, 2025, 130: 103610.
[34] Biswas U, Bhattacharya S. ML-based intelligent real-time feedback system for blended classroom [J]. Education and Information Technologies, 2024, 29(4): 3923-3951.
[35] Qiu J. Voice and Beyond: Shaping the Future of Personalized Conversational Agents[D]. OCAD University, 2025.