Exploring the effects of AI feedback on affective-motivational outcomes in language learning: A Meta-analysis

Shuo Chen [1], Zhenzhen Chen[1*]
[1]School of Humanities, Beijing University of Posts and Telecommunications
[*] czz@bupt.edu.cn

Abstract: *Despite the growing interest in the application of artificial intelligence (AI) in language learning, there is still a lack of research on the effects of AI feedback on language learning, particularly affective-motivational outcomes in language learning. Through a meta-analysis based on 14 empirical studies, this study aims to explore the impact of AI feedback on learners' affective-motivational outcomes in language learning and identify relevant moderating variables. The study reports a large and significant effect size ($g = 0.850$, $p < 0.001$), and educational level, intervention duration, as well as affective-motivational factor are the moderating factors that explain the variance of the effect size. The research findings provide insights for educators on integrating AI feedback into teaching design and also offer a reference for future directions in related research.*

Keywords: feedback, AI in education; language learning; AI feedback in language learning

1. Introduction

Feedback is an essential component of the educational process, playing a crucial role in encouraging and reinforcing learning (Hyland & Hyland, 2006). Especially in language learning, learners rely more on feedback to identify and correct language errors, adjust their learning strategies, and strengthen their grasp of language rules (Hyland & Hyland, 2006). However, in the context of evolving educational needs, the limitations of traditional teacher feedback have become increasingly apparent. In large classes, teachers face challenges in delivering in-depth, personalized and timely feedback to all students (Mohsen, 2022), which may not only lead to the failure to support students' long-term learning and development but also add substantial pressure to teachers' workloads (Ferris, 2007).

With the rapid development of natural language processing (NLP) and artificial intelligence (AI) technologies, the enormous potential of AI technologies in assisting language learning has drawn widespread attention. AI feedback can provide learners with instant, reliable, and personalized feedback, alleviating the heavy burden on teachers regarding students' academic performance while improving students' learning outcomes (Escalante et al., 2023; Rad et al., 2023), effectively addressing the limitations of traditional teacher feedback. However, despite the growing interest in AI feedback, research on its effects on language learners' affective-motivational outcomes remains limited (Wiboolyasarin et al., 2024). Therefore, this study aims to synthesize existing research on AI feedback in language learning, explore its impact on the affective-motivational development of language learners, and investigate potential variables that moderate these effects.

2. AI-powered feedback in language learning

Feedback plays a critical role in language learning. Effective feedback can narrow the gap between learners' current performance and their goals by gradually enhancing their ability to complete specific tasks (Kluger & DeNisi, 1996). AI has revolutionized the forms and functions of feedback considering its capability to offer learners more immediate and personalized guidance, overcoming the spatial and temporal limitations of traditional feedback (Huang et al., 2022; Cheng & Zhang, 2024). As the most studied AI-powered feedback systems, Automated Writing Evaluation (AWE) systems are widely used in writing instruction. Research shows AWEs can significantly enhance the efficiency of feedback and lead to significant improvements in learners' writing performance (Ngo et al., 2024). Researchers attributed the pedagogical benefits of AWE tools to their capacity to offer adaptive metalinguistic explanations, enhance learner engagement and foster self-directed learning (Barrot, 2023a).

Additionally, the recent emergence of generative AI tools, such as ChatGPT and other large language models (LLMs) has introduced novel feedback experiences for language learners. Through NLP, these GenAI tools simulate authentic conversational scenarios, offering interactive feedback (Li et al., 2024). They have played a significant role in helping learners improve their speaking skills (Yuan, 2023) and vocabulary acquisition (Zhang & Huang, 2024), enhancing the effectiveness of language learning and boosting learners' confidence. Barrot (2023b) highlighted several uses of ChatGPT as a feedback tool in writing, including giving automated grades, offering quality feedback on content, organization, clarity of purpose, etc. However, research on the application of GenAI in language education is still in its infancy, and more in-depth studies are warranted to further reveal the impact of GenAI on different facets of language learning (Wu & Yu, 2024).

3. Previous review studies of AI feedback in language learning

With the proliferation of studies of AI feedback tools in language education, quite a number of review studies have been conducted to analyze the development status and research trends. In an early critical review by Stevenson & Phakiti (2014), they found that there was only limited evidence to support the positive effects of AWE feedback on writing, and they called for more studies employing different AWE systems to explore the effects of AWE feedback on the quality of writing. Recently, Fu et al. (2022) found in their review that AWE can improve students' writing skills to some extent, though less effectively than human feedback. However, students generally hold a positive attitude toward AWE, perceiving it as useful and reporting being motivated during its use. Shi and Aryadoust (2024) reviewed 83 SSCI-indexed journal articles from 1993 to 2022 and their study lent support to the favorable outcomes of automated-written feedback (AWF). They ended their analysis by advocating for further investigations into how AWF influences motivational and affective aspects, including anxiety, motivation, and self-regulation.

Apart from studies that concentrate on AWE, there are also systematic reviews that broaden their focus by synthetically analyzing how AI-driven feedback impacts language learning. Evenddy (2024) explored the role of AI in enhancing feedback mechanisms for English language learning and concluded that AI feedback typically outperforms traditional methods in speed, accessibility, and personalization, significantly improving learning outcomes in grammar, vocabulary, and pronunciation. Nevertheless, the review also pointed out notable challenges of AI feedback, including issues of accuracy and data dependency. In a systematic review of ten empirical articles published between 2019 and 2023, Lee and Moore (2024) discovered that GenAI, with its variety of feedback formats and technological benefits, not only improves educational outcomes by making learning environments more effective and encouraging for students, but also reduces the workload of teachers.

In summary, existing review studies highlight the broad application potential of AI feedback in language education, yet there remain some notable research gaps. First, in terms of research methodology, most existing review studies are narrative reviews, lacking quantitative data-driven meta-analytic studies to comprehensively synthesize and examine the effects of AI feedback in language learning. Second, the literature is overly concentrated on AWE, with insufficient focus on emerging technologies represented by GenAI. Finally, while most studies focus on AI feedback's impacts on learning outcomes such as writing quality and grammatical accuracy, the effects of AI on learners' affective-motivational states such as anxiety, self-efficacy, and motivation, etc are under-explored.

In order to fill the gap in the research, the current meta-analysis investigates how AI feedback affects affective-motivational outcomes in language learning. Two specific research questions are intended to be addressed:

Q1: What is the overall effectiveness of AI feedback on affective-motivational outcomes in language learning?

Q2: What potential moderating variables significantly modulate the effects of AI feedback on affective-motivational outcomes in language learning?

4. Method

4.1. Study inclusion criteria

The following criteria were applied to evaluate a study's eligibility for inclusion in order to guarantee the inclusion of high-quality research that aligned with the goals of this meta-analysis:

(1)The study should examine the impact of AI feedback on affective-motivational outcomes in language learning such as motivation, self-efficacy, and anxiety, etc. (2) The study should adopt an experimental or quasi-experimental design that includes a control group without AI feedback application. (3) The study should report the quantitative data that are necessary for the calculation of effect sizes, such as means, SD, sample sizes, p-values, or F-values.

4.2. Search strategy and study selection

We conducted an electronic search for peer-reviewed journal papers published in English between Jan 2004 and Dec 2024 in the Web of Science (WOS) and ERIC databases with the following four keyword categories:

(1)AI-related keywords: AI, artificial intelligence, machine learning, deep learning, Large Language Models, LLMs, ChatGPT, chatbot, automated writing evaluation; (2)language-learning-related keywords: language learning, second language learning, foreign language learning, ESL, EFL, second language acquisition, SLA; (3)affective-motivational-factor-related keywords: willingness to communicate, WTC, anxiety, motivation, engagement, self-efficacy, self-regulation, satisfaction; (4)feedback-related keywords: feedback.

In WOS, 390 articles were retrieved, while ERIC returned 62 articles. After removing 25 duplicates, two researchers conducted an independent screening of 427 titles and abstracts. Full papers were subsequently retrieved and downloaded to assess their eligibility. Any issues or discrepancies encountered were resolved through discussion. Ultimately, 15 studies were included in the meta-analysis.

4.3. Coding scheme

Based on previous meta-analyses (Ngo et al., 2024; Zhai & Ma, 2023), we developed a coding scheme in order to investigate how various moderator variables affected the effectiveness of AI feedback: (1) educational level: university, others; (2) intervention duration: less than 10 weeks, 10 or more than 10 weeks, unspecified; (3) type of AI systems: GenAI, non-GenAI; (4) affective-motivational factor: engagement, motivation, anxiety, WTC, self-regulation, self-efficacy, enjoyment etc.; (5) feedback source: AI, AI with other sources.

Twenty percent of the studies included in the analysis were first individually coded by the two researchers as a trial. Any inconsistencies in the initial coding were then addressed and resolved. Following additional clarification of the criteria, the remaining eligible studies were coded separately by both researchers. Any disagreements that arose throughout the coding process were settled through discussion.

4.4.Calculation of effect size

In calculating the effect sizes, Comprehensive Meta-Analysis 3.0 software was used, with Hedge's g as the measure for reporting effect sizes. When calculating the overall effect size, the choice between a random or a fixed effects model depends on study heterogeneity. For high-heterogeneity studies, a random-effects model is often better (Borenstein et al., 2009). Cochran's Q test and the I^2 statistic can check study homogeneity. A random-effects model is preferred when the Q-test p-value is less than 0.05, which indicates substantial differences (Borenstein et al., 2009). Higgins et al. (2003) define I^2 values of 25%, 50%, and 75% as low, moderate, and high heterogeneity respectively.

5.Results and Discussion

5.1. Overall effect

One eligible study (Rad et al., 2023) was identified as an outlier with an exceptionally large effect size (g=44.028), and was therefore removed from the analysis (Lipsey & Wilson, 2001). According to the results based on the random-effects model ($Q = 465.571$, $p < 0.001$; I2 = 97.208%), a large overall effect size of 0.850 (p<.0.001) was found for using AI feedback (Cohen, 1988), which means AI feedback significantly improves affective-motivational outcomes in language learning.

Our results support other research findings about the impact of AI on emotional aspects (Huang et al., 2022; Wu & Yu, 2024; Lee & Moore, 2024) and show that AI feedback can generally have a favorable impact on learners' affective-motivational states. This may be explained by the alignment between the technical characteristics of AI feedback and the psychological needs of learners (Li et al., 2024). By providing learners with immediate assessments and corrective suggestions immediately after producing language output, AI feedback increases learning efficiency and boosts learners' self-efficacy (Huang et al., 2024). Additionally, AI's personalized function can offer feedback that is tailored to learners' emotional needs. Compared with teacher feedback and peer feedback, AI can help students feel less pressured to perform well, create a more comfortable learning environment, and build their confidence, all of which increase students' intrinsic motivation and engagement in studying (Neo et al., 2024). Ultimately, it forms a positive driving force for the affective-motivational outcomes in language learning.

Table 1. Overall effect size of AI feedback on language learning

					95% CI		Test of null		Heterogeneity		
	k	*g*	SE	Varianc	LL	UL	*Z*	*P*	*Q*	*p*	I^2
Fixed	14	0.747	0.047	0.002	0.654	0.839	15.836	0.000	465.571	0.000	97.208
Rando	1	0.85	0.31	0.101	0.22	1.47	2.681	0.00			

SE: standard error; CI: confidence interval; LL: lower limit; UL: upper limit.

Table 2. Effect sizes of moderator variables

Moderators			95% CI				Heterogeneity		
	K	*g*	LL	UL	*Z-*	*p-*	*Q*	*df*	*p*
Educational level							5.017	1	0.025
university	8	0.311	-	0.644	1.829	0.067			
others	6	1.620	0.524	2.715	2.898	0.004			
Intervention duration							10.765	2	0.005
⩾10 weeks	9	0.914	0.580	1.249	5.356	0.000			
＜10 weeks	3	-0.288	-	0.397	-0.823	0.410			
Unspecified	2	2.282	-	5.113	1.580	0.114			
Type of AI systems							1.909	1	0.167
GenAI	3	0.320	-	0.867	1.145	0.252			
Non-GenAI	11	0.986	0.215	1.758	2.505	0.012			
Feedback source							2.061	1	0.151
AI with other source	3	0.377	0.087	0.666	2.548	2.505			
AI feedback	11	0.974	0.212	1.736	0.011	0.012			
Affective-motivational							30.379	6	0.000
Anxiety	4	-0.106	-	1.219	-0.157	0.875			
engagement	4	0.715	0.190	1.240	2.699	0.008			
motivation	3	0.783	-	1.599	1.879	0.060			
self-efficacy	5	0.084	-	1.043	0.172	0.863			
self-regulation	3	2.018	0.142	3.894	2.108	0.035			

WTC	1	2.344	1.755	2.933	1.755	2.933
enjoyment	2	0.783	0.470	1.096	4.903	0.000

CI: confidence interval; LL: lower limit; UL: upper limit.

5.2. Educational Level

Table 2 shows effect sizes of moderator variables. We divided the educational level into two levels: university and others. Results show AI feedback produced a small and statistically insignificant effect size for university students (g= 0.311, p= 0.067) and a large and significant effect size for other educational levels (g=1.620, p=0.004). The difference of effects of AI feedback between groups reached a significant level (Q=5.017, p= 0.025).

Most previous studies that investigated the effectiveness of AI feedback on language learning achievements reported a larger effect size for university students (Ngo et al., 2024; Zhai & Ma, 2023; Wu &Yu, 2024). However, according to our research, AI feedback significantly impacts the non-university group in terms of affective-motivational effects. This might be the case because the participants in this group—which includes primary students, middle school students and others—require more affective-motivational help from AI due to their relatively weak autonomous learning abilities. Furthermore, their curiosity and reaction to the technology may be more intense. As for university students who have stronger metacognitive regulation ability and mature learning strategies, their affective-motivational states are probably less influenced by external factors such as AI feedback, thus weakening its impact.

5.3. Intervention duration

The group who received the intervention for ten weeks or longer had a large and statistically significant effect size of 0.914 (p<.0.001), while the group that received it for less than ten weeks had a negative and non-significant effect size (g=-0.288, p=0.410). Two studies which didn't report specific intervention durations were classified into unspecified group and showed a very large but insignificant effect size (g=2.282, p=0.114). The effects of AI feedback differed significantly across invention durations (Q= 10.765, p=0 .005).

Our results, which show the substantial influence of a longer intervention duration on the efficacy of AI feedback, are mostly in line with the findings of earlier studies (Ngo et al., 2024; Wu, 2024). One possible explanation might be the transformation of affective-motivational factors usually requires a relatively long time for internalization, and short-term interventions are often insufficient to have a significant impact.

5.4. Type of AI systems

For the group using GenAI, the effect size was small and insignificant (g=0.320, p=0.252), while for the non-GenAI group, the effect size was large and significant (g=0.986, p=0.012). However, there was no statistically significant difference between the effects of different AI systems (Q=1.909, p=0.167).

In contrast to Lee and Moore (2024), our study demonstrated no significant effect of generative AI feedback on affective-motivational features. Conversely, non-Gen AI tools, such as AWE, have a greater impact. The following is an explanation for the reasons. Non-Gen AI feedback focuses more on specific learning problems such as grammar and vocabulary (Ngo et al., 2024; Chen & Zhang, 2024), and provides more targeted guidance that stimulates affective-motivational aspects more effectively. Gen AI's open-ended discussions as well as its technical limitations force learners to filter information, raising their cognitive load (Evenddy, 2024; Li et al., 2024).

5.5. Feedback source

Results indicate a large, significant effect size for using AI feedback alone (g=0.974, p= 0.012) and a small and insignificant effect size for AI feedback combined with other feedback sources (g=0.377, p=2.505). There was no statistically significant difference between the impacts of different feedback sources (Q=2.061, p=0.151).

The results show that using AI feedback alone is more effective in influencing the affective-motivational aspects compared to using AI in combination with other feedback sources. This finding is consistent with the previous research results of Zhai & Ma (2023) and Ngo et al. (2024), which reported larger effect sizes for students using AWE independently than using AWE with the help of teachers. With single-source AI feedback, learners may concentrate more on the AI's instructions without being distracted by multiple feedback sources, which might lower the possibility of cognitive conflict and increase feedback trust.

5.6. Affective-motivational factor

Very large and significant effect sizes were reported for self-regulation (g=2.018; p=0.035). Medium and significant effect sizes were found for language learning engagement (g=0.715, p= 0.008) and enjoyment (g=0.783, p< 0.001). The effects of AI feedback on reducing anxiety (g=-0.106, p= 0.875), improving motivation (g=0.783, p=0.060) and self-efficacy (g=0.084, p= 0.863) were not statistically significant. Only one study was concerned with improving WTC and the effect size was large but not significant (g=2.344, p=2.933). The differences between different affective-motivational outcomes were significant (Q=30.379, p<0.001).

Significant variation can be seen in the effects of AI feedback on different affective-motivational elements. It has a very large and significant effect on improving students' self-regulation ability. Mohebbi (2025) asserts that the ability to self-regulate is a reflection of students' active engagement in the learning process and is essential to fostering and promoting their autonomy. The personalized and timely feedback provided by AI can help learners become autonomous and self-regulated learners by developing their skills and confidence (Qiao & Zhao, 2023). The fairly substantial benefits on learners' engagement and enjoyment may be explained by the fact that AI feedback can give students a relaxed learning environment, which will increase their passion and interest in learning (Lee & Moore, 2024; Neo et al., 2024). There is no discernible impact on anxiety, self-efficacy, or motivation, which runs counter to earlier studies by Huang et al. (2024) and Wu & Yu (2024). Considering the complexity of the affective-motivational system in language learning, the results of AI feedback may also be affected by factors such as research design and learner ability differences (Wu, 2024). Regarding WTC, the evidence is weak because there is only one study that addresses it and additional research is still required to support pertinent conclusions.

5.7. Publication bias

In this study, publication bias was evaluated using Classic Fail-safe N, and Orwin's Fail-safe N. The Classic Fail-safe N test revealed that 927 additional studies would be required to render the effect size non-significant, which is significantly higher than the 80-threshold (5*14 + 10) suggested by Rosenthal (1991). 1032 studies were needed to lower the impact size to 0.01 according to Orwin's Fail-safe N test. These findings imply that publication bias has little bearing on the study's conclusions.

6. Conclusion

Overall, our study reveals a very positive effect of using AI-powered feedback on affective-motivational outcomes in language learning. Educational level, intervention duration and affective-motivational factor are significant moderators that explain the variance in effect sizes.

6.1. Implication

Our study has important implications for both for research and language instruction. Firstly, our study provides evidence supporting the impact of AI feedback on the emotional and motivational outcomes in the language learning process, revealing the rationality of applying AI feedback to language learning. The analysis of potential moderating variables that affect feedback effectiveness motivates educators to implement more precise and reasonable teaching designs according to the characteristics and advantages of technologies. Secondly, our study uncovers the imbalance in the distribution of research on AI feedback in certain areas. There is relatively less research on how AI feedback affects emotional and motivational aspects. In the future, more relevant empirical studies can be conducted in this regard.

6.2. Limitations

The study has some limitations. Firstly, our data retrieval was only based on two databases, and the number of studies retrieved and included in the analysis was relatively small, which may have a certain impact on the research results. Secondly, due to the insufficient information provided by the authors, the exploration of other moderating variables that may affect the effectiveness of AI feedback was inadequate. Future studies could start from a broader range of variables to provide a more detailed and comprehensive understanding.

Acknowledgements

This work was supported by the the Humanities and Social Sciences Research Projects, Ministry of Education, China, under Grant [21YJAZH010]; *2023 Key Projects of Beijing Higher Education Society.*

References

Barrot, J. S. (2023a). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning*, *36*(4), 584–607.https://doi.org/10.1080/09588221.2021.1936071

Barrot, J. S. (2023b). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing,* 57, 100745. https://doi.org/10.1016/j.asw.2023.100745

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester: Wiley. DOI: 10.1002/9780470743386

Cheng, X., & Zhang, L. J. (2024). Examining second language (L2) learners' engagement with AWE-teacher integrated feedback in a technology-empowered context. *The Asia-Pacific Education Researcher*, *33*(4), 1023–1035. https://doi.org/10.1007/s40299-024-00877-8

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Evenddy, S. S. (2024). Investigating AI's Automated Feedback in English Language Learning. *Foreign Language Instruction Probe*, *3*(1), 76-87. https://doi.org/10.54213/flip.v3i1.401

Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing*, *16*(3), 165–193. https://doi.org/10.1016/j.jslw.2007.07.003

Fu, Q. K., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Computer Assisted Language Learning*, *0*(0), 1–43. https://doi.org/10.1080/09588221.2022.2033787

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *bmj*, *327* (7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of computer assisted learning*, *38*(1), 237-257.

Huang, X., Xu, W., Li, F., & Yu, Z. (2024). A meta-analysis of effects of automated writing evaluation on anxiety, motivation, and second language writing skills. *The Asia-Pacific Education Researcher*, *33*(4), 957-976. https://doi.org/10.1007/s40299-024-00865-y

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, *39*(2), 83-101. doi:10.1017/S0261444806003399

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Lee, S.S. & Moore, R.L. (2024) Harnessing Generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning, 28*(3), 82-104. DOI: 10.24059/olj.v28i3.4593

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage

Mohebbi, A. (2025). Enabling learner independence and selfregulation in language education using AI tools: a systematic review. *Cogent Education, 12*:1, 2433814, DOI: 10.1080/2331186X.2024.2433814

Mohsen, M. (2022). Computer-Mediated Corrective Feedback to Improve L2 Writing Skills: A Meta-Analysis. *Journal of Educational Computing Research*, *0*(0), 1–24. https://doi.org/10.1177/07356331211064066

Ngo, T., Chen, H., & Lai, K. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. *Interactive learning environments*, *32*(2), 727-744. https://doi.org/10.1080/10494820.2022.2096642

*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*, 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of wordtune application. *Interactive Learning Environments*, 1-22. https://doi.org/10.1080/10494820.2023.2208170

Rosenthal, R. (1991). Meta-analysis: A review. *Psychosomatic Medicine, 53*(3), 247–271.

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*, 187–209. https://doi.org/10.1017/S0958344023000265

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, *6*, 100228. https://doi.org/https://doi.org/10.1016/j.caeai.2024.100228

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Wu, X.-Y. (2024). Artificial Intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System*, 103498. https://doi.org/10.1016/j.system.2024.103498

Xue, Y. (2024). Towards automated writing evaluation: A comprehensive review with bibliometric, scientometric, and meta-analytic approaches. In *Education and Information Technologies* (Issue March). Springer US. https://doi.org/10.1007/s10639-024-12596-0

*Yuan, Y. (2023). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments,* 1-17. https://doi.org/10.1080/10494820.2023.2282112

Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, *61*(4), 875-900. https://doi.org/10.1177/07356331221127300

Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon, 10*(3). https://doi.org/10.1016/j.heliyon.2024.e25370