AI-ENHANCED ENGLISH LANGUAGE LEARNING: EFFECTS ON PROFICIENCY AND COMMUNICATIVE COMPETENCE

Rivojidin Fazlidinovich Tadjidinov
1st-year undergraduate student majoring in Philology and Language Teaching: English,
Urgench RANCH University of Technology
https://doi.org/10.5281/zenodo.22725725

Abstract: This review examines what AI-enhanced English language learning appears to change and where the evidence remains less conclusive. Intelligent tutoring, automated feedback, conversational AI, speech technologies, and generative AI are considered as distinct pedagogical tools rather than as a single intervention. The quantitative evidence is encouraging. Xu and Wang's (2024) meta-analysis of 40 empirical studies involving 3,290 participants and 55 effect sizes reported a large overall effect (Hedges' g = 0.812). Wu and Yu's (2024) meta-analysis of 24 randomized studies likewise identified a large overall effect for AI chatbots (ES = 0.964), although effect magnitude varied considerably across outcomes and intervention conditions. Individual studies add further detail. Mizumoto et al. (2024) found a strong correspondence between ChatGPT and human judgements of linguistic accuracy (ρ = .79), while Han and Li (2024) examined the use of AI-generated feedback within teacher-mediated EFL writing instruction involving 102 students and four instructors. These findings provide a strong basis for considering AI as a means of expanding practice, feedback, revision, and individualisation. They offer less decisive evidence concerning durable acquisition, independent proficiency, transfer, and communicative competence in socially situated interaction. Drawing on Hymes (1972) and Canale and Swain (1980), the review distinguishes grammatical development from the sociolinguistic, discourse, and strategic resources involved in communication. AI-supported performance may improve substantially while the broader development of communicative competence remains only partly demonstrated.

Keywords: AI-assisted language learning; linguistic proficiency; communicative competence; intelligent tutoring; automated feedback; conversational AI; generative AI; L2 writing

Introduction

English proficiency increasingly influences access to academic study, professional communication, and international information. At the same time, the practical conditions for second-language development remain uneven. Classroom time is limited, teacher attention must be distributed across groups of learners, and students vary in proficiency, pace, prior knowledge, confidence, and opportunities for practice. Within these conditions, AI changes one particularly consequential feature of instruction: the amount and immediacy of individualised language activity that can be made available beyond the classroom. The term "AI-assisted language learning," however, covers a heterogeneous set of technologies. Intelligent computer-assisted language learning (ICALL) environments may adapt tasks and feedback to learner responses. Automated assessment systems can evaluate selected dimensions of performance. Speech technologies support pronunciation and oral practice, whereas conversational agents provide opportunities for interactional rehearsal. Generative AI extends the range further by explaining language, producing examples, reformulating texts, generating tasks, and sustaining dialogue in response to learner input. The quantitative literature provides stronger evidence for beneficial effects than some sceptical accounts imply, but the estimates require careful interpretation. Xu and Wang (2024) synthesised 40 empirical studies from ten countries involving 3,290 participants and 55 effect sizes and

reported an overall Hedges' *g* of 0.812 for AI-supported English language-learning achievement. Their moderator analysis indicated that sample size, learning phase, and students' major influenced effect magnitude. The pooled estimate therefore represents a heterogeneous literature rather than a universal measure of AI effectiveness. The present review addresses a narrower issue**:** To what extent, and through which mechanisms, does AI-enhanced English language learning influence linguistic proficiency and communicative competence? The analysis distinguishes performance on technologically assisted tasks from independent language proficiency and from communicative competence in socially situated interaction. The empirical case for AI as a means of increasing linguistic practice and feedback is comparatively strong. Evidence for durable, context-sensitive communicative competence remains more limited.

Personalisation is often identified as one of the most important advantages of AI in language education. The more consequential issue, however, concerns what the adaptation actually changes in the learning process. A system may identify recurring errors, alter task difficulty, vary prompts, and increase practice in areas of apparent weakness. Such fine-grained adjustment is difficult to sustain in a conventional classroom when one teacher must respond to many learners simultaneously. Weng and Chiu (2023), in their systematic review of 83 ICALL papers, examined instructional design through Merrill's First Principles of Instruction. They reported that demonstration and application were particularly relevant to automatic feedback, activation was associated with intelligent tutoring, and problem-centred and integration principles appeared in personalised learning environments. The result directs attention toward instructional design rather than technological sophistication alone. An adaptive system still depends on the quality of the learning sequence in which its adaptations occur. The quantitative evidence supports this more qualified interpretation. Xu and Wang (2024) reported Hedges' $g = 0.812$ across 40 empirical studies involving 3,290 participants and 55 effect sizes. Their moderator analysis showed that sample size, learning phase, and student major affected effect magnitude. The observed benefits therefore emerged under varied conditions rather than from a single standardised form of AI instruction.

Wu and Yu (2024) found an overall effect of ES = 0.964 across 24 randomized studies, alongside substantial heterogeneity indicated by $Q = 686.59$, $df = 41$, $I^2 = 94\%$, and $p < .001$. Their outcome-specific estimates included ES = 1.028 for learning performance, ES = 1.020 for motivation, ES = 1.206 for self-efficacy, ES = 1.084 for interest, and ES = 1.397 for perceived learning value. Anxiety showed an effect estimate of ES = −0.715. The breadth of these outcomes suggests that chatbot-based learning may influence both performance and learner-related variables, but the very high heterogeneity makes simple generalisation difficult.

Duration introduces another interpretive issue. Wu and Yu reported stronger effects in shorter interventions than in longer ones. Novelty may partly account for that pattern: a new chatbot can attract attention, encourage experimentation, and make repetitive activity feel more varied. Whether those motivational and behavioural advantages survive longer exposure remains uncertain.

Generative AI expands the scope of personalisation further. A learner can request an explanation at a particular level, ask for additional examples, generate vocabulary exercises, rehearse a dialogue, or seek another formulation of a sentence almost immediately. Such responsiveness has clear practical value, particularly when learners require repeated explanation or additional practice outside scheduled instruction. The pedagogical consequences depend on what learners do with that access. When students interrogate AI output, compare alternatives, revise language, and test their own hypotheses, the interaction can support active learning. When the system supplies solutions without requiring much learner processing, the same accessibility may reduce the amount of linguistic work performed by the

student. Personalisation and learner autonomy can therefore reinforce one another, but they do not necessarily do so.

AI-supported instruction is particularly well suited to linguistic features that can be identified and represented with relative precision. Grammar is the clearest example. A system can flag an error, propose a revised form, explain a rule, and generate further examples within seconds. For learners who otherwise receive delayed or limited feedback, the resulting increase in feedback density can be educationally valuable. The learning consequence, however, depends on what follows the correction. The learner must notice the discrepancy, understand the explanation, retrieve the relevant form again, and eventually use it without having the answer supplied. An AI-generated correction can support these processes, but the correction itself does not demonstrate that they have occurred.

Vocabulary involves a similar distinction. AI systems can provide definitions, examples, paraphrases, collocations, and retrieval activities quickly and can generate materials that appear tailored to individual needs. These functions can increase exposure and opportunities for retrieval. Vocabulary knowledge nevertheless involves more than recognition of a definition. It includes the ability to retrieve lexical items, combine them with other words, recognise register, and deploy them appropriately within discourse. Successful completion of an AI-supported vocabulary task thus provides evidence about task performance; it does not, by itself, establish durable lexical development.

Writing offers a more tractable context for examining these issues because learner texts can be stored and compared systematically. Mizumoto et al. (2024) analysed 232 writing samples from the Cambridge Learner Corpus First Certificate in English dataset and compared ChatGPT's assessment of linguistic accuracy with human evaluation and Grammarly. ChatGPT's accuracy judgements showed a strong correspondence with human ratings (ρ = .79), while Grammarly showed ρ = .69. The corresponding correlations with learners' writing scores were ρ = −.63 for ChatGPT, ρ = −.58 for the human accuracy measure, and ρ = −.55 for Grammarly. These findings establish something specific: ChatGPT performed credibly in the assessment of linguistic accuracy. They do not establish that learners improved their writing because of that assessment. The study tested correspondence among judgements rather than longitudinal acquisition, so the evidence should remain within that boundary. Han and Li (2024) examined a different instructional arrangement involving 102 Chinese university students, divided across two undergraduate classes of 47 and 55 students, together with four instructors. Students completed an argumentative essay and an expository essay. ChatGPT generated corrective feedback across 15 error categories and also produced broader rhetorical feedback; instructors then reviewed, adapted, and incorporated this material into the final feedback delivered to students. Learners revised their essays within five days. Teacher mediation is consequential in interpreting the findings. Students did not receive unreviewed model output as an autonomous instructional source. Instructors evaluated and reshaped the feedback before presenting it to learners. Han and Li therefore provide evidence for an AI-supported teacher-feedback model rather than for an instructional process in which ChatGPT independently produced the observed learning conditions. Barrot's work raises related issues. His (2023) review identifies substantial potential for ChatGPT in L2 writing while also documenting concerns about inaccurate or unintelligible output, over-reliance, and changes in the learner's role during writing. His later pedagogical treatment (Barrot, 2024) likewise situates ChatGPT within writing instruction rather than as a substitute for instruction itself. The most defensible conclusion from this body of research is fairly specific. AI expands the availability, speed, and granularity of linguistic feedback and can facilitate revision. The extent to which these benefits develop into stable independent writing ability remains less firmly established, particularly when learners are later required to perform without technological assistance.

The practical appeal of AI is especially visible in oral language learning. Learners can speak repeatedly without depending on the availability of a teacher or peer. Conversational systems create space for rehearsal, while speech-recognition technologies provide mechanisms for examining and responding to spoken production.

Lai and Lee's (2024) systematic review of conversational AI in English language teaching examined 32 studies published between 2013 and November 2023. All 32 studies were conducted in Asian EFL contexts. Quasi-experimental designs accounted for 45% of the studies, while cross-sectional designs represented 41%; mixed-method approaches appeared in half of the reviewed studies. Affective outcomes were reported in 43% of the studies and cognitive outcomes in 41%. This literature points toward beneficial effects, but it is also geographically concentrated and methodologically weighted toward designs that are better positioned to document short- and medium-term outcomes than durable acquisition.

Increased speaking practice is a plausible mechanism. Learners may produce more language, rehearse difficult structures repeatedly, and experiment with responses without the immediate social pressure associated with speaking before classmates. For hesitant EFL learners, a substantial increase in practice volume may be educationally meaningful. The interactional environment, however, differs from ordinary human conversation. Human interlocutors interrupt, hesitate, misunderstand, shift topics, infer intentions, and respond to meanings that may differ from those the speaker intended. Communication develops through repair and negotiation as much as through accurate sentence production. Conversational AI can reproduce some aspects of these exchanges, but the learner may still be operating within a system whose interactional conditions are more predictable than those of face-to-face communication.

Pronunciation presents a related case. Speech technologies can prompt repetition and provide immediate feedback on spoken forms, creating opportunities for intensive practice. Yet intelligible oral production involves more than segmental and suprasegmental accuracy. Speakers must manage turn-taking, timing, emphasis, repair, and responsiveness. Improvement in pronunciation may therefore coexist with limited ability to manage spontaneous interaction. Listening is similarly broader than signal decoding. AI systems can control input, repeat passages, change speech rate, generate comprehension questions, and provide exposure to varied topics. Human listening frequently involves reduced forms, hesitation, variable accents, interruptions, overlapping speech, and incomplete information. The learner must construct meaning as the interaction unfolds rather than process an input sequence whose conditions can be adjusted at will.

Conversational AI thus has a clear practical role in increasing the availability and density of oral practice. Whether that practice develops into durable interactional competence depends on transfer beyond the simulated environment and remains an empirical question.

The distinction between proficiency and communicative competence has a long history in applied linguistics. Hymes (1972) challenged accounts of competence that focused primarily on formal grammatical possibility and situated language knowledge within social use. Communicative ability involves knowing what can be said, what is appropriate, what is feasible, and what is meaningful in particular circumstances. This theoretical position places a useful constraint on interpretations of AI language-learning outcomes. A model may determine whether a sentence is grammatically well formed, while communicative competence raises questions about its appropriateness for a particular interlocutor, setting, relationship, and purpose.

Canale and Swain (1980) developed a framework for communicative competence in second-language teaching and testing that distinguished grammatical, sociolinguistic, and strategic competence.

Grammatical competence concerns lexical items and the rules governing morphology, syntax, sentence-grammar semantics, and phonology. Sociolinguistic competence concerns the social meaning and appropriateness of utterances and includes rules of discourse involving cohesion and coherence. Strategic competence encompasses verbal and nonverbal strategies used to compensate for communication difficulties arising from performance limitations or incomplete competence. These dimensions are interconnected rather than reducible to a simple distinction between accuracy and pragmatics. A request, for example, can be grammatically impeccable yet socially miscalibrated because it is too direct, overly formal, insufficiently deferential, or poorly suited to the relationship between interlocutors. Formal correctness alone does not determine communicative appropriateness.

AI systems can model some of these differences. A learner may request alternative formulations appropriate for a professor, a close friend, a client, or a colleague. A generative model can explain differences in register and produce role-play scenarios designed around contrasting social relationships. Such activities have obvious pedagogical value. The more difficult issue concerns how pragmatic knowledge develops. Learners encounter social expectations through interaction, observe reactions, experience misunderstandings, and adjust their linguistic choices in response. One expression may work effectively with a friend but sound inappropriate with a supervisor; another may appear polite in one setting while being evasive in another. These relationships between linguistic choice and social meaning are not fully captured by isolated examples.

Discourse competence raises a related problem. Canale and Swain's framework recognises that communication extends across connected language, where cohesion and coherence shape meaning. AI-generated texts can be highly coherent because they are produced as completed discourse. Human interaction develops incrementally. Topics shift unexpectedly, speakers interrupt one another, misunderstandings trigger repair, and relevance is negotiated during the exchange. Strategic competence is especially revealing because communication routinely requires participants to manage breakdowns. Speakers paraphrase, simplify, gesture, repeat, request clarification, or substitute another expression when a desired word is unavailable. AI can provide opportunities to practise some of these strategies. The unresolved issue is whether learners subsequently deploy them spontaneously in interactions with human interlocutors, particularly in unfamiliar situations where the social and linguistic demands are less predictable.

For this reason, improvements in linguistic accuracy should not be treated as a direct measure of communicative competence. Grammar and vocabulary remain foundational resources, but the ability to use them effectively also depends on discourse management, social interpretation, and strategic responses to interactional difficulty. The reliability problem in generative AI is more specific than the observation that models "sometimes make mistakes." One concern is the relationship between fluent presentation and perceived credibility. A model can generate an elegant linguistic explanation that is nevertheless inaccurate. Learners with limited metalinguistic knowledge may be poorly positioned to identify such errors because the language of the explanation itself conveys confidence.

Barrot (2023) documents this tension in L2 writing. ChatGPT can support revision and aspects of language form, but its output may also be inaccurate, unclear, or pedagogically unsuitable. Mizumoto et al. (2024) provide an important qualification: in the specific task of linguistic accuracy assessment, ChatGPT correlated strongly with human judgments (ρ = .79). The two findings concern different tasks and therefore should not be collapsed into a general judgement about the reliability of generative AI. Reliability is task-specific. Han and Li's (2024) study illustrates one pedagogically plausible response. ChatGPT generated feedback, after which teachers inspected, modified, and delivered the resulting comments to learners. The study's 102 students and four instructors were therefore situated within a

mediated instructional process rather than a fully automated one. Teacher mediation also preserves an element of pedagogical judgement that is difficult to reduce to automated feedback. Teachers decide which problems are worth addressing, how much explanation a learner can usefully process, what should be practised next, and whether a model-generated suggestion fits the instructional context.

Learner agency introduces another complication. When students ask an AI system to rewrite a paragraph, the resulting text may improve while the amount of linguistic work performed by the learner decreases. A better product can therefore conceal limited development in independent language processing. Studies concerned with proficiency should account for this possibility through assessments in which learners must perform without AI assistance. Temporal design matters as well. Wu and Yu's finding of stronger effects in shorter interventions raises the possibility that novelty and initial engagement contribute to some observed benefits. Xu and Wang's Hedges' $g = 0.812$ is substantial, yet it represents a heterogeneous body of interventions rather than a longitudinal demonstration of stable acquisition. These distinctions are central to the interpretation of AI-related gains.

The evidence base is also geographically uneven. Lai and Lee (2024) found that all 32 conversational-AI studies in their review were conducted in Asian EFL contexts. That concentration does not invalidate the findings, but it constrains the extent to which they can be generalised across educational, linguistic, and sociocultural settings. A blended pedagogical model therefore offers the most defensible position. AI can extend opportunities for practice, feedback, and rehearsal; teachers can retain responsibility for instructional judgement, contextualisation, and validation; and learners can remain active interpreters of AI output rather than treating it as an unquestionable authority.

Conclusion

The evidence reviewed here supports a differentiated account of AI-enhanced English language learning. Such findings establish that AI-supported learning can generate sizeable measurable effects. They do not, by themselves, establish that those gains persist over time or appear in the same form when learners perform independently. The individual studies sharpen this interpretation. Mizumoto et al. (2024), using 232 learner writing samples, found a strong correspondence between ChatGPT and human judgments of linguistic accuracy (ρ = .79). The result supports the viability of ChatGPT for that assessment function. It does not demonstrate longitudinal improvement in learners' writing. Han and Li (2024), working with 102 Chinese university students and four instructors, examined a different arrangement in which AI-generated feedback entered a teacher-mediated revision process. Their findings therefore support an instructional configuration in which AI complements human pedagogical judgement.

Speaking and interaction raise a parallel set of questions. Conversational AI increases access to oral practice, but Lai and Lee's (2024) review shows that the evidence base consists of 32 studies conducted in Asian EFL contexts and is dominated by quasi-experimental and cross-sectional designs. These studies provide useful evidence about the consequences of AI-mediated practice, yet they offer less direct evidence about long-term transfer to unfamiliar human interactions. The resulting research agenda is correspondingly specific. Future studies need to determine whether observed gains remain after AI support is withdrawn, whether learners transfer language knowledge to unfamiliar communicative settings, whether short-term chatbot effects diminish as novelty fades, and whether improvements in AI-supported writing remain visible under unassisted assessment. Pragmatic and interactional outcomes also require direct measurement rather than inference from general proficiency scores. The evidence does not warrant either technological optimism or broad scepticism. AI has demonstrated real instructional value in expanding opportunities for practice, accelerating feedback, supporting revision, and increasing access to individualised language activity. Its strongest role at present is complementary. The technology can

alter the conditions under which learning takes place; communicative competence, however, must ultimately be demonstrated in language use that extends beyond the technological support through which that learning was developed.

References

1. Barrot, J. S. (2024). Leveraging ChatGPT in the writing classrooms: Theoretical and practical insights. *Language Teaching Research Quarterly,*
2. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics,*
3. Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 126*, 103502.
4. Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.
5. Lai, W. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence,*
6. Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics,*
7. Weng, X., & Chiu, T. K. F. (2023). Instructional design and learning outcomes of intelligent computer assisted language learning: Systematic review in the field. *Computers and Education: Artificial Intelligence,*
8. Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology,*
9. Xu, T., & Wang, H. (2024). The effectiveness of artificial intelligence on English language learning achievement. *System,*