USING ARTIFICIAL INTELLIGENCE TOOLS IN ENGLISH LANGUAGE TEACHING: OPPORTUNITIES, CHALLENGES, AND PEDAGOGICAL IMPLICATIONS

Turdialiyeva Dildora Khurshidali kizi

Republic of Uzbekistan, Shahrisabz district, Technical school № 1

dildorastudent@gmail.com

Annotation: The rapid proliferation of artificial intelligence (AI) technologies has brought transformative shifts across numerous domains, and education—particularly English language teaching (ELT)—stands among the most profoundly affected fields. This study investigates the extent to which AI-powered tools, including automated writing evaluation systems, intelligent tutoring systems, natural language processing applications, and conversational AI platforms, contribute to language learning outcomes among tertiary-level English as a Foreign Language (EFL) learners.

Keywords: *artificial intelligence, English language teaching, EFL learners, natural language processing, autonomous learning, intelligent tutoring systems, mixed-methods research*

1. INTRODUCTION

The intersection of technology and language education has been a site of continuous pedagogical negotiation for decades. From audio-lingual laboratories of the mid-twentieth century to computer-assisted language learning (CALL) platforms of the 1980s and 1990s, each successive wave of technological innovation has prompted educators to reconsider how languages are most effectively taught and acquired. The emergence of artificial intelligence as a sophisticated, accessible, and increasingly context-sensitive technology marks what many scholars regard as a qualitative—rather than merely quantitative—***departure from prior educational technologies (Zawacki-Richter et al., 2026).***

Artificial intelligence encompasses a spectrum of computational approaches—including machine learning, deep learning, neural networks, and natural language processing—that enable machines to perform tasks conventionally requiring human cognitive abilities. Within English language teaching, AI manifests in diverse forms: grammar and writing feedback tools such as Grammarly and Quillbot, conversational agents and chatbots, automated essay scoring systems, pronunciation trainers, vocabulary acquisition platforms driven by spaced repetition algorithms, and most recently, large language models such as those underlying ChatGPT and similar applications (Warschauer & Healey, 1998; Godwin-Jones, 2022).

The present study sets out to investigate three interrelated questions. First, to what extent do AI-assisted learning practices improve EFL learners’ grammatical and writing competencies compared to conventional instruction? Second, how do learners perceive the utility, usability, and potential drawbacks of AI tools within their language learning experience? Third, what pedagogical and ethical concerns do English language instructors identify when engaging with AI-enhanced teaching environments? By addressing these questions through a carefully designed mixed-methods study, the research aims to generate both statistically reliable findings and contextually grounded insights that can inform policy and practice.

2. LITERATURE REVIEW

Scholarly engagement with AI in language education has expanded considerably since the turn of the millennium, accelerating in pace following breakthroughs in deep learning. In the domain of writing instruction, automated writing evaluation (AWE) systems have attracted sustained scholarly attention. A meta-analysis by Stevenson and Phakiti (2014) examined seventeen studies on AWE and concluded that learners using automated feedback demonstrated measurable gains in grammatical accuracy, though improvements in higher-order writing qualities such as argumentation and voice were more variable. A subsequent study by Warschauer and Grimes (2008) found that secondary EFL

students who used AWE tools over a full academic year produced texts rated significantly higher on measures of syntactic complexity and vocabulary range.

Natural language processing applications have similarly been explored as tools for vocabulary acquisition and reading comprehension. Research by Xu and Peng (2017) demonstrated that learners using adaptive vocabulary platforms retained significantly more target items at delayed post-tests than peers relying on non-adaptive digital flashcards. The emergence of conversational AI and chatbots has opened new possibilities for speaking and interaction practice. Critical perspectives have equally gained prominence, with researchers raising concerns about ecological validity of AI feedback, risks of algorithmic bias, and the intensifying challenge of academic integrity in the generative AI era (Hyland & Hyland, 2006; Chapelle & Sauro, 2017).

3. METHODS

This study employed a convergent parallel mixed-methods design, in which quantitative and qualitative data were collected simultaneously, analyzed separately, and subsequently integrated during interpretation (Creswell & Plano Clark, 2018). Quantitative participants comprised 120 undergraduate students enrolled in compulsory English language courses at a state university in Uzbekistan during the 2023–2024 academic year. Participants were randomly assigned to either an experimental group (n = 60) that incorporated designated AI tools into weekly study activities, or a control group (n = 60) that followed the standard curriculum without AI tool access.

Qualitative participants included eight English language instructors purposively selected to represent variation in teaching experience, familiarity with digital tools, and departmental roles. Writing proficiency was measured through a pre-test and post-test essay task scored using a modified version of the IELTS analytical writing rubric (Cohen’s $\kappa = 0.81$). Experimental group participants were introduced to three AI-based tools: Grammarly for grammar and writing feedback, Quillbot for paraphrasing practice, and a designated ChatGPT-based

conversational interface for interactive speaking and reading comprehension practice.

4. RESULTS

At baseline, no statistically significant difference existed between experimental and control groups in overall writing proficiency scores ($t(118) = 0.74$, $p = .46$). By the end of the fourteen-week period, the experimental group demonstrated substantially higher mean writing scores ($M = 6.31$, $SD = 0.78$) compared to the control group ($M = 5.64$, $SD = 0.82$). ANCOVA confirmed a significant between-group effect ($F(1,117) = 24.18$, $p < .001$, $\eta^2 = 0.17$), indicating a medium-to-large effect of AI tool integration on writing proficiency development.

The experimental group’s advantage was most pronounced in grammatical accuracy and lexical resource. Post-test ANCOVA for grammar revealed a significant effect ($F(1,117) = 19.43$, $p < .001$, Cohen’s $d = 0.62$). Questionnaire data revealed generally positive attitudes: approximately 78% of respondents agreed that AI tools made grammar feedback more accessible and timely, 71% reported increased awareness of recurring errors, and 66% expressed greater confidence in independent writing. However, 42% acknowledged concerns about dependency on AI feedback, and 38% expressed uncertainty about the accuracy of AI-generated explanations.

Thematic analysis of instructor interviews yielded four primary themes: (1) pedagogical affordances—instructors recognized AI tools as valuable supplements for formative assessment given large class sizes; (2) concerns about authenticity and integrity—instructors experienced difficulty distinguishing AI-assisted from student-authored work; (3) learner autonomy and dependency—a nuanced tension between fostering independence and creating reliance; and (4) the irreducible human element—instructors’ conviction that affective, relational, and culturally responsive dimensions of teaching cannot be replicated by AI systems.

5. DISCUSSION

The results contribute meaningfully to the existing literature on AI in language education. The finding that AI tool integration produced statistically significant improvements in writing proficiency and grammatical accuracy over fourteen weeks is broadly consistent with the meta-analytic evidence reported by Stevenson and Phakiti (2014). However, the pattern of results suggests that AI tools are not uniformly beneficial across all dimensions of language learning. The pronounced gains in surface-level features contrast with non-significant differences in task achievement scores, implying that higher-order qualities such as argumentative structure and rhetorical awareness require forms of pedagogical engagement that current AI tools do not adequately provide (Hyland & Hyland, 2006).

The theme of authenticity and integrity reflects one of the most pressing challenges confronting ELT in the generative AI era. Unlike earlier AI tools that corrected specific grammatical errors, current large language models can reformulate or entirely generate argumentative prose, fundamentally destabilizing conventional assessment practices. Several instructors had already introduced oral follow-up tasks, process-oriented portfolios, and restricted in-class writing assessments in response—adaptive practices that align with recommendations in the broader educational integrity literature.

6. CONCLUSION

This study has investigated the effects of AI tool integration on EFL learners’ writing and grammatical development, as well as the perceptions of both learners and instructors. The findings support a carefully qualified optimism: AI tools can meaningfully enhance surface-level writing quality and grammatical accuracy, offer scalable formative assessment capabilities, and support learner autonomy under well-designed conditions. At the same time, their limitations in addressing higher-order rhetorical skills, the risks of dependency and academic integrity violations, and the inherent limitations of machine-generated feedback counsel against uncritical adoption.

Three recommendations are advanced. First, AI tools should be integrated within structured pedagogical frameworks that scaffold their use and progressively reduce AI support as learner competence develops. Second, institutional assessment practices require redesign that prioritizes process-oriented, multimodal, and performance-based evidence of language development. Third, professional development for language instructors should address both the practical affordances of AI tools and the pedagogical and ethical dimensions of their use, equipping teachers to serve as informed guides. Future research should pursue longer-term longitudinal designs and multi-institutional comparisons to build a more robust empirical foundation for AI-informed ELT practice.

REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Chapelle, C. A., & Sauro, S. (Eds.). (2017). The handbook of technology and second language teaching and learning. Wiley-Blackwell.

Coniam, D. (2014). The linguistic accuracy of chatbot technology. Computer Assisted Language Learning, 27(5), 422–440.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. Language Learning and Technology, 26(2), 5–24.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.

Hyland, K., & Hyland, F. (2006). Interpersonal aspects of response. In Feedback in second language writing (pp. 206–224). Cambridge University Press.

Khamidova, N. (2021). Challenges and prospects of English language education in Uzbekistan. Journal of Language and Linguistic Studies, 17(2), 1138–1149.

Nakata, T. (2011). Computer-assisted second language vocabulary learning. Computer Assisted Language Learning, 24(1), 17–38.

Selwyn, N. (2016). Is technology good for education? Polity Press.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. Assessing Writing, 19, 51–65.

Warschauer, M., & Grimes, D. (2008). Audience, authorship, and artifact. Annual Review of Applied Linguistics, 27, 1–23.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. Language Teaching, 31(2), 57–71.

Xu, Q., & Peng, H. (2017). Investigating mobile-assisted oral feedback. Computer Assisted Language Learning, 30(3–4), 173–182.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on AI applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 1–27.