



AI and English Language Teaching Research Gaps: A Systematic Review

Dr. D. Kausalya¹ & Mr.M.Sabareedharan²

¹Associate Professor, Department of English, PSG College of Arts and Science, Coimbatore, Tamil Nadu

²Ph.D Research Scholar, Department of English, PSG College of Arts and Science, Coimbatore, Tamil Nadu



Open Access

Manuscript ID: BIJ-2025-OCT-077

Subject: English

Received : 29.09.2025

Accepted : 15.10.2025

Published : 31.10.2025

DOI:10.64938/bijri.v10n1.25.Oct077

Copy Right:



This work is licensed under
a Creative Commons Attribution-
ShareAlike 4.0 International License.

Abstract

Artificial Intelligence (AI) has emerged as a central driver of innovation in English Language Teaching (ELT), yet despite widespread adoption of AI-enabled tools, the field of research remains scattered and under-theorized. This systematic review synthesizes peer-reviewed literature published between 2015 and 2025 to identify the state of current research and reveal the key gaps that continue to limit knowledge development. Guided by PRISMA principles, 112 studies were initially retrieved from Scopus, Web of Science, ERIC, and Google Scholar. Following a rigorous inclusion and exclusion process, 48 studies were analyzed using thematic synthesis. The review identifies six major gaps in existing research: the lack of pedagogical integration of AI tools with established second language acquisition (SLA) frameworks, the overconcentration of research in technologically advanced nations with limited evidence from Global South contexts, the neglect of cognitive and affective variables such as cognitive load, motivation, and learner anxiety, the scarcity of longitudinal and experimental research designs, the insufficient focus on equity and accessibility in diverse learning environments, and the absence of robust investigations into ethical challenges such as plagiarism, data privacy, and algorithmic bias. The results indicate that although AI holds great promise for revolutionizing second language acquisition, the current body of research falls short in offering thorough or fair insights. The article ends with a research agenda for the future that prioritizes theory-driven, longitudinal, and contextually inclusive methods, especially in high-stakes test preparation contexts like India's IELTS.

Keywords: Artificial Intelligence, English Language Teaching, Systematic Review, Research Gaps, IELTS, SLA.

Introduction

One of the most important advances in applied linguistics over the past ten years has been the convergence of artificial intelligence (AI) and English language teaching (ELT). In L2 learning environments, AI-powered tools—from chatbot-based conversational partners to intelligent tutoring systems and automated writing evaluation (AWE) tools—have started to take center stage. In high-stakes exam preparation environments like the TOEFL and IELTS, where effectiveness, feedback quality, and learner autonomy are crucial, these technologies are being used more and more (Kohnke & Jarvis,

2021). Since the demand for international mobility has increased in India, AI-enhanced platforms have entered every sector of education, from universities to local private coaching centres, to reflect the global trend. (Kukulska-Hulme et al., 2022).

Even though there are promising advancements, the literature on AI and ELT is still inconsistent and inadequate. Most of the studies are tool-centred, focused on the learners' feedback or satisfaction, with very little regard for how these tools are related to SLA frameworks. Importantly, a majority of the published research articles did not focus on the Global South and did not fully examine (Zawacki-



Richter et al., 2019). Particularly in understudied contexts like India, these imbalances raise questions regarding the pedagogical, ethical, and cognitive implications of AI for language learning.

The purpose of this systematic review is to provide a consolidated understanding of the state of AI research in ELT, identify thematic research gaps, and articulate a forward-looking agenda. By systematically analyzing published work over the past decade, this review seeks to guide researchers, practitioners, and policymakers toward a more balanced, rigorous, and inclusive approach to AI in L2 learning.

Methodology

This review followed proper guidelines to ensure transparency and reproducibility with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). The search was conducted across four major databases—Scopus, Web of Science, ERIC, and Google Scholar—covering publications between January 2015 and March 2025. Keywords were carefully selected to capture the breadth of relevant studies, including “Artificial Intelligence in English Language Teaching,” “AI and SLA,” “AI in IELTS preparation,” “intelligent tutoring systems,” and “AI-driven writing and speaking assessment.”

The initial search produced 112 records. After duplicate removal, 93 studies remained for screening. Abstracts and full texts were then examined for relevance. The inclusion criteria required studies to be peer-reviewed journal articles or book chapters, empirical or conceptual in design, published in English, and directly focused on AI applications in ELT or L2 learning. Exclusion criteria eliminated conference abstracts, opinion pieces, and studies that addressed AI in general education without a language-specific focus. Following this process, 48 studies were retained for final analysis.

The coding and analysis process involved extracting key information from each study, including author, year, AI tool type, methodology, learner population, geographical context, and key findings. Using thematic synthesis, the studies were then organized into conceptual categories to identify both converging patterns and gaps in the

literature. Inter-coder reliability was established by independently coding a subset of studies and cross-checking categories. This method enabled both descriptive mapping of the field and interpretive identification of research gaps.

Results

The final set of 48 studies revealed a highly uneven distribution of research activity. Geographically, a disproportionate number of studies were conducted in advanced regions such as North America, Western Europe, and East Asia, with China being a major contributor technologically. Only a handful of studies originated in India or other South Asian contexts, despite the region’s significance in global English learning. Methodologically, the majority of studies relied on exploratory or small-scale case study designs, often involving fewer than 50 participants. Surveys and qualitative interviews were common, while large-scale experimental or longitudinal studies were rare.

In terms of focus, writing-related applications dominated the research field, with 40 percent of the reviewed studies centered on automated writing evaluation tools such as Criterion, Grammarly, or Turnitin. Speaking applications comprised roughly 30 percent of studies, typically involving speech recognition software or AI-powered chatbots. Reading and listening applications were less represented, while conceptual discussions of AI frameworks accounted for the remaining share. Overall, the literature demonstrated a strong emphasis on evaluating technical accuracy and learner perceptions, but relatively less attention was given to cognitive, affective, or ethical dimensions.

Thematic Analysis of Research Gaps

The synthesis of findings revealed six major categories of research gaps, each of which undermines the ability of current scholarship to provide a holistic understanding of AI in ELT.

One of the most salient gaps is pedagogical integration. While numerous studies have assessed AI tools for their functional benefits, few have situated their findings within established SLA frameworks such as Task-Based Language Teaching, sociocultural theory, or cognitive load theory. This



disconnect results in research that is often descriptive rather than explanatory. For instance, several studies reported improvements in learner writing accuracy following the use of Grammarly, yet rarely did they analyze whether such improvements translated into deeper linguistic competence or whether they simply reflected surface-level corrections (Ellis, 2020). Without embedding AI use into pedagogical frameworks, the sustainability of these improvements remains questionable.

A second gap concerns the geographic and contextual imbalance in the evidence base. The overwhelming majority of studies are conducted in high-income nations with robust technological infrastructure, creating an implicit assumption that learners globally experience similar conditions. This assumption obscures the realities of learners in India or rural regions, where internet access may be unstable, devices may be limited, and classrooms may lack adequate teacher training to integrate AI effectively. Kukulska-Hulme et al. (2022) note that such disparities perpetuate a “digital colonialism,” wherein research is exported from technologically advanced contexts and then generalized globally without consideration of local realities.

Cognitive and affective dimensions represent a third major gap. Despite the centrality of motivation, anxiety, and cognitive load in SLA, few studies systematically measure these variables in relation to AI tools. For example, while speech recognition applications may offer immediate feedback on pronunciation, they may simultaneously increase learner anxiety due to fear of machine misinterpretation. Similarly, automated writing evaluators may reduce cognitive load by simplifying corrections, yet they may also impose extraneous load if learners are overwhelmed by large volumes of feedback (Sweller, 2010). The limited empirical focus on these psychological factors leaves an incomplete picture of learner experiences.

Methodological limitations form a fourth category of gaps. The predominance of exploratory case studies and surveys has led to a descriptive literature that lacks predictive power. Longitudinal designs that track learner progress over extended periods, randomized controlled trials that test causal relationships, and large-scale cross-cultural

comparisons are notably absent. As a result, the field lacks strong evidence to determine whether AI-mediated interventions have durable impacts on language acquisition.

Equity and accessibility emerge as a fifth underexplored area. While some studies briefly acknowledge digital divides, very few investigate how socioeconomic status, gender, or rural-urban disparities affect learners’ ability to engage with AI-mediated ELT. Selwyn (2020) argues that without confronting these issues, AI risks amplifying rather than mitigating existing educational inequalities.

Finally, ethical concerns represent a sixth gap. Issues such as data privacy, algorithmic bias, plagiarism detection, and overreliance on AI-generated outputs are often mentioned but rarely explored in depth. Williamson and Piattoeva (2022) highlight that the growing use of generative AI for writing assistance poses significant challenges to academic integrity, yet the ELT literature has not yet caught up with these debates.

Discussion

The review’s conclusions support the idea that AI in ELT is at a turning point. Although the abundance of tools has increased opportunities for both teachers and students, the body of research is still too focused on privileged contexts and is not rigorous enough. A skewed perception of AI’s function in L2 learning could result from the disregard of pedagogical, cognitive, and ethical factors. Particularly, in a country like India, where millions of students prepare for global exams like IELTS, the potential research of AI-integrated research is missing. Without addressing these gaps, the potential of AI to democratize language learning will not be realized.

Conclusion

This systematic review highlights six important gaps in AI and ELT research. It includes a weak pedagogical integration, geographic inclusion, neglect of cognitive load and other affective factors, lack of methodological rigor, lack of proper ethical frameworks, and unexamined ethical issues. To overcome these constraints, future research needs to be methodologically sound, theory-driven, and contextually inclusive studies should be conducted.



In countries like India, where many students prioritize IELTS preparation, studies must take into account viewpoints on cognitive load, examine socioeconomic barriers, and develop ethical frameworks for responsible AI use. A more equitable and globally representative research agenda is required to ensure that AI is a tool for genuine educational advancement rather than a means of widening the existing gaps.

References

1. Ellis, R. (2020). Task-based language teaching: Theory and practice. Cambridge University Press.
2. Kohnke, L., & Jarvis, A. (2021). Artificial intelligence in English language teaching: Pedagogical implications. *Computer Assisted Language Learning*, 34(7), 892–915. <https://doi.org/10.1080/09588221.2020.1846568>
3. Kukulska-Hulme, A., Beirne, E., & Nygaard, C. (2022). Mobile AI for language learning: Challenges and opportunities. *Language Learning & Technology*, 26(2), 1–17. <https://doi.org/10.10125/44741>
4. Selwyn, N. (2020). Should robots replace teachers? AI and the future of education. Polity.
5. Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press.
6. Williamson, B., & Piattoeva, N. (2022). Education governance and datafication: Critical perspectives on AI and learning. *Learning, Media and Technology*, 47(1), 1–15. <https://doi.org/10.1080/17439884.2021.1970439>
7. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>