Contributions of AI Tools to Critical Thinking Development in EFL Learning: A Systematic Review and Meta-Analysis

Marcela Vildósola[1], Pedro Salcedo-Lagos[2], Gabriela Kotz[3], Cristian Sanhueza-Campos[4]*

[1] Doctorado en Lingüística, Facultad de Humanidades y Artes, Universidad de Concepción, Chile.

[2] Departamento de Metodología de la Investigación e Informática, Facultad de Educación, Universidad de Concepción, Chile.

[3] Departamento de Idiomas Extranjeros, Facultad de Humanidades y Artes, Universidad de Concepción, Chile.

[4] Departamento de Teoría, Política y Fundamentos de la Educación, Escuela de Educación, Universidad de Concepción - Campus Los Ángeles, Chile.

Corresponding author's email: crsanhuezac@udec.cl

*ORCID: https://orcid.org/0000-0002-7924-5910

ABSTRACT

The use of Artificial Intelligence tools within language learning environments is an emerging trend in educational settings. Existing literature highlights the potential of AI applications to enhance English as a Foreign Language (EFL) learners' competencies, such as listening, speaking, vocabulary acquisition, and cross-cultural understanding. Despite these promising outcomes, systematic reviews exploring how AI technologies contribute to the development of critical thinking in EFL students remain limited. This systematic review examines the contributions of AI tool implementation to the development of critical thinking in EFL students, identifying the methodologies used and findings obtained across different educational contexts. Articles from major databases including ERIC, Scopus, Scielo, WOS, and Science Direct were screened by reviewing titles, abstracts, and full texts. Of the 1,319 articles analysed, only 30 published between 2015 and 2024 were found relevant to the research topic. The findings suggest the potential of AI tools in fostering critical thinking when thoughtfully integrated with pedagogical strategies in the EFL classroom. Future research could explore the long-term effects of AI tools on critical thinking and address ethical considerations and cultural sensitivities that may affect fair and inclusive integration.

Introduction

The rise of artificial intelligence (AI) tools in education marks a shift in how learning is approached, particularly in English as a Foreign Language (EFL) contexts. These tools promise to change traditional methods by enabling adaptive, interactive, and personalized learning experiences. While their potential for improving linguistic competencies, such as vocabulary acquisition and writing skills, is well-documented, their role in fostering higher-order cognitive

abilities, like critical thinking, remains less explored. The integration of AI into EFL education thus presents a dual opportunity: to enhance language proficiency and to cultivate essential analytical and reflective skills needed by learners at present.

AI is rapidly transforming education, where it holds significant potential for improving learning outcomes. Within EFL instructional settings, AI tools have been able to address some challenges EFL learners often face, like limited exposure to authentic English environments, lack of confidence in oral communication, and struggles with linguistic and cultural differences (Shen & Chiu, 2019). To that, AI tools present a promising solution, creating individualized, interactive, and effective learning experiences.

Critical thinking is essential for language learners since it enables them to evaluate information, construct arguments, and express ideas clearly. While widely emphasized in education (Paul & Elder, 2003), its integration into EFL remains underexplored (Bag & Gursoy, 2021). Recent advancements in AI tools, like chatbots, intelligent tutoring systems, and automated feedback, offer opportunities to enhance critical thinking through reflective tasks (Li et al., 2024; Teng, 2024). Yet, most research focuses on linguistic gains, overlooking how these tools explicitly support critical thinking development (Deng & Yu, 2023; Zhai & Wibowo, 2023).

A growing body of literature has examined AI in education, highlighting its benefits and challenges. Studies demonstrate its ability to personalize feedback, facilitate collaborative learning, and boost engagement (AlTwijri & Alghizzi, 2024; Lai & Lee, 2024). However, the interplay among the use of AI tools, the enhancement of critical thinking skills and the EFL context remains underexplored. This gap emphasizes the need to investigate AI's potential to foster higher-order cognitive skills in EFL.

This systematic review seeks to address this gap by analysing the contributions of AI tools to the development of critical thinking in EFL learners. By examining 30 studies that incorporate various AI-based tools, this review aims to provide a comprehensive understanding of their impact and identify strategies for their effective integration. The findings will not only contribute to the growing discourse on AI in education but also offer actionable insights for educators and researchers striving to enhance critical thinking in EFL learning environments.

Theoretical framework

Artificial Intelligence in Education

Artificial intelligence refers to computational systems capable of executing functions conventionally associated with human cognition, including learning, reasoning, and decision-making. While the term "artificial intelligence" has existed for decades (McCarthy, 2007), recent advances in machine learning, natural language processing, and user-friendly AI-powered tools have accelerated its transformative impact across multiple domains, including education (Chiu et al., 2022; Huang et al., 2023; Salcedo-Lagos et al., 2024).

In education, AI aligns with theories from cognitive psychology and self-regulated learning (SRL) by providing personalized, adaptive support that helps learners manage cognitive load and monitor their progress (Sweller et al., 2011; Zimmerman & Schunk, 2011). By tailoring content and feedback to individual proficiency levels and learning styles (Agarwal &

Chakraborty, 2019), AI supports both lower-order skills (e.g., grammar accuracy) and higher-order skills like analysis and critical evaluation (Bahroun et al., 2023; Ma & Chen, 2024). Automated feedback systems, intelligent tutoring, and conversational agents allow real-time interaction, reinforcing students' cognitive and metacognitive engagement (Deeva et al., 2021; Seraj et al., 2021).

AI tools in language education can be broadly categorized into generative and non-generative systems, each serving distinct roles. Generative AI tools use machine learning models to create new content based on user prompts and pre-trained data (X. Yan et al., 2024). Examples include ChatGPT—which rapidly gained widespread popularity upon its launch (Chartr, 2022; Teubner et al., 2023)—and other tools like Bard, Elicit, Consensus, and Perplexity (Garg et al., 2024; Whitfield & Hofman, 2023; Pawar et al., 2024). These tools facilitate dynamic, human-like language practice, debate, and reflective exploration. Conversely, non-generative AI tools focus on analyzing and improving existing content through pattern recognition and rule-based corrections (Pan et al., 2023). Popular examples include Grammarly, which provides real-time grammar, spelling, and style feedback (Abu Qub'a et al., 2024); Quillbot for paraphrasing and refinement; and Automatic Writing Evaluation (AWE) tools that guide writing improvement. Platforms like Mosoteach integrate AI-driven writing support with features that promote self-regulated learning behaviors (Li & Jianbin, 2019). Together, these tools personalize and enrich language learning experiences, enhancing learner confidence, accuracy, and autonomy (Macinska & Vinkler, 2023).

EFL and AI Integration

English as a Foreign Language (EFL) involves learning and teaching English in contexts in which it is not used as the dominant means of communication (Wu & Shafait, 2024). Learners often encounter challenges related to receptive and productive skills, alongside linguistic and cultural nuances distinct from their native languages (Hossain, 2024). To address these challenges, AI tools have emerged as powerful innovations providing individualized support and resources tailored to learners' specific needs (Chang et al., 2023). These tools tackle various facets of language learning, including vocabulary development, grammar correction, conversational practice, and pronunciation improvement (Dai & Liu, 2024). For instance, pronunciation tools deliver real-time feedback to refine accent and spoken clarity (W. Yan et al., 2024)), while conversational AI applications like Replika and ChatGPT simulate real-world scenarios, enabling learners to practice speaking and listening without fear of judgment (Belda-Medina & Calvo-Ferrer, 2022). Gamified language learning apps enhance engagement (Fuentes-Riffo et al., 2023), and tools like Grammarly and Quillbot provide automated grammar and style feedback suited to EFL learners' needs (Abu Qub'a et al., 2024). Interactive vocabulary platforms further adapt exercises to learners' proficiency levels, ensuring personalized experiences (Macinska & Vinkler, 2023).

However, while AI integration offers these benefits, it also presents challenges that must be addressed to maximize positive outcomes. Excessive dependence on AI tools may result in passive learning and reduce learners' active engagement in critical analysis and problem-solving tasks (Ali et al., 2024). Ethical concerns, including data privacy, the potential misuse of sensitive information, and biases or inaccuracies in AI-generated content, further complicate

responsible usage (Bozkurt & Sharma, 2023). Additionally, unequal access to advanced AI tools in low-resource settings risks widening existing educational disparities (Marzuki et al., 2023). Therefore, successful EFL and AI integration requires thoughtful, pedagogically sound, and ethically responsible implementation that complements—rather than replaces—traditional teaching. When well-managed, AI can enhance both language proficiency and critical thinking; when mismanaged, it may hinder the development of essential skills and autonomy.

Critical Thinking in the EFL Context

Critical thinking is a multifaceted cognitive skill central to education, described as purposeful, self-regulatory judgment (Facione, 1990). For EFL learners, developing critical thinking means moving beyond rote memorization to analysing information, questioning sources, constructing arguments, and articulating clear, reasoned opinions (Lipman et al., 1980; Norris & Ennis, 1989; Tang et al., 2020). As shown in recent corpus-based analyses of reflective writing, epistemic and effective stance strategies reveal learners' evaluative reasoning and agency (Domínguez Romero et al., 2025), which parallels the metacognitive and attitudinal dimensions of critical thinking.

Advances in AI-mediated instruction offer new pathways for fostering these skills. AI-powered platforms simulate debates and problem-solving scenarios (Hu & Škultéty, 2024), guide text analysis and rhetorical critique (Glaser, 2023), and provide reflective feedback loops that nurture metacognitive awareness (Teng, 2024; Lu & Xie, 2022). This aligns with findings that AI, when integrated thoughtfully, supports the development of critical reflection and analytical reasoning in second language contexts (Bezanilla et al., 2019).

Connecting AI to Critical Thinking: Conceptual Model

Building on SLA, cognitive load, and SRL theories, this review conceptualizes AI tools as catalysts that foster critical thinking through three interrelated pathways:

- Cognitive Pathway: Tools like ChatGPT and intelligent tutoring systems scaffold tasks that demand analysis, evaluation, and synthesis, directly exercising critical thinking subskills.

- Metacognitive Pathway: Feedback tools and AWE systems prompt learners to plan, monitor, and revise their work critically, strengthening self-regulated learning behaviours.

- Affective Pathway: By lowering anxiety and increasing confidence (Lai & Lee, 2024), AI creates a safe environment for students to take intellectual risks and engage deeply with ideas.

In sum, this framework positions AI not merely as a linguistic aid but as an instructional mediator that can transform how EFL learners develop critical thinking—provided its use is carefully guided to avoid over-reliance and ethical pitfalls (Ali et al., 2024; Bozkurt & Sharma, 2023).

Literature review

A growing body of systematic reviews has explored the diverse roles and limitations of AI in

EFL education. For example, Deng and Yu (2023) examined chatbot-assisted language learning, showing benefits for personalized feedback and scaffolding but limited direct improvement in critical thinking due to a primary focus on factual recall. AlTwijri and Alghizzi (2024) investigated how tools like ChatGPT, SIRI, and MOCA impact affective factors—finding that these tools enhance motivation and reduce speaking anxiety through interactive dialogues, although small sample sizes and brief interventions limit generalizability.

Li et al. (2024) reviewed ChatGPT's first year of integration, highlighting its effectiveness in supporting self-regulated learning and enhancing teacher productivity and student writing. However, they also noted persistent issues around content reliability, student over-dependence, bias, and ethical challenges such as plagiarism—calling for longitudinal research to clarify its true impact on critical thinking. Similarly, Teng (2024) found that while ChatGPT boosts motivation, self-efficacy, and writing structure, it can hinder creativity and deeper reasoning if learners rely too heavily on its outputs, underscoring the need for AI literacy.

Zhai and Wibowo (2023) analysed AI dialogue systems' effects on EFL interactional competence, showing gains in vocabulary and cultural awareness but flagging limitations in systems' reasoning abilities, which constrains critical thinking development. Lai and Lee (2024) provided a broader synthesis, confirming AI's positive impacts on language skills, motivation, and anxiety reduction, but warned that the lack of EFL-specific AI design and limited focus on reasoning tasks weaken their contribution to fostering higher-order thinking.

Taken together, these reviews show that while AI tools have clear potential to support EFL learning and engagement, there remains a notable gap: most research focuses on linguistic gains—vocabulary, grammar, writing fluency—while direct empirical evidence on how AI shapes critical thinking remains scarce or indirect. Addressing this gap, the present systematic review aims to synthesize and clarify how different AI tools contribute specifically to critical thinking development in EFL contexts, identifying both opportunities and limitations to guide more effective integration.

Research Question

The following research question guided this systematic review:

How has the implementation of AI tools in learning English as a foreign language contributed to the development of critical thinking?

Methods

This review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Data were retrieved from five major academic databases: the Educational Resources Information Centre (ERIC), Web of Science (WOS), Scopus, Scielo, and Science Direct. The search strategy, selection criteria and data coding and analysis in this review are presented below. All authors collaboratively designed and refined each stage of the review process.

Search Strategy

Inclusion and exclusion criteria were applied to eliminate irrelevant studies. To be included in this review the study had to be (1) published in English or Spanish, (2) dated from 2015 to 2024

inclusively, (3) empirical research or case studies conducted in educational contexts, and (4) studies that integrated one or more AI applications relevant to EFL learning and/or critical thinking. The studies were excluded if they (1) were published prior to 2015, (2) were theses, systematic reviews, book, conference proceedings, editorials or opinion articles, (3) were unrelated to education or not involving AI tools, and (4) were written in languages other than English or Spanish. The initial database search yielded 1,388 records. After removing duplicates, 223 records remained for title and abstract screening by two independent reviewers. Disagreements were discussed, with a third author mediating as needed. Full-text screening resulted in 30 studies meeting all criteria, as summarized in Figure 1. Before full coding began, all four authors held a calibration session to refine the coding scheme and ensure consistent application across studies. Discrepancies during coding were resolved through discussion or, when necessary, by consensus among all authors.

Figure 1.

PRISMA flowchart for the systematic review
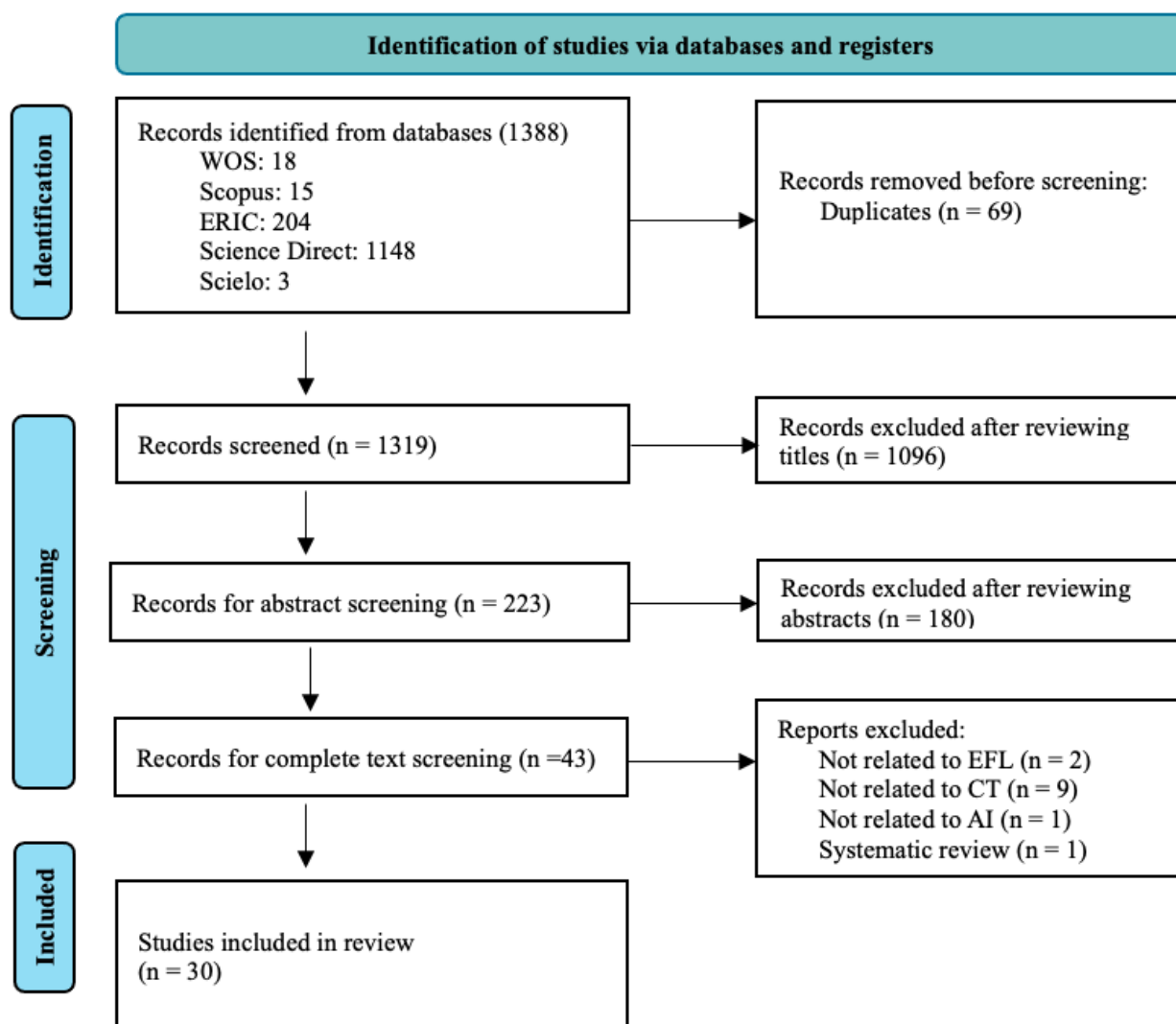
Coding procedures and inter-rater reliability

A coding protocol was developed collaboratively by all authors. It included study characteristics (author, year, country), participant details, AI tool types, research design, outcome measures, and main findings. Two authors independently coded a random sample of 20% of the studies. Cohen's kappa was calculated, yielding $\kappa = 0.86$, indicating high agreement. The remaining studies were coded by the first and second authors and verified by the third and fourth authors to ensure accuracy.

Quality assessment

To appraise the methodological rigor of included studies, the Mixed Methods Appraisal Tool (Hong et al., 2018) was used. Each study was assessed according to criteria suitable for its design (qualitative, quantitative, or mixed methods). MMAT scores were not used to exclude studies but informed the critical interpretation of methodological strengths and limitations during synthesis.

Operational definitions

To ensure consistent categorization of outcomes and studies, the authors defined the following key constructs a priori:

Critical Thinking. Following Facione (1990), critical thinking was defined as purposeful, self-regulatory judgment that includes core subskills such as analysis, inference, evaluation, explanation, and self-regulation. Outcomes were classified as direct (measured with validated critical thinking tests or rubrics) or indirect (proxied by argument quality, depth of reflection, logical reasoning tasks, or similar indicators).

Artificial Intelligence (AI) Tools. AI tools were defined as digital applications that perform tasks traditionally requiring human intelligence, including natural language processing, automated feedback, predictive analytics, generative content creation, or personalized tutoring. For this review, tools were categorized as generative AI (e.g., ChatGPT, Bard, Elicit, Consensus) or non-generative AI (e.g., Grammarly, QuillBot, Automatic Writing Evaluation systems).

English as a Foreign Language (EFL). EFL contexts were defined as learning environments in which English is taught as a subject and not used as the primary medium of everyday communication or official societal interaction (Wu & Shafait, 2024). This includes primary, secondary, and tertiary educational settings in non-English-dominant countries.

Other outcomes (e.g., self-efficacy, confidence, motivation, writing performance) were coded according to established frameworks in second language acquisition.

Results

This section examines the background information of the selected documents (Table 2), the kinds of studies, most used AI tools, as well as the main findings of the studies selected, including the distribution of countries according to where the studies were conducted (Table 1).

Table 1.

Distribution of selected articles based on the countries where they were conducted

Country	Percentage (%)	Continent	Percentage (%)
China	40.00		
Indonesia	13.30		
Iran	3.30		
Taiwan	3.30		
Vietnam	6.60		
Thailand	3.30	Asia	86.60
Yemen	3.30		
Singapore	3.30		
Saudi Arabia	3,30		
Turkey	6.60		
Czech Republic	3.30	Europe	6.60
Scotland	3.30		
United States	6.60	America	6.60

Table 1 reveals that scholars in China produced more articles than those in other countries. Notably, an 86.60% of the articles were conducted in Asian countries like Indonesia, Iran, Vietnam, among others. Additionally, 6.60% of the studies were conducted in America, specifically in the United States. Also, 6.60% of the studies were produced in Europe, specifically in Scotland and Czech Republic. Notably, the majority of studies were conducted in East Asian contexts, particularly China and Indonesia. This geographic concentration suggests that while the findings provide valuable insights for similar EFL contexts, caution should be exercised when generalizing results to other cultural or educational settings where AI adoption and pedagogical norms may differ.

Table 2.

Reviewed articles

Nº	Author / year	Kind of study	Objective	AI used	Sample size	Main findings
1	Cheng, 2017	Mixed-methods approach	To investigate the effect of online automated feedback on the quality of EFL students' reflective journals.	Web-based automatic classification system	138 university students.	Students receiving automated feedback outperformed the control group, showing significant improvement across journals, despite some concerns about its limitations.
2	Koltovskaia, 2020	Qualitative approach	To explore how ESL students behaviorally, cognitively, and affectively engage with automated written feedback.	Grammarly	2 ESL students.	Students showed varying engagement: one critically questioned feedback, while the other relied on it. Both achieved moderate draft improvement but demonstrated limited deeper learning.
3	Hapsari and Wu, 2022	Qualitative approach	To design an AI chatbot learning model to reduce speaking anxiety, boost enjoyment, and foster critical thinking in EFL speaking classes.	AI Chatbots (speaking)	Number not specified.	The model effectively supports speaking development by reducing anxiety, enhancing enjoyment, and fostering critical thinking through structured chatbot interactions and peer discussions.
4	Muthmainnah et al., 2022	Mixed-methods approach	To determine the impact of AI friend apps on critical thinking skills and English proficiency among EFL learners.	Chatbots like Mitsuku and Replika	453 non-native English students.	AI applications enhanced critical thinking, fostering trust, confidence, and open-mindedness. Students improved in English proficiency, motivation, and real-life communication skills through AI interactions.
5	He et al., 2023	Quantitave approach	To assess the effectiveness of blended learning with an intelligent cloud teaching platform in English courses.	Intelligent Cloud Teaching Platform developed by Beijing Lanmo Big Data Research Institute.	110 university students.	The blended learning mode improved self-efficacy, autonomy, and critical thinking, leading to better academic performance through big data and cloud-based integration.
6	Liu et al., 2023	Mixed-methods approach	To evaluate the effectiveness of peer assessment and automated evaluation in improving EFL students' writing.	Automatic Writing Evaluation (AWE) tool.	64 university students.	The integrated approach improved writing performance, critical thinking, and reduced anxiety. Students reflected more deeply on feedback and outperformed those using conventional automated evaluation.
7	Zhang et al. 2023	Mixed-Methods approach	To examine the effect of chatbot-based training on logical fallacies to enhance self-efficacy and proficiency in EFL argumentative writing.	Custom-developed chatbot hosted on Facebook Messenger	15 university students.	Chatbot-based training enhanced argumentative writing by improving reasoning and logical fallacy awareness but slightly lowered self-efficacy as learners recognized their limitations.
8	Mohamed, 2023	Qualitative approach	To explore EFL faculty perceptions of ChatGPT's effectiveness in supporting students' English learning and enhancing teaching methods.	ChatGPT	10 EFL faculty members.	Faculty acknowledged ChatGPT's potential for real-time feedback, personalized learning, and engagement but raised concerns about accuracy, overreliance, and ethics, urging further studies.
9	Liu et al., 2023	Mixed methods approach	To enhance EFL writing performance, self-efficacy, and self-regulated learning through reflective thinking in AI-supported writing environments.	Mosoteach, a widely used app in China, with AI-based automated feedback capabilities.	104 university students.	The approach improved writing performance, reduced cognitive load, and boosted self-efficacy. Reflective thinking enhanced understanding and fostered autonomy through AI feedback.
10	Darwin et al., 2024	Qualitative approach	To explore EFL students' perceptions of AI's role in developing critical thinking.	AI tools, such as Natural Language Processing (NLP) systems and data analytics platforms.	7 master's degree students.	AI enhances critical thinking by supporting research and theory analysis. However, students highlight its limitations, including lack of personalization, echo chambers, and challenges with nuanced understanding.

11	Klimova et al., 2024	Mixed-methods approach	To examine the potentials and pitfalls of using ChatGPT for academic purposes among university undergraduates.	ChatGPT	90 university students.	Students appreciated ChatGPT for idea generation, research support, and concept clarification but raised concerns about overreliance, accuracy, and ethics, emphasizing the need for responsible use guidelines.
12	Ngo et al., 2024	Qualitative approach	To examine EFL teachers' perceptions of AI-based academic dishonesty and their strategies for mitigation.	ChatGPT	31 university lecturers.	Teachers linked AI dishonesty to low motivation, academic pressure, and skill gaps. They advocated stricter regulations, AI detection tools, and pedagogical reforms to uphold academic integrity.
13	Koh et al., 2024	Mixed-Methods approach	To explore the effectiveness and pedagogical value of generative AI word clouds in English and Chinese teaching contexts.	ChatGPT (GPT-3.5)	75 university students.	Generative AI word clouds enriched discussions by highlighting key ideas. Instructors found them useful but expressed concerns about accuracy and transparency.
14	Yuan et al., 2024	Qualitative approach	To analyze ChatGPT's benefits, challenges, and impact on EFL learners' academic writing.	ChatGPT	7 participants.	ChatGPT improves writing fluency, content quality, and efficiency but poses risks of plagiarism, overreliance, and authenticity concerns, requiring learners' critical reflection for responsible use.
15	Kartal, 2024	Qualitative approach	To examine ChatGPT's impact on EFL student teachers' thinking skills and creativity.	ChatGPT	12 student teachers.	ChatGPT enhanced creative lesson planning and critical thinking but posed originality risks due to overreliance. Balanced human-AI collaboration proved essential for effective integration.
16	Mohammad H and Al-khresheh, 2024	Qualitative approach.	To examine teachers' perspectives on the benefits and challenges of integrating ChatGPT into English Language Teaching globally.	ChatGPT.	46 English language teachers.	Teachers appreciated ChatGPT for personalized feedback and interactive learning but noted challenges like overreliance, cultural insensitivity, and limited speaking and listening support. They stressed curriculum adaptation and professional training.
17	Shen and Feng Teng, 2024	Quantitative approach	To examine temporal relationships between critical thinking, self-directed learning, and AI-assisted writing among EFL learners.	Quillbot	204 university students.	Critical thinking and self-directed learning were reciprocally related, both predicting AI-assisted writing. Self-directed learning moderated critical thinking's impact, though inconsistently.
18	Hieu and Thanh Thao, 2024	Qualitative approach	To examine EFL teachers' perceptions of using ChatGPT for fairy tale retelling, focusing on challenges and opportunities.	ChatGPT	9 EFL teachers.	Teachers noted ChatGPT improved engagement, creativity, and language proficiency. Challenges included overreliance, cultural misalignment, ethical issues, and technical barriers. Balanced integration and teacher training were emphasized.
19	Tseng and Lin, 2024	Qualitative approach	To examine ChatGPT's integration into EFL writing courses using ADDIE and TPACK frameworks to enhance writing quality and learning efficiency.	ChatGPT (GPT-3.5)	15 university students.	ChatGPT boosted writing efficiency, provided instant feedback, and improved organization. Students actively refined drafts, balancing AI use with critical thinking and creativity.
20	Zhi, and Wang, 2024	Quantitative approach	To examine the interplay between EFL students' attitudes toward AI, teachers' immediacy, rapport, and willingness to communicate.	AI-mediated learning tools (not specified)	1,090 students at undergraduate, Master's, and doctoral levels.	Positive teacher immediacy and rapport significantly influence students' willingness to communicate. AI-related attitudes improve WTC, but interpersonal teacher-student relationships remain more influential.
21	Hao et al., 2024	Quantitative approach	To explore critical thinking as a mediator between motivation and creativity in business English learners through cognitive flexibility theory.	Chat GPT	153 university students.	Motivation significantly influenced critical thinking, which in turn enhanced creativity. Critical thinking fully mediated this relationship, underscoring its key role in fostering innovation.
22	Koltovskaia et al., 2024	Qualitative approach	To explore graduate ESL students' behavioral, cognitive, and affective engagement with ChatGPT when revising academic research proposals.	ChatGPT-3.5 Turbo	6 university students.	Students used ChatGPT mainly for lower-order concerns, accepted style feedback for clarity, and rejected inaccurate mechanical suggestions. Cognitive engagement showed critical evaluation, while affective engagement reflected satisfaction with its role in enhancing professionalism.
23	Yan et al.,	Qualitative	To explore EFL learners' perceptions of	ChatGPT	281 university	Students viewed GenAI as a human-like assistant, tool, brain, and

	2024	approach	Generative AI in language learning using metaphor analysis.		students.	resource. While positive attitudes dominated, concerns included overreliance and diminished critical thinking skills.
24	Tafazoli, 2024	Qualitative approach	To investigate how Generative AI addresses educational challenges and promote inclusivity in English language education.	Generative AI (GenAI) systems, especially ChatGPT	23 EFL school teachers.	GenAI improves access to tailored learning, fosters personalization, minimizes biases, overcomes tech barriers, and connects educators and learners to global trends.
25	Darmawansah et al., 2024	Mixed- methods approach	To assess the impact of ChatGPT-supported collaborative argumentation on EFL students' speaking performance, critical thinking, and collaboration.	ChatGPT	67 university students.	ChatGPT-supported argumentation enhanced students' speaking performance, critical thinking, and collaboration. Argument quality varied by proficiency, emphasizing the need for prompt training.
26	Alkamel et al., 2024	Quantitative approach	To explore EFL students' perceptions of ChatGPT's benefits and challenges in enhancing writing skills.	ChatGPT	144 EFL students.	ChatGPT enhanced writing fluency, grammar, and coherence but raised concerns about academic integrity, plagiarism risks, and the need for critical evaluation of AI outputs.
27	Kurt and Kurt, 2024	Qualitative approach	To explore pre-service teachers' perceptions of ChatGPT as an automated writing feedback tool compared to peer and teacher feedback.	ChatGPT (GPT-3.5)	52 pre-service teachers.	ChatGPT feedback was valued for practicality, interactivity, and depth, but inconsistencies and overreliance were concerns. Teacher feedback remained most trusted, with a combined AI, peer, and teacher approach recommended.
28	Liang and Wu, 2024	Mixed-methods approach	To examine how ChatGPT enhances EFL learners' critical thinking skills using pedagogical frameworks and a post-humanist perspective.	ChatGPT	58 university students.	ChatGPT broadened students' thinking, offered diverse perspectives, and supported analysis, but concerns about overreliance and bias emerged. Students emphasized balancing AI with independent thought for effective learning.
29	Waluyo and Kusumastuti, 2024	Mixed-Methods approach	To investigate student and teacher perceptions of generative AI in English learning in higher education and its impact on outcomes.	Grammarly, Quillbot, ChatGPT, Bard, BingChat, Reverso	25 university students.	Students embraced AI for learning efficiency and engagement, while teachers raised ethical concerns and warned against overreliance, advocating for responsible AI integration in curricula.
30	Ngo and Hastie, 2025	Mixed-methods approach	To develop and implement an AI-integrated EAP module to improve AI literacy and enhance academic skills among international students.	ChatGPT Elicit Consensus Perplexity Quillbot	39 international students.	AI literacy boosted students' confidence, critical evaluation skills, and ethical awareness. They utilized AI for research, writing, and personalized learning, expanding its academic applications.

Most studies included in this review were published from 2022 onward, reflecting the surge in AI research following the emergence of ChatGPT and other advanced tools. Specifically, 66.6% of studies appeared in 2024, 20% in 2023, 6.6% in 2022, and only a small fraction (3.3%) in 2017 and 2020. This trend underscores AI's growing prominence in EFL research contexts. Although many studies were conducted during the ChatGPT era, a variety of tools were explored, including AI chatbots, AI-mediated feedback systems, and integrated AI learning platforms.

Regarding research design, 43.4% of studies employed qualitative approaches, 40% used mixed-methods, and 16.6% were quantitative. Common qualitative methods included semi-structured interviews and thematic analyses to capture students' and teachers' perceptions and experiences. Quantitative studies typically adopted quasi-experimental or correlational designs. Mixed-methods studies combined scales, tests, interviews, and open-ended surveys to provide a comprehensive understanding of AI's impact on learning.

In terms of language skills targeted, several studies focused on writing, employing tools like ChatGPT, Grammarly, AWE, Quillbot, and the Intelligent Cloud Teaching Platform to enhance grammatical accuracy, idea organization, and paraphrasing (Zhang et al., 2023; Koltovskaia, 2020; Kurt & Kurt, 2024; He et al., 2023). Others addressed speaking skills, using chatbots such as Mitsuku and Replika to facilitate oral practice and reduce anxiety (Muthmainnah et al., 2022; Hieu & Thao, 2024). Additional tools provided real-time analysis, personalized learning, and visualizations to support critical analysis and classroom discussions (Koh et al., 2024; Tseng & Lin, 2024; Darwin et al., 2024).

Findings indicated that AI tools positively influenced critical thinking in EFL contexts, both directly and indirectly. Direct effects included improved analysis, evaluation, and inference skills through interactive chatbots and AWE platforms (Muthmainnah et al., 2022; Liu et al., 2023a; He et al., 2023). Indirectly, tools like Grammarly and ChatGPT fostered reflective revision and self-assessment (Koltovskaia, 2020; Kurt & Kurt, 2024), while automated feedback encouraged students to engage critically with content (Zhang et al., 2023).

However, over-reliance on AI emerged as a significant challenge. Studies reported that excessive dependence on AI-generated responses may reduce learners' initiative and hinder critical thinking and creativity (Klimova & Luz de Campos, 2024; Liang & Wu, 2024; Yuan et al., 2024; Darwin et al., 2024). Concerns about plagiarism and unethical use were also prevalent (Ngo Cong-Lem et al., 2024; Al-Khresheh, 2024). Nonetheless, some research highlighted that guided, reflective AI use, coupled with group discussions and co-creation, can mitigate these risks and enhance creativity (Kartal, 2024; Hieu & Thao, 2024).

Studies further emphasized ethical considerations, pointing to gaps in students' AI literacy and citation practices, as well as risks related to accuracy, privacy, and bias (Ngo Cong-Lem et al., 2024; Darwin et al., 2024; Yuan et al., 2024). Structured interventions and critical evaluation of AI outputs were found to help foster more responsible use (Ngo & Hastie, 2025).

In addition to cognitive and ethical aspects, AI tools were shown to help reduce emotional barriers such as anxiety and stress. Chatbots and AWE platforms supported motivation and created more supportive learning environments (Liu et al., 2023a; Darwin et al., 2024). Many

studies also noted efficiency gains: tools like ChatGPT, Grammarly, and Quillbot provided immediate feedback and boosted learners' confidence and engagement, especially in resource-constrained contexts (Ngo & Hastie, 2025; Hieu & Thao, 2024; Waluyo & Kusumastuti, 2024).

Finally, comparisons with traditional methods highlighted AI's complementary role. Blended learning approaches combining AI and teacher or peer feedback outperformed conventional approaches in supporting motivation, self-efficacy, deep learning, and critical thinking (Liu et al., 2023b; Kurt & Kurt, 2024; Koltovskaia et al., 2024; Ngo & Hastie, 2025).

Meta-analysis

the 30 studies selected for this systematic review, only 7 met the design criteria required to conduct a meta-analysis; therefore, this section focuses on those studies. The meta-analysis was conducted using Python (version 3.10), employing Pandas for data organization, NumPy for statistical computations, and SciPy for effect size and confidence interval estimation. Matplotlib was used to visualize results, including forest plots provided in the supplementary materials to enhance transparency.

The analyses followed standard meta-analytic procedures, including a random-effects model to account for anticipated heterogeneity in study designs, interventions, and outcome measures. This meta-analysis synthesized quantitative evidence on the impact of AI tools on critical thinking and related cognitive, metacognitive, and affective outcomes in EFL contexts. Diverse outcomes (e.g., critical thinking skills, writing performance, confidence, self-regulated learning) were included in a single model because they collectively represent interconnected constructs relevant to critical thinking development, consistent with established frameworks in cognitive psychology and SLA (Facione, 1990; Zimmerman & Schunk, 2011).

The combined effect size was 0.638 (Cohen's d), indicating a moderate to strong impact of AI tools on the outcomes studied (Table 3). The 95% confidence interval ranged from 0.580 to 0.695, suggesting robustness. However, substantial heterogeneity was observed (I^2 = 96.04%), underscoring considerable variability among studies. The Q statistic (429.30) and Tau^2 (0.360) confirmed this. Therefore, this pooled estimate should be interpreted cautiously as an indication of general trends rather than precise effects for each outcome.

Table 3

Overall meta-analysis results

Metric	Value
Combined Effect Size (Cohen's d)	0.64
95% CI Lower	0.58
95% CI Upper	0.70
Q Statistic	429.30
Degrees of Freedom (df)	17.00
I^2 (%)	96.04
Tau^2	0.36

To explore sources of heterogeneity, subgroup analyses were conducted based on outcome type, AI tool, intervention duration, and sample size. By type of outcome (Table 4), confidence-

related measures (e.g., "Confidence in Elicit", "Confidence in Consensus") showed the largest effect sizes, suggesting that AI tools are particularly effective in enhancing learners' self-confidence when interacting with academic and technological resources. In contrast, critical thinking outcomes showed moderate effect sizes, indicating positive but less pronounced impacts, while self-efficacy and self-regulated learning demonstrated smaller effects. This can be visually represented in Figure 2.
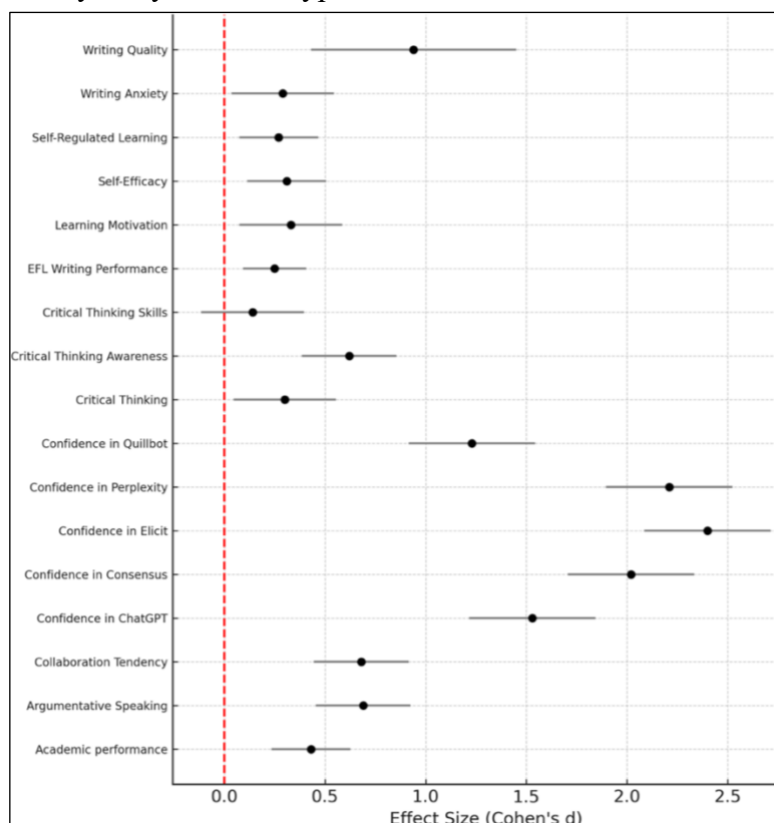
Table 4.

Meta-analysis by type of outcome

Research outcome	Combined E. S. (d)	S.E.	95% CI Lower	95% CI Upper	Q Statistic	Df
Academic performance	0.43	0.10	0.24	0.61	0	0
Argumentative Speaking	0.69	0.12	0.45	0.93	0	0
Collaboration Tendency	0.68	0.12	0.44	0.92	0	0
Confidence in ChatGPT	1.53	0.16	1.21	1.84	0	0
Confidence in Consensus	2.02	0.16	1.71	2.34	0	0
Confidence in Elicit	2.40	0.16	2.09	2.72	0	0
Confidence in Perplexity	2.21	0.16	1.89	2.52	0	0
Confidence in Quillbot	1.23	0.16	0.92	1.54	0	0
Critical Thinking	0.30	0.13	0.06	0.55	0	0
Critical Thinking Awareness	0.62	0.12	0.38	0.86	0	0
Critical Thinking Skills	0.14	0.13	-0.12	0.40	0	0
EFL Writing Performance	0.25	0.08	0.10	0.40	0.567	1
Learning Motivation	0.33	0.13	0.08	0.57	0	0
Self-Efficacy	0.31	0.10	0.12	0.51	0	0
Self-Regulated Learning	0.27	0.10	0.08	0.47	0	0
Writing Anxiety	0.29	0.13	0.04	0.53	0	0
Writing Quality	0.94	0.26	0.43	1.45	0	0

Note. $I^2=0$, $Tau^2=0$

Forest plots for each subgroup and outcome category are provided as supplementary figures, allowing for visualisation of individual study effects, confidence intervals, and pooled estimates.

Figure 2.

Forest Plot. Meta-analysis by outcome type

The meta-analysis by AI tool (Table 5) showed that generative tools like Elicit, Consensus, and Perplexity produced high effect sizes for confidence and research support tasks. Tools such as ChatGPT and Quillbot demonstrated consistent positive effects, particularly in developing writing and critical thinking. Educational tools like Mosoteach and PA-AWE showed targeted effects for self-regulated learning and writing performance (See Figure 3). These findings highlight the distinct contributions of different AI tools.
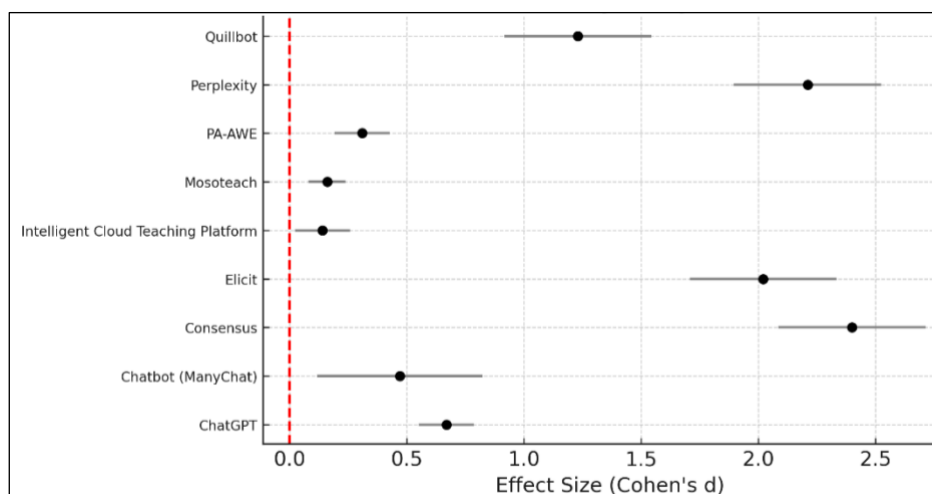
Table 5.

Meta-analysis by specific tool

AI tool	Combined E. S. (d)	S.E.	95% CI Lower	95% CI Upper	Q Statistic	Df	I² (%)	Tau²
ChatGPT	0.67	0.06	0.56	0.79	45.12	4	91.13	0.14
Chatbot (ManyChat)	0.47	0.18	0.11	0.83	3.31	0	100.00	0.11
Consensus	2.40	0.16	2.09	2.72	0.00	0	0.00	0.00
Elicit	2.02	0.16	1.71	2.34	0.00	0	0.00	0.00
Intelligent Cloud Teaching Platform	0.14	0.06	0.03	0.25	8.83	0	100.00	0.03
Mosoteach	0.16	0.04	0.07	0.24	4.06	2	50.75	0.00
PA-AWE	0.31	0.06	0.19	0.43	0.07	3	0.00	0.00
Perplexity	2,21	0.16	1.89	252	0.00	0	0.00	0.00
Quillbot	1.23	0.16	0.92	1.54	0.00	0	0.00	0.00

Figure 3.

Forest Plot. Meta-analysis by AI tool

Regarding intervention duration, short-term implementations (six weeks or less) yielded a combined effect size of 0.64 (Table 6). Due to limited data, long-term effects could not be comprehensively evaluated, highlighting a gap for future research on sustainability.

Table 6.

Meta-analysis by intervention duration

Duration	Combined E. S. (d)	S.E.	95% CI Lower	95% CI Upper	Q Statistic	Df	I² (%)	Tau²
Short-term	0.64	0.03	0.58	0.70	429.30	17.00	96.04	0.36

Finally, analysis by sample size (Table 7) revealed that medium sample sizes (31–60 participants) reported the largest combined effect size (1.48), with high heterogeneity. Small samples also showed moderate effects with high variability, while large samples yielded more conservative estimates, likely capturing greater real-world variation.

Table 7.

Meta-analysis by sample size

Sample size	Combined E. S. (d)	S.E.	95% CI Lower	95% CI Upper	Q Statistic	Df	I² (%)	Tau²
Small (≤30)	0.47	0.18	0.11	0.83	3.31	0.00	100.00	0.11
Medium (31-60)	1.48	0.06	1.36	1.60	171.93	5.00	97.09	0.66
Large (>60)	0.26	0.03	0.21	0.32	36.85	10.00	72.86	0.02

Discussion

This systematic review highlights the significant contributions of artificial intelligence (AI) tools to developing critical thinking in English as a Foreign Language (EFL) learning. The results show that AI tools, such as ChatGPT, Grammarly, QuillBot, and chatbots like Mitsuku and Replika, positively impact EFL learners' abilities to analyze, evaluate, and synthesize information. Specifically, tools that provide immediate feedback, personalized learning experiences, and structured activities have demonstrated effectiveness in fostering critical reflection and metacognitive skills. These findings align with the growing recognition of AI as a transformative force in education, as evidenced by AlTwijri and Alghizzi's review (2024), which reported that AI tools similar to those examined in the present work effectively foster collaboration and reduce speaking anxiety through interactive dialogues and question-answering activities.

An interesting trend observed in the review is the geographic concentration of studies, with 86.6% conducted in Asia, particularly in China, Indonesia, and Vietnam. This suggests that AI integration in EFL education is gaining significant traction in these regions, likely driven by advancements in AI technologies and the prioritization of English language education. However, this highlights a gap and presents opportunities for research, as similar studies do not appear to have been conducted in Latin American countries by the time the studies were selected for this review. Additionally, the temporal distribution of studies, with 66.6% published in 2024, reflects the recent surge in AI tool adoption, likely influenced by the introduction of generative AI tools like ChatGPT. This aligns with Chartr (2022) and Teubner et al. (2023), who reported the impact of ChatGPT and the broader revolution caused by generative AI tools.

While the findings highlight the potential of AI tools, challenges related to over-reliance and ethical use must be addressed. Teng's review (2024) concluded that excessive dependence on ChatGPT could hinder reflective writing and independent analytical skills. Similarly, several studies in the present work noted that students often accept AI-generated outputs without critical evaluation, risking diminished independent thought and creativity, which aligns with findings by Lai and Lee (2024). Moreover, issues such as plagiarism, academic dishonesty, and data privacy underscore the importance of fostering AI literacy among learners and educators—challenges also noted by Li et al. (2024) in their review of ChatGPT's first year of implementation. Structured interventions, such as evaluating AI outputs and engaging in group discussions, have shown promise in mitigating these challenges and promoting ethical AI usage. These findings are consistent with Chang et al. (2023), who demonstrated that coupling AI tools with detailed step-by-step procedures, self-regulated learning frameworks, and judgments of learning can optimize outcomes.

The meta-analysis results showed a moderate to strong overall effect size (Cohen's d = 0.638) of AI tools on critical thinking development. However, this was not evident in Deng and Yu (2023), who, despite explicitly addressing the topic, found minimal improvements in critical thinking. Their outcome was likely influenced by the challenges of measuring such a complex skill and the chatbot's primary emphasis on knowledge retention rather than reflective thinking. In the present work, subgroup analyses reveal that confidence-building outcomes, such as confidence in eliciting and synthesizing information, exhibited the most substantial effects. This

suggests that AI tools are effective in creating environments where learners feel empowered to engage with academic and technological resources, as also observed by Ma and Chen (2024), AlTwijri and Alghizzi (2024), and Lai and Lee (2024), who studied engagement and other affective factors.

While this review highlights the overall positive influence of AI tools on learners' critical thinking, it is important to distinguish between direct and indirect effects reported in the included studies. A subset of studies measured critical thinking explicitly, using structured argument analysis, rubric-based assessments, or specific critical thinking tasks targeting skills such as analysis, inference, and evaluation. In contrast, many other studies inferred potential gains in critical thinking indirectly through related outcomes such as increased confidence, enhanced self-regulated learning behaviours, and more reflective revision practices. Although these factors can support the development of critical thinking, they do not fully capture its multifaceted nature and should therefore be interpreted with caution.

Additionally, this review acknowledges the inherent challenges in operationalizing and measuring critical thinking consistently across EFL contexts. Critical thinking is a complex, multidimensional construct influenced by task design, cultural context, and learner background. Variability in assessment tools, indirect proxies, and differing pedagogical goals can contribute to the high heterogeneity observed in this meta-analysis. To strengthen future research, studies should prioritize the use of robust, validated instruments and consider triangulating multiple data sources (e.g., direct tests, written tasks, interviews) to better capture the breadth and depth of learners' critical thinking development.

The heterogeneity observed in the meta-analysis ($I^2 = 96.04\%$) suggests significant variability in the effectiveness of AI tools across studies. Factors such as study design, intervention duration, sample sizes, and the specific AI tools used contributed to this variability. Regarding intervention duration, the studies reviewed in this work did not provide sufficient data to comprehensively evaluate long-term interventions. This represents a limitation of the current study, as well as of studies by AlTwijri and Alghizzi (2024), Lai and Lee (2024), Deng and Yu (2023), Zhai and Wibowo (2023), and Teng (2024), emphasizing the need for future research to investigate the sustainability of AI tools' impact over extended periods. Regarding sample size, studies with medium sample sizes (31–60 participants) reported the largest effect sizes, while those with larger sample sizes showed more conservative estimates. This indicates that while AI tools are effective, their impact may vary depending on contextual and methodological factors.

Conclusion

This systematic review highlights the potential of AI tools for nurturing critical thinking when used thoughtfully alongside sound teaching practices. Teachers have a key role in helping students see AI not as a replacement for their own thinking, but as a helpful partner that can spark reflection and deeper understanding. Combining AI-generated feedback with guidance from teachers and peers can encourage students to question ideas, refine their work, and collaborate more effectively. To make this possible, professional development should prepare educators to confidently use AI in the classroom while also fostering students' independent

reasoning and creativity.

At a broader level, clear policies and ethical guidelines are needed to ensure that AI use in language learning respects data privacy, upholds academic integrity, and fits local cultural contexts. Policymakers can help by supporting fair access to AI tools in all schools and funding training that builds both technical and pedagogical literacy. Researchers should continue exploring how AI affects learning over longer periods and in diverse cultural settings, and develop clear ways to measure its true impact on students' thinking skills. Finally, institutions should back these efforts by providing practical support, updating curricula, and encouraging teachers to share their experiences and insights. By working together, teachers, policymakers, researchers, and institutions can ensure that AI becomes a valuable ally in language education, helping learners not only master a language but also think critically about the world around them.

Limitations and Future Research

This systematic review offers valuable insights into the contributions of AI tools to the development of critical thinking in EFL learning; however, several limitations must be acknowledged. First, the quality and methodological rigor of the included studies varied considerably. Although only peer-reviewed articles were selected, differences in research design, sample size, and reporting practices may have affected the reliability and comparability of the findings. Second, there is a notable lack of longitudinal data, as most interventions were short-term, limiting understanding of the sustained impact of AI tools on learners' critical thinking over time. Third, the review shows a marked geographic concentration, with the majority of studies conducted in Asian countries such as China, Indonesia, and Vietnam, and very few studies from other regions. This restricts the generalizability of the findings to global EFL contexts. Fourth, the meta-analysis revealed substantial heterogeneity, indicating significant variability in effect sizes due to differences in study design, participant characteristics, intervention duration, and types of AI tools employed. Lastly, potential publication bias cannot be ruled out, as only published studies in English or Spanish were included, and studies with positive results may have been more likely to be published, potentially overestimating the true effects.

Future research should address these limitations in two ways. First, researchers should design primary studies that adopt rigorous and standardized methodologies, use larger and more diverse samples, and implement longer intervention periods to better capture the long-term impact of AI tools on critical thinking development in the EFL context. Investigating pedagogical strategies that balance AI use with fostering independent analysis, reflection, and creativity is also essential. Cross-cultural studies in underrepresented contexts, including Latin America, Africa, and other regions, would enhance the global relevance and cultural adaptability of findings. Second, future systematic reviews and meta-analyses should broaden their scope by including grey literature such as dissertations, conference papers, and institutional reports to mitigate potential publication bias and present a more comprehensive evidence base. Employing formal publication bias tests and conducting sensitivity analyses will further strengthen the robustness of future syntheses. Additionally, developing and applying standardized quality appraisal tools and transparent reporting guidelines will improve the

methodological consistency and comparability of studies in this rapidly evolving field.

References

Abdallah, M. M. S. (2025). Using contextualised instruction through Reverso Context to develop EFL student teachers' translation skills. *Sohag University International Journal of Educational Research (SUIJER), 11*(January 2025). DOI: 10.21608/suijer.2024.311861.1011

Abu Qub'a, A., Abu Guba, M. N., & Fareh, S. (2024). Exploring the use of Grammarly in assessing English academic writing. *Heliyon, 10*(2024), e34893. https://doi.org/10.1016/j.heliyon.2024.e34893

Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer-aided pronunciation training (CAPT) in English. *Education and Information Technologies, 24*(6), 3731–3743. https://doi.org/10.1007/s10639-019-09955-7

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and strategies. *Technological Forecasting and Social Change, 199*, 123076. https://doi.org/10.1016/j.techfore.2023.123076

Al-Khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence, 5*, 100218. https://doi.org/10.1016/j.caeai.2024.100218

AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the Integration of Artificial Intelligence in English as Foreign Language Classes for Enhancing Learners' Affective Factors: A Systematic Review. *Heliyon*, 10, e31053. DOI: 10.1016/j.heliyon.2024.e31053.

Bag, H. K., & Gursoy, E. (2021). The effect of critical thinking embedded English course design to the improvement of critical thinking skills of secondary school learners. *Thinking Skills and Creativity, 41*, 100910. https://doi.org/10.1016/j.tsc.2021.100910

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability, 15*(17), 12983. https://doi.org/10.3390/su151712983

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences, 12*(17), 8427. https://doi.org/10.3390/app12178427

Bezanilla, M. J., Fernández-Nogueira, D., Poblete, M., & Galindo-Domínguez, H. (2019). Methodologies for teaching-learning critical thinking in higher education: The teacher's view. *Thinking Skills and Creativity, 33*, 1–10. https://doi.org/10.1016/j.tsc.2019.100584

Bozkurt, A., & Sharma, R. C. (2023). Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning. *Asian Journal of Distance Education, 18*(1). Retrieved June 4, 2023, from http://asianjde.com/ojs/index.php/AsianJDE/article/view/714

Chang, D. H., Lin, M. P. -C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability, 15*(17), 12921. https://doi.org/10.3390/su151712921

Chartr. (2022). ChatGPT: The AI bot taking the tech world by storm. https://www.chartr.co/stories/2022-12-09-1-chatgpt-taking-the-tech-world-by-storm

Chen, X., Xie, H., & Hwang, G. J. (2020). A multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers & Education: Artificial Intelligence*. https://doi.org/10.1016/j.caeai.2020.100005

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2022). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Dai, K., & Liu, Q. (2024). Leveraging artificial intelligence (AI) in English as a foreign language (EFL) classes: Challenges and opportunities in the spotlight. *Computers in Human Behavior, 159*, 108354. https://doi.org/10.1016/j.chb.2024.108354

Darwin, R., Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education, 11*(1), 1–18. https://doi.org/10.1080/2331186X.2023.2290342

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education, 153*, 104094. https://doi.org/10.1016/j.compedu.2020.104094

Deng, X.; Yu, Z. (2023). A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education. *Sustainability*, *15*, 2940. DOI: 10.3390/su15042940

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D.C. Heath and Company.

Domínguez Romero, E., Sanhueza-Campos, C., Díaz-Vargas, C., Vildósola Campos, M., & Sáez Carrillo, K. (2025). Exploring the Use of Epistemic and Effective Stance Strategies in the EFL Classroom: Evidence from Reflective Writing. *Sage Open*, *15*(4). https://doi.org/10.1177/21582440251378837

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. American Philosophical Association. https://www.insightassessment.com/wp-content/uploads/ia/pdf_files/Consensus.pdf

Fitria, T. N. (2024). Using chatbot-based artificial intelligence (AI) for writing an English essay:

The ability of ChatGPT, Perplexity AI, and ChatSonic. *Journal of Language Intelligence and Culture, 6*(2), 103–128. https://doi.org/10.35719/jlic.v6i2.139

Fuentes-Riffo, K., Salcedo-Lagos, P., Sanhueza-Campos, C., Pinacho-Davidson, P., Friz-Carrillo, M., Kotz-Grabole, G., & Espejo-Burkart, F. (2023). The Influence of Gamification on High School Students' Motivation in Geometry Lessons. *Sustainability, 15*(21), 15615. https://doi.org/10.3390/su152115615

Garg, S., Ahmad, A., & Madsen, D. Ø. (2024). Academic writing in the age of AI: Comparing the reliability of ChatGPT and Bard with Scopus and Web of Science. *Journal of Innovation & Knowledge, 9*, Article 100563. https://doi.org/10.1016/j.jik.2024.100563

Glaser, N. (2023). Exploring the potential of ChatGPT as an educational technology: An emerging technology report. *Technology, Knowledge and Learning, 28*(4), 1945−1952. https://doi.org/10.1007/s10758-023-09684-4

He, J., Ma, T., & Zhang, Y. (2023). Design of blended learning mode and practice community using intelligent cloud teaching. *Education and Information Technologies, 28*, 10593–10615. https://doi.org/10.1007/s10639-023-11606-x

Hieu, H. H., & Thao, L. T. (2024). Exploring the impact of AI in language education: Vietnamese EFL teachers' views on using ChatGPT for fairy tale retelling tasks. *International Journal of Learning, Teaching and Educational Research, 23*(3), 486–503. https://doi.org/10.26803/ijlter.23.3.24

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. McGill.

Hossain, K. I. (2024). Reviewing the role of culture in English language learning: Challenges and opportunities for educators. *Social Sciences & Humanities Open, 9*, 100781. https://doi.org/10.1016/j.ssaho.2023.100781

Hsu, Y. C., & Ching, Y. H. (2023). Generative artificial intelligence in education, part one: The dynamic frontier. *TechTrends, 67*, 603–607. https://doi.org/10.1007/s11528-023-00863-9

Hu, W.-C., & Škultéty, R. (2024). Unlocking the learning potential: ChatGPT as a virtual platform for cross-interaction in English language learning. *Engineering Proceedings, 74*(1), 59. https://doi.org/10.3390/engproc2024074059

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society, 26*(1), 112–131. https://www.jstor.org/stable/48707971.

Kartal, G. (2024). The influence of ChatGPT on thinking skills and creativity of EFL student teachers: A narrative inquiry. *Thinking Skills and Creativity, 51*, 101332. https://doi.org/10.1016/j.tsc.2024.101332

Klimova, B., & Luz de Campos, V. P. (2024). University undergraduates' perceptions on the use of ChatGPT for academic purposes: Evidence from a university in Czech Republic. *Cogent Education, 11*(1), 2373512. https://doi.org/10.1080/2331186X.2024.2373512

Koh, E., Zhang, L., Lee, A. V. Y., & Wang, H. (2024). Revolutionizing word clouds for teaching and learning with generative artificial intelligence: Cases from China and Singapore. *IEEE Transactions on Learning Technologies, 17*, 1390–1401. https://doi.org/10.1109/TLT.2024.3385009

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 45*, 100450. https://doi.org/10.1016/j.asw.2020.100450

Kuhail, M. A., Alturki, N., Alramlawi, S., et al. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technology, 28*, 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research, 23*, Article 24. https://doi.org/10.28945/5370

Lai, W. Y. 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, 7*, 100291. DOI: 10.1016/j.caeai.2024.100291.

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence, 7*(2024), 100266. DOI: 10.1016/j.caeai.2024.100266

Li, C., & Jianbin, H. (2019). Application of mobile teaching mode based on MosoTeach: Benefits and challenges. *International Journal of Applied Linguistics and English Literature, 8*(3), 138–146. https://doi.org/10.7575/aiac.ijalel.v.8n.3p.138

Liang, W., & Wu, Y. (2024). Exploring the use of ChatGPT to foster EFL learners' critical thinking skills from a post-humanist perspective. *Thinking Skills and Creativity, 52*, 101645. https://doi.org/10.1016/j.tsc.2024.101645

Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society, 23*(1), 78–92.

Lipman, M., Sharp, A. M., & Oscanyan, F. S. (1980). *Philosophy in the classroom* (2nd ed.). Temple University Press.

Liu, C.-C., Liu, S.-J., Hwang, G.-J., Tu, Y.-F., Wang, Y., & Wang, N. (2023). Engaging EFL students' critical thinking tendency and in-depth reflection in technology-based writing contexts: A peer assessment-incorporated automatic evaluation approach. *Education and Information Technologies, 28*, 1–26. https://doi.org/10.1007/s10639-023-11697-6

Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments, 31*(9), 5614–5632. https://doi.org/10.1080/10494820.2021.2012812

Lu, D., & Xie, Y. N. (2022). Critical thinking cultivation in TESOL with ICT tools: A systematic review. *Computer Assisted Language Learning, 37*(1–2), 222–242. https://doi.org/10.1080/09588221.2022.2033788

Ma, Y., & Chen, M. (2024). AI-empowered applications effects on EFL learners' engagement in the classroom and academic procrastination. *BMC Psychology, 12*, Article 739. https://doi.org/10.1186/s40359-024-02248-w

Macinska, S., & Vinkler, A. (2023). The role of AI in language learning: Research evidence and strategies. *Cambridge University Press*. https://www.cambridge.org/sites/default/files/media/documents/The_Role_of_AI_in_Language_Learning_Research_Evidence_and_Strategies.pdf

Marzuki, W., Rusdin, U., Darwin, D., & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), 1–17. https://doi.org/10.1080/2331186X.2023.2236469

McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence, 171*(18), 1174–1182.

Mosqueda, E. (2024). La inteligencia artificial como aliada del aprendizaje y el pensamiento crítico. *Revista Mexicana de Bachillerato a Distancia, 32*(16). https://doi.org/10.22201/cuaieed.20074751e.2024.32.89555

Muthmainnah, M., Seraj, P. M. I., & Oteir, I. (2022). Playing with AI to investigate human-computer interaction technology and improving critical thinking skills to pursue 21st century age. *Education Research International, 2022*, Article 6468995. https://doi.org/10.1155/2022/6468995

Newell, A., & Simon, H. (1956). The logic theory machine: A complex information processing system. *IEEE Transactions on Information Theory, 2*(3), 61–79.

Ngo Cong-Lem, N., Tran, T. N., & Nguyen, T. T. (2024). Academic integrity in the age of generative AI: Perceptions and responses of Vietnamese EFL teachers. *Teaching English with Technology, 24*(1), 28–47. https://doi.org/10.56297/FSYB3031/MXNB7567

Ngo, T. N., & Hastie, D. (2025). Artificial intelligence for academic purposes (AIAP): Integrating AI literacy into an EAP module. *English for Specific Purposes, 73*, 101288. https://doi.org/10.1016/j.esp.2024.09.001

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.

Pan, M., Wang, J., & Wang, J. (2023). Application of artificial intelligence in education: Opportunities, challenges, and suggestions. In *2023 13th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 623–627). IEEE. https://doi.org/10.1109/ITME60234.2023.00130

Patamaporn-Thaiposria, P.W. (2015). Enhancing students' critical thinking skills through teaching and learning by inquiry-based learning activities using social network and cloud computing. *Procedia - Social and Behavioral Sciences, 174*, 2137–2144. https://doi.org/10.1016/j.sbspro.2015.01.875

Paul, R., & Elder, L. (2003). *The thinker's guide to analytic thinking: How to take thinking apart and what to look for when you do*. Foundation for Critical Thinking. https://www.criticalthinking.org/files/SAM_Analytic_Think2007b.pdf

Pawar, A. M., Desai, R., & Thakur, B. (2024). From tedious to targeted: Optimizing oral cancer research with Consensus AI. *Oral Oncology Reports, 10*, Article 100383. https://doi.org/10.1016/j.oor.2024.100383

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*(1), 22. https://doi.org/10.1186/s41039-017-0062-8

Salcedo-Lagos, P., Pinacho-Davidson, P., Pinninghoff, J. M. A., Contreras, R. A., Fuentes-Riffo, K., & Carrillo, M. F. (2024). Optimizing didactic sequences with artificial intelligence: Integrating Bloom's taxonomy and emotion in the selection of educational technologies. In J. M. Ferrández Vicente, M. Val Calvo, & H. Adeli (Eds.), *Artificial intelligence for neuroscience and emotional systems* (IWINAC 2024) (*Lecture Notes in Computer Science*, Vol. 14674). Springer, Cham. https://doi.org/10.1007/978-3-031-61140-7_48

Seraj, P. M. I., Habil, H., Hasan, M. K., & Sharmin, F. (2021). Exploring EFL teachers' perception on readiness to use smartphones for teaching oral English communication skills at tertiary level. *Malaysian Journal, 45*(4). https://doi.org/10.61871/mj.v45n4-3

Shen, M., & Chiu, T. (2019). EFL learners' English speaking difficulties and strategy use. *Education and Linguistics Research, 5*(2), 88. https://doi.org/10.5296/elr.v5i2.15333

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Tang, T., Vezzani, V., & Eriksson, V. (2020). Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity, 37*, 100696. https://doi.org/10.1016/j.tsc.2020.100696

Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies, 6*(3), 36–57. DOI: 10.58304/ijts.20240304.

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of ChatGPT et al.: The prospects of large language models. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-023-00795-x

Tseng, Y.-C., & Lin, Y.-H. (2024). Enhancing English as a Foreign Language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of e-Learning, 21*(5), 558–570. https://doi.org/10.34190/ejel.21.5.3329

Waluyo, B., & Kusumastuti, S. (2024). Generative AI in student English learning in Thai higher education: More engagement, better outcomes? *Social Sciences & Humanities Open, 10*(1), 101146. https://doi.org/10.1016/j.ssaho.2024.101146

Whitfield, S., & Hofmann, M. A. (2023). Elicit: AI literature review research assistant. *Public Services Quarterly, 19*(3), 201–207. https://doi.org/10.1080/15228959.2023.2224125

Wu, S., & Shafait, Z. (2024). Assessing EFL (English as foreign language) education for sustainable development: Exploring the cultural teaching literature. *European Journal of Investigation in Health, Psychology and Education, 14*(8), 2282–2299. https://doi.org/10.3390/ejihpe14080152

Yan, W., Lowell, V. L., & Yang, L. (2024). Developing English language learners' speaking skills through applying a situated learning approach in VR-enhanced learning experiences. *Virtual Reality, 28*, 167. https://doi.org/10.1007/s10055-024-01061-5

Yan, X., Xiao, Y., & Jin, Y. (2024). Generative large language models explained. *IEEE Computational Intelligence Magazine, 19*(4), 45–46. https://doi.org/10.1109/MCI.2024.3431454

Yin, J., Goh, T. T., Yang, B., & Xiaobin, Y. (2020). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research, 58*(1), 1–24. https://doi.org/10.1177/0735633120952067

Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. *CercleS – European Journal of Language Policy and Practice, 14*(1), 1–18. https://doi.org/10.1515/cercles-2023-0006

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence, 4*, 100134. DOI: 10.1016/j.caeai.2023.100134

Zhang, R., Zou, D., & Cheng, G. (2023). Chatbot-based training on logical fallacy in EFL argumentative writing. *Innovation in Language Learning and Teaching, 17*(5), 932–945. https://doi.org/10.1080/17501229.2023.2197417

Zimmerman, B. J., & Schunk, D. H. (2011). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Routledge.

Biodata

Marcela Vildósola Campos is an English teacher based in Chile. She holds a Master's degree in Innovation in English Teaching, Learning, and Assessment, and is currently pursuing a Ph.D. in Applied Linguistics at the University of Concepción. Her research explores vocabulary acquisition and critical thinking in EFL contexts.
ORCID: https://orcid.org/0000-0001-5826-4736

Pedro Salcedo Lagos is Full Professor at the Universidad de Concepción, where he directs the Department of Research Methodology and Educational Informatics. His research focuses on Artificial Intelligence, Affective Computing, ICT in Education, Digital Transformation, and Neurocomputing. He also serves on multiple doctoral committees and the Fondecyt study group.
ORCID: https://orcid.org/0000-0002-1741-714X

Gabriela Kotz holds a PhD in Linguistics and teaches German and interculturality at Universidad de Concepción. She also lectures in the Linguistics Postgraduate Program. Her research focuses on lexicology, discourse, and AI. Recent publications explore emotional language and lexical studies from a linguistics and intercultural communication perspective.
ORCID: https://orcid.org/0000-0001-5300-7669

Cristián Sanhueza Campos, PhD in Linguistics, is Associate Professor at Universidad de Concepción. His research in applied linguistics focuses on L2 vocabulary and pronunciation with technology. He also studies initial teacher education from intercultural perspectives. He teaches in postgraduate programs and coordinates the Interdisciplinary Research Group in Education (GIIE).
ORCID: https://orcid.org/0000-0002-7924-5910