Contemporary Educational Technology, 2026, **18**(3), ep676
ISSN: 1309-517X (Online)

OPEN ACCESS

Review Article

Mapping research on ChatGPT and LLMs in EFL writing: A scoping review with bibliometric analysis

Georgia Andreou [1]
0000-0001-6433-2301

Pinelopi Christani [1*]
0009-0003-8158-5695

[1] Department of Special Education, University of Thessaly, Volos, GREECE
[*] Corresponding author: pchristani@uth.gr

Citation: Andreou, G., & Christani, P. (2026). Mapping research on ChatGPT and LLMs in EFL writing: A scoping review with bibliometric analysis. *Contemporary Educational Technology, 18*(3), Article ep676. https://doi.org/10.30935/cedtech/19080

ARTICLE INFO

Received: 4 Dec 2025

Accepted: 6 Jul 2026

ABSTRACT

The present paper investigates the current trends of second language learning and teaching mediated by generative artificial intelligence (GenAI) tools, like ChatGPT and other large language models within the English as a foreign language (EFL) writing context. Scopus, which is a major database, was carefully researched using targeted keywords. All the inclusion and exclusion criteria were implemented consistently, and a total set of 43 articles were finally analyzed to synthesize the present review. The PRISMA method was implemented to ensure transparency and reliability, whereas the bibliometrix R-tool was utilized to generate clear and vivid visual depictions of the bibliometric data. This study adopts a complementary dual-method design, combining a scoping review approach with a bibliometric analysis to investigate the use of GenAI in EFL writing instruction. The scoping review serves as the primary analytical component of the study, identifying major pedagogical themes, applications, and challenges related to artificial intelligence-supported writing enhancement. The bibliometric analysis is used to contextualize the structure and evolution of the field by mapping its intellectual structure, publication evolution, influential publications, productive authors, institutional affiliations and keyword trends. The synthesis of the reviewed studies indicates that AI-assisted feedback is frequently associated with improvements in grammar, vocabulary, coherence, and learner engagement. Several studies also highlight challenges, including concerns about academic integrity, students' overreliance on AI, and the necessity for teacher guidance. By systematically synthesizing these results, this review offers a clear picture of current practices, emerging trends, and research gaps, providing insights for researchers, educators and policymakers in the EFL domain. This study has certain limitations such as filtering constraints and strict time limits, but there are some useful guidelines that researchers can utilize in their future research.

Keywords: generative artificial intelligence, large language models, ChatGPT, English as a foreign language, language learning, writing skills, scoping review

INTRODUCTION

The launch of ChatGPT in November 2022 fundamentally transformed language teaching and learning (Kasneci et al., 2023; Kohnke et al., 2023). Language acquisition has been on the threshold of major changes that are bound to cause a breakthrough in the way lessons are conducted, and how learning is facilitated in second language (L2) environments. Rooted in Vygotsky's (1978) socio-cultural theory of learning and constructivism frameworks, the use of artificial intelligence (AI) tools provides practical implementation for targeted English as a foreign language (EFL) teaching (Zawacki-Richter et al., 2019). In this respect, the present study aims to provide insights into the current trends in English as a second/foreign language written instruction, focusing primarily on the affordances of ChatGPT, the most recognized and accessible generative artificial intelligence (GenAI). This can be attributed to the fact that ChatGPT is accessible via the students'

personal devices, and its capabilities can be both utilized within structured institutional classrooms and in informal, independent learning environments (Yu, 2024).

Given the rapidly expanding body of literature and growing academic interest in this particular domain, a scoping review design was selected, because it allows for a broader scope of literature to be explored, and it provides a comprehensive overview of all the multifaceted issues raised by the researchers (Munn et al., 2018). The PRISMA methodology was adopted to ensure a reliable, transparent as well as a thorough and detailed analysis of the topic (Page et al., 2021). Furthermore, the bibliometrix R-package tool (Aria & Cuccurullo, 2017), which is an R-based package, was applied on the final results of the PRISMA dataset, to generate precise, reproducible scientific visualizations such as thematic maps, scientific production trends, and global citation networks.

Although the study integrates both scoping review and bibliometric approaches, the primary objective is to synthesize pedagogical evidence related to the use of GenAI in EFL writing. The bibliometric component is employed as a complementary method to map publication trends, influential contributions, and thematic developments within the field. Conversely, the scoping review approach is employed to examine how AI has been applied to support EFL learners' writing development and to identify the main pedagogical themes, challenges, and research gaps reported in literature. By merging these two approaches, the study offers both a macro-level mapping of the scientific landscape and a micro-level synthesis of educational insights regarding GenAI-supported EFL written instruction.

Writing skill development was selected as the primary objective of this study, as it represents one of the most extensively researched areas in EFL learning, with a substantial body of literature suggesting that GenAI tools may contribute to the development of writing proficiency. Although situated within the broader field of GenAI, the specific study focuses specifically on ChatGPT and it is closely related to large language models (LLMs), as they currently constitute the most prominent and widely investigated applications of GenAI in EFL writing research. Among these tools, ChatGPT has emerged as one of the most accessible and frequently used platforms in educational settings. Empirical studies indicate that students tend to perceive AI-based writing tools as effective support mechanisms for writing tasks. For instance, large-scale evidence suggests that a substantial proportion of EFL learners report a preference for tools such as ChatGPT, Grammarly, and QuillBot, citing their usefulness in improving accuracy and facilitating writing processes (Alkhawaldeh & Khasawneh, 2023).

Recent studies have further explored the impact of ChatGPT on scholarly communication, student engagement, and EFL writing proficiency (Adhikari et al., 2025; Kawinkoonlasate, 2025; Mohammad et al., 2025; Pum, 2025), highlighting both the opportunities and challenges of integrating GenAI into higher education. The most well documented gains of ChatGPT on writing skills are reported to be grammar precision, clarity of expression and content organization (Adhikari et al., 2025; Poláková & Ivenz, 2024). ChatGPT's feedback has also been reviewed in accordance with its reliability and effectiveness compared to human raters, proving to be a high-rate reliable equivalent (Jmaiel et al., 2025; Li et al., 2024).

Driven by the rapid growth of research interest in GenAI and LLMs within educational contexts, several bibliometric studies have recently examined the role of AI within educational technology. These studies have primarily explored the use of intelligent systems across broad educational contexts rather than focusing specifically on EFL education. In addition, recent bibliometric analyses have identified major themes in AI-assisted educational processes, including automated feedback, autonomous learning, and intelligent tutoring systems that can support learners without constant instructor intervention. While these studies provide valuable insights into the integration of GenAI technologies within educational environments, they tend to address general educational practices rather than examining the distinctive pedagogical characteristics and challenges of EFL programs. As a result, the specific contribution of LLMs to language learning, particularly their role in supporting the development of learners' writing skills, remains insufficiently explored.

To address these research gaps, the present study conducts a comprehensive scoping and bibliometric analysis of LLMs in EFL writing education. Specifically, the study aims to map the evolution of this emerging research field, identify major publication trends and research themes, and outline future directions for the use of GenAI in EFL writing education. The research questions (RQs) that the present paper addresses are:

RQ1: *What are the main pedagogical applications, benefits, and challenges of using LLMs/ChatGPT to support EFL learners' writing skills as reported in the literature?*

RQ2: *What are the bibliometric characteristics of research on LLMs/ChatGPT in EFL writing in terms of influential publications, authors, affiliations, and keyword trends?*

THEORETICAL BACKGROUND

The integration of GenAI in EFL education is increasingly examined through established learning theories that explain how digital technologies support language acquisition and learner engagement. Constructivist, socio-cultural and cognitive learning theories offer foundational frameworks for understanding how learning is or should be facilitated. This can be reinforced through the use of digital tools and AI apps in the learning environment.

Constructivism and Differentiated Scaffolding

Specifically, constructivism regards learning as an active construction of knowledge and interaction resulting from learners being actively engaged in various meaningful activities. In the contemporary world of AI-supported learning processes, constructivist theories are accomplished through the use of tools and apps that promote personalization and self-centered experiences (Luckin et al., 2016). Recent research suggests that AI-supported writing tools promote constructivist learning by enabling learners to actively construct knowledge through drafting and feedback processes (Huang & Mizumoto, 2024; Li et al., 2024).

Vygotsky's (1978) zone of proximal development (ZPD) is another theory that can be supported by the use of GenAI technologies in the learning environment. LLMs, such as ChatGPT can be adapted to each learner's specific needs and pace of learning, thereby providing customized scaffolding. As a conversation chatbot, ChatGPT responds to prompts imposed by the user, by differentiating its responses based on the inserted inquiries. It offers immediate personalized feedback through an interactive, collaborative dialogue which serves as a foundation for differentiated instruction. Consequently, the main characteristics of GenAI tools that support the constructivist theories of learning is that learners themselves assume greater control over their instructional means, taking the lead in decision-making processes. as they seek information, and construct knowledge in collaboration with peers and platforms.

The Socio-Cultural Theory and Mediated Interaction

The socio-cultural theory of Vygotsky (1978) further contextualizes the incorporation of AI into language learning processes by asserting that learning is mainly a social process that presupposes interaction with other people, intertwined with cultural and linguistic elements. Within a person's ZPD, learning is co-constructed through interaction with others who have acquired a higher level of learning. This interaction is jointly connected to the cultural context of the people that come into contact, creating the need for the educational process to be underlined with the norms of social discourse. In this process, mediation can be possible through tools, i.e., intelligent systems, applications and platforms that can be adapted to the learners' specific needs, and conversational agents, like intelligent chatbots (i.e., ChatGPT) (Holmes et al., 2019). Acting as a medium, those tools can provide guidance to individual learners by providing comprehensible feedback and simulating collaboration.

Similarly, socio-cultural perspectives highlight the role of mediated interaction in language development, which AI conversational agents such as ChatGPT can facilitate through dialogic feedback and scaffolding mechanisms (Holmes et al., 2019; Luckin et al., 2016). Ultimately, since language learning relies heavily on immersion with a responsive, social, dialogic and cultural environment, these intelligent systems serve as a vital cognitive partner that mimic the interactive norms of social discourse.

These theoretical perspectives provide a conceptual framework for understanding how AI systems may support EFL learning processes. Within this context, the present study employs a dual scoping and bibliometric review methodology to map how recent research has examined AI-assisted learning, classifying the main themes in literature, and the emerging trends across the field.

METHODOLOGY

RQ1. Scoping Review – PRISMA Method

This study employs a two-stage methodological approach that integrates a scoping review and a bibliometric analysis. First, a scoping review was conducted to qualitatively and quantitively examine how LLMs, with a particular focus on ChatGPT, have been deployed to support L2 writing development (RQ1). This stage also synthesizes core pedagogical issues, instructional conditions and possible risks. Secondly, bibliometric techniques were applied to map the intellectual and structural characteristics of the field, including publication trends, citation distributions, institutional affiliations, and keyword networks. The integration of these two approaches allows both a macro-level overview of the research landscape and a thematic understanding of GenAI applications in EFL writing instruction.

This scoping review follows the preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews (PRISMA-ScR) guidelines (Tricco et al., 2018), to ensure reliability, accuracy, transparency and replication of the present analysis. Following the PRISMA guidelines this review was based on the main elements of the RQ, theoretical background, selection criteria, table of the selection process, and analysis of the final results.

To secure peer-reviewed literature and rigid academically accepted results, the database of Scopus was selected. A strict publication timeline constraint for the years 2022-2025 was applied. This boundary captures the rapid surge in research following the launch of ChatGPT in November 2022. The search strategy was designed to align with the study's focus on GenAI applications in EFL writing. Specifically, keywords were selected to capture three main conceptual domains of AI and GenAI tools, L2 learning contexts, and writing skills. The search was executed across the TITLE-ABS-KEY fields between 10th and 15th May 2025, using the following exact query: (TITLE-ABS-KEY ("Artificial intelligence" OR "AI" OR "Generative AI" OR "GenAI" OR "ChatGPT" OR "chat gpt") AND TITLE-ABS-KEY ("language learning" OR "language acquisition" OR "foreign language" OR "language education") AND TITLE-ABS-KEY ("writing")).

The initial database search yielded a set of 489 results. Systematized automation tools were employed in order for irrelevant data to be excluded. Specifically, a total of 255 articles met the criteria of the English language and the document type (journal articles) for eligibility reasons. The following subset included 105 entries referring to arts & humanities and social sciences topics excluding any computer science or medical entries, ensuring the relevance to the topic. To ensure a focused and coherent dataset, an additional filtering step was applied whereby the scope was restricted to an explicit reference to ChatGPT or similar LLMs in the title, abstract, or author keywords. This step reflects the study's emphasis on LLM-based GenAI tools, with ChatGPT serving as the most prominent and widely investigated example in literature. Studies examining non-LLM GenAI tools were excluded because such tools are often rule-based or domain-specific and do not generate natural language (Liu et al., 2025; Minaee et al., 2025) in a way that aligns with our review's focus. This distinction ensures a coherent and meaningful mapping of evidence regarding ChatGPT/LLMs in EFL writing. Finally, six more entries had no valid DOIs, making the full text inaccessible. After the automation exclusion was completed a total of 74 articles were screened.

During the screening phase, each researcher worked independently. To ensure inter-rater-reliability a pilot data extraction scheme was implemented, using 5 sample papers per researcher. Any discrepancies were resolved through consensus or the intervention of a third independent researcher. Generally, we focused our research on study characteristics (i.e., authors, year of publication, country the study took place), participants (i.e., number or participants and level of English), intervention (i.e., type of GenAI tool, length of the intervention), main results (accuracy, reliability of the results, scores in writing, basic limitations), methodological features (i.e., study design, methods and types of tests to lead the analyses). If there was any of this information missing or misleading, we discussed the options carefully and reached a consensus after pre-defining some criteria. In case of complete lack of information, we did not use the equivalent data in our analysis.

Screening the full texts resulted in further exclusions. Three entries that referred to the acquisition of a L2 other than English were eliminated since they did not align with the scope of the present study. In addition, two more articles with reference to GenAI tools, and not specifically to ChatGPT had to be excluded from the

Table 1. Summary of inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Published English	Documented in any other languages
Peer-reviewed primary journal articles	Literature reviews, book chapters, etc.
Quantitative, mixed-methods, quasi-experimental design	Purely qualitative methodology
English as second/foreign language	Non-English second language acquisition
ChatGPT	Non LLM digital tools
Writing skills	Other skills

final dataset. Literature reviews were further excluded so as not to affect the results and the discussion, and only primary sources were included. To maintain a focused analytical scope, the study included quantitative, quasi-experimental, and mixed-methods designs that report measurable aspects of EFL writing performance and AI-supported interventions. Purely qualitative studies were excluded, as the primary aim of the scoping synthesis was to identify patterns in reported writing-related outcomes and intervention effects. This decision does not diminish the importance of qualitative research, which offers valuable insights into classroom practices, learner experiences, and ethical considerations. Rather, it reflects a methodological choice to ensure comparability across studies with respect to reported outcomes.

The inclusion and exclusion criteria are summarized in **Table 1**. Following this rigorous process, a final sample of 43 articles met the eligibility criteria and were selected for full synthesis.

It is important to acknowledge that qualitative studies constitute a significant part of the literature on AI in EFL education, particularly in relation to classroom implementation, learner experiences, and ethical considerations. Although such studies fall outside the present study's analytical scope, they remain essential for a comprehensive understanding of AI-supported language learning and should be considered in future research.

RQ2. Bibliometric Analysis

To complement the qualitative synthesis, we conducted a bibliometric analysis using the bibliometrix R-package and its graphical interface Biblioshiny within the R statistical environment. Bibliometrics provides a comprehensive framework for quantitative evaluation of scientific literature, allowing for the systematic exploration of publication patterns, citation structures, and thematic relationships across the selected studies. We utilized the most recent available version of the bibliometrix R-package at the time of the study.

The bibliometric workflow followed several stages. First, the bibliographic data retrieved from the selected database was exported in compatible formats and imported into the R-package bibliometrix. Data cleaning procedures were applied to remove duplicates and standardize author names, keywords, and source titles where necessary. Subsequently, we calculated descriptive bibliometric indicators, including publication trends, source productivity, and citation metrics. For network analyses, we generated co-occurrence and co-authorship networks to examine relationships among authors, keywords, and thematic clusters within the literature.

Regarding the analytical parameters, we applied full counting when calculating occurrences and link strengths to ensure that each author or keyword occurrence contributed equally to the network structure. For keyword co-occurrence analysis, we established a minimum occurrence threshold to focus on the most relevant and frequently appearing terms. Normalization procedures were applied to network matrices to reduce potential bias caused by highly prolific sources and to improve the interpretability of the resulting maps. The generated networks and thematic structures were visualized using Biblioshiny's built-in plotting functions, which graphically represent relationships among the key elements of literature**.**

RESULTS

Findings From the Scoping Review (RQ1) – PRISMA ScR Flow Diagram

All this screening and selection process is carefully imprinted in the following PRISMA flow diagram (**Figure 1**), representing all the processes that were implemented in the final data set selection.

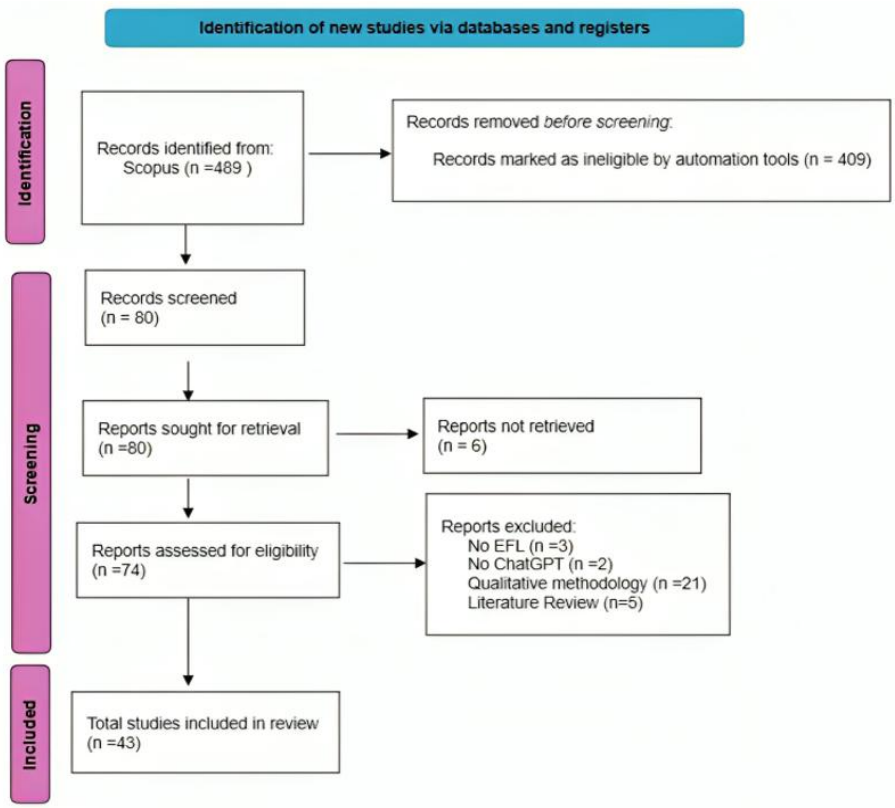


Figure 1. PRISMA flow diagram (Page et al., 2021)

Table of results

In this section the dataset of the 43 results included in the present review is presented. **Table 2** details key characteristics aligned with the PICO framework, which are outlined in the subsequent rows, namely the context, number of participants, type of task(s) employed, the GenAI tool utilized, research design and a summary of the major outcomes. To organize and analyze the extracted data, a structured data extraction matrix was developed, allowing for the systematic comparison of study characteristics and outcomes. It should be noted that a comprehensible table with more detailed information on the final dataset is presented in **Appendix A**.

Table 2. Results – Concise table with the key characteristics included in the review with PICOS

No	Context	Participants	Task type	GenAI tool	Research design	Key outcome
1	Indonesian higher education	288 students	Text comprehension	ChatGPT	Quantitative	Students demonstrated lower comprehension compared to AI-generated responses
2	Saudi university EFL	48 students	SLA language skills	ChatGPT	Quantitative	ChatGPT useful for SLA topics but occasional inaccuracies observed
3	Saudi university EFL	112 students	Academic writing feedback	ChatGPT	Mixed methods	AI feedback immediate but teacher feedback perceived as more useful
4	Japanese EFL learners	33 students	Editing and proofreading	ChatGPT	Mixed methods	Improved clarity and cohesion in writing
5	Jordan higher education	Not reported	General writing support	ChatGPT	Quantitative	Benefits include time-saving and accessibility; concerns about critical thinking and misinformation
6	Saudi EFL context	Not reported	Writing error correction	ChatGPT	Mixed methods	High accuracy in grammar correction but weaker performance in connectors and structure
7	Saudi university EFL	52 students	Writing productivity	ChatGPT	Quantitative	AI supported idea generation and clarity but raised overreliance concerns

Table 2 (Continued).

No	Context	Participants	Task type	GenAI tool	Research design	Key outcome
8	Saudi EFL context	546 students	Creativity and writing support	ChatGPT	Descriptive	AI enhanced creativity and vocabulary but raised concerns about dependency
9	Iranian university EFL	68 students	Writing feedback integration	ChatGPT	Mixed methods	Personalized AI feedback improved writing proficiency
10	Kazakhstan TEFL students	334 students	Academic writing	ChatGPT	Mixed methods	High perceived usefulness but concerns about plagiarism and reliability
11	Spanish teacher education	58 students	Language learning tasks	ChatGPT	Mixed methods	Positive perceptions of usefulness and behavioral intention
12	Bangladesh EFL learners	159 students	Academic writing	ChatGPT	Quantitative	AI functioned as cognitive and emotional support
13	Spanish CLIL courses	205 students	AI feedback for writing	ChatGPT	Mixed methods	Feedback perceived as immediate and useful but sometimes complex
14	Chinese pre-service teachers	134 students	Language teaching tasks	ChatGPT	Quantitative	Improved grammar and writing support but concerns about academic integrity
15	Indonesian EFL learners	67 students	Argumentative speaking	ChatGPT	Quasi-experimental	Improved argumentation skills and collaboration
16	Vietnamese university EFL	31 students	Academic writing	ChatGPT	Mixed methods	Improved writing skills but risk of shallow learning with overuse
17	Asia-Pacific ENL context	91 students	Essay feedback	ChatGPT	Quasi-experimental	AI feedback is feasible but blended AI-teacher model recommended
18	Chinese EFL context	102 students + 4 teachers	Writing feedback	ChatGPT	Controlled experiment	"AI + teacher" approach recommended
19	Japanese EFL learners	35 students	Technology acceptance	ChatGPT	Mixed methods	Positive relationship between motivation and AI use
20	Japanese EFL learners	80 students	Writing activities	ChatGPT	Mixed methods	AI usage associated with increased motivation
21	Saudi undergraduate EFL	142 students	Writing tasks	ChatGPT	Quantitative	Improved organization and vocabulary; mixed results for anxiety reduction
22	Korean EFL learners	79 students	Self-revision and error correction	ChatGPT	Controlled experiment	Reduced grammatical errors but prepositional errors remained
23	Chinese EFL learners	104 students	Model text feedback	ChatGPT	Quasi-experimental	Improved text quality but concerns about stylistic naturalness
24	Macedonian university EFL	180 students	Language learning tasks	ChatGPT	Mixed methods	Useful for writing and grammar but skepticism about creativity
25	Asia-Pacific ELL writing	119 essays	Automated essay scoring	Multiple LLMs	Quantitative	Model performance varies across systems
26	Czech university EFL	110 students	Writing feedback	ChatGPT	Mixed methods	Improved writing skills among Gen Z learners
27	Turkish EFL learners	11 students	Writing support	ChatGPT	Controlled experiment	Useful for formal writing but less effective for informal writing
28	Indonesian university EFL	106 students	Writing tasks	ChatGPT	Mixed methods	Practical benefits but need for institutional policies
29	Indonesian EFL course	95 students	Writing assistance	ChatGPT	Mixed methods	Increased acceptance of AI in language learning
30	Thai EFL context	50 students	Writing development	ChatGPT	Mixed methods	Teachers remain essential for critical thinking and ethics
31	Korean EFL college	44 students	Narrative writing	ChatGPT	Quantitative	Improved writing fluency and complexity
32	Egyptian EFL assessment	100 essays	Writing assessment	ChatGPT	Quantitative	Teacher scoring more accurately than AI
33	Chinese EFL learners	45 students	Writing feedback	ChatGPT	Mixed methods	Increased motivation, engagement, and self-efficacy

Table 2 (Continued).

No	Context	Participants	Task type	GenAI tool	Research design	Key outcome
34	Chinese EFL learners	40 students	Writing feedback	ChatGPT	Mixed methods	AI provided cognitive and emotional support
35	Chinese EFL learners	83 students	Writing engagement	ChatGPT	Quantitative	Improved behavioral and affective engagement
36	Croatian ESP learners	79 students	Writing assistance	ChatGPT	Mixed methods	Helpful for overcoming writer's block but with limitations
37	EFL English majors	44 students	AI-assisted writing	ChatGPT	Quantitative	Writing quality improved but fairness concerns raised
38	Japanese EFL university	79 students	Academic writing	ChatGPT, Google Translate	Mixed methods	Writing improved but teachers and peer learning remain essential
39	Thai EFL learners	39 students	Collaborative writing	ChatGPT	Quasi-experimental	AI-supported collaboration improved writing performance
40	Hong Kong secondary EFL	4 students	Writing task completion	ChatGPT	Controlled experiment	Prompt training improved strategic AI use
41	Chinese L2 Chinese learners	153 students	Corrective feedback	ChatGPT	Quantitative	GPT-4 produced more accurate feedback than GPT-3.5
42	Pakistani EFL learners	30 students	Argumentative writing	ChatGPT	Quasi-experimental	Improved argumentative writing performance
43	Chinese EFL learners	20 students	Feedback comparison	ChatGPT	Mixed methods	Students preferred teacher feedback but valued AI support

Scoping Review Findings (RQ1) – Discussion

Characteristics of included studies

The scoping analysis of the final dataset reveals a relatively compact but expanding body of literature distributed among a global network of researchers. This topic is predominantly featured in journal articles with a high annual growth rate of 379.58 %, suggesting a burgeoning research domain. Considering the strict timeline constraint (2023-2025), the citation impact (11.28) can be regarded reasonable, relatively low but yet promising. Notably, while the initial search framework spanned 2022-2025, the screening exclusion process naturally constrained the final dataset's timeline to 2023-2025. Although this compressed timeframe significantly limits the citation counts and averages of the present paper, the relatively high annual growth rate signals an emerging subject for widespread academic production. Furthermore, the total set of 43 documents are distributed across 29 journals and publication outlets. This ratio suggests a consolidated research effort within well-defined academic channels. The annual growth rate observed is the most striking metric in this data set, since it underscores an extremely rapid expansion of the field. This rate can further signalize the scientific community's urgency to explore the driving forces of GenAI systems adoption, including rapid technological advancements, evolving socio-educational needs, and the pressing need for updated institutional policies.

An average of 11.28 citations per document represents a substantial impact, especially considering the recent publication timeline within this dataset. The metric indicates that the literature retrieved for this study is well-recognized and cited by scholars with the same scientific interest. Consequently, the identification of the most highly cited articles within the dataset, that was a core objective of our RQ, provides critical insight into the specific contributions that have made the greatest scientific impact.

This analysis enhanced understanding of the specific contributions that have had the greatest impact. In comparison with the table representing the annual publication rate of each of these authors, Huang, J.'s three publications across 2024 and 2025 signify an emerging research focus on the effect of GenAI on EFL. Furthermore, analyzing publication titles and their equivalent summaries reveals that this author has increasingly focused on the domain of student's motivation regarding ChatGPT integration into EFL environments. The second most influential researcher, Mizumoto, A., centers their work on the impact of ChatGPT on the L2 motivation self-system, complementing the thematic focus of Huang, J. to the relevant field. Notably, Mizumoto, A.'s research also investigates Japanese EFL learners' preferences for AI-assisted editing, focusing on cultural and contextual considerations. Teng, M. F., represents the third most influential author in this dataset, demonstrating a growing presence in the field with publication in 2024 and 2025. Teng,

M. F.'s research primarily explores how EFL learners perceive and develop metacognitive awareness when using ChatGPT for written feedback, specifically focusing on the learner's experiences throughout this procedure.

Observations into the most commonly used keywords and the most influential publication titles confirm that research clusters around ChatGPT and its impact on EFL writing, feedback, student motivation and perceptions. The findings align precisely with the pre-established inclusion and exclusion criteria of the present study. Overall, there is a burgeoning interest in embedding ChatGPT within EFL classroom environments as a supplementary tool to enhance proficiency of all four skills. ChatGPT is an easily accessible, free of charge, complementary tool that can be easily integrated within the curriculum. Because the bibliometric analysis identified "ChatGPT" and "Artificial Intelligence" as the most frequently occurring words across the dataset, it is evident that the integration and implementation of GenAI tools in foreign language acquisition represents a rapidly expanding and promising domain. Researchers are actively investigating ways to implement these emerging tools and methods to maximize the students' benefits and to foster skill development (Asadi et al., 2025; Song & Song, 2023; Thi et al., 2025; Tram et al., 2025). Writing skills remain the primary focus of empirical scientific attention, largely because AI integration offers substantial pedagogical benefits, including reduced teacher workloads, timely feedback, and optimized management of large-scale classrooms.

Applications of GenAI in EFL writing

A deeper synthesis of the results of these tables offers several critical insights into the landscape of GenAI in language education. Because ChatGPT stands out as the most prominent of all GenAI tools within EFL contexts, a substantial body of literature explores its pedagogical implementation and direct effects on teaching and learning procedures. Broadly, there are studies that discuss the effect of GenAI incorporation to boost proficiency in writing skills targeting specific areas. Furthermore, several studies have focused on the effect of these newly inserted technologies on students' motivation, willingness to participate and thus achievement levels. There is also a growing interest in the exploration of students' psychological and emotional states towards achievement in L2 proficiency. According to the data presented in the tables, the reviewed literature spans a diverse spectrum, ranging from quantitative reviews to quasi-experimental designs. Notably, there is a distinctive prevalence of mixed method approaches that combine qualitative and quantitative means of research.

LLMs, such as ChatGPT, are increasingly utilized to support learners across multiple stages of the writing process by offering automated and instantaneous corrective feedback. Empirical evidence demonstrates that these systems offer learners substantial support for content generation, linguistic accuracy checks, and vocabulary expansion. By analyzing user-generated texts, LLMs can simultaneously generate detailed comprehensive feedback reports addressing several aspects of writing, including grammar, spelling, lexical choice, syntax, cohesion, coherence, and idea development. The rapid and comprehensive feedback enables learners to identify specific weaknesses and revise their texts accordingly, often within seconds.

Consequently, learners benefit from repeated opportunities for revision and self-correction, which fosters autonomous learning practices and allows individuals to regulate their own instructional progress. This continuous engagement with automated feedback not only facilitates the internalization of linguistic structures and enhances writing proficiency, but also stimulates broader cognitive development (Aljasser, 2025; Lu & Zeng, 2025).

In addition to general corrective feedback, research indicates positive outcomes regarding discourse-level aspects of writing, particularly cohesion and coherence. LLM-generated feedback effectively assists learners in organizing arguments logically and improving the overall rhetoric flow of their texts (Allen & Mizumoto, 2024). For instance, automated prompts regarding the deployment of linking devices and the formulation of topic sentences may guide learners toward more structured paragraph development. Nevertheless, scholars emphasize that these tools should serve as complement to, rather than a replacement for human instruction, as educators remain indispensable for interpreting feedback and scaffolding writing development (Tsai et al., 2024). Furthermore, improvements in discourse cohesion are occasionally less pronounced than improvements in surface-level grammar, given that structural organization represents a relatively advanced stage of writing proficiency (Alwasidi & Al-Khalifah, 2025). Despite these limitations, empirical evidence

confirms that LLMs integration yields measurable improvements in writing quality. Wang (2024), for example, demonstrated that ChatGPT-mediated feedback successfully mitigated errors in spelling, grammar, vocabulary, and sentence cohesion, ultimately leading to higher scores on standardized writing assessments.

Scholarly attention has likewise centered on the impact of LLM-assisted instruction on specific linguistic components such as grammar, lexis, and syntax. Alsaweed and Aljebreen (2024) observed improvements in student performance across multiple grammatical subcategories following a ChatGPT-based instructional intervention, noting significant progress in verb forms, word order, passive constructions, conditional sentences, subject-verb agreement, singular and plural noun usage, word forms, and lexical selection, alongside moderate gains in article usage, modal verbs and verb tenses. Similarly, lexical acquisition stands out as one of the most widely documented benefits of LLM integration. Alzubi et al. (2025) reported that learners' active vocabulary expanded substantially during LLM-supported writing tasks, while Kanwal (2025) identified long-term vocabulary retention as a recurring advantage of these digital interventions. Grammatical accuracy shows parallel trajectories of improvement. Chung and Jeong (2024) highlighted the efficacy of ChatGPT in minimizing grammatical errors in learners' written output, echoing Kanwal's (2025) observation that LLM-scaffolded feedback drastically reduces overall grammatical mistakes. However, certain linguistic features may remain resistant to change, for example, prepositional errors frequently persist even after targeted ChatGPT-assisted interventions.

Beyond accuracy metrics, LLMs positively influence writing fluency and overall writing competence. Seo (2024) noted that sustained interaction with LLM-based systems enhances learners' writing fluency, thereby elevating the complexity and global quality of written production. Crucially, the pedagogical utility of these tools increases when learners receive explicit training on tool interaction. Woo et al. (2025) report that following personalized and collaborative instruction on effective prompting techniques, students deployed highly sophisticated strategies when engaging with ChatGPT. This heightened strategic competence fostered a more collaborative interaction with AI tools, culminating in improved writing outcomes. Collectively, these insights underscore that while LLMs cannot substitute for the pedagogical expertise of teachers, their systematic integration into EFL writing curricula offers multi-dimensional support for writing development, particularly when learners are explicitly trained in how to use these systems effectively.

Pedagogical benefits

Most of the aforementioned studies underscore the utility of ChatGPT in EFL classrooms for delivering detailed assessment and feedback without demanding excessive instructor time or labor. Consequently, GenAI technology is highly advantageous for enabling educators to provide detailed and in-depth analysis of students' compositions within compressed timeframes. This process allows instructors to offer individualized assistance tailored to the distinct needs of each learner. Another primary benefit documented in the current analysis is the enhancement of target-specific writing sub-skills. Namely, coherence and cohesion have been positively influenced by chatbot interventions, with student texts demonstrating marked improvements in structural flow (Alwasidi & Al-Khalifah, 2025). While literature predominantly emphasized the role of chatbots in correcting surface-level mechanics, like spelling and grammar to optimize morphosyntactic accuracy (Noy & Zhang, 2023), contemporary GenAI applications offer more intermediate, isolated corrections. By generating customized examples extracted directly from the students' own output, these tools expose learners to well-structured language and sophisticated prompt engineering, thereby facilitating exposure to authentic, native-like English discourse.

This section examines how the pedagogical integration of LLMs such as ChatGPT translates to authentic classroom practices and student learning outcomes, with a specific focus on motivation, self-efficacy, and engagement in EFL writing contexts. Broadly, the reviewed literature suggests that learners perceive AI writing tools primarily as supportive, mediational resources that can elevate the immediate quality of their written texts, rather than tools that substantially cultivate their long-term writing competence. Furthermore, while automated feedback is often characterized as rapid, accessible, and convenient, students continue to place a premium on human teacher feedback. For instance, the participants in Aljasser's (2024) study acknowledged the promptness and accessibility of ChatGPT-generated feedback but still deemed instructor feedback more qualitatively useful overall. Similarly, students voiced a clear preference for instructor-led guidance while recognizing ChatGPT as a complementary feedback resource (Zou et al., 2025). These insights indicate that AI

interventions achieve optimal efficacy when deployed as supplementary instructional aids rather than absolute replacements for teacher-mediated feedback.

Several studies further indicate that LLMs yield pronounced benefits for lower-proficiency or novice EFL learners by reinforcing foundational aspects of writing. For example, Alwasidi and Al-Khalifah (2025) note that these systems effectively scaffold idea generation, eliminating redundancy, and enhance written expression clarity. Similarly, Asadi et al. (2025) argue that ChatGPT possesses transformative potential for delivering personalized feedback, which addresses localized challenges encountered during the writing process while simultaneously driving global writing proficiency. Positive user perceptions toward AI tools have also been documented through technology acceptance models. Bekturova et al. (2025) determined that learners perceive LLM-based tools as highly functional and intuitive, reporting high levels of user satisfaction and a robust behavioral intention (BI) to continue utilization. Consistently, Belda-Medina and Kokošková (2024) observed favorable metrics regarding both perceived usefulness and BI, demonstrating that students are overwhelmingly receptive to incorporating of AI-assisted writing platforms within formal educational settings.

Beyond mere usability and perceived effectiveness, recent empirical research highlights the broader cognitive and affective roles that AI tools provide during academic writing tasks. Bristi et al. (2025), for instance, conceptualize ChatGPT not only as a cognitive support mechanism but also as an effective support tool that reduces writing anxiety while bolstering student confidence. However, Dinh (2025) emphasizes that the realization of academic writing gains depends heavily on meticulous, pedagogically sound curriculum incorporation. To this end, scholars increasingly advocate for a balanced, hybrid instructional framework wherein AI complements, rather than replaces, human expertise. Escalante et al. (2023) and Han (2024) emphasize the necessity of a blended approach that combines the analytical and generative capabilities of AI systems with the interpretive expertise and pedagogical oversight of teachers.

Concurrently, positive motivational outcomes linked to AI writing software are also widely reported. Huang and Mizumoto (2025) identify increased learner motivation as a pivotal advantage of embedding AI tools within EFL writing instruction. Similarly, Lu and Zeng (2025) observe that learners express favorable perceptions regarding the use of ChatGPT-generated model texts as a form of feedback and structural support. Other investigations suggest that these technologies may be particularly appealing to younger learners. Specifically, Poláková and Ivenz (2024) highlight the potential of ChatGPT as an innovative educational medium capable of fostering writing proficiency among Generation Z students. In addition, Salam (2025) points that prolonged exposure to AI-scaffolded writing environments encourages more positive attitudes toward technological shifts and promotes more frequent adoption of emerging digital tools.

Finally, the synthesized literature demonstrates that AI-mediated writing environments deeply influence distinct dimensions of learner engagement. Teng (2025) similarly describes the dual cognitive and emotional roles of ChatGPT academic writing contexts, while Teng (2024) documents significant positive effects of AI assistance on overall writing performance, writing motivation, self-efficacy, task engagement, and collaborative writing inclinations. More nuanced parameters are introduced by Teng and Huang (2025), whose empirical findings reveal that while AI-supported writing tasks significantly enhance effective and behavioral engagement among EFL learners, cognitive engagement shows no statistically significant variance. Collectively, these insights demonstrate that the pedagogical value of LLM-based tools extends far beyond their capacity for automated linguistic correction, lying fundamentally in their potential to enhance learners' motivation, build self-efficacy, and engagement when purposefully incorporated into instructional designs.

Challenges and limitations

A recurring theme across literature underscores the necessity for a cautious and pedagogically balanced integration of AI tools, specifically LLMs such as ChatGPT, into EFL writing instruction. Although numerous studies report improvements in student writing performance, scholars frequently emphasize that AI deployment must be carefully regulated to mitigate user overreliance and safeguard the development of independent writing competencies. Interestingly, empirical data reveal that students themselves articulate concerns regarding excessive dependence on automated assistance and the broader ethical implications of AI utilities. These insights highlight the imperative need for educators to adopt a hybrid instructional approach that balances traditional writing practices, such as pen-and-paper drafting, with AI-scaffolded feedback and revision processes (Alwasidi & Al-Khalifah, 2025; Teng, 2024).

Beyond overreliance concerns, students frequently question the complexity, relevance, and reliability of LLM responses. Campos (2025), for instance, notes that learners occasionally struggle with the complexity of the feedback produced by AI tools and may find it difficult to map automated feedback onto their specific writing tasks. Similarly, Bekturova et al. (2025) report that widespread student skepticism regarding the accuracy of AI content, alongside explicit anxieties surrounding plagiarism, the potential loss of foundational writing skills, and the need for greater feedback precision. These apprehensions are echoed by Alzubi et al. (2025), who identify overreliance, diminished critical thinking, and intellectual property vulnerabilities as core challenges inherent to AI-assisted writing environments. Furthermore, Afandi (2024) demonstrates that learners often exhibit comprehension gaps when reviewing texts generated by AI systems, suggesting that surface-level text manipulation does not inherently signify a deeper cognitive understanding of the produced content.

Consequently, the literature affirms that the pedagogical role of the classroom instructor and the collaborative benefits of peer interaction remain entirely irreplaceable. For example, Wang (2024) found that while AI tools optimized individual text quality, they could not replicate the instructional and collaborative dynamics traditionally facilitated by teacher guidance and peer learning. Sawangwan (2024) similarly argues that crucial aspects of writing pedagogy, such as internalizing assessment criteria, cultivating individual critical thinking skills, and providing ethical guidance, remain firmly outside the capabilities of current AI systems. Moreover, the utility of AI platforms appears highly dependent on the writing task. For instance, Punar Özçelik and Yangın Ekşi (2024) questioned the necessity of deploying ChatGPT for informal writing genres and raised doubts regarding its efficacy in teaching neutral linguistic register.

Scholars also raise significant concerns regarding the broader cognitive and psychological impacts of extensive AI interaction in academic contexts. While certain studies position AI tools as potential buffers against writing-related anxiety, the evidence remains fragmented. Kanwal (2025) reports mixed outcomes regarding the extent to which ChatGPT alleviates writing-related anxiety. Concurrently, Dinh (2025) warns that excessive exposure to AI writing utilities can induce surface-level learning, increased technological dependency, and difficulties in interpreting automated feedback. Issues of academic integrity and ethical use are equally widely documented. Chung and Jeong (2024) highlight the risks related to overreliance, a decline in critical engagement, the authenticity of AI-facilitated interactions, threats to data privacy and academic integrity. In a broader systemic critique, Almahasees et al. (2024) identify additional risks, including reduced critical engagement, fragmented subject knowledge, technological dependency, the propagation of algorithmic misinformation, data privacy vulnerabilities, and biased AI responses.

While LLMs exhibit substantial promise for addressing diverse aspects of second language acquisition (SLA) and writing development, occasional inaccuracies and limitations call for ongoing empirical scrutiny and rigorous pedagogical implementation. Alhazmi and Muftah (2025) argue that although ChatGPT demonstrates utility across diverse SLA domains, frequent inaccuracies necessitate continuous investigation to optimize its educational value. Furthermore, the extensive use of AI tools complicates writing assessment and evaluation metrics. Tsai et al. (2024) note that disproportionate, AI-mediated improvements in writing quality challenge the validity and fairness of traditional evaluation methods. Similarly, Lu and Zeng (2025) caution that AI-generated texts frequently lack contextual appropriateness and occasionally display unnatural phrasing, which severely limits their effectiveness in authentic communication tasks.

Given these challenges, literature increasingly demands structured institutional interventions and standardized pedagogical frameworks. While AI writing tools offer significant pragmatic affordances, their effective and ethical execution requires the immediate establishment of institutional policies. Rahayu et al. (2025) emphasize the urgent need for such regulations in order to preserve responsible student usage, academic integrity, and maximize the educational potential of AI-assisted writing platforms. Collectively, the evidence indicates that while LLM-based systems can enhance specific dimensions of writing pedagogy, their integration must be meticulously orchestrated to support rather than supplant the development of independent writing abilities, critical thinking, and responsible academic practices.

It should be noted that many pedagogical, experiential, and ethical dimensions of AI integration in EFL writing are extensively explored in qualitative studies, which fall outside the scope of the present review. While such studies provide valuable insights into classroom practices, learner perceptions, and teacher

Table 3. Summary table with themes and number of studies

Major themes	Sub-themes	n
Linguistic accuracy and grammar	Syntax, spelling, mechanical corrections, and error tracking	21
Vocabulary development	Lexical diversity, register choices, and synonym expansion	8
Cohesion, coherence, and organization	Discursive flow, paragraph structuring, and structural clarity	9
Learner engagement and autonomy	Emotional/cognitive support, self-efficacy, and mitigating anxiety	12
Instructional conditions and contexts	Guided prompting, blended "AI + Teacher" structures	9
Challenges and risks	Over-reliance, shallow learning, bias, plagiarism, and grading fairness	13

Note. n: Count of studies

perspectives, they were not included due to the study's strict focus on research reporting measurable outcome-based writing achievements.

To complement these findings, the following section presents a bibliometric analysis of the field, providing a structural overview of publication trends, highly influential contributions, and emergent thematic developments.

Thematic synthesis of writing outcomes (RQ1)

To complement the descriptive presentation of results, a thematic synthesis was conducted to identify the most consistently reported outcomes, instructional conditions, and challenges associated with the integration of LLMs in EFL writing. Across the reviewed studies (n = 43), several key themes emerged:

Linguistic accuracy and grammar (most frequently reported): The most frequently reported outcome across literature (n = 21 studies) involves improvements in grammatical accuracy, spelling, and sentence structure following the use of AI-assisted feedback. These improvements are particularly evident in error correction tasks and revision activities, where learners receive immediate and detailed feedback.

Vocabulary development: A subset of papers (n = 8) highlights gains in lexical diversity and vocabulary acquisition. Reviewed AI tools were frequently found to support word choice, synonym generation, and lexical expansion, particularly among intermediate and lower-proficiency learners.

Cohesion, coherence, and organization: Discourse-level features, including paragraph organization, logical flow, and coherence, were addressed in 9 studies. However, these improvements appear less consistent than those related to grammar and vocabulary, especially during more advanced writing tasks.

Learner engagement, motivation, and autonomy: A substantial proportion of studies (12 studies) emphasize increased learner engagement, motivation, and self-directed learning. The immediacy and accessibility of AI feedback appear to encourage iterative revision and greater learner autonomy.

Instructional conditions and use contexts: Nine studies evaluated specific pedagogical parameters, emphasizing that the effectiveness of AI-dependent models without guidance yield weaker outcomes compared to blended setups. Better outcomes were consistently observed in guided tasks featuring clear prompts and structured revision activities. The most prevalent pedagogical frameworks in the reviewed literature are blended models that combine AI and teacher feedback. In contrast, fully AI-dependent implementation without pedagogical guidance is associated with weaker learning outcomes.

Challenges and risks: Several recurring concerns were identified across 13 studies, including mainly over-reliance on AI tools. Secondary concerns involve reduced critical thinking, threats to academic integrity and plagiarism issues. Furthermore, both students and educators raised issues regarding the limited depth of learning ("shallow learning"), concerns regarding accuracy, bias, and data privacy. These challenges underscore the necessity of controlled, pedagogically informed implementation.

Overall, the synthesis suggests that AI tools are most effective as supportive instruments within structured instructional environments, particularly for improving lower-level linguistic features. Conversely, their capacity to develop higher-order writing skills and critical thinking remains limited and context-dependent.

Table 3 is a summary table with themes and number of studies.

Figure 2. Descriptive bibliometric summary of the dataset generated through bibliometrix R-package (Biblioshiny interface) (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

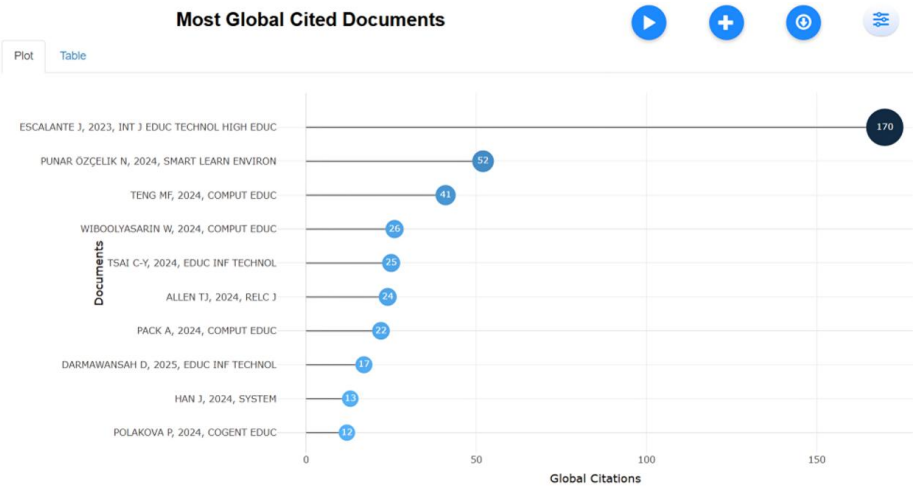


Figure 3. Most global cited documents plot (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

Bibliometric Analysis of the Results (RQ2)

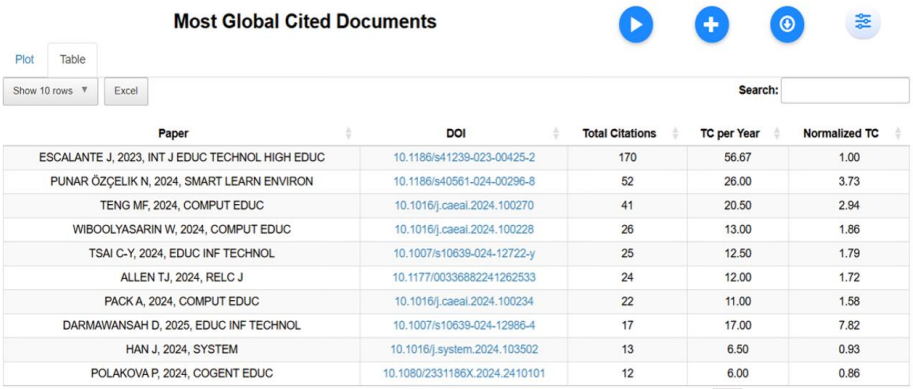
Publication trends

While the scoping synthesis provides a thematic overview of pedagogical findings, the bibliometric analysis offers a macro-level perspective on the evolution and structure of the research field. This component was executed using the bibliometrix R-package tool (Aria & Cuccurullo, 2017), a comprehensive tool for quantitative science mapping and data visualization. This software is widely utilized in bibliometric research because it enables investigators to track the development of specific research scopes, evaluate the citation impact of a field, and map structural elements within the literature (Chen, 2016; van Eck & Waltman, 2010). Consequently, bibliometrix was employed in the present study to analyze the demographics and structural characteristics of the selected database. Upon importing the bibliographic data, the software generates a comprehensive overview of the dataset's descriptive statistics (**Figure 2**).

As illustrated in **Figure 2**, the final dataset comprises 43 documents spanning the years 2023 to 2025. Even though the initial literature search was conducted within the time span of 2022-2025, no results published in 2022 met the predefined inclusion and exclusion criteria, hereby confining the final analytical dataset to 2023-2025 threshold. Within the data summary, the sources metrics denote the journals in which the articles were published, while the remaining descriptive categories represent standard bibliometric indicators generated by the software to profile the dataset's composition.

Most influential publications

To address the first component of the RQ2, it is necessary to identify the publications that have exerted the greatest impact on the domain of GenAI integration in EFL writing. Accordingly, the most highly cited and influential documents within the dataset were mapped and visualized (**Figure 3** and **Figure 4**).



Most Global Cited Documents

Plot Table

Show 10 rows Excel Search:

Paper	DOI	Total Citations	TC per Year	Normalized TC
ESCALANTE J, 2023, INT J EDUC TECHNOL HIGH EDUC	10.1186/s41239-023-00425-2	170	56.67	1.00
PUNAR ÖZÇELIK N, 2024, SMART LEARN ENVIRON	10.1186/s40561-024-00296-8	52	26.00	3.73
TENG MF, 2024, COMPUT EDUC	10.1016/j.caeai.2024.100270	41	20.50	2.94
WIBOOLYASARIN W, 2024, COMPUT EDUC	10.1016/j.caeai.2024.100228	26	13.00	1.86
TSAI C-Y, 2024, EDUC INF TECHNOL	10.1007/s10639-024-12722-y	25	12.50	1.79
ALLEN TJ, 2024, RELC J	10.1177/00336882241262533	24	12.00	1.72
PACK A, 2024, COMPUT EDUC	10.1016/j.caeai.2024.100234	22	11.00	1.58
DARMAWANSAH D, 2025, EDUC INF TECHNOL	10.1007/s10639-024-12986-4	17	17.00	7.82
HAN J, 2024, SYSTEM	10.1016/j.system.2024.103502	13	6.50	0.93
POLAKOVA P, 2024, COGENT EDUC	10.1080/2331186X.2024.2410101	12	6.00	0.86

Figure 4. Most global cited documents table (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

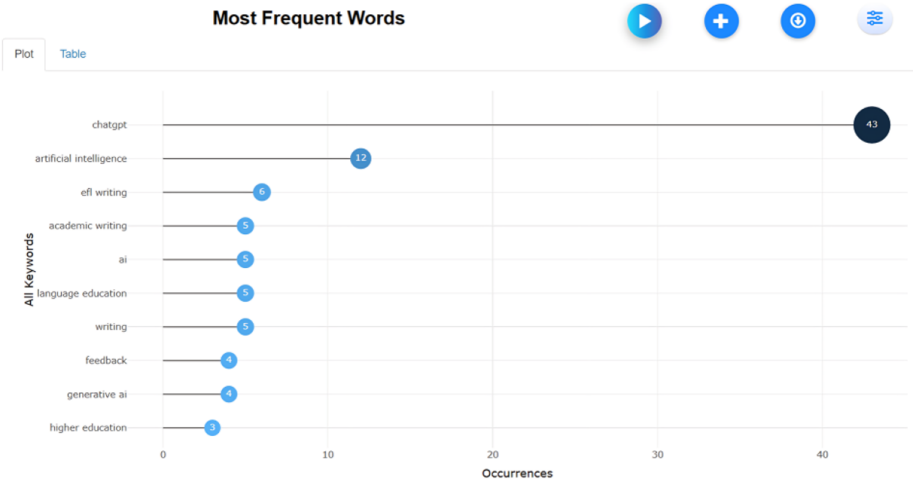


Figure 5. Most frequent words plot (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

A compilation of the top ten most globally cited documents indicates a clear hierarchical impact within the corpus, prominently led by Escalante et al. (2023). Their study, entitled "*AI-generated feedback on writing: insights into efficacy and ENL student preference*", published in the "International Journal of Educational Technology in Higher Education", investigated the impact of AI-assisted feedback on written texts and its contribution to overall writing skill development compared to a hybrid model of AI+teacher version (**Figure 4**). Ranked second is the work by Punar Özçelik, N. and Yangın Ekşi, G. (2024), titled "*Cultivating writing skills: the role of ChatGPT as a learning assistant—A case study*" (in the journal *Smart Learning Environment),* which represents another highly influential investigation exploring student experiences and the perceptions of automated, AI generated feedback. The third most globally cited document is Teng (2024), titled *"ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing" (Computer Education)*, which similarly examines students' attitudes toward AI-driven formative feedback.

Overall, the most influential literature in this domain converges on two primary dimensions, evaluating the empirical effectiveness of these systems in improving EFL writing skills, and assessing learner reactions to automated intervention. The remaining publications within this top ten (**Figure 3** and **Figure 4**) reinforce this trend, underscoring a rapidly growing research interest in the systematic integration of LLMs into pedagogical frameworks.

Keywords analysis and research themes

To map the conceptual structure of the field, the bibliometrix R-package tool was used to generate a keyword co-occurrence map (**Figure 5**) and a word cloud (**Figure 6**). As illustrated across both visualizations, "ChatGPT" emerges as the most frequently occurring keyword in this dataset. This concentration reflects

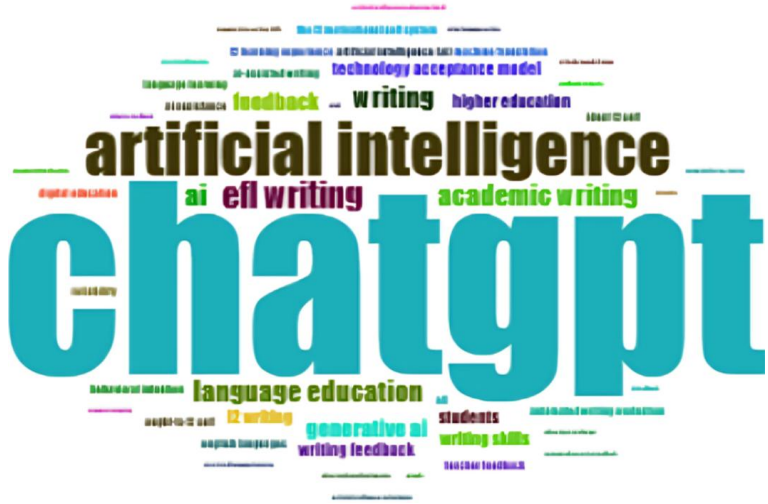


Figure 6. Word cloud (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

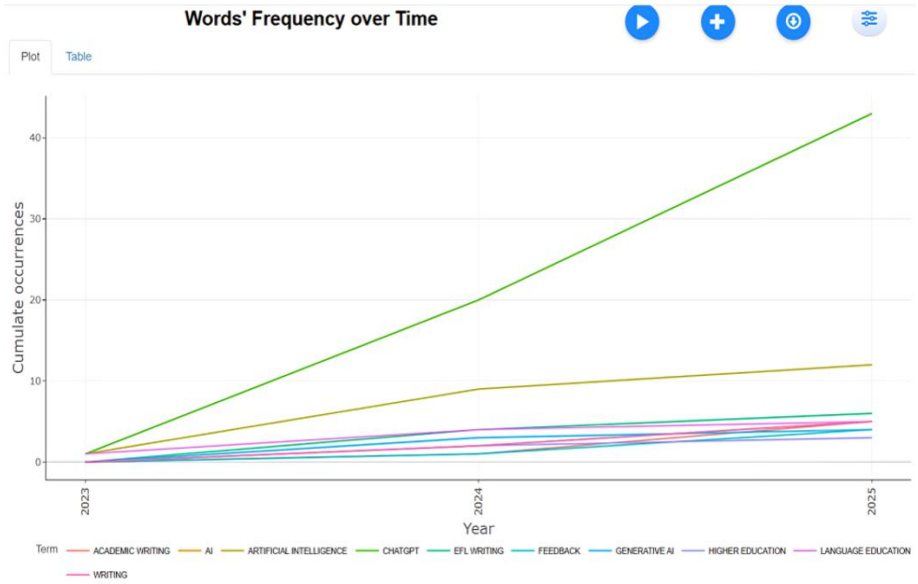


Figure 7. Words' frequency overtime plot (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

researchers' operational preference for this specific platform, driven largely by its open-access availability, which enhances the generalizability and cross-contextual applicability of empirical findings.

The next most frequent keyword is "Artificial Intelligence", serving as a broader indexing term for this research domain. Interestingly, the more specific term "Generative AI" (GenAI) has rapidly gained popularity over the broader "AI" descriptor in recent literature, as it more precisely captures the text-generation capabilities relevant to educational and instructional contexts. Collectively, these keyword frequencies underscore the recognition of ChatGPT as a primary instrument of interest when appropriately integrated into the EFL classroom.

Lastly, to examine the temporal distribution of the most frequent terms, a cumulative keyword plot was generated (**Figure 7**). Mirroring the previous analyses, 'ChatGPT" emerges as the dominant term, demonstrating the steepest growth trajectory since 2023 and reflecting an exponential upward trend in scholarly interest. The pronounced visibility of "ChatGPT" relative to macro-level terms such as "Artificial Intelligence" and "Generative AI" indicates a clear paradigm shift, research is increasingly focusing on the domain-specific application of LLMs within language learning contexts, rather than on generalized educational AI applications.

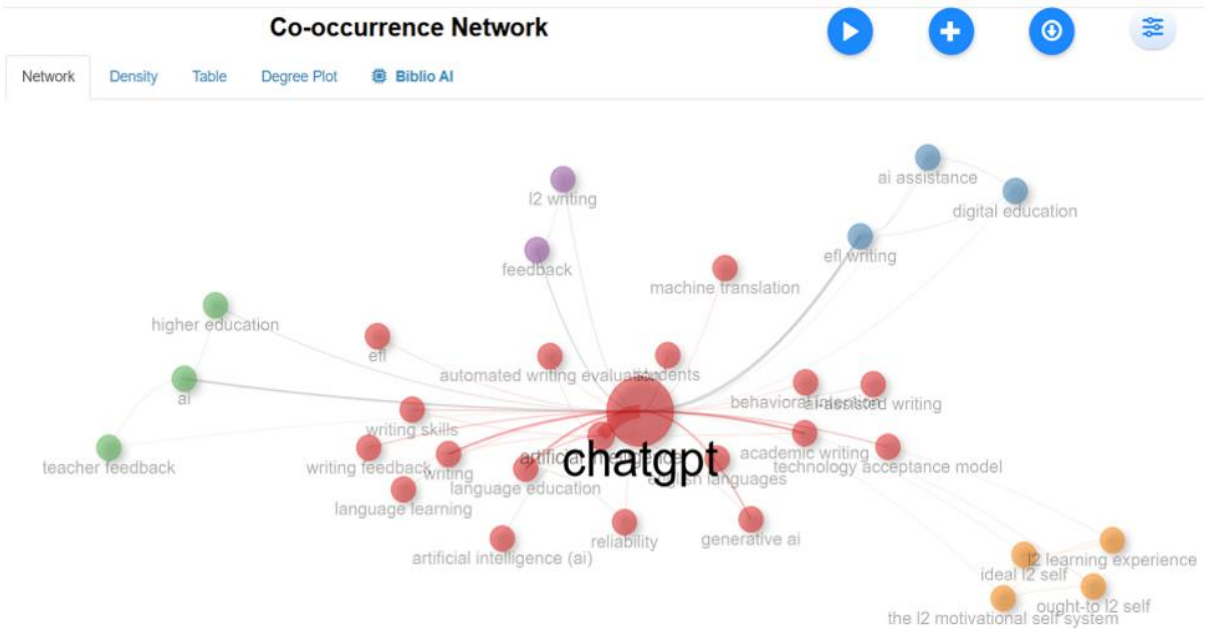


Figure 8. Co-occurrence network of terms (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

Concurrently, pedagogical and contextual keywords, including "writing", "EFL writing", "feedback", "language education", and "higher education", exhibit a moderate but steady growth rate, signaling sustained academic interest in these core areas. These chronological patterns reinforce the thematic clusters identified in the keyword co-occurrence network, particularly the strong empirical focus on AI-assisted writing instruction, learner perceptions and the systematic integration of generative tools into established educational structures. Synthetically, this temporal analysis demonstrates that research surrounding ChatGPT in EFL writing has expanded rapidly since 2023, driven by the sudden global proliferation of GenAI and its recognized pedagogical affordances.

To map the conceptual intersections within the literature, a keyword co-occurrence network diagram was generated (**Figure 8**). **Figure 8** demonstrates the co-occurrence of terms throughout the dataset. There is a pronounced thematic convergence, where "ChatGPT" frequently links with terms such as "academic writing", "artificial intelligence" and "EFL writing", thereby confirming that current research is overwhelmingly centered on the writing support that AI automated tools generate.

Notably, the prominent clustering of "academic writing" and "higher education" suggests a distinct demographic concentration within the selected literature, indicating that empirical investigations of automated feedback and AI assistance are predominantly situated in tertiary education rather than among young learners or beginner-level proficiency ones. Furthermore, the "feedback" cluster underscores the operational role of these AI technologies within modern language curricula, highlighting their documented utility in providing productive, comprehensible, and individualized feedback to learners.

Most relevant authors

To map the leading contributors to this domain, the dataset was analyzed to identify the most prolific authors based on the selected inclusion criteria. The analysis reveals that Huang, J. represents the most prominent author in this field, followed by Mizumoto, A., and Teng, M. F. (**Figure 9**). Examining author prominence provides critical insights into the structural leadership driving the intersection of AI integration and SLA. To supplement these findings, the annual scientific production of the top ten authors was mapped to illustrate the temporal distribution of their scholarly output (**Figure 10**). A comparative analysis of these metrics indicates that the publication trajectories of both Huang, J. and Teng, M. F. have increased steadily over the last five years. This upward trend suggests a sustained research agenda focused on AI integration within writing instruction programs, with particular emphasis on motivational aspects of AI-generated contribution, student engagement and automated support.

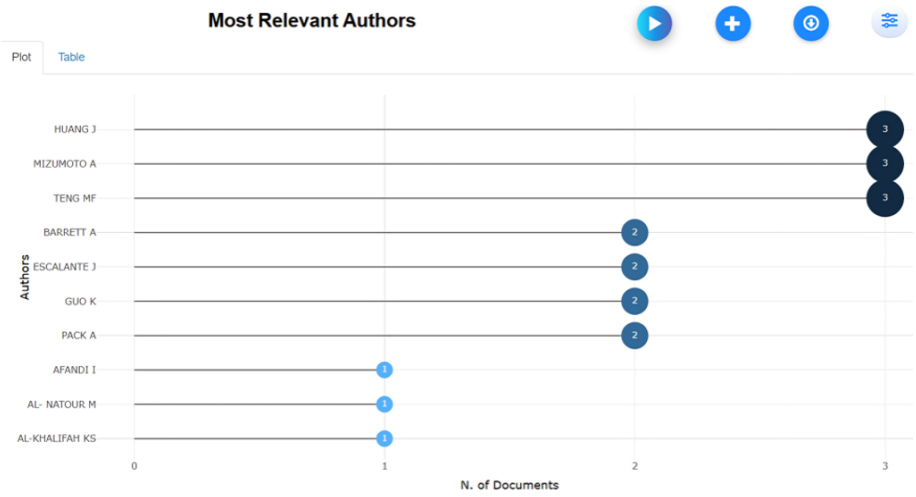


Figure 9. Most relevant authors plot (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

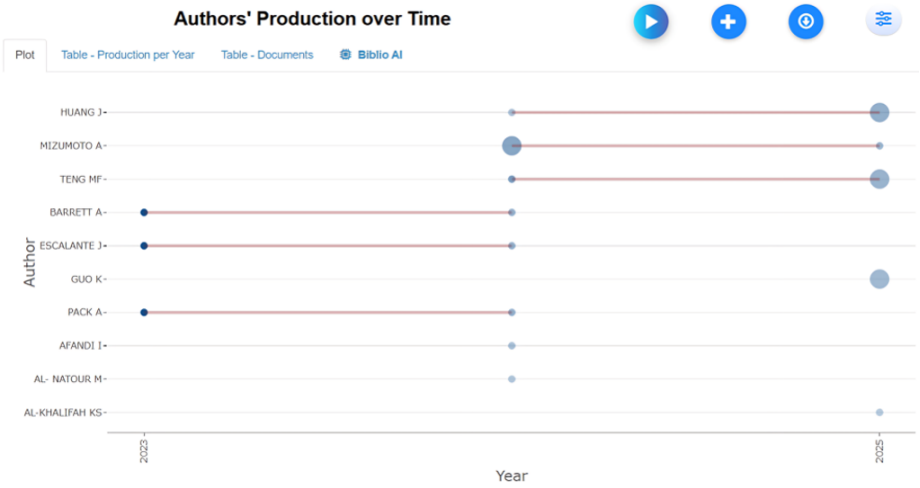


Figure 10. Annual authors' production (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

Conversely, Mizumoto, A.'s publication frequency within this specific dataset exhibits a decline over the same five-year period, potentially reflecting a shift of research focus toward alternative dimensions of AI-assisted language learning other than writing. The remaining highly ranked authors demonstrate more sporadic publication patterns, indicating localized, highly targeted contributions addressing more targeted publications within the literature.

Most influential affiliations

The subsequent analysis evaluates the institutions responsible for producing the most influential literature, mapped across their production trajectories over time. As illustrated in **Figure 11**, all universities exhibited a sharp increase in research output during 2024 and 2025. Because data collection concluded in May 2025, the mid-year surge strongly indicates that the field experienced unprecedented annual growth by the conclusion of this study was conducted in May 2025, the mid-year surge strongly indicates that the field experienced unprecedented annual growth by the conclusion of 2025. Notably, Kansai University overtook Brigham Young University – Hawaii in 2024 regarding the volume of publications centered on integrating AI tools into EFL education, a trend that persisted through 2025. This shift likely reflects strategic research investments or the establishment of specialized research clusters dedicated to AI-assisted language learning. Ultimately, these trends underscore both the intensifying competitive landscape among leading institutions and the rapidly expanding scholarly interest regarding AI integration into EFL teaching practices.

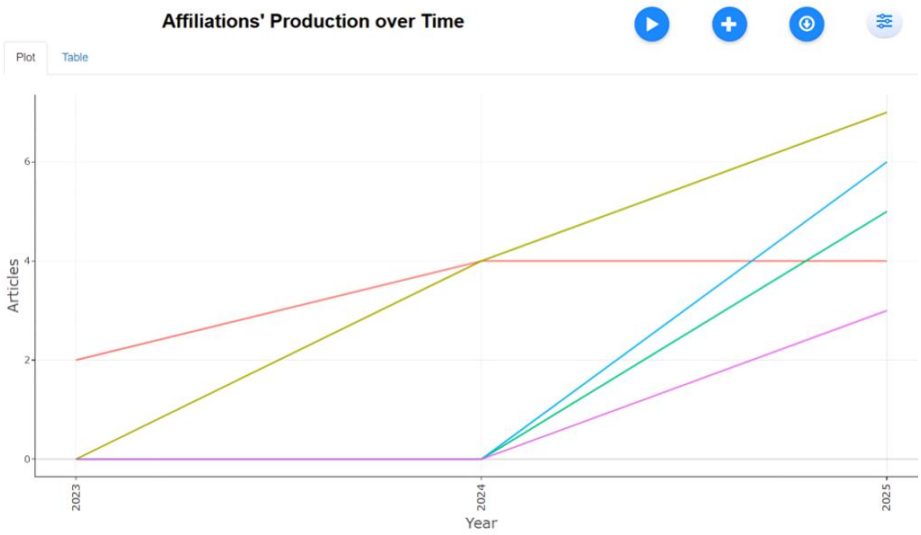


Figure 11. Affiliations production over time (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

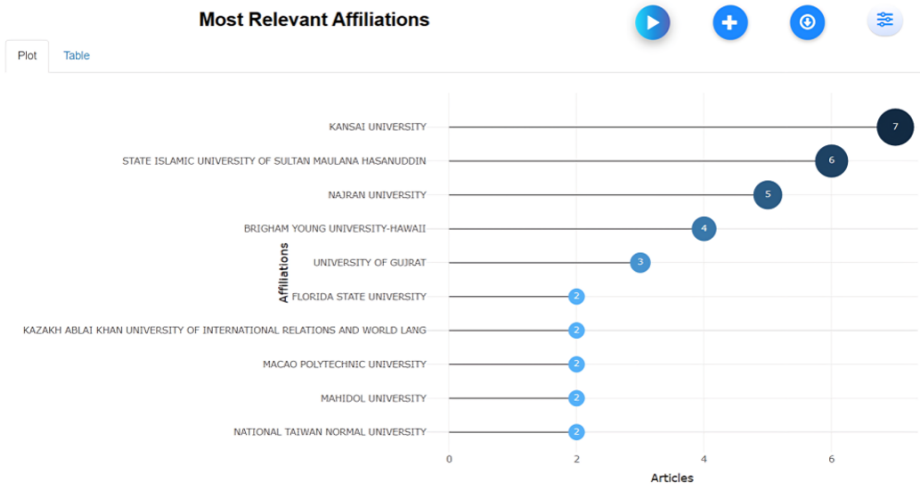


Figure 12. Most relevant affiliations plot (generated by the authors using the bibliometrix R package [Aria & Cuccurullo, 2017] based on the analyzed dataset)

Figure 12 provides a cumulative depiction of the preceding analysis, illustrating the aggregate output of the institutions that have contributed the most substantial research to this designated domain. This longitudinal representation highlights the total contribution of each institution over time, thereby facilitating a more robust comparison for their sustained research activity in AI integration within EFL education. The visualization confirms a prominent role of leading institutions, specifically Kansai University and Brigham Young University – Hawaii, both of which demonstrate consistent, accelerated growth in publications. This cumulative perspective underscores not only the broadening institutional engagement with AI-supported language learning but also the emergence of specific universities as foundational contributors to the evolution of this research area.

PRACTICAL IMPLEMENTATION

Utilizing the PRISMA framework in conjunction with the bibliometrix R-package, this study systematically reviewed the implementation of GenAI technologies within a foreign language environment, with a specific focus on written language production. The bibliometric data reveals a prevailing reliance on ChatGPT within EFL writing instruction across diverse cultural backgrounds, a trend largely driven by its global universality and open-access model. This finding aligns with recent syntheses by Huang and Mizumoto (2024) and

Mizumoto and Eguchi (2023), who identified ChatGPT as the most frequently evaluated AI application in EFL contexts due to its widespread accessibility, user-friendly interface and rapid global adoption by educators and researchers worldwide. Furthermore, keyword cluster analysis indicated a concentrated scholarly focus on written output. This corroborates the assertions of Teng (2023) that automated writing evaluation and AI-supported revision processes have matured into primary domains of inquiry within computer-assisted language learning.

These thematic trends are further contextualized by the distribution of productive authors and institutional affiliations. Prolific contributors such as Huang, J., Mizumoto, A., and Teng, M. F. anchor the primary research clusters by driving scholarly output in this field (Teng, 2023). At the institutional level, the ascendance of Kansai University, which surpassed Brigham Young University – Hawaii in cumulative output from 2024 onward, points to targeted institutional investments and dedicated research groups specializing in educational AI. The pronounced surge in publication volume throughout 2024 and 2025 reflects the rapid integration of GenAI into higher education practice and research agendas, a macro-trend documented in recent surveys on AI adoption in higher education (Kasneci et al., 2023).

Beyond identifying central themes, the longitudinal data captures distinct evolutionary shifts in the research landscape over time. Early inquiries primarily emphasized the theoretical possibilities and speculative capabilities of AI tools. In contrast, recent publications demonstrate a definitive shift toward empirical, applied studies exploring active classroom practices, measurable learner outcomes, and concrete implementation strategies. Notably, while ethical concerns regarding academic integrity and plagiarism persist, they are no longer the exclusive focal point of literature. Instead, recent scholarship prioritizes practical pedagogical applications, signaling a departure from earlier debates that foregrounded ethical challenges at the expense of educational utility.

From a pedagogical perspective, integrating GenAI tools like ChatGPT into EFL writing instruction requires structured and responsible implementation (Tseng & Lin, 2024). Teachers may adopt classroom workflows in which students first produce an initial draft independently and subsequently use AI tools for guided revision, focusing on aspects such as grammar, vocabulary enrichment, and organizational coherence (Huang & Wang, 2025). Rather than positioning AI as a replacement for human instruction, current scholarship advocates for its integration as a supplementary mechanism, provided educators are adequately trained in its pedagogical affordances and constraints. AI-generated feedback should complement rather than replace teacher feedback, allowing instructors to focus on higher-order skills such as argumentation, critical thinking, and discourse development (Mizumoto & Eguchi, 2023). Additionally, to safeguard academic integrity, institutions must establish transparent parameters. These should include explicit policies on permissible AI utility, standardized citation practices for AI assisted work, and reflective tasks that require students to meta-cognitively justify how they utilize AI feedback (Benke & Szőke, 2024). Ultimately, successful implementation relies on institutional professional development initiatives designed to cultivate comprehensive AI literacy, effective prompting strategies, and ethical classroom practices among educators (Chen & Delaney, 2025; Park, 2025).

CONCLUSION

This study provided a comprehensible scoping overview with bibliometric mapping on the rapidly evolving research landscape surrounding GenAI integration, specifically ChatGPT, within writing instruction globally. By utilizing the bibliometrix R-package to evaluate keyword trends, prolific authors, and institutional networks, this paper successfully mapped current research hotspots, structural shifts, and latent gaps in the literature.

The empirical findings directly address our core RQs, confirming that ChatGPT stands as the architectural center of current GenAI-focused language research, with a predominant application toward enhancing written language production. The thematic architecture of the field is heavily anchored in AI-assisted feedback mechanisms, individualized language learning support, and practical educational workflows. Our analysis confirms a pronounced escalation in research productivity during the 2024-2025 period, aligning with recent scholarship that highlights ChatGPT's capacity to deliver customized writing feedback and optimize student learning outcomes (Kawinkoonlasate, 2025; Mohammad et al., 2025), while simultaneously prompting essential dialogues regarding student critical agency and cognitive trust in automated systems (Pum, 2025).

In sum, this study synthesizes emerging trends to provide actionable insights for both educational practitioners and researchers. The framework derived from this bibliometric mapping advocate for the structured, scaffolded implementation of GenAI to complement rather than replace human instruction, thereby safeguarding learner autonomy and critical thinking. By categorizing the contributions or pioneering authors and ascendant institutions, this paper establishes a benchmark for future longitudinal investigations, advancing the collective understanding of how LLMs can be effectively and ethically managed in language education. In accordance with these observations, the most frequently occurring keywords highlight the main research topics addressed in the literature, particularly the integration of GenAI tools such as ChatGPT in EFL writing instruction, as well as themes related to AI-assisted feedback, language learning support, and pedagogical implementation. These thematic patterns may guide future researchers in identifying key areas of interest and emerging directions in the field.

Specifically, the study explores RQs regarding the prevalence and focus of AI applications in EFL writing, revealing that ChatGPT is the most central tool in current research, primarily applied to writing skills. The analysis of authors and institutions highlights the main contributors driving the field and shows a clear increase in research productivity during 2024-2025, reflecting the growing academic interest in integrating AI into language education. The study contributes to the field by synthesizing emerging trends, identifying gaps, and providing actionable insights for both researchers and practitioners. Pedagogical recommendations derived from the findings suggest structured implementation of AI tools in classrooms, complementing teacher feedback, and supporting learner autonomy and critical thinking. Furthermore, the study highlights ethical considerations and the need for teacher professional development to ensure responsible AI use. Overall, this research not only maps the current landscape of AI in EFL writing instruction but also provides guidance for future studies in underexplored areas, thereby advancing understanding of how GenAI tools can be effectively and responsibly integrated into language learning.

Given the restricted data collection timeframe and the relatively concentrated pool of analyzed literature, definitive conclusions can still be confidently drawn regarding the burgeoning scholarly interest in this domain. To optimize these potential advantages while carefully balancing inherent limitations, educators and academic institutions must deliberately and methodically implement these tools within their curricula. Rather than replacing the indispensable human factor, current scholarship advocates for the systematic integration of such platforms into active classroom practices, presupposing on instructors being comprehensively informed of both their pedagogical affordances and operational constraints. Concurrently, students must be meticulously prepared for responsible tool utility, because such training is vital to foster ethical engagement, prevent academic misconduct or systemic overreliance, and effectively support the authentic development of sustainable writing skills.

Limitations and future research guidelines

Despite the substantive insights generated by this study, several boundaries must be acknowledged to contextualize the findings and guide future inquiries. Firstly, the review exclusively evaluated the single skill of written language production. Furthermore, data collection was restricted strictly to the acquisition of EFL. To enhance generalizability, future research should expand into the productive skill of speaking, the receptive skills of listening and reading, and explore GenAI applications across non-English target languages.

Due to specific eligibility criteria, for example focusing heavily on quantitative or mixed-methods designs measuring tangible writing outcomes, the final dataset comprised a highly concentrated pool of 43 entries. While this decision allowed for a more focused analysis of studies reporting measurable writing-related outcomes, by explicitly omitting purely qualitative research design, this study lacks deeper, narrative insights into classroom dynamics, lived learner experiences, and teacher perspectives. Future scoping reviews should adopt broader inclusion criteria that incorporate qualitative evidence to construct a more holistic view of AI-supported CALL environments.

In addition, the limitation of the utilization of only one database, namely Scopus, is another obstacle that restricts the generalizability of the results of this study. In other words, researchers should be cautious to avoid overgeneralization given the restricted research that was conducted. Finally, while our review focuses on ChatGPT and similar LLMs, it does not cover other GenAI tools, which are often rule-based or narrowly

domain-specific. Future researchers could explore these tools to provide a broader perspective on AI-supported writing in EFL contexts.

It is also important to acknowledge certain limitations associated with the bibliometric component of this study. The dataset primarily consists of relatively recent publications, reflecting the emerging nature of research on AI tools in language learning. As a result, citation-based indicators should be interpreted with caution, since recently published studies have had limited time to accumulate citations. Citation counts and related metrics may therefore fluctuate significantly as the field continues to develop. Consequently, the bibliometric results presented in this study should be viewed primarily as indicative of emerging research trends and thematic patterns, rather than definitive measures of scholarly impact. Future studies conducted over longer time spans and with larger datasets may provide more stable citation-based indicators and allow for more robust bibliometric interpretations.

Author contributions: AG: conceptualization, formal analysis, methodology, project administration, validation, writing – review & editing; **PC: c**onceptualization, data curation, formal analysis, investigation, methodology, resources, validation, writing–review & editing. Both authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: This review does not require any ethics committee approval, since no human subjects or any other kind of subject participated in this research. The research includes a comprehensive literature review based on published material.

AI statement: During the preparation of this work the authors have made use of no AI or GenAI tool.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Adhikari, C., Das, P. K., & Biswas, S. D. (2025). AI in higher education: An analysis of ChatGPT's impact on scholarly communication. *Educational Point, 2*(2), Article e137. https://doi.org/10.71176/edup/17641

Afandi, I. (2024). Assessing text comprehension proficiency: Indonesian higher education students vs ChatGPT. *XLinguae, 17*(1), 49-68. https://doi.org/10.18355/XL.2024.17.01.04

Alhazmi, A. A., & Muftah, M. (2025). Evaluating ChatGPT's reliability in second language acquisition (SLA): Insights on language skills and technology's role. *Electronic Journal of E-Learning, 23*(1), 96-112. https://doi.org/10.34190/ejel.23.1.3762

Aljasser, A. (2025). Investigating EFL students' perceptions of feedback: A comparative study of instructor and ChatGPT-generated responses in academic writing. *Education and Information Technologies, 30*, 22967-22988. https://doi.org/10.1007/s10639-025-13649-8

Alkhawaldeh, M. A., & Khasawneh, M. A. S. (2023). Language learning tools and improvements in speaking and writing skills. *Journal of Southwest Jiaotong University, 58*(5), 428-444. https://doi.org/10.35741/issn.0258-2724.58.5.33

Allen, T. J., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-foreign-language learners' preferences for editing and proofreading strategies. *RELC Journal, 57*(1), 89-106. https://doi.org/10.1177/00336882241262533

Almahasees, Z., Al-Natour, M., Bataineh, K. A., & Aminzadeh, S. (2024). Students' perceptions of the benefits and challenges of integrating ChatGPT in higher education. *Pakistan Journal of Life and Social Sciences, 22*(2), 3479-3494. https://doi.org/10.57239/PJLSS-2024-22.2.00256

Alsaweed, W., & Aljebreen, S. (2024). Investigating the accuracy of ChatGPT as a writing error correction tool. *International Journal of Computer-Assisted Language Learning and Teaching, 14*(1). https://doi.org/10.4018/IJCALLT.364847

Alwasidi, M. A., & Al-Khalifah, K. S. (2025). Assessing the impact of ChatGPT on EFL students' writing productivity and proficiency. *Journal of Language Teaching and Research, 16*(3), 986-995. https://doi.org/10.17507/jltr.1603.29

Alzubi, A. A. F., Nazim, M., & Alyami, N. (2025). Do AI-generative tools kill or nurture creativity in EFL teaching and learning? *Education and Information Technologies, 30*, 15147-15184. https://doi.org/10.1007/s10639-025-13409-8

Aria, M., & Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity, 56*, Article 101766. https://doi.org/10.1016/j.tsc.2025.101766

Bekturova, M., Tulepova, S., & Zhaitapova, A. (2025). Predicting Kazakhstani TEFL students' continuance intention towards using ChatGPT in academic writing. *Education and Information Technologies, 30*(9), 11889-11909. https://doi.org/10.1007/s10639-024-13306-6

Belda-Medina, J., & Kokošková, V. (2024). ChatGPT for language learning: Assessing teacher candidates' skills and perceptions using the technology acceptance model (TAM). *Innovation in Language Learning and Teaching, 20*(3), 446-461. https://doi.org/10.1080/17501229.2024.2435900

Benke, E., & Szőke, A. (2024). Academic integrity in the time of artificial intelligence: Exploring student attitudes. *Italian Journal of Sociology of Education, 16*(2), 91-108. https://doi.org/10.14658/PUPJ-IJSE-2024-2-5

Bristi, N. L., Shafie, L. A., Hajimia, H. B., & Hasan, K. (2025). Decoding EFL Learners' intention to use ChatGPT for academic writing: Cognitive and emotional drivers. *International Journal of Learning, Teaching and Educational Research, 24*(3), 1-28. https://doi.org/10.26803/ijlter.24.3.1

Campos, M. (2025). AI-assisted feedback in CLIL courses as a self-regulated language learning mechanism: Students' perceptions and experiences. *European Public and Social Innovation Review, 10*, 1-14. https://doi.org/10.31637/epsir-2025-1568

Chen, C. (2016). Grand challenges in measuring and characterizing scholarly impact. *Frontiers in Research Metrics and Analytics, 1*. https://doi.org/10.3389/frma.2016.00004

Chen, J., & Delaney, V. (2025). Leveraging AI to enhance children's learning: Anchoring policy and practice in equity, AI literacy, and ethics for education leaders and teachers. *Early Childhood Education Journal, 54*, 2655-2665. https://doi.org/10.1007/s10643-025-02036-0

Chung, J. Y., & Jeong, S.-H. (2024). Exploring the perceptions of Chinese pre-service teachers on the integration of generative AI in English language teaching: Benefits, challenges, and educational implications. *Online Journal of Communication and Media Technologies, 14*(4), Article e202457. https://doi.org/10.30935/ojcmt/15266

Darmawansah, D., Rachman, D., Febiyani, F., & Hwang, G.-J. (2025). ChatGPT-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance EFL students' argumentation skills. *Education and Information Technologies, 30*(3), 3803-3827. https://doi.org/10.1007/s10639-024-12986-4

Dinh, C. T. (2025). Undergraduate English majors' views on ChatGPT in academic writing: Perceived vocabulary and grammar improvement. *FWU Journal of Social Sciences, 19*(1). http://doi.org/10.51709/19951272/Spring2025/1

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

Han, J. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 126*, Article 103502. https://doi.org/10.1016/j.system.2024.103502

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Huang, J., & Mizumoto, A. (2024). Examining the relationship between the L2 motivational self-system and technology acceptance model post ChatGPT introduction and utilization. *Computers and Education: Artificial Intelligence, 7*, Article 100302. https://doi.org/10.1016/j.caeai.2024.100302

Huang, J., & Mizumoto, A. (2025). The effects of generative AI usage in EFL classrooms on the L2 motivational self-system. *Education and Information Technologies, 30*(5), 6435-6454. https://doi.org/10.1007/s10639-024-13071-6

Huang, Y., & Wang, D. (2025). Can ChatGPT serve as a writing collaborator? Insights from Chinese EFL learners. *System, 133*, Article 103775. https://doi.org/10.1016/j.system.2025.103775

Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., Al-Khresheh, M. H., & Al-Qadri, A. H. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: An experimental study of gender and major differences. *Discover Education, 4*, Article 240. https://doi.org/10.1007/s44217-025-00700-6

Kanwal, A. (2025). Exploring the impact of ChatGPT on psychological factors in learning English writing among undergraduate students. *World Journal of English Language, 15*(3), 404-413. https://doi.org/10.5430/wjel.v15n3p404

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kawinkoonlasate, P. (2025). A comparative study of Google Gemini and ChatGPT in enhancing English language learning for EFL learners: A case study of the English research writing course. *Pedagogical Research, 10*(4), Article em0251. https://doi.org/10.29333/pr/17670

Kim, M., & Chon, Y. V. (2025). The impact of self-revision, machine translation, and ChatGPT on L2 writing: Raters' assessments, linguistic complexity, and error correction. *Assessing Writing, 65*, Article 100950. https://doi.org/10.1016/j.asw.2025.100950

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications, 11*, Article 1268. https://doi.org/10.1057/s41599-024-03755-2

Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments, 12*, Article 48. https://doi.org/10.1186/s40561-025-00406-0

Lu, D., & Zeng, Y. (2025). Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2025.2525341

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025). *Large language models: A survey*. arXiv. https://doi.org/10.48550/arXiv.2402.06196

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), Article 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mohammad, N. M., Demers, M., McCubbin, E., Mitchell, J., & Fulmer, S. (2025). How college students use ChatGPT. *Pedagogical Research, 10*(4), Article em0250. https://doi.org/10.29333/pr/17428

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*, Article 143. https://doi.org/10.1186/s12874-018-0611-x

Neshkovska, S., Ivanovska, L., Trajkovska, V., & Kasaposka-Chadlovska, M. (2025). ChatGPT in foreign language acquisition: A close ally or a distant concept for Macedonian university students? *Respectus Philologicus, 47*(52), 50-65. https://doi.org/10.15388/RESPECTUS.2025.47.4

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence, 6*, Article 100234. https://doi.org/10.1016/j.caeai.2024.100234

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Park, J. (2025). A systematic literature review of generative artificial intelligence (GenAI) literacy in schools. *Computers and Education: Artificial Intelligence, 9*, Article 100487. https://doi.org/10.1016/j.caeai.2025.100487

Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1), Article 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Pum, M. (2025). The paradox of augmentation and erosion: Navigating utility, critical agency, and trust deficit in AI writing tools among Cambodian English major undergraduates. *Interdisciplinary Educational Technology, 1*(1), Article e105. https://doi.org/10.71176/interedtech/17573

Punar Özçelik, N., & Yangın Ekşi, G. (2024). Cultivating writing skills: The role of ChatGPT as a learning assistant—A case study. *Smart Learning Environments, 11*, Article 10. https://doi.org/10.1186/s40561-024-00296-8

Rahayu, A., Tarihoran, N., Rahmawati, E., Muslihah, E., & Ma'mur, I. (2025). Navigating the digital writing landscape: EFL students' perspectives on ChatGPT utilization. *International Journal of Learning, Teaching and Educational Research, 24*(5), 727-752. https://doi.org/10.26803/ijlter.24.5.37

Salam, U. (2025). The integration of ChatGPT in English for foreign language course: Elevating AI writing assistant acceptance. *Computers in the Schools, 42*(2), 145-165. https://doi.org/10.1080/07380569.2024.2446239

Sawangwan, S. (2024). ChatGPT vs teacher roles in developing EFL writing. *International Journal of Computer-Assisted Language Learning and Teaching, 14*(1). https://doi.org/10.4018/IJCALLT.361235

Seo, J.-Y. (2024). Exploring the educational potential of ChatGPT: AI-assisted narrative writing for EFL college students. *Language Teaching Research Quarterly, 43*. https://doi.org/10.32038/ltrq.2024.43.01

Shabara, R., ElEbyary, K., & Boraie, D. (2024). Teachers or ChatGPT: The issue of accuracy and consistency in L2 assessment. *Teaching English with Technology, 24*(2), 71-92.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1260843

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270

Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education, 60*(1), Article e12811. https://doi.org/10.1111/ejed.12811

Teng, M. F., & Huang, J. (2025). Incorporating ChatGPT for EFL writing and its effects on writing engagement. *International Journal of Computer-Assisted Language Learning and Teaching, 15*(1). https://doi.org/10.4018/IJCALLT.367874

Thi, X. H. N., Thien, H. V. H., Vuong, K. N., & Nguyen, T. T. (2025). Enhancing writing skills through AI-powered tools: Perceived benefits and challenges among Vietnamese EFL students. *Discover Education, 4*, Article 472. https://doi.org/10.1007/s44217-025-00905-9

Tica, L., & Krsmanović, I. (2024). Overcoming the writer's block? Exploring students' motivation and perspectives on using ChatGPT as a writing assistance tool in ESP. *ELOPE: English Language Overseas Perspectives and Enquiries, 21*(1), 129-149. https://doi.org/10.4312/ELOPE.21.1.129-149

Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System, 127*, Article 103528. https://doi.org/10.1016/j.system.2024.103528

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467-473. https://doi.org/10.7326/M18-0850

Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies, 29*(17), 22427-22445. https://doi.org/10.1007/s10639-024-12722-y

Tseng, Y., & Lin, Y. (2024). Enhancing English as a foreign language (EFL) learners' writing with ChatGPT: A university-level course design. *The Electronic Journal of e-Learning, 2024*, 78-97. https://doi.org/10.34190/ejel.21.5.3329

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*, 523-538. https://doi.org/10.1007/s11192-009-0146-3

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, Y. (2024). Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. *System, 126*, Article 103505. https://doi.org/10.1016/j.system.2024.103505

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

Woo, D. J., Guo, K., & Susanto, H. (2025). EFL secondary students' use of ChatGPT for writing task completion pathways. *Journal of Educational Research, 118*(6), 596-609. https://doi.org/10.1080/00220671.2025.2510382

Yang, C. T.-Y., & Chen, H. H.-J. (2025). ChatGPT and L2 Chinese writing: Evaluating the impact of model version and prompt language on automated corrective feedback. *Computer Assisted Language Learning, 39*(4), 1119-1147. https://doi.org/10.1080/09588221.2025.2453205

Yasmin, M., Fatima, W., & Irshad, I. (2025). Evaluating ChatGPT's effectiveness in enhancing argumentative writing: A quasi-experimental study of EFL learners in Pakistan. *Sustainable Futures, 10*, Article 100809. https://doi.org/10.1016/j.sftr.2025.100809

Yu, H. (2024). The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. *Heliyon, 10*(2), Article e24289. https://doi.org/10.1016/j.heliyon.2024.e24289

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher-and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning, 39*(4), 1062-1091. https://doi.org/10.1080/09588221.2024.2447279

APPENDIX A

Table A1. Main findings and the full dataset employed for the present paper

Article	Title	Country	Sample size	Methodology	AI tool(s)	Outcomes summary
Afandi (2024)	Assessing text comprehension proficiency: Indonesian higher education students vs ChatGPT	Indonesia	288	Quantitative	ChatGPT	Indonesian students have a gap in understanding texts compared to ChatGPT
Alhazmi and Muftah (2025)	Evaluating ChatGPT's reliability in second language acquisition (SLA): Insights on language skills and technology's role	Saudi Arabia	48	Quantitative	ChatGPT	ChatGPT shows promise in addressing diverse SLA topics/occasional inaccuracies underscore the necessity for ongoing research to optimize its utility in educational contexts
Aljasser (2025)	Investigating EFL students' perceptions of feedback: A comparative study of instructor and ChatGPT-generated responses in academic writing	Saudi Arabia	112	Mixed methods	ChatGPT	ChatGPT's feedback was immediate and accessible/the instructors' feedback was more useful
Allen and Mizumoto (2024)	ChatGPT over my friends: Japanese English-as-a-foreign-language learners' preferences for editing and proofreading strategies	Japan	33	Mixed methods	ChatGPT	Technology provides effective feedback, improving clarity and cohesion
Almahasees et al. (2024)	Students' perceptions of the benefits and challenges of integrating ChatGPT in higher education	Jordan		Quantitative	ChatGPT	Saving time, getting immediate answers, writing for projects and homework, aiding in language learning, unified feedback, further learning material, accessibility of information/ less critical thinking, only partial language capacity, incomplete knowledge of subjects, dependency on technology, misinformation, data privacy, biased responses
Alsaweed and Aljebreen (2024)	Investigating the accuracy of ChatGPT as a writing error correction tool	Saudi Arabia		Mixed methods	ChatGPT	High-performance accuracy in encompassing verb forms, word order, passive voice, conditional sentences, subject-verb agreement, singular-plural nouns, word form, and word choices/moderate performance in articles, modals, and verb tense, low performance in unclear messages, connectors, sentence structure, and non-idiomatic expressions
Alwasidi et al. (2025)	Assessing the impact of ChatGPT on EFL students' writing productivity and proficiency	Saudi Arabia	52	Quantitative	ChatGPT	Helpful tool for beginning EFL learners, assisting with idea generation, reducing redundancy, and fostering writing clarity/ need for judicious application, overreliance and independent development writing skills
Alzubi et al. (2025)	Do AI-generative tools kill or nurture creativity in EFL teaching and learning?	Saudi Arabia	546	Descriptive	ChatGPT	Enhancing creativity, conversation practice, idea generation, feedback on writing, vocabulary building, and collaborative learning/over-reliance, critical thinking, accuracy, and intellectual property
Asadi et al. (2025)	The impact of integrating ChatGPT with teachers' feedback on EFL writing skills	Iran	68	Mixed methods	ChatGPT	Transformative potential to deliver personalized feedback, advancing writing proficiency and addressing practical challenges
Bekturov et al. (2025)	Predicting Kazakhstani TEFL students' continuance intention towards using ChatGPT in academic writing	Kazakhstan	334	Mixed methods	ChatGPT	More useful, easier to use, satisfied, greater intention to continue using it / doubts about reliability, plagiarism, loss of writing ability, desire for enhanced accuracy
Belda-Medina and Kokošková (2024)	ChatGPT for language learning: Assessing teacher candidates' skills and perceptions using the technology acceptance model (TAM)	Spain	58	Mixed methods	ChatGPT	Positive perceptions regarding perceived usefulness (PU) and behavioral intention (BI)/concerns on accuracy and reliability

Table A1 (Continued).

Article	Title	Country	Sample size	Methodology	AI tool(s)	Outcomes summary
Bristi et al. (2025)	Decoding EFL learners' intention to use ChatGPT for academic writing: cognitive and emotional drivers	Bangladesh	159	Quantitative	ChatGPT	ChatGPT in academic writing as a cognitive and emotional support tool
Campos (2025)	AI-assisted feedback in CLIL courses as a self-regulated language learning mechanism: Students' perceptions and experiences	Spain	205	Mixed methods	ChatGPT	Beneficial feedback for improving writing skills, immediacy, specificity/ concerns complexity and relevance
Chung and Jeong (2024)	Exploring the perceptions of Chinese pre-service teachers on the integration of GenAI in English language teaching: Benefits, challenges, and educational implications	China	134	Quantitative	ChatGPT	Enhance teaching efficiency, particularly in grammar and writing/ concerns about over-reliance, its impact on critical thinking, authenticity of AI-facilitated interactions, data privacy, academic integrity
Darmawansah et al. (2025)	ChatGPT-supported collaborative argumentation: Integrating collaboration script and argument mapping to enhance EFL students' argumentation skills	Indonesia	67	Quasi-experimental	ChatGPT	Enhanced EFL students' argumentative speaking performance, critical thinking awareness, and collaboration tendency
Dinh (2025)	Undergraduate English majors' views on ChatGPT in academic writing: Perceived vocabulary and grammar improvement	Vietnam	31	Mixed	ChatGPT	Enhance academic writing skills when integrated thoughtfully into curricula/overuse risks shallow learning, overdependence on the tool/ difficulty in interpreting feedback
Escalante et al. (2023)	AI-generated feedback on writing: Insights into efficacy and ENL student preference	Asia-Pacific	48+43	Quasi-experimental	ChatGPT	AI-generated feedback can be incorporated into ENL essay evaluation without affecting learning outcomes-recommend a blended approach
Han (2024)	Exploring ChatGPT-supported teacher feedback in the EFL context	China	102 students +4 instructors	Controlled experiment	ChatGPT	The importance of an "AI + Teacher" model that leverages the analytical strengths of AI while maintaining essential teacher-student interactions
Huang and Mizumoto (2024)	Examining the relationship between the L2 motivational self-system and technology acceptance model post ChatGPT introduction and utilization	Japan	35	Mixed methods	ChatGPT	A notable correlation between the L2 motivational self-system and the technology acceptance model, ought-to L2 self-positively predict actual usage
Huang and Mizumoto (2025)	The effects of GenAI usage in EFL classrooms on the l2 motivational self-system	Japan	80	Mixed methods	ChatGPT	Enhancing motivation and writing skills
Kanwal (2025)	Exploring the impact of ChatGPT on psychological factors in learning English writing among undergraduate students	Saudi Arabia	142	Quantitative	ChatGPT	Enhanced cognitive skills, thought organization, vocabulary retention, and analytical abilities in English writing, increases motivation and engagement, supports adaptation to new writing challenges, aids in managing writing-related stress/ effectiveness in reducing writing-related anxiety is inconsistent
Kim and Chon (2025)	The impact of self-revision, machine translation, and ChatGPT on L2 writing: Raters' assessments, linguistic complexity, and error correction	Korea	79	Controlled experiment	ChatGPT	Significantly reduced grammatical errors, but did not affect prepositional errors
Lu and Zeng (2025)	Exploring the use of ChatGPT-generated model texts as a feedback instrument: EFL students' text quality and perceptions	China	104	Quasi-experimental	ChatGPT	Positive perceptions of using ChatGPT-generated model texts as a feedback tool/ concern about the potential lack of contextual relevance and stylistic unnaturalness

Table A1 (Continued).

Article	Title	Country	Sample size	Methodology	AI tool(s)	Outcomes summary
Neshkovska et al. (2025)	ChatGPT in foreign language acquisition: A close ally or a distant concept for Macedonian university students?	Macedonia	180	Mixed methods	ChatGPT	Benefits for writing, translation, grammar, research/skepticism in generating high-quality writing ideas, improving speaking skills, fostering critical thinking, and creativity
Pack et al. (2024)	Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability	Asia-Pacific	119	Quantitative	Google's PaLM 2, Anthropic's Claude 2, ChatGPT-3.5 and GPT-4	Some models perform better than others, all models are subject to fluctuations in performance
Poláková and Ivenz (2024)	The impact of ChatGPT feedback on the development of EFL students' writing skills	Czech Republic	110	Mixed methods	ChatGPT	Potential of ChatGPT as an innovative educational tool for fostering writing proficiency among Gen Z students
Punar Özçelik and Yangın Ekşi (2024)	Cultivating writing skills: The role of ChatGPT as a learning assistant—A case study	Turkey	11	Controlled experiment	ChatGPT	Beneficial for acquiring formal register knowledge/unnecessary for informal writing, the effectiveness of ChatGPT in teaching neutral register was questioned
Rahayu et al. (2025)	Navigating the digital writing landscape: EFL students' perspectives on ChatGPT utilization	Indonesia	106	Mixed methods	ChatGPT	Need for institutional policies, practical benefits
Salam (2025)	The integration of ChatGPT in English for foreign language course: Elevating AI writing assistant acceptance	Indonesia	95	Mixed methods	ChatGPT	Fostered a positive attitude toward technological change and encouraged students to adopt it more frequently
Sawangwan (2024)	ChatGPT vs teacher roles in developing EFL writing	Thailand	50	Mixed methods	ChatGPT	Teachers' roles in areas ChatGPT cannot replicate are understanding writing criteria, developing individual critical thinking, ethical guidance
Seo (2024)	Exploring the educational potential of ChatGPT: AI-assisted narrative writing for EFL college students	South Korea	44	Quantitative	ChatGPT	Writing fluency, overall performance, clause complexity, clausal complements per clause/ syntactic complexity was lower
Shabara et al. (20245)	Teachers or ChatGPT: The issue of accuracy and consistency in L2 assessment	Egypt	100	Quantitative	ChatGPT	Teachers' scores exhibited a higher level of accuracy, a moderate, yet questioned level of intra-rater reliability, concerns about the accuracy and consistency of ChatGPT's scoring
Teng (2025)	Metacognitive awareness and EFL learners' perceptions and experiences in utilizing ChatGPT for writing feedback	China	40	Mixed methods	ChatGPT	ChatGPT in academic writing as a cognitive and emotional support tool
Teng (2024)	"ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing	China	45	Mixed methods	ChatGPT	Significant positive effects of assistance on writing, writing motivation, self-efficacy, engagement, and collaborative writing tendency/ concerns about the use of ChatGPT
Teng and Huang (2025)	Incorporating ChatGPT for EFL writing and its effects on writing engagement	China	83	Quantitative	ChatGPT	Enhance specific aspects of student engagement, like affective and behavioral engagement, in EFL writing/ Cognitive engagement did not differ significantly
Tica and Krsmanović (2024)	Overcoming the writer's block? Exploring students' motivation and perspectives on using ChatGPT as a writing assistance tool in ESP	Croatia	79	Mixed methods	ChatGPT	Versatility, speed, accuracy/ concerns about limitations

Table A1 (Continued).

Article	Title	Country	Sample size	Methodology	AI tool(s)	Outcomes summary
Tsai et al. (2024)	Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges		44	Quantitative	ChatGPT	Significant improvements in all four dimensions of writing quality, vocabulary, grammar, organization, content, effective in providing timely feedback/ disproportionate improvement raises concerns about fairness in evaluation
Wang (2024)	Cognitive and sociocultural dynamics of self-regulated use of machine translation and GenAI tools in academic EFL writing	Japan	79	Mixed methods	ChatGPT, Google Translate	Writing outcomes improved/the role of instructors and peer learning cannot be replaced by the AI tools
Wiboolyasarin et al. (2024)	Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments	Thailand	39	Quasi-experimental	ChatGPT	Positive impact of a collaborative writing intervention improved by AI feedback on specific aspects of writing
Woo et al. (2025)	EFL secondary students' use of ChatGPT for writing task completion pathways	Hong Kong	4	Controlled experiment	ChatGPT	With personalized and collaborative instruction on prompt content, sequence and capabilities, students can become more strategic, sophisticated collaborators with ChatGPT on writing tasks
Yang and Chen (2025)	ChatGPT and L2 Chinese writing: evaluating the impact of model version and prompt language on automated corrective feedback	China	153	Quantitative	ChatGPT	ChatGPT-4.0 superior performance in generating more accurate, detailed, and understandable corrective feedback compared to ChatGPT 3.5, the model version significantly influences effectiveness
Yasmin et al. (2025)	Evaluating ChatGPT's effectiveness in enhancing argumentative writing: A quasi-experimental study of EFL learners in Pakistan	Pakistan	30	Quasi-experimental	ChatGPT	Progress in students' argumentative writing
Zou et al. (2025)	Investigating students' uptake of teacher-and ChatGPT-generated feedback in EFL writing: A comparison study	China	20	Mixed methods	ChatGPT	General preference for teacher feedback, the potential of using ChatGPT feedback as a complementary resource