

# The impact of using ChatGPT for feedback in EFL writing: a systematic review

RAQUEL PARRA NÚÑEZ

Universidad Nacional de Educación a Distancia (UNED), España  
rparra52@alumno.uned.es

<https://orcid.org/0000-0002-5503-3817>

MARIA DOLORES CASTRILLO DE LARRETA-AZELAIN

Universidad Nacional de Educación a Distancia (UNED), España

mcastrillo@flog.uned.es

<https://orcid.org/0000-0002-0713-2351>

## Abstract

The integration of generative artificial intelligence (GenAI) tools, such as ChatGPT, into English as a Foreign Language (EFL) writing instruction has garnered significant attention in recent years. This systematic review synthesizes 30 empirical studies published between 2023 and 2025 that explore the impact of ChatGPT feedback on students' writing performance, examining which dimensions of writing (e.g. linguistic accuracy, content development, organization or vocabulary use) experience more improvements as a result of ChatGPT usage. Results indicate that ChatGPT is particularly effective in enhancing surface-level writing features related to the accurate use of language, as well as in the overall increase in writing quality and the enhancement of text organization. However, the analysed studies also report challenges related to students' overreliance on the tool, and the need for a student-centred integration of ChatGPT within instructional frameworks. Implications for pedagogy suggest that the effectiveness of ChatGPT as a writing feedback tool is maximized when coupled with the teaching of effective prompting skills, students' critical engagement with AI feedback as well as teacher input and supervision. Further research is required to explore long-term effects and the sustainability of GenAI-driven writing improvements.

## Keywords

ChatGPT; artificial intelligence; writing; feedback.

# Impacto de ChatGPT como herramienta de retroalimentación en la escritura en inglés como lengua extranjera: revisión sistemática

## Resumen

La integración de herramientas de inteligencia artificial generativa (GenAI), como ChatGPT, en la enseñanza de la escritura en inglés como lengua extranjera ha recibido una atención considerable en los últimos años. Esta revisión sistemática sintetiza 30 estudios empíricos publicados entre 2023 y 2025 que exploran el impacto de la retroalimentación de ChatGPT en el desempeño de los estudiantes en la escritura, e investiga qué dimensiones de esta (por ejemplo, precisión lingüística, desarrollo del contenido, organización del texto o uso del vocabulario) experimentan mayores mejoras como consecuencia del uso de ChatGPT. Los resultados indican que ChatGPT es especialmente eficaz para mejorar aspectos superficiales de la escritura relacionados con el uso preciso de la lengua, así como en el aumento general de la calidad de la escritura y la mejora en la organización textual. Sin embargo, los estudios analizados también reportan desafíos relacionados con la dependencia excesiva de los estudiantes de la herramienta y la necesidad de integrar ChatGPT en los distintos marcos educativos con una perspectiva centrada en el alumno. Las implicaciones pedagógicas sugieren que la efectividad de ChatGPT como herramienta de retroalimentación en la escritura se maximiza cuando se combina con la enseñanza de habilidades efectivas para generar *prompts*, así como con la contribución y supervisión del profesor y la capacidad de los alumnos de ser críticos con la retroalimentación generada por la herramienta. Se requieren más estudios para explorar los efectos en la escritura a largo plazo, así como la sostenibilidad de las mejoras derivadas del uso de herramientas de GenAI.

## Palabras clave

ChatGPT; inteligencia artificial; escritura; retroalimentación.

Recibido el 20/05/2025

Aceptado el 23/06/2025

## Cómo citar/how to cite

Parra Núñez, R., Castrillo de Larreta-Azelain, M. D. (2025). The impact of using ChatGPT for feedback in EFL writing: a systematic review. *Revista Internacional De Lenguas Extranjeras / International Journal of Foreign Languages*, (23), 95-119. <https://doi.org/10.17345/rile23.4221>

## 1. Introduction

The development of proficient writing skills remains a cornerstone of English as a Foreign Language (EFL) and English as a Second Language (ESL) education. Central to this development is the provision of effective, timely, and constructive feedback, which serves to motivate learners and guide them in identifying errors, understanding expectations, and refining their textual production (Hyland & Hyland, 2006). However, traditional feedback methods, often reliant on teacher input, face persistent challenges, including significant time and workload demands on educators (McMartin-Miller, 2014).

As a result, teachers and researchers have long been looking at ways to find tools that can help them automatise some of the feedback providing process (Teng, 2024b). Over the decades, Computer-Assisted Language Learning (CALL) has introduced various tools to support writing, from basic grammar checkers to more sophisticated Automated Writing Evaluation (AWE) systems. More recently, the advent of Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLMs) like ChatGPT, has brought about a paradigm shift. These models have the ability to provide real-time, individualized support and feedback that addresses the diverse needs of learners (Barrot, 2023). This is particularly important in ESL and EFL contexts, where students often face challenges related to diverse aspects of writing and do not have immediate access to support from the teacher or peers.

Recent systematic reviews on the use of ChatGPT in writing (Alsaedi, 2024; Teng, 2024a) have identified several recurring benefits and challenges associated with the use of ChatGPT in EFL/ESL writing. Reported benefits in both reviews include increased learner motivation, enhanced writing performance, as well as the ability to receive immediate and personalized feedback without fear of being judged.

However, challenges persist. Both reviews highlight the superficial nature of ChatGPT's feedback, which tends to focus on surface-level errors rather than deeper content or rhetorical structure. Additionally, concerns about overreliance on AI tools and threats to academic integrity are cited, along with the need for teacher mediation and digital literacy training (Alsaedi, 2024; Teng, 2024a). These findings underscore the importance of integrating ChatGPT strategically and ethically within writing instruction, rather than using it as a standalone tool. Additionally, it is necessary to teach students prompt engineering skills and how to critically evaluate AI-generated feedback in order to help them optimize their interactions with these tools (Barrot, 2023).

The existing reviews by Alsaedi (2024) and Teng (2024a) provide valuable overviews of the general advantages, challenges, potentials, and pitfalls of using ChatGPT in EFL/ESL writing. However, as research in this area matures, there is a growing need for a more focused synthesis that specifically examines the empirical evidence concerning the measurable impact of ChatGPT-facilitated feedback on distinct dimensions of student writing performance. While previous reviews touch upon performance improvements in general, they often group various outcomes together under broad categories like "advantages" or "opportunities."

This systematic review aims to address this gap by concentrating specifically on empirical studies that investigate and report changes in student writing performance across key sub-skills. By analysing findings related to linguistic accuracy, organization, content development, vocabulary usage, and overall writing quality, this review seeks to provide a more granular understanding of how and in what specific areas ChatGPT feedback demonstrably influences EFL/ESL writing outcomes, as reported in the primary research literature. This focused approach will complement existing reviews by offering a detailed synthesis of performance-related evidence, thereby providing educators and researchers with a clearer picture of the empirically supported effects of using ChatGPT as a feedback tool.

In this vein, the following research question is posed:

- (1) What is the impact of using ChatGPT for feedback on student writing performance in empirical studies in an EFL/ESL context?

## **2. Literature review**

### **2.1 Writing and feedback in EFL/ESL learning**

Developing writing proficiency is a fundamental goal in EFL and ESL education, recognized as a complex cognitive and linguistic task involving several sub-tasks, such as brainstorming, planning, drafting, and revising (Flower & Hayes, 1981). Moreover, in the case of writing in a second or foreign language, there is an added layer of complexity, as learners usually experience additional difficulty in accessing the necessary linguistic resources (Godwin-Jones, 2022). Research has shown that effective feedback is crucial to this process, helping learners notice gaps in their current abilities, test hypotheses, and ultimately improve their skills (Hyland & Hyland, 2006). Moreover, learners benefit from a personalized approach in which teachers provide them with tailored feedback and assign them writing tasks connected to their interests and experiences (Hyland, 2022). However, large class sizes, heavy workload and time constraints can have potential negative influence on teachers' provision of timely, consistent, and personalized feedback that effectively facilitates uptake while it addresses diverse learner needs (Warschauer & Ware, 2006).

### **2.2 Technological advancements in writing support: from automated writing evaluation systems to ChatGPT**

Technology has long been explored as a means to support writing instruction and alleviate some feedback burdens. Early CALL integrated word processors and basic grammar or spell checkers (Warschauer & Healey, 1998). More sophisticated AWE systems, such as e-Rater or Criterion, emerged later, offering holistic scores and feedback on specific writing traits based on computational analysis (Zupanc & Bosnić, 2015). While beneficial for providing rapid feedback on certain aspects, particularly surface-level errors like

grammar and mechanics, these earlier tools lacked the interactivity, contextual understanding, and ability to provide nuanced, explanatory, or dialogic feedback characteristic of human instructors (Stevenson, 2016; Wang, 2015; Ware, 2011). Their feedback was often perceived as generic, vague, sometimes difficult to understand, and limited in addressing higher-order concerns like argumentation or rhetorical effectiveness (Li et al., 2015; Ware, 2011).

The launch of ChatGPT by OpenAI in late 2022 marked a significant advancement in AI, specifically in the capabilities of LLMs for natural language processing and generation. Its ability to engage in seemingly coherent conversational interaction, generate diverse forms of human-like text, understand context, synthesize information, and perform a wide range of language-related tasks quickly captured the attention of the education sector globally. In the context of EFL and ESL writing, ChatGPT was immediately recognized for its potential to assist learners and instructors throughout the writing process—from brainstorming and outlining to drafting, translating, revising, and providing interactive feedback, even though some limitations have also been acknowledged (Barrot, 2023; Kohnke et al., 2023; Teng, 2024a).

### 3. Methodology

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021), which provide a standardized framework for the transparent and replicable reporting of evidence syntheses.

#### 3.1 Search strategy and databases

To identify relevant literature, a comprehensive search was conducted across three major academic databases commonly used in applied linguistics and education research: Scopus, Web of Science, and Ebsco Host. Scopus and Web of Science have often been used in combination in previous studies (Dizon & Gayed, 2024; López-Sánchez et al., 2023) and Ebsco Host was added to make the research more comprehensive. The final search was carried out on March 24, 2025, and it focused on articles published from 2022 onward, following the public release of ChatGPT. In order to find out studies relevant to our research question, the following combinations of keywords were used, adapted to the syntax required in the different databases:

- *“ChatGPT” OR “artificial intelligence” OR “AI”*
- *“EFL” OR “ESL” OR “L2” OR “English as a Foreign Language”*
- *“feedback”*
- *“writing”*

### 3.2 Inclusion and exclusion criteria

The results of the search were screened using the following inclusion and exclusion criteria:

- (1) Language: the review includes only studies written in English.
- (2) Population: EFL/ESL learners. Studies that deal with interventions in other L2 contexts are excluded.
- (3) LLM chatbot: only studies that focus on the use of ChatGPT are included, excluding those examining other tools (e.g. Gemini, custom-generated LLMs).
- (4) Empirical studies: The included studies describe an intervention with a group of ESL/EFL learners concerning use of ChatGPT for writing feedback, and they include an analysis of its impact on their performance. Studies that rely only on questionnaires or surveys about ChatGPT use preferences or students' perceptions, without any analysis of writing samples, are excluded.
- (5) Access: articles that were not accessible in full text at the time of the search through institutional access or open-access databases were excluded.

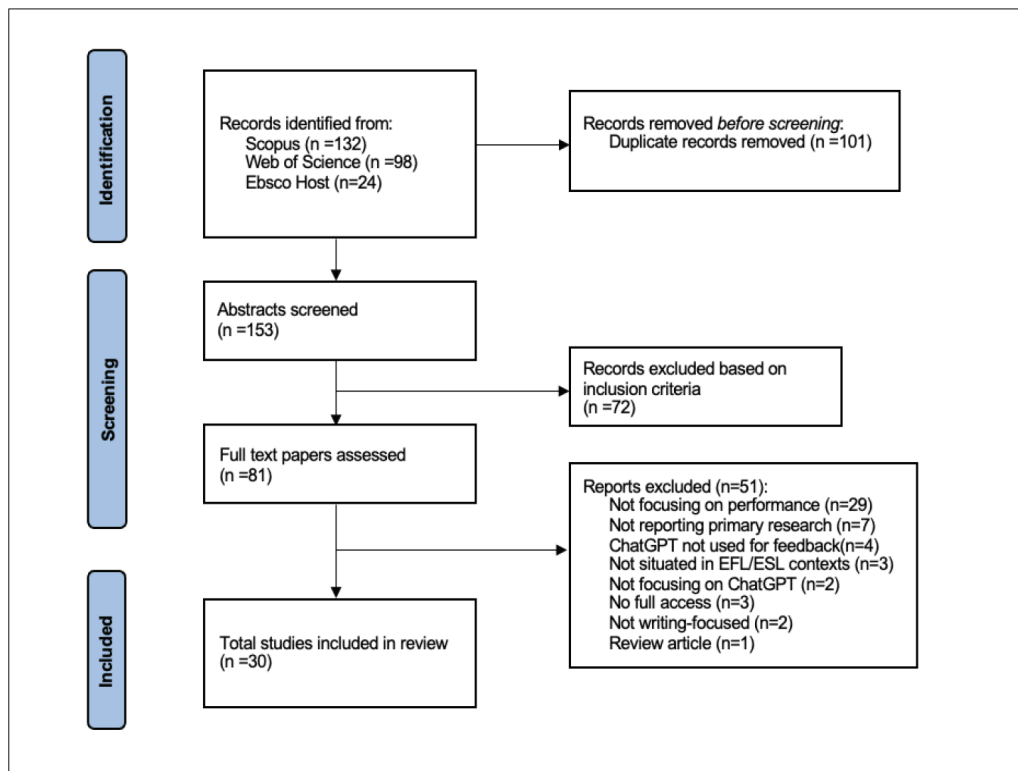
### 3.3 Screening and selection process

The initial search yielded 254 records (*Scopus=132, WoS=98, Ebsco Host=24*). After removing duplicates ( $n=101$ ), 153 articles were screened based on titles and abstracts. Those studies that did not comply with the inclusion criteria were excluded ( $n=72$ ), and full texts of the 81 potentially eligible studies were retrieved and assessed against the inclusion criteria. A total of 51 studies were excluded from this review based on the inclusion criteria. The most common exclusion reason was the lack of data on writing performance ( $n=29$ ). These studies addressed ChatGPT use in EFL contexts or writing-related tasks but did not report measurable outcomes related to learners' writing skills (e.g., accuracy, fluency, organization or content development). A second major exclusion category was the absence of a clear instructional or intervention-based context ( $n=7$ ). These studies explored either students' perceptions or developed theoretical implications without describing how ChatGPT was used by students as a learning tool and without assessing any writing-related outcomes or examining writing samples. A number of studies were also excluded because they did not focus on the use of ChatGPT for feedback ( $n=4$ ), despite examining ChatGPT in writing contexts. Similarly, studies that were not situated in EFL/ESL contexts ( $n=3$ ) were excluded to maintain consistency with the target population. Articles for which full-text access was unavailable ( $n=3$ ) were also excluded, despite attempts to locate them via institutional and open-access databases, as were studies which focused on AI tools other than ChatGPT ( $n=2$ ), such as custom-generated LLMs or alternative platforms. Other exclusions included studies that were not

writing-focused ( $n=2$ ) and a review article ( $n=1$ ). These criteria were applied rigorously during the full-text screening phase to ensure that the final set of included studies provided relevant and analysable data directly aligned with the review’s research questions. Following this process, a total of 30 studies were included in the final review.

A PRISMA flow diagram (Figure 1) outlines the selection process, including the number of records identified, screened, and excluded at each stage.

Figure 1. PRISMA flow diagram of selection process



## 4. Results

The included studies varied in several key aspects related to their research contexts, participant profiles, and design. In terms of educational context, the vast majority ( $n=28$ ) were conducted with adult learners, typically at the university or college level. Only two studies (Jamshed et al., 2024; Meyer et al., 2024) focused on upper secondary school students, suggesting that most current research is situated in adult or higher education contexts. The predominance of studies involving adult learners likely reflects both practical and ethical considerations surrounding the use of ChatGPT in educational contexts. Adults, particularly university students, tend to possess greater digital literacy and academic autonomy, making them more capable of using GenAI tools like ChatGPT with minimal supervision. In contrast, implementing such tools with younger learners—especially in secondary or primary education—raises ethical concerns related to data privacy or critical evaluation skills of AI-generated content. Moreover, minors require closer teacher monitorisation to ensure appropriate and pedagogically sound

use of GenAI, which presents logistical and regulatory challenges for researchers and institutions. Probably, as a result of some or all of the previous factors, current research in this field has primarily centred on adult populations.

The number of participants across studies ranged widely, from 11 (Hwang et al., 2024) to 459 (Meyer et al., 2024), reflecting diverse research designs—from small-scale exploratory case studies to larger quantitative or qualitative analyses. Similarly, the duration of the studies varied, with some limited to a single-session task, while others extended across several weeks or an entire academic term. None of the studies included in this review adopted a longitudinal design, which likely reflects the recent emergence of ChatGPT, as well as the exploratory nature of current research, which is still in the early phase of testing feasibility and short-term effects.

To provide further context, country of origin of the first author and the year of publication are presented in Figures 2 and 3, respectively. Figure 1 shows that although the use of ChatGPT in EFL/ESL writing has gained global attention, the geographic distribution of research in this area is far from uniform. As illustrated in Figure 2, the majority of studies were authored by researchers based in Asia and the Middle East, with China alone contributing a third of the total studies. Countries such as Saudi Arabia, Japan, Taiwan, South Korea, and Iran also appear relatively prominently, while contributions from Europe and the Americas are notably scarce. This geographic concentration may influence the findings in terms of educational practices, teacher roles, or attitudes toward AI integration. For instance, test-based accountability systems, which are prevalent in some Asian countries like China (Wang & McLaughlin, 2024), may shape how innovative tools like ChatGPT are implemented and perceived. Therefore, caution is needed when generalizing these results to other regions, particularly those with different pedagogical cultures.

Figure 2. Distribution of studies per first author's country of affiliation

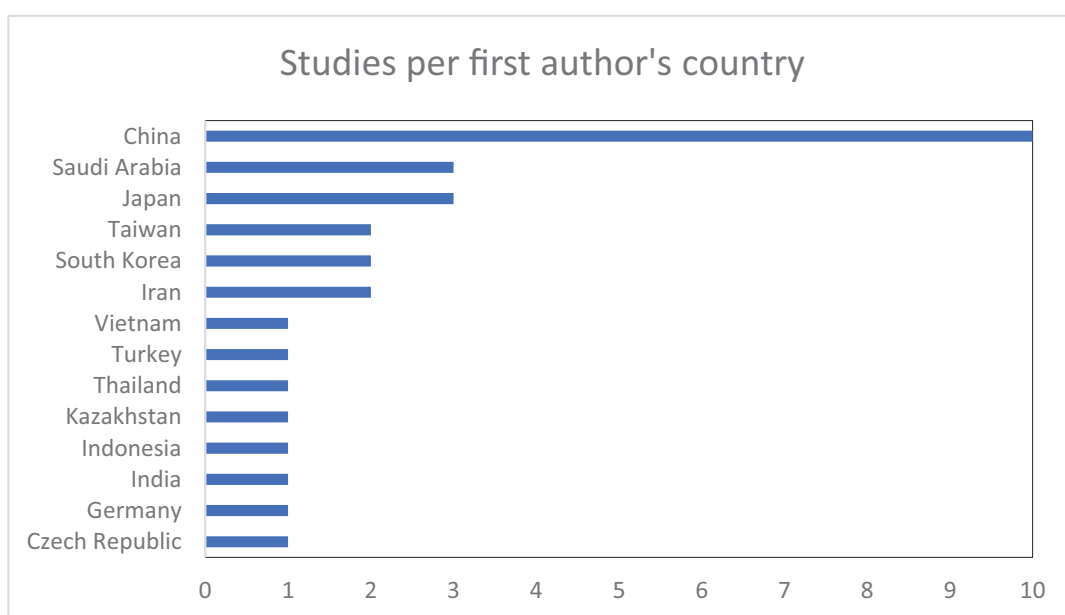
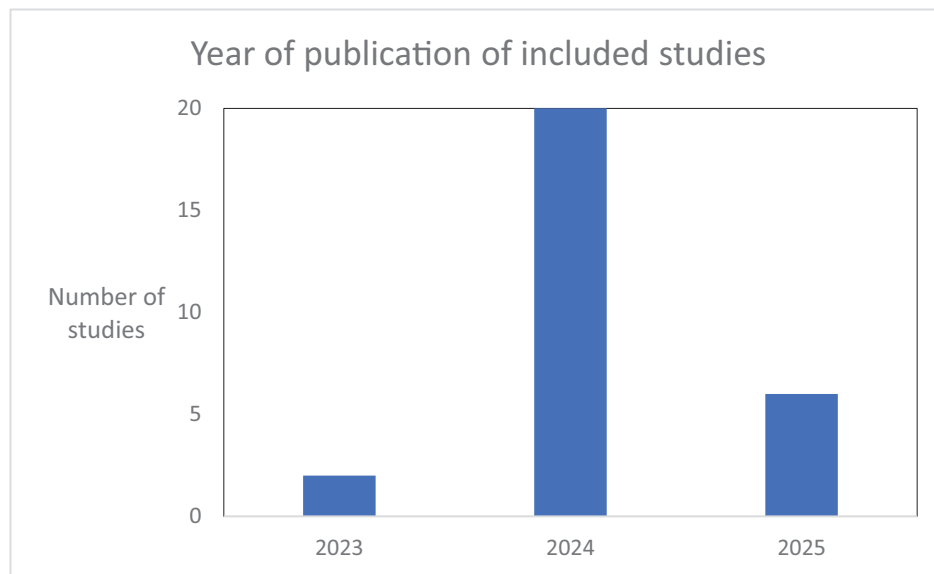


Figure 3. Number of included studies published per year



The publication timeline of the included studies reflects the rapidly emerging nature of research on ChatGPT in EFL/ESL writing. As shown in Figure 3, most of studies were published in 2024, with a small number appearing in late 2023 and the remainder emerging in early 2025. The surge of articles in 2024 corresponds closely with the public release and increasing accessibility of ChatGPT in late 2022, indicating that the field is still in its early stages of empirical exploration. The concentration of publications in a single year suggests a strong and immediate academic response to the tool's potential in language education, while also highlighting the need for ongoing research to evaluate its sustained impact and evolving use over time.

In order to identify the specific areas of writing supported by ChatGPT, all included studies were systematically coded based on their reported outcomes. A thematic analysis was conducted on the extracted data, allowing for the identification of common patterns. Each article was reviewed in detail and categorized according to the main area(s) of improvement it addressed. As shown in Figure 4, the most frequently reported area of improvement was linguistic accuracy, followed by overall writing quality and text organization, while fewer studies showed gains in content development or vocabulary use. Table 1 presents the studies grouped by the specific area(s) of writing improvement they addressed, showing the main areas of improvement reported in the articles examined.

While this systematic review included 30 studies aligned with the research questions, only 25 of these are reflected in the main results presented above. The remaining five studies, although meeting the inclusion criteria and offering valuable insights, focus on less frequently reported aspects of writing performance. To maintain clarity and focus, this review emphasizes the most recurrent and broadly relevant domains identified in the literature. Nonetheless, the excluded studies are acknowledged as important contributions that highlight emerging or context-specific dimensions of ChatGPT's impact on EFL writing, deserving further investigation.

Figure 4. Main areas of improvement reported in included studies

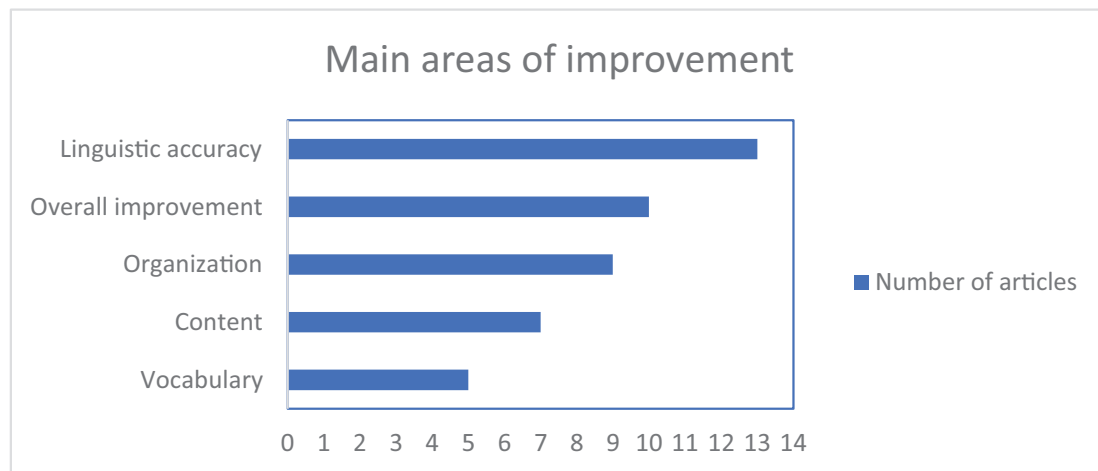


Table 1. Studies per reported area of improvement

Area of improvement	Studies
<i>Linguistic accuracy</i>	(Abduljawad, 2024; Asadi et al., 2025; Bacon & Maneerutt, 2024; Campos, 2025; Hou et al., 2024; Hwang et al., 2024; Jamshed et al., 2024; Polakova & Ivenz, 2024; Seo, 2024; Song & Song, 2023; Tsai et al., 2024; Tseng & Lin, 2024; Xiao, 2024)
<i>Overall improvement</i>	(Ghafouri et al., 2024; Han & Li, 2024; Mahapatra, 2024; Meyer et al., 2024; Oktarin et al., 2024; Özdere, 2025; Shi et al., 2025; Song & Song, 2023; Su et al., 2024; Zheldibayeva, 2025)
<i>Organization</i>	(Al Fraidan, 2025; Asadi et al., 2025; Bacon & Maneerutt, 2024; Hou et al., 2024; Seo, 2024; Song & Song, 2023; Tsai et al., 2024; Tseng & Lin, 2024; Xiao, 2024)
<i>Content</i>	(Al Fraidan, 2025; Nguyen et al., 2024; Polakova & Ivenz, 2024; Seo, 2024; Song & Song, 2023; Tsai et al., 2024; Xiao, 2024)
<i>Vocabulary</i>	(Asadi et al., 2025; Bacon & Maneerutt, 2024; Hou et al., 2024; Nguyen et al., 2024; Tsai et al., 2024)

Among the 25 studies included in the main analysis, linguistic accuracy emerged as the most frequently reported area of improvement (52%, 13 studies), followed by overall improvement (40%, 10 studies), organization (36%, 9 studies), content development (28%, 7 studies), and vocabulary (20%, 5 studies). These figures reflect the thematic priorities within the selected literature. It is important to note that some studies were coded under multiple categories, as they reported improvements across more than one area; consequently, the sum of the percentages exceeds 100%.

## 5. Discussion

The studies reviewed were generally positive about the influence of ChatGPT as a feedback tool in writing. The main areas of improvement identified in the included studies—

linguistic accuracy, overall improvement, organization, content, and vocabulary—are discussed below, noting both converging patterns and contradictions where relevant.

## 5.1 Linguistic accuracy

Linguistic accuracy encompassing grammar, punctuation, spelling and syntactic correctness emerged as one of the most frequently reported areas of improvement. Most of the studies in this group reported significant enhancement in this domain. For instance, Xiao (2024) found that the experimental group of Chinese EFL learners encouraged to use ChatGPT during the pre-writing and revision phases showed better writing proficiency than the control group in terms of language use.

Similarly, in their study with senior secondary ESL learners in India, Jamshed et al. (2024) found that the experimental group receiving feedback via the ChatGPT mobile application showed a significant reduction in common grammatical errors (including third-person singular present, past tense, plurals, etc.) compared to the control group receiving traditional teacher feedback. However, despite this general overall improvement in grammar, specific grammatical features—namely comparative and superlative forms—were better acquired under traditional teacher feedback. This suggests that while ChatGPT may provide reliable surface-level corrections, it may fall short in delivering rule-based grammatical instruction, even when explicitly prompted to do so, particularly for forms that require explicit explanation, contextual differentiation, or recursive practice. The authors conclude that AI-generated feedback, although effective in many respects, should be viewed as a supplement rather than a substitute for human instruction and teacher input remains essential for supporting nuanced grammatical learning and ensuring that learners fully understand the underlying rules behind automated corrections.

In the same vein, Hwang et al. (2024) observed that Korean first-year university learners predominantly used ChatGPT mainly for surface-level revisions, and their improvements in higher-order aspects (e.g., content, organization) were minimal. This suggests that while linguistic accuracy benefits significantly, these improvements may not translate into more substantive writing development without targeted instructional scaffolding and efficient prompt crafting techniques. Similarly, Bacon and Maneerutt (2024), in their study with Thai undergraduates integrating the use of ChatGPT with peer mentoring, observed significant advancements in grammar and vocabulary use, which the authors attributed mainly to the immediacy of the feedback. However, it was observed that English proficiency levels played an important role, as higher proficiency learners showed greater gains in both grammar and vocabulary. These students were able to engage more critically with the AI chatbot and perform higher-order revisions, whereas lower-proficiency learners focused merely on surface-level aspects and struggled with using the feedback from ChatGPT for more complex revisions.

Adding another dimension to this discussion, Abduljawad (2024) reported improved grammatical accuracy in their study conducted in Saudi Arabian higher education. However, the authors observed that learners tended to adopt a cautious writing

approach, since they often prioritized correctness over creativity or individuality in expression, raising concerns about the potential of AI tools to limit personal voice and risk-taking in writing.

Likewise, Campos (2025), in a study conducted with Japanese university students enrolled in a business-focused CLIL course, reported high correction rates for grammar (79.1%) and spelling (80.3%) errors after students revised their drafts using ChatGPT feedback. Students appreciated the tool's immediacy and clarity, with over 85% reporting it was helpful for improving grammatical accuracy. However, the study also highlighted challenges related to the complexity of some AI-generated feedback and the potential for over-reliance on the tool, which could undermine learners' critical thinking and revision autonomy.

These findings point to ChatGPT as a powerful tool for enhancing grammatical accuracy and encouraging iterative revisions. The predominance of improvements in this domain may be explained by its strength in offering clear, form-focused corrections, which are more readily actionable by learners. However, its effectiveness is maximized when it is strategically embedded within a learner-centred instructional framework that promotes metacognitive awareness, critical engagement with feedback, and balanced teacher mediation.

## 5.2 Overall improvement

Several studies in this review reported overall improvement in student writing performance following ChatGPT-supported interventions. Rather than isolating gains in discrete areas, these studies assessed overall writing quality, typically using aggregate measures of multiple subskills. The results suggest that when integrated thoughtfully into instruction, ChatGPT can serve as a multi-functional support tool, leading to meaningful development in EFL learners' writing competence.

Shi et al. (2025) present a comprehensive comparison between ChatGPT, an AWE system, and traditional feedback systems in an EFL context. Using an eleven-week intervention with pre- and post- testing, they found that Chinese university learners in the ChatGPT group achieved significantly greater gains in overall writing performance compared to those in the AWE and control groups. These gains included improvements in coherence, vocabulary, grammar, and organization, which the authors attributed largely to the dialogic and immediate nature of ChatGPT feedback. However, despite these performance benefits, the study revealed a seemingly counterintuitive outcome regarding students' ideal L2 writing self—a construct rooted in Dörnyei's (2009) L2 Motivational Self System, which conceptualizes the learners' aspirational identity as proficient L2 writers. Post-test scores on this measure were significantly lower for the ChatGPT group than the AWE group. Qualitative data suggested that overreliance on AI-generated content led some students to feel detached from their own writing process, raising concerns about their creativity and authorship. This finding points to a possible paradox: while ChatGPT may enhance the quality of writing output, it may simultaneously challenge students' sense of ownership over their learning trajectory if not critically

mediated. In the same vein, Nguyen et al. (2024) found that learners reporting high levels of affective engagement with ChatGPT also demonstrated significantly better writing outcomes. The authors identified a positive correlation between emotional engagement and overall writing performance, suggesting that emotional connection with the writing process can enhance learning. Together, these studies highlight a critical tension: while positive emotional experiences can support short-term gains, long-term alignment with one's writer identity may falter if the tool is not used critically. This could suggest that optimal learning with ChatGPT occurs when emotional engagement is paired with a reflective and autonomous use that helps both the development of performance and the maintenance of self-concept.

Moreover, Zheldibayeva (2025) presents a nuanced view of ChatGPT's impact on writing development, showing that while overall improvement is evident in the short term, these gains may not be fully sustained over time. In a quasi-experimental study with non-English majors, the experimental group that engaged in weekly AI-mediated writing tasks using a ChatGPT-based chatbot demonstrated significantly higher writing scores at the immediate post-test compared to a traditionally taught control group. However, at the delayed post-test—conducted several weeks after the intervention—this performance advantage had notably diminished, with no significant difference between groups. The results suggest that although ChatGPT can effectively produce short-term writing gains through its immediate, personalized feedback, these improvements may regress without continued use or reinforcement. This highlights a potential challenge in sustaining AI-driven learning benefits and underscores the importance of longitudinal support strategies to maintain writing proficiency gains over time.

In their study with Chinese university learners, Su et al. (2024) highlight that overall improvement in EFL writing is closely linked to how learners process and respond to ChatGPT-generated feedback. Using learning analytics, the study identified three distinct learner profiles—thinking-alienator, tool-inspired, and comprehension-focused—each representing varying degrees of cognitive and affective engagement with AI feedback. *Thinking-alienators* showed minimal interaction with the feedback, often passively accepting the suggestions received, which resulted in the weakest writing improvements. *Tool-inspired* learners actively used ChatGPT's feedback, often experimenting with rephrasing or revising their work based on suggestions; while engaged, their revisions were more surface-level. In contrast, *comprehension-focused* learners—who achieved the most substantial gains in writing quality—took a reflective and analytical approach: they critically examined ChatGPT's suggestions, assessed their relevance, and thoughtfully integrated them into their revisions. While tool-inspired learners also engaged actively—often rephrasing feedback or using it to spark deeper revisions—they were slightly outperformed by comprehension-focused peers who prioritized understanding and accurate application over experimentation. Once more, these findings suggest that overall writing improvement is not just about using AI tools, but about how thoughtfully learners engage with and internalize the feedback provided. These findings reinforce the importance of fostering feedback literacy and strategic interaction in AI-enhanced writing pedagogy.

In a complementary line of inquiry, Ghafouri et al. (2024) introduced a pedagogically structured intervention that differs notably from other studies reviewed in this section. Conducted with Iranian university students and their EFL instructors, the study implemented a three-phase protocol—planning, instruction, and assessment—in which ChatGPT was primarily used by the teachers to support lesson design, provide scaffolded feedback, and simulate IELTS-style evaluations. While students did not engage directly with ChatGPT during in-class activities, they were encouraged to use the tool autonomously at home to revise and improve their writing assignments. This hybrid model aimed to enhance student learning outcomes while also supporting teacher development. The results showed significant gains in overall writing performance for students in the experimental group, as well as an increase in motivation and engagement. Additionally, the participating teachers reported higher levels of self-efficacy by the end of the program, suggesting that structured teacher-mediated AI integration can benefit both learners and instructors in EFL writing contexts.

Further support for the overall improvement of student writing performance comes from other studies conducted in varied EFL tertiary education contexts. Mahapatra (2024) conducted a mixed-methods intervention with Indian science and engineering undergraduates. The study confirmed that ChatGPT, when used for self- and peer-assessment activities, contributed to significant writing gains. Focus group discussions revealed that learners particularly valued its support in idea generation, textual focus, and grammatical accuracy, though some voiced concerns about overreliance and diminished motivation to think independently, highlighting the need for guided and critical integration into writing instruction. In an Indonesian university setting, Oktarin et al. (2024) reported overall improvements in writing performance following a sixteen-week ChatGPT-supported intervention. Their findings emphasized the role of ChatGPT in promoting feedback literacy and learner autonomy, as they showed that students used the tool not only to revise texts but also to engage in metacognitive reflection and collaborative practices. Özdere (2025), working with Turkish EFL undergraduates, combined AI-generated feedback with targeted training sessions—teacher-led workshops specifically designed to address common writing issues identified in the AI feedback. Students used ChatGPT autonomously to revise their drafts, and the study reported significant gains in writing scores across multiple versions. The author emphasizes the importance of structured pedagogical support provided by the teacher in order to help learners interpret and apply AI feedback effectively. Finally, Han and Li (2024) examined an *AI + teacher* feedback model in Chinese undergraduate writing classes, where instructors adapted ChatGPT-generated suggestions and delivered them as personalized written feedback. The study reported a notable reduction in error rates across successive writing tasks and highlighted improvements in both specific and global aspects of student writing, including rhetorical development and organization. The authors argue that combining AI input with teacher mediation offers a scalable and pedagogically sound model for enhancing EFL writing instruction.

To conclude, while multiple studies in this review affirm the potential of ChatGPT to enhance overall EFL writing performance, they collectively underscore that its

effectiveness depends on how learners interact with the tool. Gains in writing quality are most pronounced when students use ChatGPT in reflective, emotionally engaged, and autonomous ways, especially when supported by teacher mediation or structured pedagogical guidance. However, the risk of overreliance and reduced learner agency cautions against uncritical adoption of AI-generated feedback. Furthermore, the contrast between short-term performance gains and longer-term sustainability concerns highlights the need for pedagogical frameworks that promote strategic and critical engagement with AI feedback. Ultimately, ChatGPT should not be positioned as a replacement for student thinking or teacher input, but as a catalyst for deeper learning, always paired with metacognitive support, affective investment, and guided critical use.

The following section explores how ChatGPT interventions affect organizational quality in learner writing and considers the extent to which this tool can support structural and logical development beyond surface-level accuracy.

### 5.3 Organization

Improvement in organizational quality—referring to the logical structure, cohesion, and progression of ideas within written texts—was another recurring benefit observed across the reviewed studies. This theme reflects ChatGPT’s affordance as a structural scaffold for learners grappling with how to sequence arguments, introduce transitions, and develop ideas.

Tsai et al. (2024), in their study with Taiwanese EFL majors, found that although vocabulary showed the greatest improvement, organization still ranked among the areas of significant enhancement. While the study did not analyse organizational revisions in depth, the authors suggested that ChatGPT’s feedback often addressed accuracy and vocabulary choices, which extended into improving both essay content and organization.

Similarly, Hou et al. (2024) and Tseng & Lin (2024) observed gains in organizational aspects of university students’ writing. The participants in Hou’s study reported that ChatGPT helped them enhance structural clarity and reorganize content more effectively. In the same vein, Tseng & Lin’s course-integrated intervention further emphasized ChatGPT’s role as a peer reviewer and cognitive partner during idea generation, outlining, and revision. Through iterative prompting and structured revision tasks, learners used ChatGPT to build more logically sequenced and coherently developed texts, particularly benefiting from its support in handling the mechanical aspects of writing, thus giving them more resources to focus on higher-order aspects, such as reasoning or structuring.

Offering a slightly different perspective, Asadi et al. (2025) employed a feedback approach similar to that of Ghafouri et al. (2024), where teachers used ChatGPT to generate initial feedback which they then supplemented, validated, and adapted before delivering it to students. This hybrid model was implemented with intermediate Iranian EFL learners preparing for the IELTS exam. The intervention resulted in significant improvements across multiple dimensions of writing, including coherence, cohesion,

and overall organization, as well as vocabulary, grammar, and task achievement. Organization was a key area of enhancement, but it was part of a broader pattern of comprehensive writing gains observed in the experimental group. The study highlighted the critical role of teacher mediation in interpreting AI feedback to facilitate meaningful improvements in text structure and writing quality.

However, some studies (Hwang et al. 2024; Nguyen, 2024) pointed out that the extent of improvement in organization depended heavily on learner agency and prompt specificity. Hwang et al. (2024) noted that while learners often aimed to improve structure or other higher-order aspects, their prompts to ChatGPT were overly general (e.g., “rewrite this essay”), leading to minimal feedback on organization. The authors stressed the importance of prompt literacy—the ability to articulate precise goals when interacting with AI—to fully leverage the tool’s potential in supporting higher-order writing features such as logical flow and cohesion.

These findings were echoed by Nguyen (2024). Participants in this study, Vietnamese university EFL learners, used ChatGPT during the pre-writing stage, and although they experienced clear improvements in vocabulary and content, there were no significant gains in organization. In examining learners prompting behaviours, the authors noted that they sought primarily feedback in idea generation and lexical issues, therefore, the tool’s responses were largely confined to those dimensions, leaving higher-order structural features like paragraph flow, logical sequencing, and global coherence under-addressed. This pattern underscores the importance of prompt specificity and learner intention in shaping the type of feedback ChatGPT delivers. Nguyen’s study illustrates that ChatGPT’s impact on writing is shaped as much by how learners interact with it as by the tool’s inherent capabilities. When learners do not perceive organization as an area needing feedback—or lack the skills to seek help in that domain—they may receive little to no support in improving structural coherence, despite the tool’s potential to assist with it.

Similar results were reported by Seo (2024), who found that organization was one of the aspects of writing that improved in the ChatGPT-assisted narrative writing intervention with Korean university students, alongside content and language use. However, it was not necessarily the most pronounced area of gain compared to overall fluency and general writing performance. However, in the analysis of the types of prompts students used when interacting with ChatGPT, the authors found that the majority focused on linguistic aspects such as grammar and vocabulary (65.6%), followed by revision requests beyond language use, including style, tone, and formality (11.9%). Requests specifically related to organizational aspects were less frequent but present (8.2%), indicating some student awareness of the need to improve text structure. This pattern suggests that while learners are generally aware of how to use ChatGPT to support surface-level linguistic accuracy, they should be encouraged to make more frequent and targeted prompts related to textual organization, since doing so could further enhance the development of higher-order writing skills.

In conclusion, the findings consistently highlight ChatGPT’s potential to enhance the organizational quality of written texts. However, the degree of improvement in

organizational aspects is closely tied to the specificity of the learner's prompts and their understanding of the structural requirements of a written text. A key issue appears to be that many learners may not fully recognize organization as a central component of writing quality; as a result, they are less likely to request feedback on it. This may partly explain why gains in linguistic accuracy tend to be more prominent, since students are more aware of grammatical correctness as a marker of writing quality and therefore prompt ChatGPT more frequently in that domain. To address this imbalance, it is crucial for educators to foster students' awareness of textual organization as a core aspect of academic writing, alongside developing their prompt literacy, so they can make informed and strategic use of ChatGPT to enhance all dimensions of their written work.

## 5.4 Content

Content development—referring to the depth and relevance of ideas in writing—was positively influenced by ChatGPT in several of the reviewed studies (e.g. Al Fraidan, 2025; Song & Song, 2023; Tsai et al., 2024). In contexts where students used ChatGPT during the pre-writing stage, the tool served as a scaffold for generating ideas, elaborating arguments, and clarifying points, ultimately improving the conceptual richness of their texts (Nguyen et al., 2024).

While Nguyen et al. report immediate gains, Al Fraidan (2025) suggests that content development may require sustained engagement with ChatGPT across multiple writing tasks. The author had a group of Saudi university learners use ChatGPT to improve their argumentative writing following the Toulmin model of argumentation along four consecutive writing tasks. They found that although students' initial use of ChatGPT had little effect on essay depth, significant improvements emerged in later tasks (these were seen in task 3, the third of a set of 4 tasks), where learners demonstrated more nuanced arguments, clearer claims, and stronger integration of supporting evidence. These improvements were attributed to students' growing familiarity with both the model of argumentation applied and the iterative use of ChatGPT as a tool for expanding and refining ideas. Correlational analysis revealed a strong positive relationship between ChatGPT use and content depth in the later stages of the study, suggesting that the tool's effectiveness in promoting deeper writing is cumulative rather than immediate. Taken together, these findings point to the importance of task sequencing and repeated AI interaction, which allow learners to gradually develop the necessary confidence and strategic awareness to use ChatGPT as a content development scaffold.

Another contribution to this theme comes from Polakova and Ivenz (2024) who demonstrate that ChatGPT feedback can enhance several aspects of content quality, including conciseness and inclusion of key information. Their quasi-experimental study with EFL students in a Czech university showed significant improvements following ChatGPT-assisted revisions, with qualitative data indicating students' positive perceptions of the tool's detailed and personalized feedback. However, the authors also note the importance of complementing AI feedback with teacher guidance to ensure sustained and meaningful content development.

To sum up, the evidence suggests that ChatGPT can support content development in EFL writing, but this potential is highly contingent on how learners interact with the tool. To promote authentic content gains, educators should provide scaffolding in prompt design and opportunity to practice and experiment with AI tools, so that learners are better equipped to use ChatGPT for idea generation and conceptual expansion.

## 5.5 Vocabulary

Vocabulary development was also reported as an area of improvement across the reviewed studies, with several sources highlighting gains in lexical variety, precision, and appropriateness. Most notably, Tsai et al. (2024) reported that vocabulary was the most significantly improved dimension in their evaluation of ChatGPT-revised essays, exceeding grammar, content, and organization. The authors attribute this outcome to the fact that EFL learners often exhibit limited lexical range, and ChatGPT's feedback tends to focus on precisely this area. Learners frequently received suggestions that enhanced their word choices, replaced repetitive terms, and improved the overall naturalness and fluency of their writing. As Tsai et al. (2024) note, these improvements are immediate, visible, and easy to integrate, especially within time-constrained revision tasks—factors that likely contributed to the disproportionately high gains in vocabulary scores.

Nguyen et al. (2024) further emphasized that vocabulary was the domain in which students focused the most when using ChatGPT during the pre-writing stage, to explore new ways of expression or find better suited lexical items for their essays. This proactive use of the tool enabled students to embed richer vocabulary more organically into their compositions, which led to increased scores in this area compared to the control group not using ChatGPT.

All in all, the studies reviewed indicate that ChatGPT can be a highly effective tool for improving vocabulary in EFL writing. Its strength lies in immediately offering natural, fluent, and varied language, and its feedback is particularly accessible for learners to adopt.

## 5.6 Limitations of the current review

This study reviewed existing literature regarding the use of ChatGPT for feedback in writing in EFL/ESL contexts published in three databases using the keywords previously outlined. Three limitations were noticed by the author. Firstly, unpublished literature and papers in other databases were not included. Secondly, the search was limited to literature published until March 2025. Thirdly, the study included only articles which examined students' writing performance, excluding papers which dealt exclusively with students' perceptions on the use of ChatGPT, which may have added valuable information to this review. The authors acknowledge these limitations and present this paper as a first attempt at providing a preliminary overview of the aspects of writing which are most clearly impacted by students' use of ChatGPT as a feedback tool.

## **6. Conclusions, pedagogical implications and future research directions**

On the whole, the studies reviewed provide compelling evidence of the positive impact of ChatGPT on students' writing performance in ESL/EFL contexts, especially in areas like linguistic accuracy, organization, content or vocabulary. The overall improvement observed in student writing performance is noteworthy, even though these benefits are not universal, and the success of ChatGPT as a writing tool depends significantly on how students engage with the technology.

Despite the generally positive outcomes reported across studies, it is difficult to draw firm conclusions about which specific factors account for these improvements. The reviewed interventions vary considerably in terms of learner proficiency levels, instructional design, and intervention duration, making it challenging to isolate consistent patterns. Some studies involved brief, tool-focused activities, while others adopted more structured or teacher-mediated approaches, and both types have reported gains in writing performance. Similarly, improvements were documented among both lower-intermediate and advanced learners. This heterogeneity suggests that the experience with ChatGPT is highly user-dependent, so its use may offer benefits across a wide range of contexts, but it also points to a lack of systematic comparison in the literature.

One of the key findings of this review is the enhancement of linguistic accuracy. Several studies highlight the tool's ability to correct surface-level errors, though it may fall short when it comes to teaching complex grammar rules or fostering deeper understanding of linguistic structures. The effectiveness of ChatGPT in grammar correction also underscores the importance of blending AI-driven feedback with traditional teaching methods. While the tool provides immediate and reliable feedback, teacher input remains crucial for addressing more complex language challenges that require explicit instruction or recursive practice. Moreover, the predominance of linguistic accuracy as the most frequently improved aspect may also reflect learners' heightened awareness of grammatical correctness as a key indicator of writing quality. Students are more likely to seek feedback in this area, prompting ChatGPT to focus its suggestions accordingly—an interaction that reinforces and amplifies gains in surface-level accuracy.

In terms of organization, ChatGPT's role as a structural scaffold is apparent. Learners have benefited from its ability to suggest improvements in text cohesion, logical flow, and idea progression. However, this improvement is often contingent on the learner's ability to craft specific prompts and engage with the feedback meaningfully. Therefore, it is important to teach students how to interact with AI in ways that foster higher-order writing skills, since without guidance, students may focus too heavily on surface-level corrections, leaving the more complex aspects of writing underdeveloped.

Content development, while positively impacted by ChatGPT, requires sustained engagement to see significant improvements. As learners repeatedly use the tool for idea generation, refinement, and argumentation, their ability to produce deeper, more complex content increases. A single interaction with ChatGPT may not lead to substantial changes in content quality, but over time, students can develop the necessary prompting

and critical skills for using the tool to develop more sophisticated ideas and stronger arguments. This reinforces the importance of providing guided opportunities for learners to practice with AI tools to refine their content generation skills.

The development of vocabulary has been another area of success, as ChatGPT's capacity to suggest more varied and precise vocabulary has proven particularly useful for learners with limited lexical range. The tool's suggestions help students diversify their word choice, thus improving the fluency and naturalness of their writing, and given that the feedback is immediate, students can readily incorporate ChatGPT's suggestions into their texts.

While the studies affirm ChatGPT's effectiveness in enhancing various writing skills, they also point to critical challenges, particularly regarding learner engagement and long-term sustainability. Overreliance on AI, especially if not coupled with reflective and critical engagement, may undermine students' sense of ownership over their writing. Furthermore, the potential regression in writing improvements after the cessation of AI use highlights the need for the development of critical skills and the provision of scaffolded learning environments to ensure that short-term gains lead to long-term retention.

All in all, ChatGPT holds considerable promise as a tool to enhance students' writing performance in ESL/EFL contexts, particularly in improving linguistic accuracy, vocabulary, and organizational quality. However, its effectiveness is maximized when students are trained to interact with the tool thoughtfully and autonomously. Pedagogically, instructors should encourage metacognitive awareness, critical reflection, and strategic prompting of AI tools.

As regards the development of metacognitive skills, teachers could encourage learners to reflect on how they use ChatGPT as part of their writing process. This could involve asking students to keep a journal or log of their interactions with the tool, where they keep a record of the types of feedback received, and the ways in which they chose to revise (or not) their writing based on that feedback. By doing this, students may become more autonomous learners, and therefore able to utilize AI tools effectively in their language learning journey.

Moreover, instructors should guide students in critically evaluating the feedback provided by ChatGPT. Rather than accepting AI-generated corrections at face value, students should ask themselves whether the suggestions are appropriate for their writing context. For instance, students can be taught in class to question why certain vocabulary choices are suggested or how a particular reorganization of a sentence enhances clarity. Teachers could incorporate reflective exercises, such as asking students to explain why they chose to accept or reject certain feedback, helping them develop a deeper understanding of language use.

Additionally, one of the key factors that influences the quality of feedback from ChatGPT is how students frame their requests. Therefore, teachers can provide instruction on how to craft precise and clear prompts to maximize the effectiveness of AI-generated feedback. For example, instead of simply asking, "Can you check my grammar?", students could be encouraged to ask, "Can you help me with subject-verb

agreement in my second paragraph?” By fostering skills in prompt engineering, educators ensure that students get more targeted and relevant feedback, which can lead to greater improvements in their writing.

Finally, integrating ChatGPT into a balanced instructional framework—where human feedback complements AI suggestions—will help mitigate potential drawbacks, such as over-reliance and disengagement. As AI continues to evolve, educators must remain vigilant, ensuring that ChatGPT serves as a tool for facilitating learning rather than a crutch that replaces students’ cognitive and creative processes.

Nevertheless, in order to build on the promising results found in the studies reviewed, it is crucial to conduct longer-term studies that assess the sustainability of ChatGPT’s impact on writing performance over time. Additionally, it would be beneficial to expand the scope of research to include a more diverse range of populations beyond university students, which could offer valuable insights into the tool’s versatility and potential limitations across varying learner profiles. These two strands of research could help refine the pedagogical approaches for integrating ChatGPT in EFL and ESL writing contexts and provide a more comprehensive understanding of the consequences of its adoption in the long-term.

## Bibliography

- ABDULJAWAD, Samah Abdulhadi (2024)\*. Investigating the Impact of ChatGPT as an AI Tool on ESL Writing: Prospects and Challenges in Saudi Arabian Higher Education. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1-19. <https://doi.org/10.4018/IJCALLT.367276>
- AL FRAIDAN, Abdullah (2025)\*. AI and Uncertain Motivation: Hidden allies that impact EFL argumentative essays using the Toulmin Model. *Acta Psychologica*, 252. Article 104684. <https://doi.org/10.1016/j.actpsy.2024.104684>
- ALSAEDI, Najah (2024). ChatGPT and EFL/ESL Writing: A Systematic Review of Advantages and Challenges. *English Language Teaching*, 17(5), 41-50. <https://doi.org/10.5539/elt.v17n5p41>
- ASADI, Marjan, EBADI, Saman y MOHAMMADI, Laleh (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>
- BACON, Edward Devere y MANEERUTT, Gessanee (2024)\*. Enhancing English as a Foreign Language academic writing through AI and peer-assisted learning. *Journal of Institutional Research South East Asia*, 22(3), 282–314.
- BARROT, Jessie S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- CAMPOS, Miguel (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>

---

\* References marked with an asterisk (\*) are the articles included in the systematic review.

- DIZON, Gilbert y GAYED, John M. (2024). A systematic review of Grammarly in L2 English writing contexts. *Cogent Education*, 11(1), Article 2397882. <https://doi.org/10.1080/2331186X.2024.2397882>
- DÖRNYEI, Zoltán (2009). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 9–42). Multilingual Matters. <https://doi.org/10.21832/9781847691293-003>
- FLOWER, Linda y HAYES, John R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365-387. <https://doi.org/10.2307/356600>
- GHAFOURI, Mohammad, HASSASKHAH, Jaleh y MAHDAVI-ZAFARGHANDI, Amir (2024)\*. From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*, 1-23. <https://doi.org/10.1177/13621688241239764>
- GODWIN-JONES, Robert (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning*, 26(2), 5–24. <http://doi.org/10.125/73474>
- HAN, Jining, Y LI, Mimi (2024)\*. Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, Article 103502. <https://doi.org/10.1016/j.system.2024.103502>
- HOU, Xiaolan, HE, Shuyan, y CUIGONG, Rongxiu (2024)\*. Learner Use of AI-Generated Feedback for Written Corrective Feedback in L2 Writing: Usefulness, User Proficiency and Attitude. *Proceedings of the 2024 8th International Conference on Education and Multimedia Technology*, 70–76. <https://doi.org/10.1145/3678726.3678767>
- HWANG, Myunghwan, JEENS, Robert, Y LEE, Hee-Kyung (2024)\*. Exploring Learner Prompting Behavior and Its Effect on ChatGPT-Assisted English Writing Revision. *Asia-Pacific Education Researcher*, 34, 1157–1167. <https://doi.org/10.1007/s40299-024-00930-6>
- HYLAND, Ken (2022). *Teaching and researching writing* (4th ed.). New York: Routledge. <https://doi.org/10.4324/9781003198451>
- HYLAND, Ken, y HYLAND, Fiona (2006). *Feedback in Second Language Writing: Contexts and Issues* (1st ed.). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524742>
- JAMSHED, Mohammad, AHMED, Abu Saleh Md Manjur, SARFARAJ, Md, Y WARDA, Wahaj Unnisa (2024)\*. The Impact of ChatGPT on English Language Learners' Writing Skills: An Assessment of AI Feedback on Mobile. *International Journal of Interactive Mobile Technologies*, 18(19), 18–36. <https://doi.org/10.3991/ijim.v18i19.50361>
- KOHNKE, Lucas, MOORHOUSE, Benjamin Luke, y ZOU, Di (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- LI, Jinrong, LINK, Stephanie, y HEGELHEIMER, Volker (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18. <https://doi.org/10.1016/j.jslw.2014.10.004>

- LÓPEZ-SÁNCHEZ, Jerri Alejandro, PATIÑO-VANEGAS, Juan Camilo, VALENCIA-ARIAS, Alejandro, y VALENCIA, Jackeline (2023). Use and adoption of ICTs oriented to university student learning: Systematic review using PRISMA methodology. *Cogent Education*, 10(2), Article 2288490. <https://doi.org/10.1080/2331186X.2023.2288490>
- MAHAPATRA, Santosh (2024)\*. Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- MCMARTIN-MILLER, Cristine (2014). How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing*, 19, 24–35. <https://doi.org/10.1016/j.asw.2013.11.003>
- MEYER, Jennifer, JANSEN, Thorben, SCHILLER, Ronja, LIEBENOW, Lucas W., STEINBACH, Marlene, HORBACH, Andrea, y FLECKENSTEIN, Johanna (2024)\*. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, Article 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- NGUYEN, Long Quoc, LE, Ha Van, y NGUYEN, Phuc Thinh (2024)\*. A mixed-methods study on the use of ChatGPT in the pre-writing stage: EFL learners' utilization patterns, affective engagement, and writing performance. *Education and Information Technologies*, 1-24. <https://doi.org/10.1007/s10639-024-13231-8>
- OKTARIN, Irene Brannita, SAPUTRI, Maria Edistianda Eka, MAGDALENA, Betty, HASTOMO, Tommy, y MAXIMILIAN, Aksendro (2024)\*. Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology*, 8(4), 2306–2319. <https://doi.org/10.55214/25768484.v8i4.1600>
- ÖZDERE, Mustafa (2025)\*. AI in Academic Writing: Assessing the Effectiveness, Grading Consistency, and Student Perspectives of ChatGPT and You.com for EFL Students. *International Journal of Technology in Education*, 8(1), 122–154. <https://doi.org/10.46328/ijte.1001>
- PAGE, Matthew J., MCKENZIE, Joanne E., BOSSUYT, Patrick M., BOUTRON, Isabelle, HOFFMANN, Tammy C., MULROW, Cynthia D., SHAMSEER, Larissa Shamseer, TETZLAFF, Jennifer M., AKL, Elie A., BRENNAN, Sue E., CHOU, Roger, GLANVILLE, Julie, GRIMSHAW, Jeremy M., HRÓBJARTSSON, Asbjørn, LALU, Manoj M., LI, Tianjing, LODER, Elizabeth W., MAYO-WILSON, Evan, MCDONALD, Steve, ... MOHER, David (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- POLAKOVA, Petra, y IVENZ, Petra (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>
- SEO, Ji-Young (2024)\*. Exploring the Educational Potential of ChatGPT: AI-Assisted Narrative Writing for EFL College Students. *Language Teaching Research Quarterly*, 43, 1–21. <https://doi.org/10.32038/ltrq.2024.43.01>

- SHI, Huawei, CHAI, Ching Sing, ZHOU, Sihan, y AUBREY, Scott (2025)\*. Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*, 1-28. <https://doi.org/10.1080/09588221.2025.2454541>
- SONG, Cuiping, y SONG, Yanping (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>
- STEVENSON, Marie (2016). A Critical Interpretative Synthesis: The Integration of Automated Writing Evaluation into Classroom Writing Instruction. *Computers and Composition*, 42, 1–16. <https://doi.org/10.1016/j.compcom.2016.05.001>
- SU, Hanyu, TONG, Yixin, ZHANG, Xinyi, y FAN, Yizhou (2024)\*. Uncovering Students' Processing Tactics Towards ChatGPT's Feedback in EFL Education Using Learning Analytics. In: Will W. K. Ma, Chen Li, Chun Wai Fan, Leong Hou U, Angel Lu (eds) *Blended Learning. Intelligent Computing in Education. ICBL 2024* (pp. 238–250). Springer. [https://doi.org/10.1007/978-981-97-4442-8\\_18](https://doi.org/10.1007/978-981-97-4442-8_18)
- TENG, Mark Feng (2024a). A Systematic Review of ChatGPT for English as a Foreign Language Writing: Opportunities, Challenges, and Recommendations *International Journal of TESOL Studies*, 6(3), 36–57. <https://doi.org/10.58304/ijts.20240304>
- TENG, Mark Feng (2024b)\*. "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>
- TSAI, Chung-You, LIN, Yi-Ti, y BROWN, Iain Kelsall (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, Yu-Ching, y LIN, Yi-Hsuan (2024)\*. Enhancing English as a Foreign Language (EFL) Learners' Writing with ChatGPT: A University-Level Course Design. *Electronic Journal of E-Learning*, 22(2), 78–97. <https://doi.org/10.34190/ejel.21.5.3329>
- WANG, Pei-Ling (2015). Effects of an automated writing evaluation program: Student experiences and perceptions. *Electronic Journal of Foreign Language Teaching*, 12, 79–100.
- WANG, Tianyi y MCLAUGHLIN, Colleen (2024). Promoting learner-centred education amid the culture of test-based accountability: insights from a cross-cultural teacher education programme. *Comparative Education*, 1–20. <https://doi.org/10.1080/03050068.2024.2423445>
- WARE, Paige (2011). Computer-Generated Feedback on Student Writing. *TESOL Quarterly*, 45(4), 769–774. <https://doi.org/10.5054/tq.2011.272525>
- WARSCHAUER, Mark, y HEALEY, Deborah (1998). Computers and language learning: An overview. *Language Teaching*, 31, 57–71.
- WARSCHAUER, Mark, y WARE, Paige (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 157–180. <https://doi.org/10.1191/1362168806lr190oa>

- XIAO, Qimin (2024)\*. ChatGPT as an Artificial Intelligence (AI) Writing Assistant for EFL Learners: An Exploratory Study of its Effects on English writing Proficiency. *Proceedings of the 2024 9th International Conference on Information and Education Innovations* (pp. 51–56), Association of Computer Machinery. <https://doi.org/10.1145/3664934.3664946>
- ZHELDIBAYEVA, Raigul (2025)\*. GenAI as a Learning Buddy for Non-English Majors: Effects on Listening and Writing Performance. *Educational Process: International Journal*, 14. Article e2025051. <https://doi.org/10.22521/edupij.2025.14.51>
- ZUPANC, Kaja, y BOSNIĆ, Zoran (2015). Advances in the Field of Automated Essay Evaluation. *Informatica*, 39, 383–395.