

THE ROLE OF ARTIFICIAL INTELLIGENCE IN ENHANCING SECOND LANGUAGE WRITING SKILLS

Urooj Gul^{*1}, Muhammad Salman Ali², Zubair Gul³

^{*1}M.Phil. Scholar, Department of Education, University of Malakand, Khyber Pakhtunkhwa, Pakistan

²M.Phil. Scholar, Department of English, University of Malakand, Khyber Pakhtunkhwa, Pakistan

³M.Phil. Scholar, Department of English, National College of Business Administration and Economics (NCBA&E), Rahim Yar Khan Campus, Pakistan

¹urjgul531@gmail.com, ²msalmanali830@gmail.com, ³zubairgulsamo95@gmail.com

Corresponding Author: *

Urooj Gul

DOI: <https://doi.org/10.5281/zenodo.21470988>

Received	Accepted	Published
01 February 2026	17 March 2026	31 March 2026

ABSTRACT

In the field of second language (L2) writing, the advent of artificial intelligence (AI) has revolutionized the pedagogical landscape, providing unprecedented opportunities for learners to receive instant feedback, real-time error correction, and generative support of the written text. The aim of the paper is to analyze the impact of the use of artificial intelligence (AI) tools such as automated writing evaluation (AWE) systems (e.g., Grammarly) and generative large language models (e.g., ChatGPT) on the writing competence of second language (L2) learners. The study follows a qualitative, interpretive research design that draws on a diverse pool of empirical and conceptual research literature published from 2017 to 2025 to answer two core questions on the pedagogical gains and continued challenges of using AI in L2 writing classrooms. The findings indicate that AI tools are consistently effective in promoting grammatical accuracy, learner motivation, and revision engagement, yet also highlight concerns about dependency, surface-level engagement with feedback, and the lack of development of higher-order composing skills, like argumentation, coherence, and authorial voice. Based on the results of the study, the author suggests that AI is best understood as a pedagogical mediator or partner instead of being a replacement for teacher mediation and that the theory-informed recommendations focus on the issues of equity and AI tools use in L2 writing pedagogy.

Keywords: artificial intelligence, second language writing, automated writing evaluation, ChatGPT, feedback, EFL/ESL learners

1. Introduction

The study of second language (L2) writing has long been acknowledged as a domain of language study, which demands that the learner control the linguistic accuracy, rhetorical organization, and content development in a non-native language all at one time. In the last decade, artificial intelligence (AI) technologies have rapidly

infiltrated the educational context and started to influence how L2 writing is taught, used and evaluated. In addition to automated writing evaluation (AWE) systems, learners can now receive immediate feedback, personalized suggestions, and even available help with drafting from tools that draw from generative large language models (LLM) like ChatGPT, which were

not previously possible in the absence of a teacher (Warschauer & Xu, 2024). The present study is placed in the dynamic technological and pedagogical landscape, a research problem is formulated and the questions, goals and relevance of the study are expressed.

1.1. Background of the Study

Traditionally, writing instruction in L2 has been based on a feedback-revision-drafting cycle involving the teacher and the students, which can be hindered by large classes, short time for teaching and writing, and time lag in receiving human feedback (Hyland & Hyland, 2006). In the limited setting, errors are often corrected after the cognitive work of writing is completed, potentially reducing the connection between error and correction. In response to some of these limitations, computer-assisted language learning (CALL) programs were developed, while automated systems for evaluating writing, like Criterion and, more recently, Grammarly, were created to provide prompt, credible, and continuous feedback on grammar, mechanics, and style (Dodigovic, 2021; Ranalli, 2018). Empirical studies of these have revealed that the tools are effective in detecting and correcting errors at the local sentence level, while questions remain about their potential to facilitate higher order aspects of writing (coherence and argumentation) (Dizon & Gold, 2023; Tambunan et al., 2022).

The introduction of generative AI systems such as ChatGPT has brought a new level of writing assistance, which is qualitatively different from what is currently available. As opposed to traditional AWE tools, which focus on error detection and correction in a learner's own texts, generative models can generate new texts, paraphrase text, generate ideas, and participate in extended, dialogic exchanges with learners concerning their writing (Barrot, 2023). The transformation has sparked a growing enthusiasm among language teachers for the potential of using AI to enhance personalized learning and alleviate writing anxiety (Song & Song, 2023), as well as raising concerns about academic integrity, authorial voice, and the possibility that students may rely on AI to perform higher-order thinking

rather than engaging in it themselves (Zou & Huang, 2024). In recent years, the widespread adoption of AI tools in L2 writing education has led to an increasing number of empirical studies that examine their role and effects (Luo et al., 2025). A recent systematic review of 39 empirical studies on AI in L2 writing education found that AI writing systems, LLM-based generators, and specialized writing assistants are now used across almost all aspects of the writing process, ranging from brainstorming to final editing, and that their pedagogical value relies on the thoughtful integration of these tools into instructional design. This is in the context of a fast-changing technology and ambiguous empirical results.

1.2. Research Problem

Although AI writing tools have emerged in language classrooms, the existing research literature is mixed in regards to what kind of value these tools offer L2 writers as a pedagogical tool. On the positive side, several studies have linked AI-driven feedback to higher levels of grammatical correctness, decreased writing anxiety, and increased motivation (Huang et al., 2024; Wei et al., 2023). On the other hand, there are several studies that have found AI-mediated feedback to be associated with greater grammatical accuracy, less writing anxiety, and more motivation (Huang et al., 2024; Wei et al., 2023). However, some studies have pointed out that automated feedback is often superficial in nature, that learners fail to meaningfully engage with the feedback they receive, and that extensive use of AI-generated feedback can diminish learners' independent composing skills and critical thinking (Fan, 2023; Koltovskaia, 2020). An additional challenge was that most of the existing studies focused on AWE tools and/or generative AI tools separately, with different methodologies and theoretical lenses, making it difficult to arrive at a coherent conclusion on how the two types of AI tools can complement each other and collectively influence L2 writing development. Thus, there is a need for a synthesis that would explore the positive and negative findings from the various fields of current L2 writing research to clarify what is known and

where teaching practice is ahead of the evidence in the field of L2 writing.

1.3. Research Questions

The study is guided by the following research questions:

1. What benefits do AI-powered tools, including automated writing evaluation systems and generative language models, provide to second language learners in the development of their writing skills?
2. What challenges and limitations are associated with the integration of AI-powered tools into second language writing instruction?

1.4. Research Objectives

The study pursues the following two objectives:

1. To examine and synthesize empirical evidence on the benefits of AI-powered writing tools for the development of second language learners' writing accuracy, fluency, motivation, and autonomy.
2. To critically analyze the challenges, limitations, and pedagogical risks associated with the use of AI tools in second language writing classrooms.

1.5. Significance of the Study

The scope of the study is of significance for a number of groups of stakeholders in the field of language education. The synthesis presented here offers an evidence-based summary of which tools in AI could be used to accomplish which writing goals and helps classroom teachers shift from anecdotal enthusiasm/skepticism to informed choices of how to integrate AI into the writing curriculum. The study provides insights for curriculum designers and institutional policy makers on how to harness the opportunities and address the equity issues with the use of AI to maintain academic integrity while providing opportunities for technological innovation. The study brings together the rapidly growing and highly fragmented literature on the subject and provides a unified account of what is supported by empirical evidence, and what is still subject to debate, highlighting promising avenues for future research. Lastly, L2 learners' awareness of the ways in which AI tools can and cannot aid in writing development might lead to more strategic,

reflective, and independent use of such technologies than to the passive and uncritical relying on the generated work.

2. Literature Review

The use of technology in L2 writing instruction is by no means a novel idea, but rather the culmination of research in computer-assisted language learning (CALL) over a long period of many decades. As early as 2006, Ware and Warschauer reported on earlier trials of the use of electronic feedback mechanisms in L2 writing classrooms, where even the simplest automated feedback devices could help to alleviate teachers' workload and provide learners with more opportunities to detect and correct errors. The early work provided the theoretical and empirical basis for the more complex technologies and tools which have since developed, yet the subject of the literature has been that technology needs to support, rather than supplant, human instructional judgment (Hyland & Hyland, 2006). Much of the research that has been conducted on AI and L2 writing so far is related to automated writing evaluation (AWE) systems such as Grammarly. In several studies, Grammarly's precision in finding various types of grammatical mistakes has been assessed. The number of errors that the platform identified was analyzed by Tambunan et al. (2022) to determine the most common errors of the EFL writers in the writing profile, which shows that the most frequent error was grammar-related error, specifically determiner error, proving that AWE tools have the potential to diagnose, and grammar problem is a dominant problem of EFL writers. Likewise, Dodigovic (2021) investigated the correctness of Grammarly's suggestions for form and found that the performance of the tool was satisfactory for many of the surface level features but not for all grammatical forms, raising the question of the reliability of AWE output. In parallel, research on feedback precision has revealed that in some cases (e.g., determiner errors), AWE systems can provide a high level of accuracy whereas in other cases they are not very accurate (e.g., errors that occur in words that do not contain a determiner). The finding supports the idea that automatic

feedback should be complemented by teacher expertise and not taken for granted.

In addition to accuracy, researchers have also examined learners' interaction with AWE feedback and their reactions to it. Koltovskaia (2020) has done a multiple case study on the extent to which students made use of the feedback grammar offers and revealed that the extent of the learner's engagement with the feedback differed significantly as some took the suggestions and used them appropriately in their writing while others only superficially used them or ignored them entirely. One reason for this according to Fan (2023) is that the students in his quasi-experimental mixed methods study of 67 EFL university students were not able to pay attention to and trust in Grammarly feedback as much as the teacher feedback. The students in the treatment group were given Grammarly feedback and teacher feedback on their writing while the students in the comparison group were only given teacher feedback on their writing. The study found that the two groups did not differ significantly in terms of syntactic complexity, lexical richness, accuracy, or fluency in the writing. Similarly, O'Neill and Russell (2019) asked university students to reflect on their use of Grammarly and found that although the students seem to have had a positive experience overall, they may have some reservations about the ability of the tool to provide feedback on organization and higher order concerns. Building on this, Dizon and Gold (2023) explored how Grammarly could affect learners' foreign language writing anxiety and learner autonomy in writing, revealing that after using Grammarly regularly, the anxiety of foreign language writing decreased and the learners' autonomy in writing increased, with modest gains in writing quality.

In addition, the literature on AWE tools also cross-cuts with other theoretical discussions of corrective feedback in SLA. Ellis (2010) has presented a framework for understanding the processing of OCF and WCF by L2 learners, which has been based on the belief that the effectiveness of OCF and WCF is highly dependent on the extent to which the learners notice the difference between what they have

produced and the target. The idea of noticing has also had an impact in AWE research, as automated feedback that is provided immediately and repeatedly is thought to provide more opportunities for noticing than teacher feedback will, as it arrives at a later time (Xu & Zhang, 2022). While scholars from a sociocultural perspective have also warned that it is not sufficient for learners to simply receive corrective feedback to learn, such as Zhang and Hyland (2018) suggest, learners must be actively involved in receiving and internalizing corrective feedback, a process which automated systems may not be able to scaffold effectively without the support of teachers.

The advent of ChatGPT in late 2022 has ushered a new and quickly growing area of study on how generative AI can be applied in L2 writing education. Barrot (2023) provided an early and ground-breaking conceptual discussion of the challenges and opportunities of using ChatGPT for L2 writing, asserting that the tool's ability to produce coherent texts and provide immediate feedback presents great promise; however, its use also brings up significant issues of authorship, overreliance, and reliability of the linguistic information it provides. But the empirical research that has followed the technology generally has been favorably inclined. In a study by Song and Song (2023), the impact of ChatGPT-based learning on the academic writing competence and motivation of EFL learners was explored, revealing that students utilizing ChatGPT to aid their writing reported higher motivation and engagement in the writing process, as well as improvements in their writing performance. In the same line of reasoning, a study conducted by Marzuki et al. (2023) on the implementation of AI writing tools to support students' text organization in the writing process was supplemented by a study conducted by Pakpati and Sri Wahyuni (2023) that showed ChatGPT's feedback positively affected EFL learners' perceived text organization while implementing the writing process with a mixed-methods design. Other research has focused more on how feedback from ChatGPT compares to, or complements, feedback from teachers and peers. Guo and Wang

(2024) examined the possibility of resisting or accepting ChatGPT as a tool for providing feedback in EFL writing, and found that ChatGPT should not replace the teacher's professional judgment, but also be viewed as a supplement to ease the burden of teacher feedback. Zeevy-Solovey (2024) investigated the comparison of peer, ChatGPT-generated and teacher corrective feedback in EFL writing and noted that teacher feedback was valued most for credibility, with ChatGPT feedback being valued for its speed and accessibility. The other strand of research that has investigated how students actually respond to AI feedback has been to compare the respective types of feedback provided by teachers and ChatGPT; in one comparison study of teacher- and ChatGPT-generated feedback, the students were found to be more likely to incorporate suggestions at the surface level rather than suggestions at the content level, indicating that AI feedback like traditional AWE feedback can be more effective in improving accuracy than higher-order composing skills.

There have also been a number of studies on learners' perceptions and attitudes towards AI writing tools in general. Al-Alami (2024) investigated the attitudes of EFL computing students towards ChatGPT as a writing skill acquisition tool and reported a generally positive attitude towards the tool, especially for higher-order thinking skills like generating ideas but with some uncertainty about the proper usage of the tool. In another study, researchers surveyed Vietnamese English Language Teaching students to determine how English learners feel about ChatGPT as a writing-assistant and found that students showed a high level of familiarity with ChatGPT as a writing-assistant tool but were heavily influenced by their prior experience with technology and their ability to critically evaluate the content created by AI. Overall, the results of these perception-based studies indicate that positive attitudes toward AI writing tools do not come naturally to learners; instead, it seems that they depend on how well learners are guided in the use of these tools purposefully and critically.

Despite these largely positive results, there has been a large number of studies that have reported serious issues and risks in the use of AI in L2

writing instruction. Some students have also been concerned with academic integrity and ethical issues regarding the use of ChatGPT, as Yan (2023) noted, when students have not received explicit guidelines from institutions. Pack and Maloney (2024) expressed concerns about the ethical and pedagogical implications of the use of AI in TESOL contexts that go beyond specific examples of misuse, and argued that there is a need for more systematic teacher development and a more precise policy regime to ensure responsible use of AI in the classroom. Similarly, when students participating in the original study by Zou and Huang (2024) about ChatGPT and academic writing felt ambivalent about its use, they appreciated the efficiency it provides for their academic writing but were concerned about its impact on their intellectual growth and personal voice as writers.

The other common issue in the literature is the lack of development of higher-order writing skills. Luo et al. (2025) have undertaken a detailed systematic review of 39 empirical studies that found that underdevelopment of critical thinking, argumentation, and originality was one of the most common issues identified when employing AI tools, with it being reported in almost half of the studies. They also noted that students using AI extensively, especially those with low proficiency, might be tempted to adopt the AI's recommendations without critically analyzing the content, bypassing the reflective writing skills that are essential for effective writing. The results align with the previous study undertaken by Al Mahmud (2023) on the use of Wordtune (AI rewriting tool) among EFL learners, which showed that the use of the tool affected learners' ability to use more vivid vocabulary and more syntactically complex sentences but did not have significant impacts on their global text organization and argumentation.

The literature on AI and L2 writing shows a field that is changing quickly, reflecting the diversity of methods and approaches, and in which there is a clear difference between the benefits that AI tools offer at the surface level (accuracy, motivation, engagement) and those that are more challenging and less agreed upon at the higher order level of

composing skills. To address the tension, some researchers have recommended considering the pedagogical environment in which AI is used. Similarly, Lo et al. (2024) found that the pedagogical effect of ChatGPT in ESL/EFL education also relies on how it is incorporated into the pedagogical design that guides teaching. Luo et al. (2025) also suggested that AI should be perceived as a mediational actor rather than a standalone instructional solution and should be used within a broader pedagogical ecology. This is the subtle and context-specific appreciation for the role of AI that interests the theoretical framework and methodology for the study.

3. Research Methodology

The section details the methodology that was used to answer the research questions used in the study. It tells about the type of study, the research design, how data was gathered and analyzed, and the theory that is applied to clarify results.

3.1. Nature of the Study

The study is an interpretive and qualitative study. Since the study does not employ new primary data collection from learners or teachers, the study is of literature-based and document-driven type, where the primary data source for the study is the published empirical and conceptual studies. This is appropriate for the research questions being addressed which aim to synthesize and critically examine an extant literature of effects rather than to produce new statistical estimates of the magnitude of these effects. A qualitative synthesis better enables a more holistic and context-sensitive interpretation of findings than a purely quantitative meta-analysis would at this relatively early stage of the field's development, particularly considering the quick evolution of the AI technologies and literature, especially since the public release of ChatGPT in late 2022.

3.2. Research Design

The study uses the design of qualitative systematic literature review, which is based on the procedures for conducting systematic and rigorous narrative reviews in the field of applied linguistics and educational technology research. The review was

structured as proposed by Luo et al. (2025), based on well-defined research questions which informed the selection, coding and synthesis of sources. The design is descriptive-analytical in nature, in that the range of AI tools and pedagogical applications described in the literature are first described and then compared and contrasted with the reported benefits and challenges of these applications. The design enables the study to progress beyond merely cataloguing individual studies to offering an integrated account of patterns, contradictions and gaps in the literature as a whole.

3.3. Data Collection Procedure

For the study, relevant peer-reviewed journal articles, systematic reviews, and conceptual papers were considered between 2017 and 2025, a timeframe that covers the previous generation of AWE research as well as the recent explosion in AWE scholarship, in particular, ChatGPT. The search strategy involved a targeted search of academic databases and search engines for suitable key words: combinations of key words such as artificial intelligence and automated writing evaluation, ChatGPT and generative AI, second language writing and EFL writing and feedback. Studies published in well known applied linguistics, educational technology, and language teaching journals were given priority: Computer Assisted Language Learning, Assessing Writing, Journal of Second Language Writing, RELC Journal, Frontiers in Psychology, and other journals. Studies were chosen when they presented empirical data relating to the use of AI/AWE tools in L2/EFL writing contexts, or when they made a significant contribution to the understanding of the role of AI in pedagogical writing contexts, in terms of either conceptually or theoretically. The studies that were limited to the context of first language writing and those that used AI in contexts not related to writing (e.g., speech recognition, reading comprehension) were not included in the analysis.

3.4. Data Analysis Procedure

The collected literature was analysed using a thematic synthesis approach, as described by

Thomas and Harden (2008) in a methodology to synthesize qualitative and mixed-methods research in systematic reviews. Each source was initially reviewed and coded based on the type of AI tool discussed (e.g., AWE systems, GPT models, or specific writing aids), the purpose assigned to the tool in the educational context (e.g., feedback, drafting, or revision), and the reported outcomes, positive, negative, or mixed. These codes were subsequently categorised into higher-order analytical themes such as grammatical accuracy gains, writing anxiety reduction, L1 motivation and autonomy gains, academic integrity concerns and higher order writing skills development. The thematic organization feeds into the organization of the Data Analysis and Discussion section that follows, which follows a gradual progression from the tools-specific findings to general conclusions regarding the pedagogical use of AI in L2 writing development.

3.5. Theoretical Framework

The theoretical framework of the study is based on Vygotsky's (1978) sociocultural theory of learning which posits that cognitive development is primarily mediated by cultural tools, such as language and by extension technology. The concept of the zone of proximal development (ZPD) is a central feature of the framework, which sees the ZPD as the conceptual space between what a learner can do on their own and what they can do with the proper help or mediation (Lantolf & Beckett, 2009). In the context, AI writing tools can be considered as a type of mediational artifact that can help learners maintain their ZPD through scaffolds and feedback that would otherwise be provided by a teacher or more advanced student. Lantolf and Thorne (2006), specifically in the context of second language development, suggested that mediation with cultural tools through the L2 context is a key way in which L2 knowledge is internalized, a perspective that can inform the interpretation of how AWE and generative AI feedback can, under these conditions, assist L2 learners to go beyond their existing capabilities.

However, the sociocultural theory reminds us that the use of a mediational device cannot ensure

developmental growth; its quality and appropriateness play a critical role in the developmental process. In his narrative of semiotic mediation, van Lier (2004) contended that the tools used by learners are not just a means to the end of what they do, but also a means to the way they think; that poorly designed or poorly used semiotic mediation can place as many constraints on the learner as it can provide support. The theoretical concern is directly connected to the literature review's tension between the potential of AI for the accuracy of the surface level of the writing process and the more limited capacity of AI for the higher-order composing skills. The sociocultural perspective is complemented by a selected theoretical framework of Ellis (2010) relating to the processing of corrective feedback which focuses on the importance of the discrepancies learners notice between their current production and the target language norms they observe. The two theoretical strands (sociocultural mediation theory and corrective feedback theory) offer a coherent lens for interpreting the empirical patterns identified in the literature: namely, AI as a genuine tool for supporting genuine writing development is most likely to happen when it is introduced and used as a mediational resource in a teacher-led pedagogical structure, rather than as an isolated substitute for human feedback.

4. Data Analysis and Discussion

This section provides an integrated analysis and discussion of the literature reviewed with reference to the two research questions of the study; that is, the advantages of AI driven writing tools, and the obstacles and constraints that temper these advantages. This is not to treat the benefits and challenges as one against the other, but an attempt to demonstrate that they are two sides of the same phenomenon: the ability of AI tools to give quick, reliable, low-stakes feedback that can be empowering and at the same time limiting, if not supported by sound pedagogy.

4.1 Benefits of AI-Powered Tools for Second Language Writing Development

When it comes to the benefits of AI writing tools (as reported in the literature), the one thing they have in common is their ability to give personalized and immediate feedback at a level that would be unmanageable for any human teacher. This is a common theme in both AWE studies and the study of generative AI tools, indicating that a key affordance of AI writing instruction is the instant feedback, irrespective of the specific AI system. Tambunan et al. (2022) evidenced that Grammarly could quickly categorize and count the linguistic errors contained in dozens of student texts, giving students and teachers a diagnostic depth of detail on spelling, grammar, punctuation, and sentence structure that would be extremely time-consuming to manually provide. This diagnostic function may be especially helpful in the large, under-resourced EFL classrooms found in many parts of the world, where teachers may not be able to give in-depth individual feedback on each assignment to all students. In these settings, AWE tools become a multiplier in providing feedback so that teachers can focus on higher order issues like content and argumentation, while automated tools can take care of more routine, rule-based error correction (Han & Sari, 2024).

In addition to the efficiency of diagnosis, AWE feedback has proven to be effective at producing measurable gains in grammatical accuracy. The randomized controlled trial (Wei et al., 2023) gives some of the strongest experimental evidence in this respect, showing that Chinese EFL learners who were given feedback on their writing based on the automated writing evaluation ones had significant gains in writing accuracy compared with a control group. Huang et al. (2024) conducted a meta-analysis of the study results and concluded that L2 writing-related anxieties and computer anxieties are significantly decreased, and that there are measurable improvements in L2 writing-related aspects with the use of automated writing evaluation, in accordance with the experimental evidence. These accuracy increases can also be demonstrated on the level of the individual error type in the feedback precision

studies, which indicate that many of the grammatical accuracy gains are achieved in specific error categories (e.g., in determining whether a determiner is included or omitted). This explains why grammatical accuracy is one of the most consistent categories of improvement in the literature on AI-mediated writing.

Another significant advantage reported in the literature is that students' motivation and involvement in the process of writing. With AI feedback on demand, and no social risk associated with submitting suboptimal content to a teacher or peer, many students say they feel more at ease trying out new expressions and revising their writing more often. However, Song and Song (2023) showed that EFL learners using ChatGPT as a writing assistant felt that it increased their motivation to write, partly because of the way the tool provided feedback, which was casual, non-critical, and enabled the learners to ask further questions to revise and deepen their understanding without embarrassment. This result aligns with Dizon and Gold's (2023) study on the effect of Grammarly on FLA, which found that students' foreign language writing anxiety levels decreased when they used Grammarly frequently. Al-Alami (2024) also found that the overall attitude of EFL computing students in using ChatGPT was mostly positive, and students especially appreciated its usefulness in dealing with the daunting blank page in the beginning of a writing task.

Learner autonomy is a theme that is frequently discussed in the research on both AWE and generative AI tools, and is closely related to motivation. The feedback and revision capabilities offered by these tools enable students to take ownership of the writing process without requiring the teacher to schedule a time to review and mark up their drafts or wait for a marked-up copy to be returned. Several studies have shown that using AI for writing practice can create a greater sense of ownership for the process of writing. Dizon and Gold (2023) reported that using Grammarly would lead to improved student autonomy, since students would start to recognize and correct their own mistakes before submission for their teacher to evaluate. Likewise, the

systematic review conducted by Luo et al. (2025) revealed that structured and reflective application of AI feedback, specifically with the aid of tools like writing logs or reflection journals, was found to motivate learners to shift from a passive reception of the AI's recommendations to an active engagement with and responsibility for their revision choices. This indicates that increasing the autonomy of AI tools is not a natural process but rather requires a conscious pedagogical approach that affords students the opportunity to carefully consider what they are being told.

A fourth, if less well-known, aspect of the literature regards the potential for generative AI tools to assist in the early stage of the writing process, namely concept generation, which is something that the traditional AWE systems, which mainly assess the existing text, cannot do. Learners, students, or researchers can use tools like ChatGPT to generate their own content in response to a prompt, which can help them with brainstorming, outlining, and finding alternative phrases or arguments prior to drafting a first copy. In a study of the effects of the AI rewriting tool Wordtune on the revision process of EFL students, Al Mahmud (2023) concluded that students' use of the tool to produce more vivid and syntactically varied language was an effective demonstration of more advanced linguistic choices that students could learn through and internalise. This modeling function is also a unique feature of generative AI tools, in a way that they can be seen as more of a linguistic resource or a writing partner, rather than a proofreading tool. Last, there are a number of studies that indicate organizational and content-related benefits, an area which was regarded as a possible weakness of the first generation of AWE tools. Furthermore, Marzuki et al. (2023) noted that AI writing tools contributed to the enhancement of the content and structure of EFL students' writing, findings that are consistent with other studies on how students use ChatGPT to assist with their writing that concluded that students perceived that their ideas were better structured and sequenced in the writing process after applying ChatGPT in their writing revision process. The finding is important in that the higher-order composing support has

historically been one of the more difficult-to-meet aspects of writing with an automated composing system, as the tools are starting to narrow the gap between surface-level correction and higher-order composing support, although it has by no means been closed, as the next section shows.

4.2 Challenges and Limitations of AI Integration in Second Language Writing Instruction

Although the evidence reviewed above is quite promising, there is also a consistent body of problems found in the literature that make a simple rosy picture of the contribution of AI to L2 writing development difficult. In spite of the positive evidence found above, there are also a set of problems that can be found in the literature that make a simple rosy picture of the contribution of AI to L2 writing development difficult. One of the most quoted issues involves issues of academic integrity and the potential for AI tools to cause confusion about authorship, particularly in the case of generative AI. Students involved in a writing practicum focused on using AI reported a significant level of moral concern with using ChatGPT, especially when trying to determine how much of the AI-generated material they could ethically include without plagiarizing. This is not just about student perception; it's about a legitimate and unresolved policy tension in language education that has been faced by language institutions in creating clear and consistent rules for acceptable use of AI in writing assignments. In essence, this vagueness puts an unfair burden on students and teachers, as they are expected to deal with vague ethical issues without accompanying institutional support, and Pack and Maloney (2024) called for systematic professional development to support language teachers in developing their AI literacy and setting up clear, pedagogically grounded usage policies.

Another significant problem is the potential for over-dependence on AI tools, especially for students who don't write confidently or with proficiency. Luo et al. (2025) synthesized the evidence for the systematic review and identified that, in many of the studies included, students were overly dependent on the AI model's recommendations and did not critically assess the

output, which is a typical aspect of skill development. The same appeal and convenience that make AI tools attractive to learners can, at least under some circumstances, hinder deeper engagement with the learning, which is a major factor in their development of writing. Koltovskaia (2020) describes this phenomenon in her case study of engagement with feedback from Grammarly: students either reflect on and selectively use automated suggestions or take many of them without thinking, assuming that the automated suggestions are the only and final source of feedback and therefore should be treated as a source of authority rather than just one source among many. The uncritical use of such writing can lead to a writing product which is technically cleaner but which represents the language that the AI system produces, not the learner's own developing voice and understanding.

The third challenge described in the literature is the constant problem AI tools have in assisting with the higher-order composing processes of argumentation, critical thinking, and rhetorical sophistication. This is not just an issue with a specific tool; it is also a problem with AWE research and generative AI, implying that it is something more fundamental about the current technology of AI and not simply a limitation of a particular tool. In a quasi-experimental study, Fan (2023) did not find that students who got feedback from Grammarly were any better off than students who got feedback from their teacher alone, which he said was partly because students weren't really "committed to the feedback or to the tool and focused mainly on the surface-level stuff." In another study, Al Mahmud (2023) discovered that the Wordtune benefits were limited to vocabulary and sentence-level syntax, but did not extend to global issues like thesis development or argument structure. Taken together, these results indicate that AI has really come a long way in its ability to assist with the organizational and content dimensions of writing (as illustrated in the previous section) but the gains have been uneven and should not be exaggerated; the higher-order rhetorical dimensions of writing still seem to rely heavily on human instruction.

A fourth challenge relates to accessibility and equity in AI tool adoption. There is a range of studies that mention that, to beneficially interact with the AI writing tools, a foundational digital literacy and knowledge of the functions and interface of the AI writing tools is required, which is not equally distributed among all learners. Marzuki et al. (2023) noted that the students who were not trained with AI feedback tended to get it wrong, and they resorted to making lexical changes only superficially instead of making substantive changes. The discovery has far-reaching impacts on educational equity, as students from less-resourced areas, who already have access to less reliable Internet service or the paid versions of some AI tools, could be further marginalized if AI-powered writing instruction is mainstreamed without support systems in place. The concern resonates with other discussions in the field of educational technology research regarding the digital divide, arguing that the positive findings of AI-supported writing instruction from this review do not necessarily apply to the entire population of learners.

The fifth and less obvious challenge is the precision and dependability of the language information that AI tools generate. AWE systems generally proved to be useful and accurate in reporting frequent errors, but Dodigovic's (2021) analysis revealed significant differences in the effectiveness of the tool in reporting errors in various grammatical structures, which implies that the learner and teacher should not take the output of the AWE systems for granted as correct answers, but rather as a point of reflection. A related, but different, concern is the novelty of generative AI tools, which can sometimes generate grammatically correct text that is factually and/or contextually inaccurate; learners must learn to critically assess the text they create in this way, which many less proficient writers may not yet have the ability to do. This is a very clear threat that Barrot (2023) highlighted in his preliminary conceptualisation of how generative AI may impact L2 writing, stating that the ease and assurance with which generative AI is able to generate text can lead to a sense of authority that could mislead young learners who are not yet able

to critically examine the text that is presented to them.

In sum, these five challenges and concerns about academic integrity, over-reliance, lack of support for higher-order skills, accessibility and equity issues, and questions of AI reliability are not the opposite of the benefits noted above, but rather they qualify and frame those benefits. The findings presented here indicate that AI resources do not necessarily signal a total replacement of human effort and are not necessarily harmful, but rather a type of instructional resources with a highly situational pedagogical potential. This finding is in line with the theoretical framework presented in the previous part of this study, which suggests that AI tools, from a sociocultural point of view, serve as mediational artifacts that help learners expand their ZPD, but can only be so if pedagogical design is carefully thought out and takes into account reflection rather than passive accepting (Lantolf & Thorne, 2006). In cases where this scaffolding is available, such as in the structured reflection routines described by Luo et al. (2025), AI tools seem to serve as a valuable facilitator of meaningful engagement with feedback and improvements in writing. In its absence, these tools can encourage shallow memorization, reliance and lack of critical thinking around writing.

The analysis also indicates a fruitful approach to bringing together the conflicting results from the literature surveyed. Research suggesting beneficial outcomes with the use of AI tools, such as Wei et al. (2023) and Song and Song (2023), tend to include teacher-led, scripted research in which AI feedback is just one part of a larger instructional design, rather than it being the intervention. In contrast, some studies with more modest or no impact, like Fan (2023), account for scenarios in which the AI feedback was given with less explicit rules or directions regarding the steps students are to take based on the feedback. This pattern helps to support the argument put forward by Guo and Wang (2024) that the question in L2 writing instruction is not whether AI tools should be used in L2 writing instruction but how and how much they can be used and for what purpose. It also recommends that future studies would be useful

to go beyond comparisons of AI-supported versus non-AI-supported conditions, to more detailed research on the impact of different AI implementation models, which differ in the amount of scaffolding from the teacher, training from the learner, and reflection from the learner, on the outcomes that AI tools can support and produce.

Last but not least, is the crucial difference between the two major categories of AI tools explored in this research. The most reliable and prominent advantages of AWE systems seem to be in the areas of grammatical correctness and mechanical precision, which are also the areas where they are most likely to help with. This is not surprising, as they are supposed to detect errors, which have to do with grammar and mechanical precision. The most consistent and noticeable advantages of AWE systems seem to be in the area of grammatical correctness and mechanical precision, which is not surprising because that is the area where they are most likely to help in terms of error detection. In contrast, generative AI tools like ChatGPT seem to have a wider, though not always sufficiently documented, set of benefits, including in content creation and idea generation, as well as in the exchange of dialogic feedback, due to their much more text-generating than text-evaluating architecture. It is important for the practical use of the tools in the classroom: Teachers who are interested in helping students correct specific grammatical errors in their writing will likely benefit more from AWE tools, whereas those who are interested in brainstorming and developing ideas or engaging in iterative dialogue with students about their writing choices will benefit more from the generative AI tools, if they are able to implement appropriate safeguards against over-reliance or academic integrity violations.

5. Conclusion and Suggestions

5.1. Conclusion

The aim of this study was to analyze the advantages and difficulties of applying AI related technologies in L2 writing classrooms and review a wide range of empirical and conceptual research on the subject from the years 2017 to 2025. It shows that

L2 writers benefit meaningfully and empirically from the use of AI tools, which range from automated systems for the evaluation of writing (e.g. Grammarly) to generative LMs (e.g. ChatGPT), in terms of their grammatical accuracy, writing anxiety, motivation, learner autonomy and, more recently, content production and organization. Concurrently, there are enduring and significant problems identified in the literature, such as threats to academic honesty, overreliance on the tools, a lack of support for higher-order language skills (e.g., argumentation, critical thinking), inequitable access to digital literacy and resources to support the effective use of these tools, and uncertainty regarding the reliability of AI-provided linguistic feedback. The study has proposed a different perspective that sees these benefits and challenges as two sides of the same coin: the pedagogical benefits of AI tools rely on the quality of the instruction in which they are used as a powerful mediational resource. From a Vygotskian perspective on learning, using AI tools to extend learner's zones of proximal development may be a genuine possibility, provided that the pedagogical settings that support their use are accompanied by reflective pedagogical support that emphasizes a critical engagement with the output of the AI tools, rather than a passive consumption of the generated text. The evidence reviewed here can therefore be neither uncritical enthusiasm nor outright rejection of the use of AI in L2 writing classrooms, but rather a theory-informed approach to integration that AI is a pedagogical partner to human teaching—not a substitute.

5.2. Suggestions

The results of the study can give some suggestions to language teachers, curriculum designers and researchers. To use AI writing tools in their classrooms, it is important for teachers to focus on the models of structured, reflective implementation instead of incidental or unstructured use of the tools, while making sure that students are explicitly instructed on how to interpret, evaluate, and selectively act on the feedback generated by AI. This could involve using reflection journals or structured revision journals

as suggested in the literature as a way to positively enhance their engagement with automated feedback. To prevent plagiarism and ensure academic honesty in writing, curriculum designers and institutional policies must establish clear, transparent guidelines on how to use AI in writing assignments beforehand, rather than reacting to the issue, and invest in teachers' professional development so that they can become AI-literate for effective implementation across a variety of classroom settings. Finally, institutions need to also consider how they can ensure that every learner is well supported and has access to training and assistance to get the most out of AI supported writing instruction without being further disadvantaged, as the equity issues surrounding AI use have been well documented.

The present study has opened up a number of fruitful avenues for future researchers. More longitudinal studies are needed which investigate the long-term impact of regular use of AI tools on L2 writers' argumentation, coherence, and autonomous writing skills, as the majority of current research focuses on short-term or immediate results. Third, future studies should go beyond comparing writing conditions with AI versus without AI to different levels of AI implementation models (with different levels of teacher scaffolding and reflective structure) and investigate the differences in the outcomes they can produce. Third, as technology changes at such a fast pace in the field, researchers should periodically update these kinds of systematic reviews so that pedagogical practice is based on the most up-to-date and reliable evidence and not assumptions carried over from previous, less powerful technologies. Last, future research could focus more on equity-oriented research designs that investigate how AI-based writing instruction can be modified to support learners in under-resourced and linguistically diverse settings to ensure that the benefits are widely and equitably shared in the global community of second language writers.

REFERENCES

- Al Mahmud, F. (2023). Investigating EFL students' writing skills through artificial intelligence: Wordtune application as a tool. *Journal of Language Teaching and Research*, 14(5), 1395–1404. <https://doi.org/10.17507/jltr.1405.2>
- Al-Alami, S. E. (2024). EFL learners' attitudes towards utilizing ChatGPT for acquiring writing skills in higher education: A case study of computing students. *Journal of Language Teaching and Research*, 15(4), 1029–1038. <https://doi.org/10.17507/jltr.1504.01>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Dizon, G., & Gold, J. (2023). Exploring the effects of Grammarly on EFL students' foreign language anxiety and learner autonomy. *JALT CALL Journal*, 19(3), 299–316. <https://doi.org/10.29140/jaltcall.v19n3.1049>
- Dodigovic, M. (2021). Automated writing evaluation: The accuracy of Grammarly's feedback on form. *International Journal of TESOL Studies*, 3(2), 71–87. <https://doi.org/10.46451/ijts.2021.06.06>
- Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32(2), 335–349. <https://doi.org/10.1017/S0272263109990544>
- Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open*, 13(3). <https://doi.org/10.1177/21582440231181296>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Han, T., & Sari, E. (2024). An investigation on the use of automated feedback in Turkish EFL students' writing classes. *Computer Assisted Language Learning*, 37(4), 961–985. <https://doi.org/10.1080/09588221.2022.2067179>
- Huang, X., Xu, W., Fan, L., & Yu, Z. (2024). A meta-analysis of effects of automated writing evaluation on anxiety, motivation, and second language writing skills. *The Asia-Pacific Education Researcher*, 33(4), 957–976. <https://doi.org/10.1007/s40299-024-00865-y>
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Lantolf, J. P., & Beckett, T. G. (2009). Sociocultural theory and second language acquisition. *Language Teaching*, 42(4), 459–475.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, 11(1), 50. <https://doi.org/10.1186/s40561-024-00342-5>
- Luo, Y., Xia, Y., & Lu, X. (2025). A systematic review of the role of artificial intelligence in second language writing education. *Digital Studies in Language and Literature*. Advance online publication. <https://doi.org/10.1515/dsll-2025-0001>
- Marzuki, Widiati, U., Rusdin, D., Darwin, I. I., Govender, N., & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1). <https://doi.org/10.14742/ajet.3795>
- Pack, A., & Maloney, J. (2024). Using artificial intelligence in TESOL: Some ethical and pedagogical considerations. *TESOL Quarterly*, 58(2), 1007–1018. <https://doi.org/10.1002/tesq.3320>
- Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 31(7), 653–674. <https://doi.org/10.1080/09588221.2018.1428994>
- 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, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Tambunan, A. R. S., Andayani, W., Sari, W. S., & Lubis, F. K. (2022). Investigating EFL students' linguistic problems using Grammarly as automated writing evaluation feedback. *Indonesian Journal of Applied Linguistics*, 12(1), 16–27. <https://doi.org/10.17509/ijal.v12i1.46428>
- van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Springer.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Ware, P., & Warschauer, M. (2006). *Electronic feedback and second language writing*. Cambridge University Press.
- Warschauer, M., & Xu, Y. (2024). Generative AI for language learning: Entering a new era. *Language Learning & Technology*, 28(2), 1–4.
- Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, 1249991. <https://doi.org/10.3389/fpsyg.2023.1249991>
- Xu, J., & Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *The Asia-Pacific Education Researcher*, 31(4), 357–367. <https://doi.org/10.1007/s40299-021-00577-7>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28(11), 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>

- Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, 6(3), 1482.
<https://doi.org/10.29140/tltl.v6n3.1482>
- Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102.
<https://doi.org/10.1016/j.asw.2018.02.004>
- Zou, M., & Huang, L. (2024). The impact of ChatGPT on L2 writing and expected responses: Voice from doctoral students. *Education and Information Technologies*, 29(11), 13201–13219.
<https://doi.org/10.1007/s10639-023-12397-x>

