Journal of English Language Teaching and Applied Linguistics
ISSN: 2707-756X
DOI: 10.32996/jeltal
Journal Homepage: www.al-kindipublisher.com/index.php/jeltal

JELTAL
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

The Impact of ChatGPT as a Formative Feedback Tool on Moroccan Secondary School Students' Paragraph Writing Skills: A Quasi-Experimental Design

Hamza Farhane
Department of English, Faculty of Letters and Human Sciences, Dhar Elmahraz, Sidi Mohammed Ben Abdellah University, Fez, Morocco
Corresponding Author: Hamza Farhane, **E-mail**: hamzafarhane96@gmail.com

| ABSTRACT
This study investigates the impact of ChatGPT-generated formative feedback on Moroccan secondary school students' paragraph writing performance along with their perceptions of the tool's effectiveness. Adopting a quasi-experimental mixed-methods design, the study involved 87 students divided into an experimental group (n = 45) and a control group (n = 42). Over an eight-week intervention, the experimental group received AI-generated feedback on their writing tasks while the control group was provided with traditional teacher feedback. Quantitative data from pre-test, post-test, and delayed post-test assessments revealed statistically significant improvements in the experimental group's writing scores (pre-test M = 6.69, post-test M = 10.01, delayed post-test M = 10.27; $p < .001$). Paired-samples t-tests confirmed substantial within-group gains ($t = -34.38$ and -35.69) while independent-samples t-tests revealed significant between-group differences in post-test ($t = 22.71$) and delayed post-test scores ($t = 16.69$) favoring the experimental group. Effect size calculations showed a large impact (Cohen's $d = 5.13$ for pre- vs. post-test; $d = 5.40$ for pre- vs. delayed post-test). Qualitative data from semi-structured interviews with nine students in the experimental group indicated that ChatGPT feedback was perceived as clear, specific, and motivational enhancing student engagement, autonomy, and metacognitive awareness. These findings underscore the pedagogical value of integrating AI feedback into EFL writing instruction particularly in resource-constrained educational settings.

| KEYWORDS
ChatGPT, formative feedback, AI in education, writing performance, EFL instruction, student perception, mixed-methods, Morocco.

| ARTICLE INFORMATION
ACCEPTED: 15 July 2025 **PUBLISHED:** 12 August 2025 **DOI:** 10.32996/jeltal.2025.7.3.26

1. Introduction

Writing proficiency is fundamental in an EFL context, and developing strong writing skills depends heavily on formative feedback. Studies have shown that students value feedback; they report that comments on their drafts help them achieve learning goals and stay motivated (Hunnes & Olsen, 2024). However, delivering such personalized feedback is very time-consuming when class sizes are large. In fact, Mahapatra (2024) notes that in many developing-country universities, crowded classrooms are common and make it challenging to implement individualized writing feedback. Instructors often question whether the traditional feedback process is sustainable at scale even as they recognize its importance to students' learning.

Because of these constraints, educators are turning to technology-based solutions. Automated Writing Evaluation (AWE) tools and AI tutors have shown promise as supplements to teacher feedback. Labadze et al. (2023), for example, found that AI chatbots can provide immediate and personalized learning support in helping students with homework and study questions while saving teachers time and effort. This general trend applies to writing instruction as well. In particular, ChatGPT, a conversational AI powered by a large language model, can act like a virtual tutor that students can query at any time. According to Xiao et al. (2025), ChatGPT is widely acknowledged for its ability to provide real-time feedback that enhances writing quality

Copyright: © 2025 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

and efficiency for EFL learners. Likewise, Lo (2023) observes that ChatGPT can support teachers by suggesting assessment tasks and instructional materials and can serve as a virtual tutor collaborating with students on writing tasks. It has been shown to generate substantial writing exemplars (Barrot, 2023), adapt text complexity to align with learners' proficiency levels (Bonner et al., 2023), offer guidance on structural features of written content, and perform translations (Imran & Almusharraf, 2023). Moreover, it supports scaffolded writing activities (Kohnke et al., 2023). These functionalities promote learner autonomy and address students' demand for immediate formative feedback. ChatGPT is capable of delivering expert-like support in brainstorming, text organization, linguistic precision, and vocabulary selection (Tai et al., 2023). These AI-driven supports can help students brainstorm ideas, refine grammar and organization, and expand vocabulary without waiting for a teacher's input.

Despite growing interest, there remains a scarcity of empirical research specifically examining the impact of ChatGPT on students' writing abilities (Su et al., 2023). While its potential as a formative feedback tool has been widely discussed (Bonner et al., 2023), few studies have provided data-driven evidence to substantiate these claims. Moreover, within the Moroccan educational context, Ouahani and Mahraj (2025) claim, only four scholarly articles (Al-Zubaidi et al., 2024; Benfatah et al., 2024; Boubker, 2024; Bouzar et al., 2024) and one book chapter (Benali & Mak, 2024) explored the integration of ChatGPT in educational settings across Morocco. Nonetheless, none of these contributions explicitly focused on its implementation within English Language Teaching (ELT) contexts. Therefore, the present study addresses this gap by investigating the effectiveness of ChatGPT as a feedback mechanism in enhancing the paragraph writing skills of Moroccan secondary school learners within a relatively large classroom setting using an intervention. To address this objective, a mixed-methods intervention design was adopted treating ChatGPT-based feedback as the independent variable and students' writing proficiency as the dependent variable. It was hypothesized that integrating ChatGPT into the feedback process would produce a significant improvement in learners' writing performance. Given the study's implementation in crowded classrooms, the results hold potential for broader applicability in comparable ESL and English as a foreign language (EFL) contexts. Moreover, the initiative to employ ChatGPT as a feedback tool contributes valuable pedagogical insights and may inspire further global research in technology-assisted writing instruction.

2. Literature Review

2.1 The Role of Formative Feedback in EFL Writing Development

Formative feedback refers to information provided to learners about their performance relative to learning goals with the intent to close the gap between current and desired performance (Sadler, 1989; Hattie & Timperley, 2007). Unlike summative evaluation which judges final outcomes, formative feedback is an ongoing, dialogic process that guides improvements during learning (Black & Wiliam, 1998; Shute, 2008). In essence, effective feedback highlights discrepancies between a student's present writing and expected standards and offers directions for revision, thereby functioning as a bridge to higher competence (Sadler, 1989; Kluger & DeNisi, 1996). This process is widely regarded as indispensable in education as decades of research show that timely and criterion-referenced feedback can significantly enhance learning across domains (Hattie & Timperley, 2007; Wisniewski et al., 2020). In the context of language learning and English as a Foreign Language (EFL) writing in particular, formative feedback has become a cornerstone of effective pedagogy providing the individualized guidance needed for developing complex writing skills (Tran, 2025). Indeed, writing scholars emphasize that continuous feedback is often seen as one of the teacher's most important tasks in offering the kind of individual attention that is otherwise rarely possible under normal classroom conditions. (Hyland, 2006). By supplying learners with insight into their errors and weaknesses as they write, formative feedback enables them to refine ideas, reorganize content, and polish language use in subsequent drafts, and thus steadily improving the quality of their compositions (Biber et al., 2011; Scherer et al., 2024). Multiple meta-analyses and empirical studies have documented the positive impact of such feedback on EFL learners' writing development linking it to gains in writing accuracy, fluency, and overall text coherence (Biber et al., 2011; Kang & Han, 2015; Peltzer et al., 2025). Formative feedback not only scaffolds better writing performance but also contributes to second language development by prompting learners to notice linguistic problems and engage in productive self-correction (Li & Vuono, 2019; Swain, 1995). As a result, it is broadly acknowledged that high-quality feedback plays a pivotal role in fostering EFL writing skills through accelerating learners' progress by turning each writing task into a learning opportunity (Crosthwaite et al., 2022; Hyland & Hyland, 2006; Zhang, 2020).

To maximize its benefits, formative feedback must possess certain key characteristics as evidenced by extensive literature on effective feedback practices. Timeliness is frequently highlighted; feedback given promptly, while the written task is still fresh in students' minds, has a substantially greater impact on learning than delayed commentary (Shute, 2008; Daneshvar & Rahimi, 2014; Hattie & Timperley, 2007). Rapid feedback allows learners to immediately assimilate suggestions and apply them in revisions, and therefore reinforcing learning when it is most pertinent (Voerman et al., 2012). Equally important is specificity; feedback should pinpoint particular aspects of the writing that need improvement rather than offering vague praise or criticism (Hattie & Timperley, 2007; Ferris, 2010). Studies indicate that detailed and focused comments lead to more effective revisions and skill gains than general remarks like "improve your ideas" (Nicol, 2010; Saeli & Cheng, 2021). Effective formative feedback also provides actionable guidance which means it not only diagnoses problems but also suggests concrete strategies for

improvement (Shute, 2008; Patchan et al., 2016). For instance, instead of merely noting that a paragraph lacks clarity, a teacher might advise the student to revise the topic sentence or provide a specific example giving a clear direction for the next draft (Brookhart, 2017). Research has consistently found that such explanatory feedback, which clarifies the nature of errors and offers solutions, is associated with greater student uptake and learning gains in EFL writing (Zheng & Yu, 2018; Winstone et al., 2017). Moreover, supportive tone and clarity are vital features: formative feedback should be delivered in a constructive manner that encourages and motivates the learner rather than simply listing faults (Shute, 2008; Mahboob, 2015). When feedback is expressed clearly and empathetically highlighting strengths, using understandable language, and focusing on the work, students are more likely to perceive it as fair and useful, which in turn promotes engagement with the feedback (Carless & Boud, 2018; Wisniewski et al., 2020). In summary, the literature converges on the view that formative feedback is most efficacious when it is timely, specific, clear, and constructive giving EFL writers precise insights into how to improve their texts and the encouragement to do so (Hattie & Timperley, 2007; Shute, 2008; Evans, 2013). Under these conditions, feedback functions as an interactive formative assessment which guides learners through cycles of revision and progressively enhancing their writing competence (Sadler, 1989; Kulhavy & Stock, 1989).

2.2 Challenges in Providing Feedback in EFL Classrooms

EFL writing teachers face numerous logistical and pedagogical constraints that hinder effective feedback delivery. Large class sizes, limited time, and heavy workloads commonly prevent teachers from providing timely and individualized feedback to all students (Teng & Ma, 2024; Yu et al., 2021). Consequently, the scope of feedback often becomes narrowly focused on surface-level linguistic corrections (e.g. grammar and vocabulary) rather than global content and organization issues (Bitchener & Ferris, 2012; Kurzer, 2018; Lee, 2016; Truscott, 2010). Traditional teacher comments tend to be delivered in a one-way, written form with little follow-up interaction reflecting a lack of dialogic feedback; students typically have few opportunities to seek clarification or engage in feedback discussions (Carless & Boud, 2018; Lee, 2014; Nicol, 2010; Winstone et al., 2017). This one-directional feedback process contributes to low uptake of teacher comments as many learners do not meaningfully incorporate the provided feedback into improved revisions or future writing tasks (Han & Hyland, 2019; Lee, 2016; Zhang & Hyland, 2018; Zheng & Yu, 2018). Moreover, affective and motivational factors can undermine feedback effectiveness: overly critical or voluminous feedback often elicits negative emotional responses (e.g. anxiety, demotivation, or lowered self-efficacy), which can discourage students and diminish their confidence in writing (Lipnevich et al., 2021; Mahfoodh, 2017; Shrivastava & Shrivastava, 2020). In light of these challenges, researchers have advocated alternative or supplementary feedback approaches to enhance the feedback process. Approaches such as structured peer feedback and technology-assisted feedback have been proposed to alleviate teacher workload, broaden the focus beyond surface errors, and foster more interactive feedback engagement (Guo & Wang, 2024; Link et al., 2022; Yu & Hu, 2017).

2.3 Technology-Enhanced Feedback and AI Tools

The use of Automated Writing Evaluation (AWE) tools has grown significantly in recent years as a means of providing technology-enhanced feedback in writing instruction. AWE systems are software platforms designed to assess written texts and deliver instant feedback on various writing features. Originating from earlier automated essay scoring technologies, modern AWE tools like Grammarly, Criterion (by Educational Testing Service), Pigai, and similar programs employ natural language processing and machine learning to analyze student writing and offer corrective feedback (Zhai & Ma, 2021). These tools typically evaluate grammar, spelling, punctuation, vocabulary use, and even elements of style or organization generating both holistic scores and specific annotations on errors or areas for improvement (Zhai & Ma, 2022). Because AWE platforms provide immediate and individualized feedback, they allow students to identify and correct language errors in real-time and to revise their texts through multiple drafts without always waiting for teacher input (Stevenson, 2016). In this way, AWE has become a popular supplement to classroom instruction especially in second language (L2) writing contexts, by serving as an always-available digital writing tutor that reinforces learning outside of regular teacher-led feedback sessions (Karatay & Karatay, 2024; Godwin-Jones, 2018).

Researchers have extensively examined the pedagogical benefits and limitations of AWE tools in recent years. A growing body of empirical studies indicates that AWE feedback can contribute to measurable improvements in students' writing accuracy and quality. For example, students who use AWE-based feedback often show reductions in grammatical errors and higher writing scores in subsequent drafts or assignments (Stevenson & Phakiti, 2019; Parra & Calero, 2019). In classroom experiments, writing scores have been observed to improve significantly from first to final drafts when learners revise using feedback from systems like Criterion, suggesting that iterative AWE-guided revisions enhance formal writing accuracy over time (Liao, 2016; Stevenson & Phakiti, 2019). There is also evidence that the benefits of using AWE can transfer to new writing tasks after practice with AWE, students tend to carry forward some improvements (e.g. fewer errors, better use of syntax) into later, independent compositions (Liao, 2016). Importantly, these tools enable a multiple-draft writing process where learners can submit a draft, receive feedback instantly, revise, and resubmit, which reinforces a process-oriented approach to writing development (Karatay & Karatay, 2024;

Ding & Zou, 2024). The overall efficacy of AWE is supported by recent qualitative syntheses and systematic reviews, which report that most studies find a positive impact of AWE on student writing outcomes including enhanced grammatical accuracy and sometimes improved organizational clarity, especially for non-native English writers (Ding & Zou, 2024; Karatay & Karatay, 2024). In addition to objective improvements, many students and teachers have expressed favorable attitudes toward AWE tools noting that instant computer feedback is a valuable supplement to traditional instruction (Ding & Zou, 2024). Learners often appreciate the immediate corrections and the autonomy to self-edit, while instructors see potential in AWE to bolster students' basic writing skills. Overall, the literature suggests that AWE systems can effectively support writing development by addressing lower-level errors and encouraging more frequent revisions (Stevenson & Phakiti, 2019; Ranalli, 2018).

One of the key pedagogical advantages of AWE tools is their capacity to extend and enhance teacher feedback rather than replace it. Because AWE programs reliably handle micro-level concerns, such as grammar, mechanics, and lexical issues, they can reduce the time instructors spend on marking surface errors, and thus allow teachers to concentrate on higher-order writing skills like idea development, content, and organization (Wilson & Czik, 2016). Studies have shown that when AWE is integrated into writing instruction, teachers in combined feedback conditions tend to provide more feedback on global aspects of writing, presumably because the software addresses many local errors (Wilson & Czik, 2016). In practical terms, AWE can thus serve as a scaffolding tool that offloads part of the feedback workload; students receive detailed corrective feedback on language use from the system giving the instructors to the chance to give more targeted comments on content and ideas (Wilson & Czik, 2016; Link et al., 2022). Furthermore, AWE's consistency and objectivity in error marking mean that feedback is less prone to human bias or inconsistency, an advantage noted in the assessment literature (Stevenson & Phakiti, 2014). The individualized nature of AWE feedback tailored to each learner's specific errors can also address varying learner needs more efficiently than one teacher managing a large class, which is particularly useful in crowded classrooms where providing in-depth one-on-one writing feedback to every student is challenging (Karatay & Karatay, 2024; Link et al., 2022). Indeed, AWE has been recommended as a formative assessment aid in such contexts helping ensure that students get at least baseline corrective feedback even when instructor time is limited (Ranalli, 2018; Hockly, 2019).

Crucially, however, researchers caution that AWE systems should be used to complement, not substitute for, human feedback. A consistent theme in the literature is the call for a balanced approach integrating AWE with teacher or peer feedback in the writing process (Karatay & Karatay, 2024). While AWE provides speedy, objective correction, it lacks the pedagogical richness and nuanced understanding that human feedback offers. For instance, automated feedback tends to focus on rule-based errors and readily quantifiable features, and it cannot interpret meaning or assess the rhetorical quality of an argument the way a human reader can (Ranalli, 2018; Zhang & Hyland, 2018). Several studies note that AWE feedback is inherently limited to a one-size-fits-all scope; it cannot fully grasp context or writer intention nor can it provide deep content critique or creative suggestions (Ranalli, 2018; Dikli & Bleyle, 2014). Therefore, relying exclusively on AWE might leave higher-level issues like clarity of ideas, persuasiveness, or genre-specific expectations insufficiently addressed. In fact, some early research found that students can become frustrated or dissatisfied if AWE is the sole source of feedback especially when the computer's comments replace the teacher's role as audience and mentor (Chen & Cheng, 2008). More recent studies have echoed these concerns pointing out that students may not significantly improve global writing quality through AWE alone in the absence of instructor guidance (Ajabshir & Ebadi, 2023; Thi & Nikolov, 2022). Moreover, there is evidence that over-reliance on automated feedback could cause learners to pay less attention to teacher comments or to develop a false confidence in their writing without fully understanding the underlying rules (Karatay & Karatay, 2024). To mitigate these issues, researchers recommend combining teacher feedback with AWE; for example, using AWE for initial drafts to help students self-correct basic errors, then employing teacher feedback on later drafts to guide content and discourse-level improvements (Grimes & Warschauer, 2010; Wilson & Czik, 2016). This integrated approach leverages the strengths of both modalities. Empirical work supports this synergy. In this respect, one study found that students receiving hybrid feedback (AWE plus instructor input) produced writing of equal quality to those with teacher-only feedback, but with the added benefit that teachers in the hybrid model could focus more on higher-level issues without any loss in final text quality (Wilson & Czik, 2016; Ajabshir & Ebadi, 2023). In summary, the consensus is that AWE tools are most effective when used in tandem with human feedback making sure that mechanistic corrections do not come at the expense of personalized guidance and meaningful interaction (Link et al., 2020; Ding & Zou, 2024).

2.4 ChatGPT and EFL Writing: Opportunities and Limitations

Recent research highlights that ChatGPT and similar generative AI tools offer notable pedagogical benefits in EFL writing instruction. For example, ChatGPT can generate human-like responses that enhance writing efficiency, creativity, and overall proficiency (Alsaedi, 2024). By providing instantaneous, tailored feedback on drafts, ChatGPT enables students to revise iteratively. Empirical studies report significant gains in writing performance when ChatGPT is used as a formative feedback tool. Mahapatra (2024) found that ESL undergraduates who used ChatGPT alongside classroom instruction produced higher-quality essays and regarded the tool as a "reliable writing tool" for large classes. Similarly, ChatGPT improves learners' writing skills including grammar and vocabulary accuracy and fosters engagement through its personalized and accessible feedback. In a

qualitative case study, Werdiningsih et al. (2024) report that Indonesian graduate students positively valued ChatGPT's versatility in refining vocabulary, grammar, idea generation, and essay structure noting that the tool supported their writing accuracy and self-efficacy. These findings align with the notion that ChatGPT can function as an autonomous writing tutor offering scaffolded practice opportunities at any time (Qadir, 2023; Ibrahim & Kirkpatrick, 2023).

In addition to skill improvements, ChatGPT appears to enhance learner autonomy and motivation. Because students can engage with ChatGPT independently, they can experiment with language in a low-stakes environment. Qadir (2023) and Ibrahim and Kirkpatrick (2023) emphasize that ChatGPT can generate customized prompts and explanations matched to learners' proficiency, and thus supporting self-directed L2 practice. Students frequently report feeling more confident and motivated when using ChatGPT. For instance, Song and Song (2023) found that EFL students using ChatGPT were more motivated and active in writing exercises than those without the tool; they attributed this to the conversational, immediate feedback that makes writing practice more engaging and less intimidating. Mahapatra (2024) observed that ChatGPT helped learners overcome anxiety associated with seeking feedback, as they could receive guidance on demand without embarrassment. In short, the evidence suggests that ChatGPT can support an iterative learning cycle by giving immediate, explanatory feedback and suggestions for revision; it encourages students to take risks, gain confidence, and view writing as a process rather than a one-shot product (Mahapatra, 2024).

ChatGPT also offers practical advantages for teachers and classrooms. Because it can automatically detect many surface-level errors and generate model answers or writing prompts, it can reduce instructors' routine workload (Alsaedi, 2024; Ibrahim & Kirkpatrick, 2023). For example, ChatGPT can produce draft lesson plans, practice worksheets, and feedback comments, freeing teachers to focus on higher-order instruction and individualized guidance (Ibrahim & Kirkpatrick, 2023; Qadir, 2022). In crowded EFL classes where one-on-one conferencing is often infeasible, integrating ChatGPT can democratize access to formative feedback. Rather than replacing teachers, studies suggest using ChatGPT as a supplement; human instructors provide context-sensitive support and promote critical skills while ChatGPT provides consistent, on-demand error correction and idea generation (Ibrahim & Kirkpatrick, 2023; Song & Song, 2023). When deployed thoughtfully, ChatGPT's affordances can complement traditional pedagogy by extending practice opportunities and personalizing writing support to each student's pace and style (Rakhmonov & Kurbonova, 2023; Shaikh et al., 2023).

Despite these benefits, the literature also warns of significant limitations and challenges. A primary concern is student over-reliance on AI output. Over-dependence on ChatGPT can lead learners to accept suggestions without critical evaluation (Alsaedi, 2024). Alsaedi's systematic review (2024) identifies "overreliance" and even "learning loss" as potential issues when ChatGPT is used indiscriminately. Other reviews concur that excessive trust in ChatGPT risks undermining students' development of critical and analytical skills (Zhai et al., 2024). For example, Zhai, Wibowo, and Li (2024) note that students who unquestioningly follow AI-generated content may show diminished decision-making and critical thinking. Relatedly, Malik et al. (2023) report that a majority of student respondents believe heavy ChatGPT use could erode their own critical writing abilities. Qualitative analyses (Esmaeil et al., 2023) similarly warn that reliance on ChatGPT for argumentative essays can inhibit independent idea generation and reasoning. In short, while ChatGPT can assist with surface features, its use must be balanced; learners need to engage actively with ideas rather than passively consuming AI suggestions. This aligns with calls for educators to emphasize critical evaluation and supplementary instruction rather than simply outsourcing cognitive tasks to the AI.

Academic integrity is another serious challenge. ChatGPT's fluent text generation makes it easy for students to submit AI-written text as original work. Educators find it difficult to detect AI-derived writing reliably raising fears of academic dishonesty. Institutions must therefore develop clear policies and ethical guidelines for ChatGPT use (Alsaedi, 2024). Without safeguards, students may be tempted to use ChatGPT to shortcut writing assignments, which could inflate grades but impede genuine language development.

Furthermore, ChatGPT's feedback has inherent limitations. Although it can generate grammatically correct and contextually relevant suggestions, it may produce inconsistent or superficial corrections. Alsaedi (2024) points out that ChatGPT often fails to identify deep-level problems in discourse and organization. ChatGPT can also introduce errors or irrelevant information; it occasionally hallucinates details or offers inaccurate content that requires careful review by a teacher (Barrot, 2023). In practice, students may find some AI feedback overly verbose or difficult to interpret (Alsaedi, 2024; Barrot, 2023). Unlike a human instructor, ChatGPT lacks full awareness of the learner's context, background, or affective state, so its responses may not always address subtle learning needs. Consequently, educators warn that ChatGPT should not replace teacher judgment; rather, teachers should review AI suggestions and use them to prompt reflection. In fact, Alsaedi (2024) and others underscore that ChatGPT's help is most effective when teachers guide students in using AI critically and supplement AI output with personalized commentary.

Other considerations temper ChatGPT's promise. Technical access is not universal; in many EFL settings, especially in under-resourced or rural contexts, students lack reliable Internet or modern devices (Quy et al., 2023). This "digital divide" means that ChatGPT advantages may not reach all learners equally, potentially exacerbating educational inequities. Moreover, teacher readiness varies; successful integration requires instructors to be trained in AI literacy and to design appropriate tasks (Ibrahim & Kirkpatrick, 2023). If teachers are unprepared, they may underutilize ChatGPT or fail to mitigate its downsides. A related social concern is that heavy use of ChatGPT can alter the teacher–student dynamic. Some educators worry that students may rely more on AI feedback than on classroom dialogue and thus potentially diminishing opportunities for teacher scaffolding and emotional support. As a result, scholars emphasize that ChatGPT be used as a supplement to human instruction (Song & Song, 2023; Ausat et al., 2023).

2.5 Research Gap and Rationale for the Present Study

The above review reveals a clear research gap at the intersection of AI-driven feedback and EFL writing instruction, particularly in the context of secondary education in regions like Morocco. While the theoretical potential of ChatGPT as a formative feedback tool has been widely discussed, there is a lack of empirical studies investigating its actual impact on students' writing development (Mahapatra, 2024). Su et al. (2023) note that, despite the enthusiasm around ChatGPT, rigorous evidence on learning outcomes is scarce; most publications so far have been exploratory papers, commentaries, or small-scale trials. This means educators and policymakers currently have little data to rely on when deciding how (or whether) to incorporate ChatGPT into writing pedagogy. Questions such as "Does using ChatGPT feedback lead to measurable improvement in students' writing skills over time?" or "How do learners perceive and act upon ChatGPT's feedback compared to teacher feedback?" remain under-examined in the research literature.

Moreover, virtually no studies have addressed these questions in the North African or Moroccan EFL context. A recent overview by Ouahani and Mahraj (2025) confirms that within Morocco, *"no article was published on the application of ChatGPT to teach the four skills, vocabulary, or grammar"* in English classrooms. In other words, the integration of ChatGPT into English language teaching in Morocco is a new frontier that has yet to be explored by researchers. Much of the international research and discussions about ChatGPT in education come from contexts like East Asia, Europe, or North America, which have different educational infrastructures and student needs. There is a pressing need to investigate how ChatGPT might function as a feedback tool in Moroccan EFL settings, where class sizes, resource availability, and linguistic backgrounds differ from those in a high-tech East Asian classroom. Cultural and curricular factors could influence the effectiveness of AI feedback; for example, Moroccan secondary school students might have varying degrees of digital literacy or differing attitudes toward AI. Without research in the local context, educators in Morocco must extrapolate from foreign studies, which may not be fully applicable. Ouahani & Mahraj (2025) highlight this issue suggesting that findings from other contexts cannot simply be generalized to Morocco due to unique cultural and educational factors; hence they call for studies that minimize such a gap by focusing specifically on Moroccan learners

The present study is particularly concerned with formative feedback on paragraph writing essentially using ChatGPT to mimic the role of a writing tutor who gives comments on a student's draft and suggests improvements during the writing process. So far, no published study has zeroed in on ChatGPT playing this formative feedback role in an EFL classroom, especially at the secondary school level. The only somewhat related studies in the region are at the post-secondary level: for instance, Al-Zubaidi et al. (2024) conducted a survey-based study on the impact of ChatGPT on academic writing at Moroccan universities. Their findings indicated that university students and faculty saw both benefits and drawbacks in ChatGPT's role (Jaafari & Touzani, 2024). However, that study did not involve an actual classroom intervention or measure writing skill improvements; nor did it focus on formative feedback per se; it was more about general perceptions and self-reported use in higher education. Thus, while Al-Zubaidi et al.'s work provides a useful starting point, it underscores that we have virtually no empirical data on whether ChatGPT can tangibly improve students' writing when used as a feedback tool in a structured way. The secondary school context might differ significantly from universities; for example, teenage learners might need more scaffolding to use ChatGPT effectively, and the types of writing tasks (e.g., paragraph writing, narrative or descriptive writing in high school) are different from university academic writing. There is also the question of teachers' role in a secondary classroom in how will the teacher integrate ChatGPT feedback into their instruction? These nuances have not been studied yet.

Considering this lacuna in both global and local research, the present study is designed to make a valuable contribution by evaluating the effectiveness of ChatGPT-based formative feedback on Moroccan secondary students' paragraph writing skills. In doing so, it directly addresses the gaps identified: it will provide empirical evidence through a quasi-experimental design on learning outcomes, and it will do so in the Moroccan EFL context at the secondary level, where no prior research exists. By comparing a group of students who receive ChatGPT-assisted feedback with a control group receiving traditional teacher feedback, the study aims to generate data on whether ChatGPT can significantly improve writing quality in terms of grammar,

organization, and content development over the course of the intervention. It will also explore students' perceptions of using ChatGPT for feedback. Such insights will be crucial for understanding how to implement AI tools in classrooms effectively.

Guided by the theoretical insights and emerging research on formative feedback and AI-assisted writing support, the present study explores how a ChatGPT-based intervention integrated into EFL writing instruction can enhance Moroccan secondary school students' paragraph writing performance and influence their perceptions of feedback. It seeks to address the following research questions:

1. What is the impact of ChatGPT-based formative feedback on Moroccan secondary school students' paragraph writing skills?
2. How do students perceive the effectiveness and usefulness of ChatGPT as a formative feedback tool in their writing process?

3. Method

3.1 Participants and research context

This study was conducted at a large urban public high school in Sidi Bennour , Casablanca-Settat region of Morocco. It involved 87 11th grade students (second year of upper secondary education) in Morocco aged 16–19 from two intact classes. All participants had at least two years of prior English instruction. During the semester in which the intervention was implemented, the participants received four hours of English instruction per week as part of the general secondary English curriculum, which covered functions, vocabulary, listening, reading, grammar and writing. Instruction in these classrooms was largely learner-centered and in line with the communicative language teaching (CLT) approach, although individual teaching styles varied. Typical classroom activities included pair and group work, open discussions, and other interactive tasks consistent with CLT principles. One of the two intact classes was designated as the experimental group (45 students) and the other as the control group (42 students). Both classes followed the same curriculum, but the mode of feedback differed; the experimental group used ChatGPT as a formative feedback tool on their writing whereas the control group received traditional feedback from their teacher. This arrangement ensured that the only systematic difference between the groups was the source of formative feedback. Other contextual factors, including classroom environment and curriculum content, were held constant across the two groups.

Within this instructional context, the intervention was integrated into the writing component of the general English course with a focus on paragraph writing tasks. Paragraph writing was a key component of the grade-level curriculum, and the writing tasks used during the study were representative of the students' regular coursework. Students in both groups used the prescribed English textbook and supplemental worksheets for writing practice. All writing materials and tasks were aligned with the national curriculum making sure that both classes engaged with the same content and objectives. In addition, the researcher prepared targeted writing prompts designed to elicit paragraph development and to facilitate the use of the feedback tool. The prompts guided students through the intervention activities.

For the intervention lessons, the researcher served as instructor for both the experimental class by delivering the sessions and guiding students in the use of the ChatGPT-based feedback tool and the control group through giving them traditional writing formative feedback. In contrast, two experienced EFL teachers graded the participants' work in both the control class and the experimental one according to standard school procedures. The rationale behind not having the researcher grade the participants' pieces of writing was to minimize any potential conflicts of interest. Lessons were conducted primarily in English to provide an immersive L2 environment although Moroccan Arabic was occasionally used to explain the functionality of ChatGPT and to clarify the writing prompts further. All students had internet-enabled smartphones; in the experimental group these devices were used during class to access ChatGPT and receive formative feedback on their paragraph drafts. By contrast, students in the control group received analogous feedback directly from their teacher during in-class writing review. In each class, students then revised their paragraphs based on the feedback provided following the normal writing revision process in the curriculum.

3.2 Procedures

This study employed a quasi-experimental design to investigate the impact of ChatGPT-generated formative feedback on Moroccan secondary school students' paragraph writing performance. The procedures were implemented over the course of eight instructional weeks and were divided into four primary phases: orientation and ethical preparation, pre-testing, intervention implementation, and post-testing including a delayed post-test. Both quantitative and qualitative data were collected to ensure a comprehensive and triangulated understanding of the intervention's effectiveness. The independent variable in this study was the mode of formative feedback (ChatGPT-based vs. teacher-provided), and the dependent variable

was students' paragraph writing performance operationalized through analytic scores on pre-test, post-test, and delayed post-test paragraph writing tasks.

During week 1, a structured orientation session was conducted for both the experimental and control groups. The purpose of this session was to provide students with a clear explanation of the study's aims, procedures, and ethical principles including voluntary participation, confidentiality, and the non-evaluative use of their writing performance. To ensure equitable understanding, explanations were delivered in both English and Moroccan Arabic. Additionally, students in the experimental group participated in a 45-minute hands-on training workshop on how to interact with ChatGPT. In this session, the teacher researcher demonstrated how to input writing prompts, interpret feedback from the tool, and make revisions based on the suggestions. The participants were given time to practice using ChatGPT on sample paragraphs ensuring a consistent baseline of digital competence and eliminating technological unfamiliarity as a potential confounding variable.

In week 2, a diagnostic pre-test was administered to both groups to establish baseline writing performance. Students were instructed to write a 100–150 word paragraph on a topic aligned with the national curriculum (e.g., "Describe a problem you have solved and how you solved it") under supervised classroom conditions without access to feedback or assistance. The written responses were assessed by two experienced EFL teachers using a four-trait analytic rubric evaluating content relevance, organization, grammar, and vocabulary with each trait scored on a five-point scale. To reduce potential bias, writing samples were anonymized and scored blindly. To reduce potential bias, writing samples were anonymized and scored blindly. Inter-rater reliability was measured using Cohen's kappa to ensure scoring consistency, and the result of Cohen's Kappa was found to be more than 0.8, indicating a strong inter-rater reliability

The intervention phase extended from weeks 3 to 6 during which both groups engaged in four structured paragraph writing tasks with one paragraph per week. Each task followed a three-step cycle: drafting, receiving feedback, and revising. The writing prompts were thematically diverse but equally distributed in complexity and aligned with curriculum standards. Importantly, all students followed identical schedules and received the same prompts with the only distinction being the source of formative feedback.

In the experimental group, the participants wrote their initial drafts during class time and then submitted them to ChatGPT via their internet-enabled smartphones. Using a standardized prompt such as "Please provide detailed feedback on this paragraph's organization, grammar, and vocabulary, and suggest improvements," the participants received immediate written feedback from the tool. The teacher researcher supervised this activity in class to make sure there is appropriate use of the tool and offer support where necessary to help students interpret and apply the feedback. To encourage reflective revision, the participants were asked to briefly note which of ChatGPT's suggestions they used and why in order to promote metacognitive engagement with the feedback.

The control group completed the same writing tasks and engaged in feedback and revision processes, but their feedback was provided by their teacher in handwritten or oral form. Feedback addressed similar writing features including organization, grammar, vocabulary, and overall clarity. The participants revised their drafts during class in the same timeframe as their peers in the experimental group. To ensure instructional fidelity, the teacher researcher delivered lessons in both classes following consistent lesson plans and maintaining a neutral stance. This ensured that the only systematic difference between the groups was the mode of feedback.

In week 7, all participants completed a post-test writing task designed to measure improvements in paragraph writing ability. The task followed the same format and assessment criteria as the pre-test but featured a new topic of similar cognitive demand. Writing samples were again anonymized and rated independently by the same two raters using the same rubric.

To assess the retention of writing gains, a delayed post-test was administered in week 8. This test replicated the conditions of the pre- and post-tests using a new writing prompt and following the same rubric. This delayed measure provided a more robust evaluation of whether improvements in writing performance were sustained over time addressing both immediate and longitudinal effects of the intervention.

In parallel with the delayed post-test, the researcher used a post-intervention semi-structured interview with 9 participants from the experimental group to gain insights into their' perceptions of using ChatGPT as a feedback tool. The interview measured perceived usefulness, clarity, relevance, and ease of use and open-ended prompts that invited students to elaborate on the strengths, challenges, and pedagogical value of the ChatGPT. The qualitative responses were analyzed thematically to capture students' attitudes and experiences, and thereby complement and contextualize the quantitative data.

3.3 Instruments and Data Analysis

3.3.1 Analytic Writing Rubric

To evaluate students' writing performance across the pre-test, post-test, and delayed post-test tasks, a detailed analytic rubric was developed. This rubric was adapted from the well-established ESL composition profile proposed by Jacobs et al. (1981), which emphasizes discrete trait assessment in second language writing. In alignment with our study's objectives, the original five-category structure was refined into four key dimensions: content relevance, organization, vocabulary usage, and grammar accuracy. Each trait was rated on a 1–5 scale, with well-defined descriptors at each level. The four trait scores were then aggregated to yield a total writing score out of 20. The rubric was pilot-tested on a sample of student paragraphs prior to the main study and reviewed by two experienced EFL instructors to ensure content validity and face validity. Their feedback led to minor revisions in phrasing and performance level anchoring ensuring the rubric's clarity, instructional relevance, and alignment with curricular objectives. The analytic scoring approach was selected for its capacity to yield trait-specific diagnostic feedback, which is considered superior to holistic assessment in supporting writing development (Weigle, 2002). This rubric also served as the basis for quantifying the study's dependent variable , paragraph writing proficiency, thereby supporting construct validity.

3.3.2 Rater Training and Scoring Procedures

All student paragraphs were scored independently by two trained raters, both certified EFL teachers with substantial experience in writing assessment. Before the main scoring phase, the raters participated in calibration sessions using benchmark paragraphs representative of various proficiency levels. These norming sessions fostered a shared understanding of rubric criteria and score distinctions. During scoring, raters were blind to students' group assignments (experimental vs. control) and testing time point (pre, post, delayed), and thus reducing potential bias. Each paragraph was evaluated separately on all four traits by both raters. If their scores diverged by more than one point on any trait, they engaged in a consensus discussion to reconcile discrepancies. This dual-blind scoring with reconciliation protocol reflects established standards in L2 writing research and contributes to scoring reliability (Barkaoui, 2007).

3.3.3 Post-Intervention Interview

At the end of the intervention, a semi-structured interview was held with the experimental group to elicit perceptions of using ChatGPT as a formative feedback tool. The interview included open-ended prompts designed to capture students' reflections on their experiences with AI feedback. Prompts included questions such as, "What did you like most about using ChatGPT feedback?" and "What were the challenges of using it?" This combination of quantitative and qualitative items allowed for a rich, triangulated understanding of learners' attitudes and experiences, and thus enhancing the study's ecological validity.

3.3.4 Quantitative Data Analysis

All statistical analyses were conducted using SPSS. Descriptive statistics (means, standard deviations) were first calculated for writing scores at each testing point. The Shapiro–Wilk test was applied to assess normality of score distributions to assure the validity of subsequent parametric tests. To examine within-group learning gains, paired-sample t-tests were used to compare pre-test scores with post- and delayed post-test scores in each group. To evaluate between-group differences at each stage, independent-samples t-tests were conducted. In all analyses, two-tailed tests with an alpha level of .05 were employed. In addition, Cohen's d was calculated to determine effect sizes for both within- and between-group comparisons. These effect size metrics allowed for the interpretation of practical significance, complementing p-values and enhancing the analytical depth of the study (Lakens, 2013).

3.3.5 Qualitative Data Analysis

Responses from the post-intervention semi-structured interviews were analyzed using inductive thematic analysis following Braun and Clarke's (2006) six-phase model. The two researchers independently read and re-read the data to become familiar with its content, generated initial descriptive codes, and then collaboratively grouped these into broader themes representing patterns in the responses. Theme generation was iterative involving constant comparison with raw responses and refinement through discussion. Coding disagreements were resolved through negotiation until consensus was reached to enhance the trustworthiness and credibility of the analysis. This approach allowed for an in-depth exploration of students' subjective experiences and perceptions, which offered a qualitative triangulation of the quantitative findings and shedding light on how ChatGPT feedback was experienced and applied in the learning process.

4. Results

4.1 The impact of ChatGPT Feedback on Students' Writing Performance

This section presents the results of the study examining the impact of ChatGPT-generated formative feedback on Moroccan secondary school students' paragraph writing performance. The findings are reported in alignment with the study's research questions and focus on students' writing scores across three testing points: pre-test, post-test, and delayed post-test. First, descriptive statistics are provided for each group. Next, inferential statistics (paired-samples and independent-samples t-tests)

are reported to determine within- and between-group differences over time. Finally, results from the post-intervention semi-structured interview are summarized to offer insights into students' perceptions of the ChatGPT feedback.

Table 1. ***Descriptive Statistics of Writing Scores by Group across Three Testing Points***

Group		PreTest	PostTest	DelayedPostTest
Control	Mean	7.0462	8.076904761904760	8.808095238095241
	N	42	42	42
	Std. Deviation	.44435	.418198739632761	.434422734971016
Experimental	Mean	6.6918	10.007333333333333	10.274000000000004
	N	45	45	45
	Std. Deviation	.46510	.374629513956486	.384373161488582
Total	Mean	6.8629	9.075402298850577	9.566321839080455
	N	87	87	87
	Std. Deviation	.48635	1.047155398829124	.841654122937116

ALT Text for Table 1: This table displays the mean scores and standard deviations for pre-test, post-test, and delayed post-test writing assessments in both experimental and control groups, indicating greater post-intervention gains for the experimental group.

Table 1 presents the means and standard deviations for paragraph writing scores across the three testing points for both the experimental and control groups. The two groups started with relatively similar pre-test means (M = 6.69 for the experimental group; M = 7.05 for the control group) indicating a comparable baseline in writing performance. After the intervention, the experimental group showed a notable improvement (M = 10.01) compared to the control group (M = 8.08). This pattern continued in the delayed post-test where the experimental group maintained their gains (M = 10.27) while the control group showed only a modest increase (M = 8.81). These preliminary trends suggest a stronger and more sustained improvement in the experimental group following the ChatGPT-based feedback intervention.

Table 2. ***Independent-Samples t-test for Pre-Test Scores***

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PreTest	Equal variances assumed	.001	.979	-3.629	85	.000	-.35441	.09766	-.54860	-.16023
	Equal variances not assumed			-3.635	84.950	.000	-.35441	.09751	-.54829	-.16054

ALT Text for Table 2: This table presents the results of the independent-samples t-test comparing pre-test scores between experimental and control groups, revealing a statistically significant initial difference favoring the control group.

An independent-samples t-test was conducted to compare the pre-test paragraph writing scores of the experimental and control groups prior to the intervention. Results indicated a statistically significant difference between the two groups, $t(85) = -3.63$, $p < .001$, with the control group (M = 7.05, SD = 0.44) outperforming the experimental group (M = 6.69, SD = 0.47). Despite both groups beginning at relatively low levels of writing performance, this initial discrepancy highlights the importance of controlling for baseline differences when evaluating the impact of the intervention.

4.2 Within-Group Comparisons

To examine within-group changes in paragraph writing performance over time, paired-samples t-tests were conducted for each group. These tests compared students' scores between the pre-test and post-test, and between the post-test and delayed post-test.

Table 3. ***Paired Samples t-Test Results for Experimental Group: Pre-Test vs. Post-Test***

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PreTest - PostTest	-3.31555555555556	.646872038616166	.096429990072641	-3.50989743075186	-3.12121368035952	-34.383	44	.000

ALT Text for Table 3: Shows paired-samples t-test results for the experimental group between pre-test and post-test scores. The data confirm a statistically significant improvement in writing performance following ChatGPT feedback.

The experimental group demonstrated a statistically significant improvement from the pre-test (M = 6.69) to the post-test (M = 10.01), t(44) = –34.38, p < .001. This gain suggests that the ChatGPT-based feedback substantially enhanced students' writing performance in a short period. This statistically significant difference was accompanied by a very large effect size **(Cohen's d = 5.11)** indicating that the intervention had a substantial and practically meaningful impact on students' immediate writing performance.

Table 4. ***Paired Samples t-Test Results Comparing Pre-Test and Delayed Post-Test Scores in the Experimental Group***

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PreTest - DelayedPostTest	-3.582222222222224	.673221187780633	.100357889318043	-3.784480258193392	-3.379964186251055	-35.694	44	.000

ALT Text for Table 4: Contains paired-samples t-test results comparing pre-test and delayed post-test scores in the experimental group. Findings indicate a sustained and statistically significant improvement in writing performance over time.

A paired-samples t-test was conducted to examine whether the improvement in paragraph writing performance observed in the experimental group was sustained over time. The results in Table 3 above indicated a statistically significant difference between the pre-test (M = 6.69) and the delayed post-test (M = 10.27) scores, $t(44) = -35.69$, $p < .001$. The mean increase of 3.58 points suggests that students who received ChatGPT-generated formative feedback not only improved their writing performance immediately after the intervention but also maintained these gains over time demonstrating a lasting effect. The magnitude of this improvement was further supported by a very large effect size **(Cohen's d = 5.34)** which reinforces the sustained influence of ChatGPT feedback on long-term writing development.

Table 5. ***Paired Samples T-Test Comparing Pre-Test and Post-Test Scores (Control Group)***

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PreTest - PostTest	-1.030714285714288	.523573056645011	.080789076590217	-1.193871135852333	-.867557435576242	-12.758	41	.000

ALT Text for Table 5: Provides results of a paired-samples t-test for the control group comparing pre-test and post-test scores. Although statistically significant, the improvement is moderate compared to the experimental group.

The control group also showed a statistically significant improvement from the pre-test (M = 7.05) to the post-test (M = 8.08), t(41) = −12.76, p < .001. However, the magnitude of this improvement was substantially smaller than that observed in the experimental group. While statistically significant, the control group's improvement reflected a moderate-to-large effect size **(Cohen's d = 1.98)** substantially lower than that of the experimental group, which indicates more limited practical impact.

Table 6. ***Paired Samples T-Test for Control Group: Pre-test vs. Delayed Post-test Writing Scores***

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PreTest - DelayedPostTest	-1.761904761904761	.560832300030285	.086538302661618	-1.936672409642328	-1.587137114167194	-20.360	41	.000

ALT Text for Table 6: Summarizes paired-samples t-test results for the control group between pre-test and delayed post-test scores. The data reflect ongoing but relatively smaller writing gains.

Results also showed continued improvement in the control group between the pre-test and the delayed post-test, t(41) = -20.36, p < .001. Still, the mean difference was less than half that of the experimental group, which highlights the stronger long-term impact of ChatGPT feedback. The delayed gains in the control group yielded a large effect size **(Cohen's d = 3.14)**, though still smaller than the effect observed in the experimental group over the same period.

4.3 Between-Group Comparisons

To examine whether the post-intervention performance differed significantly between the experimental and control groups, independent-samples t-tests were conducted for the post-test and delayed post-test scores.

Table 7. ***Independent Samples t-test Comparing Post-Test Scores between Experimental and Control Groups***

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PostTest	Equal variances assumed	.039	.844	22.707	85	.000	1.930428571428569	.085014147331742	1.761397670607363	2.099459472249776
	Equal variances not assumed			22.620	82.363	.000	1.930428571428569	.085339793334704	1.760671746213846	2.100185396643293

ALT Text for Table 7: Presents independent-samples t-test results comparing post-test scores between experimental and control groups. Results show a significant advantage for the experimental group after the ChatGPT-based intervention.

Post-test results showed a statistically significant difference between groups, t(85) = 22.71, p < .001, with the experimental group (M = 10.01) outperforming the control group (M = 8.08). This suggests that the intervention had a considerable and immediate

impact on writing performance. The between-group difference also yielded a very large effect size **(Cohen's d = 3.67)** which results in a strong practical significance of the intervention beyond statistical results.

Table 8. ***Independent-Samples t-test Comparing Delayed Post-Test Scores Between Experimental and Control Groups***

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DelayedPostTest	Equal variances assumed	1.693	.197	16.694	85	.000	1.465904761904763	.087810965473419	1.291313037329410	1.640496486480117
	Equal variances not assumed			16.623	82.007	.000	1.465904761904763	.088184914591160	1.290477101671591	1.641332422137936

ALT Text for Table 8: Contains results of an independent-samples t-test comparing delayed post-test scores between the two groups. The experimental group continues to outperform the control group, indicating a lasting effect of the intervention.

Similarly, delayed post-test scores revealed a significant difference favoring the experimental group, $t(85) = 16.69$, $p < .001$. The maintained gap underscores the enduring benefit of ChatGPT feedback on students' writing ability. The persistent advantage was further substantiated by a very large effect size **(Cohen's d = 3.59)** which illustrates the durable educational value of AI-assisted formative feedback.

4.4 Students' Perceptions of ChatGPT-Generated Feedback

To complement the quantitative findings and provide deeper insights into students' experiences, semi-structured interviews were conducted with nine randomly selected students from the experimental group. The aim was to explore their perceptions of using ChatGPT as a formative feedback tool in their paragraph writing tasks. Thematic analysis was applied to identify recurring patterns in the students' responses. Three major themes emerged: perceived usefulness of the feedback, user experience and accessibility, and the impact on learning and motivation.

4.4.1 Perceived Usefulness of ChatGPT Feedback

Most interviewees expressed that ChatGPT's feedback was detailed, helpful, and easy to understand. Several students noted that the tool provided specific suggestions regarding grammar, vocabulary, and organization, which allowed them to identify and correct errors independently. One student remarked, *"It told me exactly where my sentence was wrong and how to fix it. That helped me more than just getting a mark."* The majority of participants found the feedback actionable and more immediate compared to waiting for teacher corrections.

4.4.2 User Experience and Accessibility

Students generally reported a positive experience interacting with ChatGPT. They described the tool as easy to access and navigate especially after the initial training session. However, a few students mentioned the need for clearer prompts or guidance in order to get the most relevant feedback. One student shared, *"Sometimes I didn't know how to ask my question the right way, but once I learned the structure, it worked well."* These responses indicate that while digital literacy played a role in user satisfaction, the learning curve was manageable.

4.4.3 Enhanced Engagement and Motivation

The immediacy and interactivity of ChatGPT appeared to foster greater engagement and intrinsic motivation among students. Several interviewees emphasized that receiving instant feedback encouraged them to revise their paragraphs more thoughtfully. One student commented, *"I used to hate rewriting, but now I enjoy fixing my writing because I understand why it's wrong."* This suggests that ChatGPT not only served as a corrective tool but also supported a more reflective and student-centered approach to learning.

5. Discussion

5.1 The Impact of ChatGPT-Based Formative Feedback on Students' Writing Skills

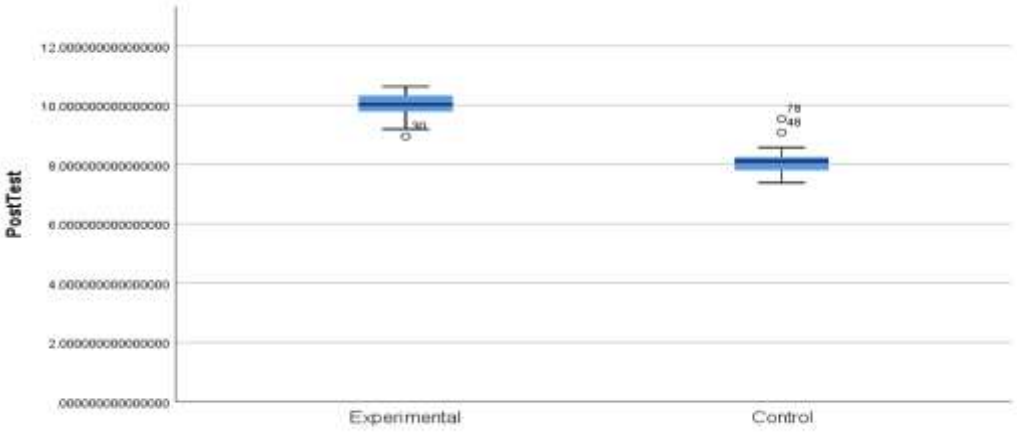
The present study found robust evidence supporting the effectiveness of ChatGPT-generated formative feedback in improving Moroccan secondary school students' paragraph writing skills. The experimental group, which received AI-assisted feedback, showed significantly higher gains than the control group, both immediately after the intervention and in the delayed post-test. Specifically, students in the experimental group improved from a mean pre-test score of 6.69 to 10.01 in the post-test and 10.27 in the delayed post-test (see Tables 3 and 4), with statistically significant t-values of -34.38 and -35.69, respectively ($p < .001$ in both cases). These large effect sizes reflect a consistent and durable enhancement in writing performance over time.

These findings align with previous research that underscores the instructional value of AI-based feedback tools. For instance, Labadze et al. (2023) reported that chatbots can provide immediate, individualized feedback that fosters student autonomy and deepens engagement. Similarly, Wu et al. (2022) highlighted the motivational effect of real-time AI support in enhancing students' writing fluency and coherence. The consistent progress across testing points in our study mirrors such conclusions and suggests that ChatGPT's scaffolding role can be especially impactful in EFL contexts marked by large class sizes and limited teacher feedback (Mahapatra, 2024).

In contrast, the control group, which received only traditional teacher feedback, showed modest improvement from 7.05 in the pre-test to 8.08 in the post-test and 8.81 in the delayed post-test (Tables 5 and 6), with t-values of -12.76 and -20.36, respectively ($p < .001$). Although statistically significant, these gains were far smaller than those of the experimental group. This disparity supports Hattie and Timperley's (2007) argument that feedback is only effective when it is timely, specific, and actionable—qualities more consistently observed in AI-generated responses.

Importantly, the independent-samples t-tests revealed significant between-group differences at both the post-test ($t = 22.71$, $p < .001$, Table 7) and delayed post-test stages ($t = 16.69$, $p < .001$, Table 8), with the experimental group outperforming the control group by 1.93 and 1.47 points, respectively. These sustained advantages echo the findings of Qadir et al. (2022), who emphasized the long-term cognitive and metacognitive benefits of interactive digital feedback in L2 writing instruction. Figure 1 displays the distribution of post-test scores between the experimental and control groups, further illustrating the significant performance gap identified in the statistical analysis.

Figure 1. ***Boxplot of Post-Test Scores by Group***



ALT Text for Figure 1: Boxplot comparing post-test writing scores between experimental and control groups. The experimental group displays a higher median and tighter spread, indicating improved and more consistent writing outcomes.

Figure 1 displays the distribution of post-test scores for the experimental and control groups. As shown, the experimental group achieved markedly higher median scores with a narrower interquartile range which indicates both stronger performance and greater consistency among students. In contrast, the control group's scores were lower and more dispersed with several outliers below the median. This visual representation reinforces the statistically significant difference between groups ($t = 22.71$, $p < .001$) and thus emphasizing the effectiveness of ChatGPT-generated feedback in enhancing writing outcomes. However, while the intervention proved effective, it is worth noting that the pre-test scores were significantly different between groups ($t = -3.63$, $p < .001$, Table 2), with the control group initially scoring higher. Although this imbalance was not ideal, the experimental

group not only closed the gap but eventually outperformed their peers which suggests the transformative potential of ChatGPT even for lower-performing students. This lends further support to the idea that AI-powered tools can serve as equalizers in resource-constrained environments (Hunnes & Olsen, 2023). Overall, these findings support the theoretical assumptions of Vygotsky's Zone of Proximal Development, where guided learning, here via ChatGPT, helps learners achieve levels of performance they would not attain alone. They also validate the view that technology-enhanced formative assessment can be a strategic pedagogical tool in the digital age (Zawacki-Richter et al., 2019).

5.2 Students' Perceived Benefits of ChatGPT-Based Formative Feedback

The interview findings offered nuanced insights into how students experienced and perceived ChatGPT as a formative feedback tool. Three recurring themes emerged from the thematic analysis and align with the literature on effective feedback and AI integration in writing pedagogy.

5.2.1 Perceived Usefulness and Clarity of Feedback

Students widely reported that ChatGPT's feedback was detailed, targeted, and actionable, especially in areas like grammar, vocabulary, and sentence structure. This resonates strongly with Shute's (2008) assertion that effective formative feedback should be specific and strategy-oriented to support student revision. In this respect, one participant stated, *"ChatGPT didn't just tell me what was wrong. It told me why and gave me examples. It was like a personal tutor."* Another added, *"I could finally understand my grammar mistakes and rewrite my sentences more confidently."* The immediacy and clarity of ChatGPT's responses allowed learners to engage in more independent error correction and reduce reliance on teacher input. This echoes findings by Stevenson and Phakiti (2019) which found that Automated Writing Evaluation (AWE) tools can lead to reductions in grammatical errors and improvements in accuracy across drafts. Moreover, the feedback provided by ChatGPT fulfilled key criteria for high-impact feedback identified in the literature being timely, specific, and clear (Hattie & Timperley, 2007; Brookhart, 2017). The ability of students to "understand why a sentence was wrong and how to fix it," as one participant described, supports Sadler's (1989) view of formative feedback as a bridge between current and desired performance.

5.2.2 User Experience and Accessibility

Students generally found ChatGPT user-friendly and accessible especially after brief orientation. While a few struggled initially with phrasing their input to receive optimal feedback, most adapted quickly. This reflects findings by Werdiningsih et al. (2024) who highlighted that while some scaffolding is necessary, students typically adapt to ChatGPT's input style and appreciate its 24/7 availability. Importantly, the responses suggest that digital literacy influenced how smoothly students navigated ChatGPT. As also noted by Ranalli (2018) and Ibrahim and Kirkpatrick (2023), student readiness plays a key role in maximizing AI tools. Nevertheless, most learners described the tool as intuitive once they understood how to frame their questions.

5.2.3 Engagement, Confidence, and Motivation

Perhaps the most striking theme was the boost in student motivation and reflective learning behaviors. Several participants noted that they were more willing to revise their work after receiving ChatGPT feedback, particularly because it allowed them to understand their errors better. This aligns with Mahapatra (2024) who found that ChatGPT fosters student engagement and reduced writing anxiety. As one student expressed, "I used to hate rewriting, but now I enjoy fixing my writing because I understand why it's wrong." This reflects not only increased motivation but also enhanced metacognitive awareness. Similarly, Song and Song (2023) note that the dialogic and responsive nature of ChatGPT supports learner autonomy and motivation.

5.2.4 Limitations and Student Challenges

Although overall perceptions were positive, some students indicated initial difficulty with formulating effective prompts to elicit relevant feedback. This concern echoes findings from Barrot (2023) and Alsaedi (2024), who argue that AI tools require prompt literacy and may deliver verbose or off-target responses if not properly queried. Moreover, a few students hinted at over-reliance on ChatGPT, a challenge also raised by Malik et al. (2023), who warned of diminished critical thinking when learners depend too heavily on AI-generated responses. However, in this study, over-reliance was not a dominant theme, likely because students continued to receive teacher guidance alongside ChatGPT feedback which confirms literature recommendations for balanced AI-human integration (Ajabshir & Ebadi, 2023; Link et al., 2020).

6. Conclusion

This study provides compelling evidence for the pedagogical potential of ChatGPT-generated formative feedback in enhancing Moroccan secondary school students' paragraph writing skills. The experimental group, which received AI-assisted feedback, exhibited significantly greater gains across all assessment stages compared to the control group, which confirms the practical value of real-time, specific, and actionable digital feedback in L2 writing contexts. These findings underscore the transformative

role of generative AI in democratizing access to individualized instruction particularly in resource-constrained educational environments. Moreover, the integration of students' qualitative insights highlighted not only the cognitive gains associated with ChatGPT but also improvements in motivation, self-regulation, and metacognitive awareness validating its role as a supportive writing assistant. However, the study is not without limitations. The initial inequality in pre-test scores between groups, the relatively small sample size, and the reliance on a single educational context may restrict the generalizability of the results. Additionally, while ChatGPT provided valuable feedback, students' overreliance on the tool and occasional difficulty in prompt formulation point to the need for digital literacy scaffolding. Future research should explore longitudinal applications of AI-based feedback, examine its efficacy across diverse language skills (e.g., speaking or reading), and assess the long-term development of learner autonomy and critical thinking. Integrating AI tools like ChatGPT into blended pedagogical frameworks that balance technological and human guidance holds promise for optimizing writing instruction in the digital age.

Acknowledgments: Not applicable.
Disclosure Statement: The authors report that there are no competing interests to declare.
Funding: This research is not funded
Artificial Intelligence Statement: AI and AI-assisted technologies were not used
Ethics Approval Statement: This study was conducted in accordance with the principles outlined in the Belmont Report and the Declaration of Helsinki. The research involved minimal risk to participants and was limited to standard educational practices within the classroom. As such, **"**Imam El Ghazali High School" determined that formal ethics approval was not required. However, informed consent from parents and student assent were obtained before participation, ensuring voluntary engagement and confidentiality of all data.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers
ORCID ID: https://orcid.org/0009-0000-7701-1128

References:

[1] Ajabshir, Z. F., & Ebadi, S. (2023). The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(1), 26.
[2] Alsaedi, N. (2024). ChatGPT and EFL/ESL writing: A systematic review of advantages and challenges. *English Language Teaching*, *17*(5), 41-50.
[3] Al-Zubaidi, K., Jaafari, M., & Touzani, F. Z. (2024). Impact of ChatGPT on Academic Writing at Moroccan Universities. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*.
[4] Ausat, A. M. A., Azzaakiyyah, H. K., Permana, R. M., Riady, Y., & Suherlan, S. (2023). The role of ChatGPT in enabling MSMEs to compete in the digital age. *Innovative: Journal Of Social Science Research*, *3*(2), 622-631.
[5] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 100745.
[6] Benali, M., & Mak, J. (2024). Perception of Preservice Moroccan Teachers Regarding the Adoption of ChatGPT in Their Teaching Practices. In *Artificial Intelligence Applications in K-12* (pp. 112-137). Routledge.
[7] Benfatah, M., Marfak, A., Saad, E., Hilali, A., Nejjari, C., & Youlyouz-Marfak, I. (2024). Assessing the efficacy of ChatGPT as a virtual patient in nursing simulation training: A study on nursing students' experience. *Teaching and Learning in Nursing*, *19*(3), e486-e493.
[8] Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *ETS Research Report Series*, *2011*(1), i-99.
[9] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.
[10] Bonner, E., Lege, R., & Frazier, E. (2023). Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas for Teaching. *Teaching English with Technology*, *23*(1), 23-41.
[11] Boubker, O. (2024). From chatting to self-educating: Can AI tools boost student learning outcomes?. *Expert Systems with Applications*, *238*, 121820.
[12] Bouzar, A., EL Idrissi, K., & Ghourdou, T. (2024). ChatGPT and Academic Writing Self-Efficacy: Unveiling Correlations and Technological Dependency among Postgraduate Students. *Arab World English Journal*.
[13] Brookhart, S. M. (2017). *How to give effective feedback to your students*.
[14] Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325.
[15] Chen, C. F. E., & Cheng, W. Y. E. C. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes.
[16] Crosthwaite, P., Ningrum, S., & Lee, I. (2022). Research trends in L2 written corrective feedback: A bibliometric analysis of three decades of Scopus-indexed research on L2 WCF. *Journal of Second Language Writing*, *58*, 100934.
[17] Daneshvar, E., & Rahimi, A. (2014). Written corrective feedback and teaching grammar. *Procedia-Social and Behavioral Sciences*, *136*, 217-221.
[18] Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback?. *Assessing writing*, *22*, 1-17.
[19] Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies*, *29*(11), 14151-14203.
[20] Esmaeil, A. A. A., Dzulkifli, D. N. A. K., Maakip, I., Matanluk, O. O., & Marshall, S. (2023). Understanding student perception regarding the use of ChatGPT in their argumentative writing: A qualitative inquiry. *Jurnal Komunikasi: Malaysian Journal of Communication*, *39*(4), 150-165.

[21] Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, *83*(1), 70-120.
[22] Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, *32*(2), 181-201.
[23] Godwin-Jones, R. (2018). Second language writing online: An update.
[24] Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6).
[25] 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.
[26] Han, Y., & Hyland, F. (2019). Academic emotions in written corrective feedback situations. *Journal of English for Academic Purposes*, *38*, 1-13.
[27] Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.
[28] Hockly, N. (2019). Automated writing evaluation.
[29] Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, *39*(2), 83-101.
[30] Ibrahim, K., & Kirkpatrick, R. (2024). Potentials and implications of ChatGPT for ESL writing instruction. *International Review of Research in Open and Distributed Learning*, *25*(3), 394-409.
[31] Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464.
[32] Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The modern language journal*, *99*(1), 1-18.
[33] Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, 103332.
[34] Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, *119*(2), 254.
[35] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, *54*(2), 537-550.
[36] Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational psychology review*, *1*, 279-308.
[37] Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL quarterly*, *52*(1), 5-33.
[38] Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 56.
[39] Lee, I. (2014). Revisiting teacher feedback in EFL writing from sociocultural perspectives. *Tesol Quarterly*, *48*(1), 201-213.
[40] Li, S., & Vuono, A. (2019). Twenty-five years of research on oral and written corrective feedback in System. *System*, *84*, 93-109.
[41] Liao, H. C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, *62*, 77-92.
[42] Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *35*(4), 605-634.
[43] Lipnevich, A. A., & Panadero, E. (2021, December). A review of feedback models and theories: Descriptions, definitions, and conclusions. In *Frontiers in Education* (Vol. 6, p. 720195). Frontiers Media SA.
[44] Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, *11*(1), 9.
[45] Mahboob, A. (2015). Understanding and providing 'cohesive'and 'coherent'feedback on writing. *Writing & Pedagogy*, *7*(2-3), 355-376.
[46] Mahfoodh, O. H. A. (2017). "I feel disappointed": EFL university students' emotional responses towards teacher written feedback. *Assessing Writing*, *31*, 53-72.
[47] Malik, A., Khan, M. L., & Hussain, K. (2023). How is ChatGPT transforming academia? Examining its impact on teaching, research, assessment, and learning. *Examining its Impact on Teaching, Research, Assessment, and Learning (April 9, 2023)*.
[48] Nicol, D. (2011). Developing students' ability to construct feedback. *QAA Scotland, Enhancement Themes*.
[49] Olsen, T., & Hunnes, J. (2024). Improving students' learning—the role of formative feedback: experiences from a crash course for business students in academic writing. *Assessment & Evaluation in Higher Education*, *49*(2), 129-141.
[50] Ouahani, N. E., & MAHRAJ, M. (2025). ChatGPT Application in ELT: the International Context, the Moroccan Context, and Future Prospects.
[51] Parra G, L., & Calero S, X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, *12*(2), 209-226.
[52] Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, *108*(8), 1098.
[53] Peltzer, K., Lira Lorca, A., Krause, U. M., Graham, S., & Busse, V. (2025). Effects of feedback on deep-level features of argumentative writing over multiple drafts: Insights from an intervention study with secondary EFL students. *Reading and Writing*, 1-32.
[54] Qadir, J. (2023, May). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE global engineering education conference (EDUCON)* (pp. 1-9). IEEE.
[55] Rakhmonov, I. U., & Kurbonova, R. S. (2023). The pedagogical principles and effectiveness of utilizing ChatGPT for language learning. *Research and education*, *2*(9), 226-243.
[56] Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, *52*, 100816.
[57] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, *18*(2), 119-144.
[58] Saeli, H., & Cheng, A. (2021). Peer Feedback, Learners' Engagement, and L2 Writing Development: The Case of a Test-Preparation Class. *Tesl-Ej*, *25*(2), n2.
[59] Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction*, *93*, 101961.

[60] Shaikh, S., Daudpota, S. M., Yayilgan, S. Y., & Sindhu, S. (2023, December). Exploring the potential of large-language models (LLMs) for student feedback sentiment analysis. In *2023 International Conference on Frontiers of Information Technology (FIT)* (pp. 214-219). IEEE.
[61] Shrivastava, S. R., & Shrivastava, P. S. (2020). Improving the feedback process in medical education. *South-East Asian Journal of Medical Education*, *14*(1).
[62] Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.
[63] 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.
[64] Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition*, *42*, 1-16.
[65] Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51-65.
[66] Stevenson, M., & Phakiti, A. (2019). Automated feedback and second language writing. *Feedback in second language writing: Contexts and issues*, 125-142.
[67] Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752.
[68] Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied linguistics*, *16*(3), 371-391.
[69] Tai, A. M. Y., Meyer, M., Varidel, M., Prodan, A., Vogel, M., Iorfino, F., & Krausz, R. M. (2023). Exploring the potential and limitations of ChatGPT for academic peer-reviewed writing: Addressing linguistic injustice and ethical concerns. *Journal of Academic Language and Learning*, *17*(1), T16-T30.
[70] Teng, M. F., & Ma, M. (2024). Assessing metacognition-based student feedback literacy for academic writing. *Assessing Writing*, *59*, 100811.
[71] Thi, N. K., & Nikolov, M. (2022). How teacher and Grammarly feedback complement one another in Myanmar EFL students' writing. *The Asia-Pacific Education Researcher*, *31*(6), 767-779.
[72] Tran, T. T. T. (2025). Enhancing EFL Writing Revision Practices: The Impact of AI-and Teacher-Generated Feedback and Their Sequences. *Education Sciences*, *15*(2), 232.
[73] Truscott, J. (2010). Some thoughts on Anthony Bruton's critique of the correction debate. *System*, *38*(2), 329-335.
[74] Voerman, L., Meijer, P. C., Korthagen, F. A., & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and teacher education*, *28*(8), 1107-1115.
[75] Werdiningsih, I., Marzuki, Indrawati, I., Rusdin, D., Ivone, F. M., Basthomi, Y., & Zulfahreza. (2024). Revolutionizing EFL writing: unveiling the strategic use of ChatGPT by Indonesian master's students. *Cogent Education*, *11*(1), 2399431.
[76] Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94-109.
[77] Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. *Studies in Higher Education*, *42*(11), 2026-2041.
[78] Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in psychology*, *10*, 487662.
[79] Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in psychology*, *10*, 487662.
[80] Xiao, F., Zhu, S., & Xin, W. (2025). Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in English as a foreign language education: A scoping review. *ECNU Review of Education*, 20965311241310881.
[81] Xuan-Quy, D., Le, N. B., Phan, X. D., & Ngo, B. B. (2023). An Evaluation of ChatGPT's Proficiency in English Language Testing of The Vietnamese National High School Graduation Examination.
[82] Yu, S., & Hu, G. (2017). Understanding university students' peer feedback practices in EFL writing: Insights from a case study. *Assessing Writing*, *33*, 25-35.
[83] Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(4), 461-493.
[84] Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, *48*, 100525.
[85] Zhai, N., & Ma, X. (2022). Automated writing evaluation (AWE) feedback: A systematic investigation of college students' acceptance. *Computer Assisted Language Learning*, *35*(9), 2817-2842.
[86] Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, *61*(4), 875-900.
[87] Zhai, X., Nyaaba, M., & Ma, W. (2024). Can generative AI and ChatGPT outperform humans on cognitive-demanding problem-solving tasks in science?. *Science & Education*, 1-22.
[88] Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439.
[89] Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90-102.
[90] Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, *37*, 13-24