

Assessing the efficacy of ChatGPT's automated corrective feedback in enhancing students' writing proficiency



Sultan M. Alanazi¹, Bilel Elmotri^{2,3}, Gamal S. Khamis¹, Abdulbasit A. Darem^{1,*}

¹Department of Computer Science, Northern Border University, Arar, Saudi Arabia

²Department of Languages and Translation, Applied College, Northern Border University, Arar, Saudi Arabia

³Department of English Language, Faculty of Letters and Humanities, University of Sfax, Sfax, Tunisia

ARTICLE INFO

Article history:

Received 16 September 2024

Received in revised form

18 January 2025

Accepted 7 February 2025

Keywords:

Writing proficiency

Automated feedback

Language models

Student perceptions

Educational tools

ABSTRACT

The rise of large language models (LLMs) like ChatGPT has sparked interest in their educational applications. This study evaluates ChatGPT's effectiveness in enhancing students' writing skills through automated feedback. Using a mixed-methods design, pre-test assessments were conducted, followed by an intervention where the experimental group used ChatGPT for writing tasks, while the control group received traditional instruction. Post-test results showed that the ChatGPT group achieved greater improvements in writing performance, with higher average scores and percentage increases compared to the control group. Surveys and discussions revealed positive student perceptions of ChatGPT, highlighting its ease of use and constructive feedback, though concerns about data privacy, bias, and occasional irrelevant suggestions were noted. The findings suggest that ChatGPT can be a valuable educational tool for improving writing proficiency, but ethical considerations and individual differences in effectiveness must be addressed. Further research should explore its long-term impact, comparison with other automated systems, and applications in diverse learning contexts.

© 2025 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The rapid advancement of technology has significantly affected various fields, and education is no exception. One of the most promising technological advancements in education is the use of Natural Language Processing (NLP) tools, particularly for writing instruction. Among these tools, ChatGPT, developed by OpenAI, stands out as a potential game-changer (Tarchi et al., 2024). This large-scale language model, trained on extensive text data, is capable of generating human-like text in response to prompts, offering valuable feedback and assistance to students. This capability can revolutionize the teaching and evaluation of writing in classrooms, particularly for students who face challenges in grammar, punctuation, and sentence structure. Writing is a critical skill for academic and professional success (Zhong et al., 2025). However, it

often poses significant challenges for students, especially those struggling with grammar, punctuation, and idea development. ChatGPT, with its advanced NLP capabilities, can provide instant, specific feedback, helping students improve their writing skills more effectively than traditional methods. This paper explores the potential benefits of using ChatGPT in educational settings, examining how this technology can be integrated into the classroom to enhance students' writing proficiency.

The integration of Natural Language Processing (NLP) tools, such as ChatGPT, into educational practices has garnered increasing attention for their potential to enhance students' writing skills by providing automated corrective feedback. OpenAI, a leading research organization founded in 2015, has made notable advancements in the development of sophisticated language models. In 2019, OpenAI introduced GPT-2, a transformer-based language model comprising 1.5 billion parameters, trained on 40GB of text data. While initial concerns regarding the potential misuse of such technology arose, the release of GPT-2 represented a pivotal moment in the evolution of language models, paving the way for subsequent innovations, including GPT-3 (Radford et al., 2019). GPT-3, launched in 2020, was a significant advancement over GPT-2, boasting 175 billion

* Corresponding Author.

Email Address: basit.darem@nbu.edu.sa (A. A. Darem)

<https://doi.org/10.21833/ijaas.2025.02.022>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0002-5650-1838>

2313-626X/© 2025 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

parameters—more than 100 times larger than its predecessor. This model demonstrates an exceptional ability to generate coherent, contextually appropriate text, establishing itself as a valuable asset in educational settings. Its capacity to understand and produce human-like responses enables it to offer comprehensive feedback on student writing, addressing frequent challenges such as grammatical inaccuracies, weak structure, and unclear expression (Brown et al., 2020).

Building on the success of its predecessors, GPT-4 was introduced in 2023, offering even greater performance improvements. GPT-4 has been described as a more robust and reliable model, excelling in complex reasoning tasks, while refining its ability to provide nuanced and context-aware feedback. Its enhanced capabilities make it particularly suitable for educational applications, where it can support not only the correction of grammar and structure but also guide students through more intricate cognitive tasks, such as critical thinking and argument development. As research continues to show the benefits of integrating AI-driven feedback alongside human instructors, tools like GPT-4 underscore the growing importance of these technologies in shaping the future of education (Mohsen, 2022).

The potential benefits of ChatGPT extend beyond just providing feedback. For example, it can help students overcome writer's block by generating prompts and ideas assisting with organizing and developing their writing, and offering suggestions for improving vocabulary and style. This can be particularly beneficial for students with learning difficulties, English language learners, and those from low socio-economic backgrounds who may not have access to additional educational resources.

This paper will explore the potential benefits and challenges of incorporating ChatGPT into educational settings to improve students' writing proficiency. It will examine current research on the use of NLP tools in education and discuss the implications of this technology for the future of writing instruction. By understanding the strengths and limitations of ChatGPT, educators can better integrate this tool into their teaching practices to support student learning.

The primary contributions of this study are to:

1. Investigate the effectiveness of ChatGPT in improving students' writing abilities as measured by standardized writing assessments.
2. Explore students' perceptions and experiences of using ChatGPT as a writing tool.

The rest of the paper begins with a literature review on AI-powered tools in education, focusing on ChatGPT and automated writing evaluation systems. The Methodology section outlines the mixed-methods approach, including data collection through pre-tests, post-tests, surveys, and focus groups. The Results section compares the performance of the experimental and control groups,

followed by a Discussion that interprets the findings. The paper concludes with recommendations for future research and considerations for the ethical use of AI in education.

2. Literature review

The rapid advancements in Natural Language Processing (NLP) have brought significant attention to the potential applications of large-scale language models in education. Among these, OpenAI's ChatGPT stands out for its ability to generate human-like text based on massive datasets. ChatGPT's capability to provide immediate and relevant feedback makes it a promising tool for enhancing writing skills, particularly in educational settings. This study aims to assess the efficacy of ChatGPT's automated corrective feedback in improving students' writing proficiency.

Writing is a critical skill essential for both academic and professional success. However, it often presents challenges for students, particularly those struggling with grammar, punctuation, and sentence structure. These difficulties are compounded for students experiencing writer's block or those who find it challenging to develop ideas. Traditional methods of providing feedback are often limited by the time and availability of instructors, which can hinder students' progress.

The integration of ChatGPT into educational practices offers a potential solution to these challenges. ChatGPT has been trained on diverse datasets, enabling it to perform various natural language tasks such as machine translation, text summarization, and question answering (Geng and Razali, 2022; Gozali et al., 2024). Its ability to generate coherent and contextually appropriate text makes it a valuable tool for providing instant feedback and assistance in the writing process. This technology can support students in overcoming common writing challenges, thereby enhancing their writing proficiency.

Several studies highlight the benefits of using automated writing evaluation (AWE) systems in education. This recent study Mahapatra (2024) explored the application of ChatGPT and similar NLP tools in improving writing proficiency, particularly in ESL settings. The study addresses the Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study.

Mayer et al. (2020) found that students with access to tools providing writing prompts and organizational suggestions demonstrated enhanced writing abilities. Moreover, Ma et al. (2024) reported that a combination of AI-generated and human feedback yielded better outcomes than either method alone. Despite these advantages, the implementation of ChatGPT and similar AWE systems is not without challenges. Critics argue that such systems may lack the personalization and nuanced understanding that human instructors provide. Concerns also arise regarding the potential for over-reliance on automated feedback, which

might diminish the role of teachers in the educational process. Ethical considerations, such as the risk of generating misleading or biased content, must also be addressed (Perkins, 2023).

Several studies have highlighted the potential benefits of using automated writing evaluation (AWE) systems and NLP tools in education. For instance, Alkhalidi (2023) found that technological tools enhance students' creative writing, improving their performance, vocabulary, and imagination. Similarly, Geng and Razali (2022) suggested that digital tools like AWE empower students to write better by offering instant scoring and personalized feedback. Chen et al. (2019) demonstrated that combining AI and human feedback leads to better writing performance than using either alone. These findings underscore the importance of integrating advanced NLP tools like ChatGPT into educational practices to enhance student learning outcomes. However, the application of NLP tools in education is still an emerging field, with ongoing debates about their effectiveness and ethical implications. Some researchers have raised concerns about the potential biases in AI-generated feedback and the need for transparent and ethical use of these technologies. For example, Zhuo et al. (2023) and Gaud (2023) highlighted issues related to data privacy, bias, and the ethical frameworks required to guide the use of AI in education. Addressing these concerns is crucial to ensuring that NLP tools like ChatGPT are used responsibly and effectively in educational settings.

The role of Automated Writing Evaluation (AWE) tools in enhancing ESL/EFL learners' writing skills has gained considerable attention. Research has shown that AWE tools can provide valuable, real-time corrective feedback, which addresses common linguistic challenges faced by non-native speakers. For example, Geng and Razali (2022) demonstrated that AWE tools, through instant scoring and corrective feedback, significantly improve writing accuracy, vocabulary, and organizational skills in ESL students. By providing continuous, structured feedback, AWE tools help students build confidence and proficiency in writing, allowing for progressive improvement.

Alkhalidi (2023) further explored how AWE tools impact creative writing in EFL contexts, showing that these tools not only support technical accuracy but also enhance students' creativity and vocabulary by offering lexical suggestions and alternative sentence structures. This dual function of AWE tools is especially beneficial for EFL learners, who often need additional support in vocabulary expansion and expression. Similarly, research by Gozali et al. (2024) found that AWE tools promote feedback literacy among EFL learners, helping them to better interpret and apply feedback in revisions. This skill is critical in language learning, as it fosters autonomy and self-assessment, empowering students to improve their writing over time. The study revealed that combining AWE tools with human feedback yielded higher writing gains than using either method alone, underscoring the potential for a balanced,

integrative approach to writing instruction for ESL/EFL students.

Moreover, Song and Song (2023) highlighted that AWE tools play a crucial role in enhancing ESL learners' motivation and engagement with writing. The study suggested that AWE tools can reduce the apprehension many ESL students feel toward writing, as immediate, formative feedback encourages continuous learning and reduces reliance on instructors for corrections. This increased engagement may lead to greater practice and improvement in writing skills.

In light of these findings, ChatGPT's unique features, such as adaptive, context-aware feedback, offer additional advantages over traditional AWE tools in the ESL/EFL context. Unlike rule-based systems, ChatGPT's ability to generate tailored feedback allows it to address complex issues in coherence, tone, and argumentation, which are often challenging for ESL/EFL learners. By providing immediate yet nuanced feedback, ChatGPT may offer more comprehensive support for students in improving both language mechanics and higher-level writing skills, making it an effective tool for advanced language learners.

3. Methodology

This section outlines the experimental platform used to address the following research questions:

1. What is the effectiveness of ChatGPT in improving students' writing abilities as measured by scores on a standardized writing assessment?
2. How do teachers and students perceive and experience using ChatGPT as a writing tool?

To answer these research questions, a mixed-methods research design was employed, incorporating both quantitative and qualitative data collection and analysis methods. This research aims to contribute to the growing body of knowledge on the use of NLP tools in education by providing empirical evidence on the effectiveness of ChatGPT in enhancing writing proficiency. Additionally, it will offer insights into students' perceptions of using AI-based tools for writing instruction and highlight the ethical considerations that need to be addressed to ensure the responsible use of such technologies in educational settings. This study will employ a mixed-methods approach to investigate the effectiveness and perceptions of using ChatGPT in educational settings. Quantitative data will be collected through standardized writing assessments to measure improvements in students' writing abilities. Qualitative data will be gathered through surveys and interviews to explore students' experiences and perceptions of using ChatGPT as a writing tool.

3.1. Experimental study

The experimental study was structured into three phases, as illustrated in Fig. 1.

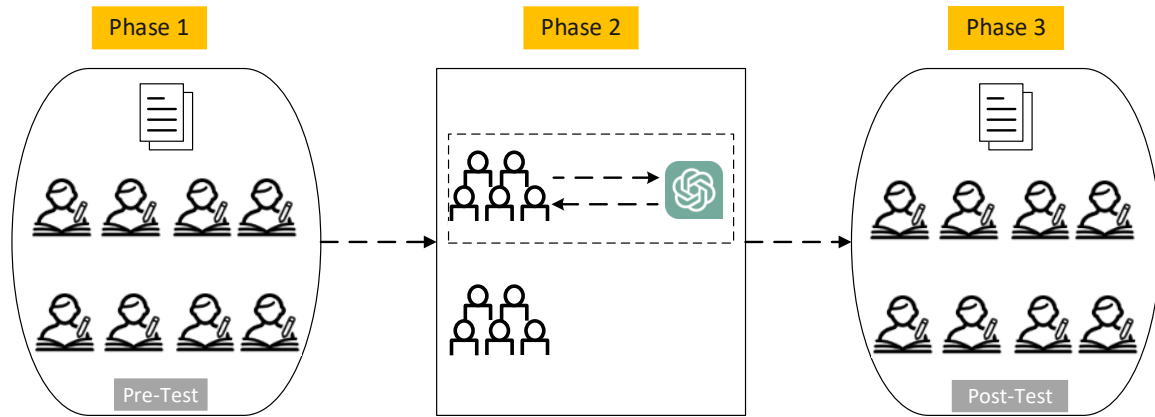


Fig. 1: Three phases of the experimental study

- Phase 1: Pre-test assessment: Initially, students were given a pre-test to gauge their writing abilities. This involved responding to a writing prompt, which was evaluated using a standardized scoring rubric assessing various aspects of writing quality, including grammar, punctuation, sentence structure, and overall coherence.
- Phase 2: Intervention with ChatGPT: Following the pre-test, students were randomly assigned to either a control group or an experimental group. The experimental group received access to ChatGPT to assist in improving their writing. Detailed instructions on effectively using ChatGPT, including formatting advice and strategies for incorporating AI suggestions into their revisions, were provided.
- Phase 3: Post-test evaluation: After a set period for revision using ChatGPT, both groups participated in a post-test that mirrored the pre-test in format and evaluation criteria. The students' performances on the pre and post-tests were compared to determine ChatGPT's effectiveness in enhancing their writing skills.

3.2. Questionnaires and interviews

Questionnaires were administered to students from various levels and departments of NBU during the first semester of 2023-2024 to gather data on using ChatGPT in writing instruction. These questionnaires included both structured and open-ended questions and were analyzed quantitatively and qualitatively. Additionally, interviews and focus groups with students in the experimental group provided further insights into their experiences and perceptions. Trained research assistants ensured consistency in data collection, and all conversations were recorded and transcribed.

3.3. Data analysis and validation

The integration of both quantitative and qualitative data provided a comprehensive framework for analyzing writing scores and understanding the nuanced patterns in student experiences. By combining these two data types, the study was able to not only assess measurable

outcomes, such as writing scores but also capture the subjective and experiential dimensions of students' writing processes, which may not be fully reflected in numerical scores alone. This mixed-methods approach enabled a richer, more in-depth analysis, offering insights into how different variables, such as writing strategies, feedback reception, and individual challenges, interact to affect overall performance.

- Quantitative data analysis: The quantitative aspect primarily focused on the statistical evaluation of writing scores. Standardized tests, rubrics, and scoring mechanisms were used to measure students' writing proficiency across various dimensions, such as grammar, structure, coherence, and argumentation. Descriptive statistics provided a snapshot of performance, while inferential statistical methods, such as t-tests, were applied to determine the significance of observed differences between groups, such as those who received AI-assisted feedback versus traditional human feedback.

b. Scoring rubrics and sample characteristics:

- Scoring rubrics: The scoring rubric applied in this study assessed writing proficiency based on five key dimensions:
 - Grammar: Evaluates the accurate use of language mechanics, including tense, agreement, and syntax. Score 1-5 Scale, ranges from "Frequent grammatical errors that impede meaning" (1) to "Nearly flawless grammatical structure with minimal errors" (5).
 - Punctuation: Assesses proper use of punctuation marks to aid clarity. Score 1-5 Scale, from "Frequent punctuation errors that disrupt flow" (1) to "Consistent, precise punctuation enhancing readability" (5).
 - Sentence structure: Measures sentence variety, clarity, and complexity. Score 1-5 Scale, spans from "Choppy or overly simple sentence structure" (1) to "Complex, well-structured sentences enhancing expression" (5).
 - Coherence: Examines logical flow and transitions between ideas. Score 1-5 Scale begins with "Disjointed flow, making text difficult to follow" (1)

and progresses to “Seamlessly cohesive structure throughout” (5).

- **Argumentation:** Looks at the strength of argument, supporting details, and overall persuasiveness. Score 1-5 Scale, ranges from “Weak argument with minimal support” (1) to “Compelling argument supported by strong, relevant evidence” (5).

The overall score for each student’s writing was computed as the average of the five dimensions, with a higher score indicating a greater proficiency level. This rubric was chosen for its ability to capture distinct aspects of writing quality, providing a well-rounded measure of improvement.

- **Sample characteristics:** The sample consisted of 50 students enrolled at Northern Border University, Saudi Arabia, divided into a control group (n=29) and an experimental group (n=21).

- **Age range:** Participants were aged 18-24, with an average age of approximately 20.5 years.
- **Language proficiency:** Students had intermediate-level English proficiency, determined through previous standardized test scores (e.g., TOEFL, IELTS), ensuring a baseline competency suitable for receiving and interpreting writing feedback.
- **Academic background:** All participants were undergraduate students, predominantly from Computer Science and Languages/Translation departments. This demographic was chosen to represent both technical and humanities disciplines, giving insights into ChatGPT’s efficacy across different academic contexts.
- **Randomization and matching:** Students were randomly assigned to control or experimental groups, ensuring comparable characteristics between groups. Efforts were made to balance both groups regarding language proficiency and prior academic exposure to English writing courses.

c. **Qualitative data analysis:** The qualitative data, collected through open-ended survey responses, interviews, and focus group discussions, aimed to uncover students’ subjective experiences and perceptions regarding the writing process. This data was essential for understanding the contextual factors influencing performance, such as motivation, perceived difficulty, and the role of feedback in shaping their writing development. Thematic analysis was employed to identify recurring themes and patterns in the qualitative responses, allowing for the categorization of common experiences and challenges.

3.4. Data validation

To ensure the accuracy and reliability of the collected data, a rigorous validation process was conducted, consisting of several key stages:

- **Expert review of questionnaires:** The questionnaires and interview protocols were

reviewed by subject matter experts to ensure that the questions were clear, relevant, and capable of capturing the intended information. This review helped refine the instruments to better align with the research objectives, ensuring that the data gathered would be both reliable and valid.

- **Categorization of open-ended responses:** The qualitative data from open-ended responses was systematically categorized using a coding framework. This framework was developed iteratively, with initial codes being refined through multiple rounds of coding to ensure that they accurately reflected the participants’ experiences. This step was crucial in reducing researcher bias and increasing the reliability of the qualitative analysis.
- **Three-stage data analysis process:**
 - **Stage 1: Preliminary data exploration**–In this initial stage, basic descriptive statistics and exploratory coding of qualitative data were conducted to identify broad trends and themes.
 - **Stage 2: Detailed statistical and thematic analysis**–The quantitative data underwent rigorous statistical testing to examine correlations, trends, and significant differences. Concurrently, the qualitative data was analyzed in more depth using thematic coding, enabling the identification of deeper patterns related to student experiences and feedback reception.
 - **Stage 3: Triangulation and integration**–In the final stage, the findings from both quantitative and qualitative analyses were integrated through triangulation. This process involved cross-verifying the results from different data sources to ensure consistency and reliability. Any discrepancies between the two datasets were examined closely to provide a holistic interpretation of the data.

Overall, this robust data analysis and validation framework ensured that the study’s findings were both comprehensive and reliable, allowing for a deeper understanding of how different factors contribute to student writing performance and experiences.

4. Results and discussion

This section presents an analysis of the data collected to evaluate the effectiveness of ChatGPT in enhancing students’ writing abilities, as well as their perceptions of using the tool. Quantitative and qualitative findings are discussed to highlight the impact of AI-assisted feedback on writing performance and student experiences.

4.1. Research question 1: Effectiveness of ChatGPT in improving writing abilities

To assess the effectiveness of ChatGPT in enhancing students’ writing abilities, a series of pre and post-tests were administered to both the experimental and control groups. These assessments

involved a standardized writing prompt, which was evaluated using a rubric focusing on key writing components such as grammar, punctuation, sentence structure, and overall coherence. The results of these

tests are illustrated in Tables 1, 2, and 3, which detail the scores of both groups in the pre-test and post-test phases.

Table 1: Experimental group (Wilcoxon signed ranks test)

	Count	Median	Mean	Standard deviation	P-value
Pre-test score	21	3.500	3.02	0.89	0.000
Post-test score	21	4.000	3.95	0.84	

Table 2: Control group (Wilcoxon signed ranks test)

	Count	Median	Mean	Standard deviation	P-value
Pre-test score	29	3.00	2.95	0.89	0.002
Post-test score	29	3.50	3.48	0.65	

Table 3: Percentage increase in test Scores from pre-test to post-test (Mann-Whitney test)

Group	Count	Median	Mean	Standard deviation	P-value
Experimental	21	25.00	39.76	43.12	0.221
Control	29	14.29	27.27	40.31	

The results presented in Tables 1, 2, and 3 indicate significant improvements in writing scores for both the experimental and control groups between the pre-test and post-test phases. Based on the Wilcoxon Signed Ranks Test, the experimental group demonstrated a highly significant improvement (P-value=0.000), while the control group also showed a statistically significant increase (P-value=0.002). These results suggest that both interventions—whether traditional instruction for the control group or AI-assisted feedback for the experimental group—contributed positively to students' writing development. However, a comparison between the two groups reveals that the experimental group, which received feedback from ChatGPT, exhibited a greater mean improvement in writing scores, with the mean increasing from 3.02 (pre-test) to 3.95 (post-test). In contrast, the control group saw a mean increase from 2.95 to 3.48. This finding highlight that the use of ChatGPT as a writing tool has a more substantial effect on improving students' writing abilities, particularly in areas like sentence structure, grammar, and coherence.

The analysis of percentage increases in writing scores (Table 3) further reinforces this finding. The experimental group showed a higher median and mean percent increase in scores (39.76%) compared to the control group (27.27%). Although this suggests that ChatGPT has a more pronounced impact on writing performance, the P-value of 0.221, based on the Mann-Whitney Test, indicates that the difference in percent increase between the two groups is not statistically significant. This suggests that, while there is a tendency for greater improvement in the experimental group, the variability in performance—evidenced by the relatively high standard deviations (43.12 for the experimental group and 40.31 for the control group)—makes it difficult to definitively assert that ChatGPT is significantly more effective in improving writing performance than traditional methods alone.

The data suggests that both interventions were effective in improving writing skills, but the use of ChatGPT had a greater impact on the experimental group, as indicated by the higher mean and percent

increase in test scores. However, the variability in improvement across individual students implies that some participants benefited more than others, possibly due to differences in how they engaged with the feedback provided by the tool.

4.2. Research question 2: Students' perceptions and experiences of using ChatGPT as a writing tool

In addition to evaluating the impact of ChatGPT on writing performance, this study explored students' perceptions and experiences of using the tool as part of their writing process. Data was collected through questionnaires and interviews, with students in the experimental group rating their experiences using a 5-point Likert scale across several dimensions, including ease of use, helpfulness in improving writing, and its ability to assist with overcoming writer's block. Table 4 summarizes the mean and standard deviations of students' responses.

Table 4: The mean and standard deviations of students' perceptions of using ChatGPT

Perception	Mean	SD
Ease of use	4.5	0.7
Helpfulness in improving writing	4.3	0.9
Helpfulness in overcoming writer's block	4.1	1.0
Helpfulness in providing ideas for writing	4.2	0.9

The high mean scores across all categories suggest that students generally found ChatGPT to be a valuable and user-friendly tool. Specifically, the tool's ease of use was rated highly (mean=4.5), reflecting that students encountered minimal difficulties in navigating and using the system. Furthermore, the mean scores related to ChatGPT's helpfulness in improving writing (4.3) and overcoming writer's block (4.1) indicate that students perceived it as effective in addressing key challenges they face during the writing process. Qualitative data from interviews and focus groups further enriched these findings. Several key themes emerged from the analysis (Table 5), offering deeper insights into students' experiences.

Table 5: The themes that emerged from the data analysis

Theme	Description
Easy to use	Most students found ChatGPT easy to use and navigate
Helpful feedback	Students appreciated the instant feedback and suggestions provided by ChatGPT, which helped them identify and correct errors in their writing
Overcoming writer's block	Students reported that ChatGPT helped them overcome writer's block and provided them with ideas for their writing
Limitations	Some students reported finding the tool's suggestions repetitive or irrelevant to their writing

The qualitative data suggests that students particularly valued the immediate feedback on grammar, punctuation, and sentence structure provided by ChatGPT, as it enabled them to quickly identify and correct errors. Additionally, the tool's ability to offer suggestions for overcoming writer's block was noted as a significant advantage, with several students reporting that it helped them generate new ideas when they were struggling to start or continue their writing. However, the analysis also revealed some limitations. A few students observed that ChatGPT's suggestions were sometimes repetitive or irrelevant to their specific writing tasks. While this did not significantly detract from their overall positive experience, it points to areas where AI-based tools like ChatGPT could be refined to offer more varied and context-sensitive feedback. In summary, the findings suggest that students had overwhelmingly positive perceptions of ChatGPT as a writing tool, particularly in terms of its ease of use, helpful feedback, and capacity to assist with writer's block. While some limitations were noted, these were relatively minor and did not significantly affect the overall utility of the tool in improving students' writing experiences.

4.3. Discussion

The evaluation of both the experimental and control groups offers several key insights into the effectiveness of the intervention applied to the experimental group, which involved the use of ChatGPT as a writing tool. While both groups demonstrated improvements in their post-test scores, suggesting that external factors—such as increased familiarity with the test format or general learning over time—might have contributed to the overall enhancement in writing performance, the experimental group outperformed the control group in several important areas. Specifically, the higher median and mean post-test scores in the experimental group suggest that the intervention had a more pronounced effect on learning outcomes, potentially offering a more targeted and impactful form of feedback compared to traditional methods.

One notable finding is the greater median and mean percentage increase in test scores from pre-test to post-test for the experimental group compared to the control group. This result indicates that students who utilized ChatGPT experienced more significant improvements in their writing abilities and at a faster rate than those who did not have access to the AI tool. This suggests that ChatGPT may facilitate more efficient learning by providing immediate, detailed feedback that enables

students to identify and address weaknesses in their writing more rapidly. The real-time corrective suggestions offered by ChatGPT likely contributed to the substantial gains observed in the experimental group, as students could promptly implement changes and refine their work during the writing process.

However, the results also reveal a higher standard deviation in both the post-test scores and the percentage increase in scores for the experimental group. This greater variability suggests that the effectiveness of ChatGPT may not be uniform across all participants. Some students appeared to benefit significantly from the intervention, while others showed more modest improvements. This inconsistency highlights the possibility that the tool may be more effective for certain learning styles or proficiency levels than others. For example, students with a strong foundation in writing might be able to leverage the detailed feedback provided by ChatGPT to make more substantial revisions, whereas students with less-developed writing skills may require additional guidance beyond what the tool can offer. Consequently, while ChatGPT shows promise as a valuable educational tool, these findings indicate that its impact may vary depending on individual learner needs and contexts.

Further research is necessary to explore the factors contributing to this variability within the experimental group. A more detailed investigation could examine how individual differences—such as prior writing proficiency, engagement with the tool, and motivation—affect the extent to which students benefit from AI-assisted feedback. Moreover, expanding the sample size and including a more diverse set of participants would help validate these findings and provide a clearer understanding of how ChatGPT influences writing performance across different demographic and educational backgrounds.

The lower median percentage increase in the control group suggests that the improvements seen in the experimental group were not solely the result of natural learning progression or external factors, but rather, were uniquely linked to the ChatGPT intervention. While the control group exhibited some gains, likely due to general learning and the passage of time, the difference in outcomes between the two groups points to the added value of AI-driven feedback. This finding reinforces the potential of ChatGPT to enhance writing instruction beyond what is typically achieved through traditional pedagogical methods. Additionally, the positive perceptions and experiences reported by students in the experimental group are significant. The favorable

feedback regarding ChatGPT's ease of use, and the helpfulness of its writing assistance underscores the tool's acceptance and usability in an educational context. These aspects likely contributed to the improved writing performance, as students who perceive a tool as easy to use and beneficial may be more inclined to engage with it and apply its feedback effectively. The fact that students found the tool useful for overcoming writer's block and generating new ideas is particularly noteworthy, as it indicates that ChatGPT can serve not only as a corrective mechanism but also as a creative catalyst in the writing process. Despite these promising results, the study highlights the importance of adopting a nuanced approach when integrating AI tools like ChatGPT into educational settings. While the tool clearly offers potential benefits, including improved writing performance and positive user experiences, it is essential to address the variability in effectiveness observed among different students. Tailoring the use of ChatGPT to accommodate individual learning needs may help to ensure that all students can fully benefit from its capabilities. For instance, more personalized interventions, such as offering additional guidance for students who struggle to apply AI feedback independently, may be necessary to enhance the tool's effectiveness for all learners. Similarly, educators might consider combining AI-driven feedback with human input to provide a more balanced and comprehensive support system. While the results of this study demonstrate the potential of ChatGPT to significantly improve writing abilities and enhance student experiences, they also underscore the need for further research to understand the underlying factors driving variability in its effectiveness. As AI tools become more integrated into educational practices, careful consideration must be given to how these technologies can be optimized to meet the diverse needs of students and to maximize their educational value.

5. Challenges and future work

This section addresses the challenges associated with using large language models (LLMs) like ChatGPT in educational settings, along with potential future research directions.

5.1. Challenges

- **Data privacy concerns:** The use of ChatGPT in educational environments raises substantial concerns regarding data privacy. Teachers and students have expressed apprehension about how their data is stored, used, and protected. Ensuring the confidentiality and security of user data remains a significant challenge in integrating AI tools into education (Song and Song, 2023).
- **Repetitive and irrelevant suggestions:** While many students found ChatGPT's feedback beneficial, some reported instances of repetitive or irrelevant suggestions, which can lead to frustration and diminish the tool's overall effectiveness. Addressing this issue is crucial for improving the user experience (Song and Song, 2023; Mahapatra, 2024).
- **Bias and fairness:** Bias is another critical challenge in AI tools like ChatGPT. Biases in the training data can result in unfair or inaccurate feedback, potentially impacting students' learning experiences negatively. It is essential to ensure that AI tools are trained on diverse and representative datasets to mitigate this risk (Gozali et al., 2024; Montenegro-Rueda et al., 2023).
- **Ethical considerations:** The ethical implications of using AI in education, including transparency, accountability, and the potential for over-reliance on technology, are critical concerns. Developing clear guidelines and policies to address these ethical challenges is necessary to ensure the responsible use of AI tools in educational contexts (Link et al., 2022).

5.2. Future work

- **Enhanced personalization:** Future research should focus on improving ChatGPT's personalization capabilities. Tailoring feedback to individual students' learning styles and needs can increase its effectiveness in enhancing writing proficiency and engagement (Song and Song, 2023).
- **Improved feedback mechanisms:** Developing more advanced algorithms to provide varied and relevant feedback will help mitigate the issue of repetitive and irrelevant suggestions. This may involve incorporating sophisticated natural language processing techniques and continuous learning from user interactions (Song and Song, 2023).
- **Integration with curriculum:** Future work should explore ways to integrate ChatGPT with existing educational curricula and assessment methods. Aligning AI-generated feedback with established educational standards and learning objectives will help maintain consistency and relevance in the educational process (Mahapatra, 2024).
- **Longitudinal studies:** Long-term studies are needed to evaluate the sustained impact of ChatGPT on students' writing skills and academic performance. Such research can identify long-term benefits and potential drawbacks of AI-assisted learning tools (Gozali et al., 2024).
- **Addressing ethical and privacy concerns:** Ongoing efforts to address ethical and privacy issues are crucial. Establishing robust privacy frameworks, transparent data usage policies, and ethical guidelines will build trust among users and promote the responsible integration of AI in educational settings (Mahapatra, 2024).

6. Conclusion

This study investigated the effectiveness of ChatGPT in enhancing students' writing abilities within a classroom setting. The findings demonstrate

that ChatGPT significantly improved writing skills, with students particularly valuing its ease of use and the constructive nature of its feedback. However, the integration of NLP tools like ChatGPT in educational environments necessitates careful attention to ethical considerations.

Teachers expressed moderate to high concerns regarding data privacy and potential biases inherent in the model's feedback. To maximize the benefits of ChatGPT, it is recommended to integrate its use with other formative feedback strategies, providing students with a more comprehensive learning experience. Additionally, training and support for educators are essential to address ethical concerns and ensure the responsible deployment of ChatGPT in the classroom.

Future research should focus on the long-term effects of using ChatGPT in educational settings, as well as comparing its effectiveness with other AI-based writing tools. Investigations should also explore the tool's impact across diverse student populations and various educational contexts. While the potential of AI-powered writing evaluation (AWE) systems in ESL classrooms is promising, it is crucial to thoroughly examine the implications for teaching, learning, and assessment. Further exploration is needed into how AWE feedback affects students' writing processes, as well as the use of action research by educators to evaluate AWE systems in specific contexts. This study affirms ChatGPT's efficacy in generating human-like text across various applications, including academic writing, creative content generation, and customer service.

The analysis revealed that while ChatGPT performs competently in producing coherent and contextually relevant outputs, human oversight is often required to ensure accuracy and appropriateness. Experimental results also suggest that ChatGPT can significantly enhance productivity by automating routine tasks and delivering rapid responses.

However, it is essential to recognize its limitations, such as occasional factual inaccuracies and a lack of deep understanding of complex topics. Future research should aim to address these limitations and explore more advanced methods for integrating ChatGPT into workflows that demand high levels of precision and reliability. Overall, ChatGPT represents a promising development in artificial intelligence, with the potential to transform various industries. By addressing its current shortcomings and continuously refining its capabilities, ChatGPT can become an even more valuable tool in the future.

Acknowledgment

The authors gratefully acknowledge the approval and the support of this research study by Grant No. SCIA-2023-12-2277 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

Compliance with ethical standards

Ethical considerations

Informed consent was obtained, data privacy ensured, and AI tool usage monitored for fairness. Risks were minimized, and participants received guidance on using AI feedback.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Alkhalidi AA (2023). The impact of technology on students' creative writing: A case study in Jordan. *Theory and Practice in Language Studies*, 13(3): 586-592. <https://doi.org/10.17507/tpls.1303.06>
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, and Agarwal S et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33: 1877-1901.
- Chen X, Fang Y, Liu X, Zhao D, Feng X, and Li P (2019). Gender differences in latent classes of sleep quality in community-dwelling adults based on the Pittsburgh sleep quality index. *Psychology, Health and Medicine*, 24(8): 901-910. <https://doi.org/10.1080/13548506.2019.1634825> **PMid:31230460**
- Gaud D (2023). Ethical considerations for the use of AI language model. *International Journal for Research in Applied Science and Engineering Technology*, 11(7): 6-14. <https://doi.org/10.22214/ijraset.2023.54513>
- Geng J and Razali AB (2022). Effectiveness of the automated writing evaluation program on improving undergraduates' writing performance. *English Language Teaching*, 15(7): 49-60. <https://doi.org/10.5539/elt.v15n7p49>
- Gozali I, Wijaya ART, Lie A, Cahyono BY, and Suryati N (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *The JALT CALL Journal*, 20(1): 1-22. <https://doi.org/10.29140/jaltcall.v20n1.1200>
- Link S, Mehrzad M, and Rahimi M (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4): 605-634. <https://doi.org/10.1080/09588221.2020.1743323>
- Ma S, Chen Q, Wang X, Zheng C, Peng Z, Yin M, and Ma X (2024). Towards human-AI deliberation: Design and evaluation of LLM-empowered deliberative AI for AI-assisted decision-making. *Arxiv Preprint Arxiv:2403.16812*. <https://doi.org/10.48550/arXiv.2403.16812>
- Mahapatra S (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11: 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mayer C, Wallner S, Budde-Spengler N, Braunert S, Arndt PA, and Kiefer M (2020). Literacy training of kindergarten children with pencil, keyboard or tablet stylus: The influence of the writing tool on reading and writing performance at the letter and word level. *Frontiers in Psychology*, 10: 3054. <https://doi.org/10.3389/fpsyg.2019.03054> **PMid:32038412 PMCID:PMC6987467**
- Mohsen MA (2022). Computer-mediated corrective feedback to improve L2 writing skills: A meta-analysis. *Journal of*

- Educational Computing Research, 60(5): 1253-1276.
<https://doi.org/10.1177/07356331211064066>
- Montenegro-Rueda M, Fernández-Cerero J, Fernández-Batanero JM, and López-Meneses E (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, 12(8): 153.
<https://doi.org/10.3390/computers12080153>
- Perkins M (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2): 07. <https://doi.org/10.53761/1.20.02.07>
- Radford A, Wu J, Child R, Luan D, Amodei D, and Sutskever I (2019). Language models are unsupervised multitask learners. OpenAI Blog. Available online at:
<https://d4mucfpksyv.cloudfront.net/better-language-models/language-models.pdf>
- Song C and Song Y (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14: 1260843.
<https://doi.org/10.3389/fpsyg.2023.1260843>
PMid:38162975 PMCID:PMC10754989
- Tarchi C, Zappoli A, Casado Ledesma L, and Brante EW (2024). The Use of ChatGPT in source-based writing tasks. *International Journal of Artificial Intelligence in Education*.
<https://doi.org/10.1007/s40593-024-00413-1>
- Zhong Y, Lian J, Huang H, and Deng H (2025). Affordances, constraints, and implications of ChatGPT in education from a social-ecological perspective: A data mining approach. *Education and Information Technologies*.
<https://doi.org/10.1007/s10639-024-13237-2>
- Zhuo TY, Huang Y, Chen C, and Xing Z (2023). Exploring AI ethics of ChatGPT: A diagnostic analysis. *Arxiv Preprint Arxiv:2301.12867*.
<https://doi.org/10.48550/arXiv.2301.12867>