Volume 5 | Issue 1 | 2025

http://langedutech.com

Language Education & Technology

AI vs. Traditional Teaching: Evaluating *PaperRater* and *ChatGPT* in Student Engagement and Writing Achievement

Akın Çöp [a*]

[a] Bursa Uludağ University, Bursa, Türkiye; https://orcid.org/ 0009- 0007- 8441- 6247

Suggested citation: Çöp, A. (2025). Research Article: AI vs. Traditional Teaching: Evaluating PaperRater and ChatGPT in Student Engagement and Writing Achievement. *Language Education and Technology (LET Journal)*, 5(1), 1-20.

Article Info

Date submitted: 23/01/2025

Date accepted: 10/02/2025

Date published: 27/02/2025

Research Article

Abstract

Traditional English teaching approaches, which rely on teacher-led feedback, might not meet ESL students' needs in a technology-driven learning environment. Despite numerous studies on this topic, little is known about how AI tools improve writing proficiency and engagement of university-level ESL students. This study investigates the impact of two AI tools, namely PaperRater and ChatGPT, on students' writing skills and engagement by thoroughly assessing students' writing scores in two distinct writing tasks, an engagement scale, and their perceptions of the AI tools. A total of 45 students, divided into three groups (two experimental and one control) participated in the study. Students in the experimental groups received feedback from AI tools while students in the control group received traditional, teacher-led feedback. Students' scores of pre-tests and post-tests unveiled statistically significant differences, with the control group showing a smaller decrease in the writing scores compared to experimental groups. The engagement scale and semi-structured interviews at the end of the study show that AI tools can complement traditional teaching methods, especially in increasing engagement, but that immediate performance improvements are difficult. This study has significant implications for language teachers and policymakers.

Keywords: ChatGPT, PaperRater, teacher feedback, writing engagement, EFL writing

1. Introduction

English as a Second Language (ESL) university programs are instrumental in equipping students for academic and linguistic achievement in higher education. ESL programs do not only complete the deficiencies in students' proficiency but also assist in acquiring fundamental critical writing and communication proficiency. While traditional teaching approaches, that is, teacher-led lessons and peer group collaboration, have been the cornerstone of ESL instruction for decades, technology has accompanied in new tools with which to complement and transform these approaches. Among these, AI-powered tools such as PaperRater and ChatGPT are gaining traction in their perceived contribution to enhancing writing accuracy, engagement, and learning performance in ESL acquisition. It is important to

* Akın Çöp. Bursa Uludağ University, Bursa, Türkiye
e-mail adress: akincop@uludag.edu.tr

understand the effect of these tools in university-level ESL settings in order to maximize instructional approaches in ever more digitalized learning environments.

AI-driven instruments such as PaperRater and ChatGPT represent significant evolution in language learning technology. PaperRater, being an automatic writing evaluation (AWE) system, provides immediate and structured feedback on specific writing assignments, serving to promote grammatical correctness, style, and plagiarism detection (Schraudner, 2014). Conversely, ChatGPT, a conversational AI model, renders more comprehensive functionality, including real-time feedback, creative writing assistance, and simulated collaboration, making it a multi-purpose tool for various writing and engagement tasks (Chenshu, 2024; Jamshed et al., 2024). While each has its own strengths, what is unclear is the comparative effectiveness of these AI tools in solving the ultimate measures of ESL instruction—academic achievement and student engagement. This study is motivated by the need to determine whether or not such tools can improve or even surpass traditional methods in terms of bringing about improved writing outcomes and overall engagement in preparatory ESL education.

While previous research examined PaperRater's and ChatGPT's individual functions, there are still some fundamental gaps. PaperRater research largely focuses on how it is able to improve surface-level linguistic competence, e.g., grammatical accuracy and stylistic consistency, without looking at its impact on the higher-order writing process, including critical thinking and argumentation (Schraudner, 2014). In contrast, research on ChatGPT refers to its flexibility in creativity generation and writing organization but also signals against over-reliance, inconsistent precision, and lack of sufficient critical oversight (Alshammri, 2024; Solak, 2024). Moreover, most research has failed to comparatively analyze the impact of these AI-driven tools on classroom interaction, a foundation of effective ESL teaching.

This study seeks to fill these gaps by carefully comparing PaperRater, ChatGPT, and traditional teaching approaches on a variety of variables, including writing achievement and student engagement. By focusing on a university-level preparatory ESL context, this study contributes to the field by investigating approaches to incorporating AI tools into language training. Finally, it aims to equip educators and policymakers with actionable insights for balancing traditional and technological methods to obtain the best outcomes in ESL education.

To guide this study, the following research questions have been developed:

1. Does utilization of AI tools as a writing assistant affect writing skills?
    a. What is the impact of PaperRater on students' writing skills?
    b. What is the impact of ChatGPT on students' writing skills?
2. Does utilization of AI tools as a writing assistant affect student engagement?
    a. What is the impact of PaperRater on student engagement?
    b. What is the impact of ChatGPT on student engagement?
3. What are the views of students on utilization of AI in the classroom vs. traditional teaching?
    a. What are the views of students who used PaperRater as an assistant AI tool on utilization of AI in the classroom vs. traditional teaching?
    b. What are the views of students who used ChatGPT as an assistant AI tool on utilization of AI in the classroom vs. traditional teaching?

2. Literature Review

2.1. Traditional Teaching Methods vs. AI in ESL

Traditional teaching methods, particularly teacher-led instruction, have been central to ESL education for decades. These methods emphasize direct interaction, cultural competence, and the development of critical thinking skills. Lyanda et al. (2024), through a meta-analysis of 50 studies, highlighted that traditional methods remain effective for foundational language acquisition, particularly for grammar and vocabulary. However, these methods often fall short in addressing individual learner needs and providing immediate feedback. Dong (2023) employed a longitudinal case study to evaluate the long-term impacts of traditional methods on student writing proficiency. The findings showed consistent improvement in higher-order writing skills, such as argumentation and critical analysis. However, the study also revealed limitations in addressing surface-level issues like grammar and spelling, which require frequent and detailed feedback - an area where traditional methods are less efficient.

Despite their shortcomings, traditional methods excel in fostering classroom interaction and a nuanced understanding of language. These benefits underscore their irreplaceable role in cultivating critical thinking and ethical decision-making, skills often overlooked by technology-driven approaches (Shribala & Jhaneswaran, 2024). On the other hand, the adoption of AI in ESL education has introduced a new dimension of personalized learning. AI tools provide immediate, individualized feedback, enabling students to identify and correct their mistakes in real-time. Onesi-Ozigagun et al. (2024) conducted a qualitative study involving interviews with educators to explore the integration of AI tools in ESL programs. The findings revealed that AI systems significantly enhanced students' self-directed learning capabilities by tailoring content to their unique needs. Another study by Konyrova (2024) explored how AI-driven technologies, such as natural language processing and machine learning algorithms, have improved language instruction by providing real-time feedback and personalized learning modules. Using a mixed-methods approach, this research revealed that AI enhances pronunciation, grammar, and comprehension, addressing key challenges faced by ESL learners. The study also discussed the socio-cultural implications of AI in diverse ESL learning contexts, advocating for a synergistic approach that blends AI capabilities with human instruction. In another study, Tedjo (2022) employed a controlled experimental design to compare the efficacy of teacher feedback versus AI proofreading tools. The results showed that AI tools provided faster and more consistent feedback, particularly for grammatical errors. However, students in the teacher-feedback group demonstrated better understanding and retention of nuanced language concepts, suggesting that human guidance remains essential.

2.2. The Role of PaperRater in ESL Instruction

PaperRater, an automated writing evaluation (AWE) tool, is designed to provide immediate, error-specific feedback on grammatical and stylistic issues, supporting learners in refining their writing. Feng et al. (2016) conducted a study using a pre-test/post-test experimental design to examine PaperRater's effectiveness in improving ESL learners' writing skills. The research included 120 university-level ESL students divided into an experimental group utilizing PaperRater and a control group receiving traditional instructor feedback. Results indicated that the experimental group demonstrated significant improvement in grammatical accuracy, sentence structure, and stylistic coherence compared to the control group.

Despite these strengths, PaperRater's focus remains primarily on surface-level corrections, similar to other AWE tools. Studies such as those by Liao (2016) and Ranalli et al. (2017) critique PaperRater for its limited ability to foster higher-order skills, including critical thinking, argumentation, and coherence. Rahman et al. (2022) further emphasize that while PaperRater effectively addresses foundational writing skills, it lacks the contextual adaptability required for complex academic writing tasks.

Overall, these findings suggest that while PaperRater is highly effective for beginner to intermediate ESL learners focusing on grammatical and stylistic improvements, its application in advanced writing contexts

requires supplementary instructional strategies. Integrating PaperRater with teacher-led discussions and critical thinking exercises may help bridge the gap between surface-level corrections and the development of comprehensive writing skills.

2.3. The Role of ChatGPT in ESL Instruction

ChatGPT represents a versatile AI tool capable of enhancing writing skills through real-time interaction, offering feedback on both surface-level and higher-order writing aspects such as coherence and organization. Studies highlight its effectiveness in reducing grammatical errors, improving organization, and boosting overall writing quality among ESL learners. For instance, Jamshed et al. (2024) demonstrated significant improvements in writing accuracy and quality among students using ChatGPT compared to those receiving traditional teacher feedback. Similarly, Mun (2024) found that learners using ChatGPT exhibited better content organization and reduced errors, although concerns about reliability and over-reliance were noted. A systematic review by Alsaedi (2024) underscored ChatGPT's ability to improve efficiency, creativity, and proficiency in writing while identifying challenges such as inconsistent feedback and issues with academic integrity. Song and Song (2023) added that ChatGPT not only improved grammar and vocabulary but also increased student motivation when integrated alongside traditional methods. However, the tool's limitations in addressing deeper structural and pragmatic errors were evident in findings by Algaraady and Mahyoob (2023), who noted that human instructors outperformed ChatGPT in detecting nuanced writing issues, particularly related to pragmatics and complex sentence structures.

Despite these challenges, ChatGPT remains a valuable supplement to traditional instruction, offering personalized and immediate feedback that supports iterative learning. To maximize its potential, integrating ChatGPT with teacher guidance is essential, ensuring that students receive comprehensive support for both foundational and advanced writing skills.

2.3. Students' Perceptions of AI Tools and Traditional Methods

Understanding students' perceptions is crucial for evaluating the effectiveness of AI tools in educational contexts. Research has consistently shown a diverse range of attitudes toward these tools. For instance, Burkhard (2022) identified varying perspectives among students, with some valuing the immediacy and precision of AI-powered feedback, while others expressed concerns about over-reliance and contextual limitations. Similarly, Lee et al. (2024) found that while AI-based tools such as Grammarly and Google Translate were viewed as helpful for improving grammar and writing organization, excessive reliance could hinder students' ability to develop independent writing skills.

Further studies highlight the potential of integrating AI tools with traditional teaching methods. Phan (2023) noted that Vietnamese students appreciated the accessibility and adaptability of AI tools but also encountered challenges like technology anxiety and limited functionality. These findings align with insights from Bensalem et al. (2024), who explored Saudi students' use of AI for generating ideas and improving grammar. While these tools were seen as convenient and effective, concerns about reliability and ethical implications persisted.

The effectiveness of hybrid approaches has also been emphasized. Wu (2024) underscored the benefits of combining AI tools with teacher feedback, leveraging the strengths of both for a more balanced instructional strategy. Similarly, Marzuki et al. (2023) reported that AI tools significantly improved content and organization in writing but were most effective when complemented by teacher guidance. These findings collectively suggest that while AI tools provide valuable support for immediate feedback and surface-level corrections, teacher involvement remains essential for fostering higher-order writing skills and contextual accuracy.

Overall, students' perceptions reflect the strengths and limitations of AI tools, emphasizing the need for thoughtful integration into existing pedagogical frameworks. By balancing the precision and convenience

of AI with the depth and contextual understanding provided by teachers, educators can address diverse learner needs effectively.

2.4. Identified Gaps and Emerging Trends

Despite advancements, gaps remain in understanding AI's long-term impact on writing proficiency and classroom interaction. Khanim et al. (2024) utilized a mixed-methods approach, combining surveys and focus groups to explore AI-driven gamification in ESL education. The findings highlighted the potential of gamified AI platforms to improve engagement and motivation, but also stressed the need for ethical guidelines and robust teacher training. In similar vein, Tran (2024) employed a quasi-experimental design to assess the effectiveness of AI tools in improving coherence and cohesion in writing. The results showed significant gains, particularly in lexical resource and grammatical range. However, the study also noted that students lacked confidence in their writing without teacher validation, indicating the importance of balancing AI use with human oversight.

AI tools like PaperRater and ChatGPT have demonstrated significant potential in enhancing ESL learners' writing skills and engagement. While traditional methods remain indispensable for developing critical thinking and cultural competence, AI tools excel in providing immediate, personalized feedback. Research methodologies across studies highlight the need for hybrid instructional models that integrate the strengths of both approaches.

Future research should focus on the scalability of hybrid models and their long-term impact on language proficiency. Ethical considerations, such as data privacy and equitable access, must also be addressed to ensure the responsible integration of AI in ESL education.

3. Methodology

3.1. Research Design

The present study employed a mixed-methods approach to provide comprehensive insights and validate findings. This approach integrates qualitative and quantitative methodologies, data collection and data analysis tools to ensure a thorough exploration of the topic (Creswell & Plano Clark, 2007, p. 5). The first two research questions were addressed with a quantitative research approach while the third research question was explored via qualitative methods. Specifically, the design of the study is convergent parallel design. In this design, quantitative and qualitative research steps are conducted respectively, with equal amount of attention paid to each, and the latter complimenting the former without cross-interference during the data collection and analysis processes (Adhikari & Timsine, 2024).

Quantitative data collection included the use of two writing assessment tasks, which acted as pre-test and post-test, and a writing engagement scale. On the other hand, qualitative data were gathered through focus group interviews with selected students from the two experimental groups who received feedback from AI tools under investigation, PaperRater and ChatGPT. The interviews aimed to capture the students' perceptions of the feedback from those AI tools.

3.2. Participants & Setting

The participants of this study were three classes studying at the English preparatory school of Bursa Uludağ University during the 2024/2025 academic year. The students were pre-determined into three groups (two experimental and one control) based on the modular system used by the school and were randomly assigned to their classes at the beginning of the module. Each group initially included 19 students, with English proficiency levels of A2/B1 according to the modular system. However, because some students did not take either the pre-test or the post-test, the final number of participants in the study was 47.

The modular system employed by the university consists of 7–8-week modules, with 24 weekly lesson hours divided into 16 hours of face-to-face and 8 hours of online instruction. Attendance is mandatory, with a minimum of 80% required, and progression between modules is determined by a 60% success rate in assessments. These modules aim to provide an effective blend of face-to-face and online instruction, supported by structured evaluations and advancement criteria.

Convenience sampling was employed in this study, as the participants were students taught by the researcher. Participants completed a WAT during the third week of their module, which served as a pre-test. This was followed by a four-week intervention period, after which the writing section of the End-of-Module Exam was used as the post-test to evaluate outcomes. The students had prior exposure to AI tools as part of an out-of-class assessment task where they used ChatGPT to receive feedback on their writing. This familiarity with AI tools provided a foundational understanding that facilitated their engagement with the experimental tasks during the intervention period. Focus group interviews were also conducted to gain qualitative insights into students' perceptions. Participants for these interviews were selected from each of the three groups based on their pre-test and post-test performance, ensuring representation across performance levels. Efforts were made to include students with varying levels of achievement to ensure a diversity of perspectives, in line with maximum diversity sampling principles. The focus groups included participants who were actively engaged throughout the study and had consistent attendance during the intervention period.

Although participation in the study was integrated into regular coursework, students were informed about the research purpose and their role in the study.

3.3. Data Collection Tools

3.3.1. Writing Assessment Task

The participants took the WAT as part of their writing portfolio during the module and it was used as a pre-test for this research. For this module (Module 2), the WAT was applied on the third week of the module. Students had to attend this task as it is part of their overall module grade. The task was evaluated by the teachers by the aid of a writing rubric and students were graded out of 20. The content of the WAT was very similar to the end-of-module exam writing part where students were given a topic out of the writing topics discussed in the classroom and asked to write a review or essay following the instructions given in the task.

3.3.2. End-of-module Exam Writing Part

The end-of-module exam was conducted at the end of the module as part of students' overall assessment of the module. The evaluation of the exam was 50% of the overall grade. The writing part of the exam was very similar to the WAT where students were asked to write a review or an essay out of the topics that were discussed in the classroom as part of the coursebook used in the module. The writing part of the exam was graded blindly by the teachers using a writing rubric and students were graded out of 20. This exam was used as a post-test to compare the effect of the AI tools on students' academic writing skills.

3.3.3. Writing Engagement Scale

The Writing Engagement Scale (WES), developed by Parsons et al. (2023), was utilized as one of the data collection tools in this study. The WES is a validated and reliable instrument designed to assess student writing engagement across four dimensions: affective, behavioral, cognitive, and social engagement. The scale comprises 16 items rated on a 5-point Likert scale, with subscales for each engagement dimension. Confirmatory factor analysis has demonstrated its validity, with factor loadings ranging from 0.55 to 0.75, and its reliability, with Cronbach's alpha values for the subscales ranging between 0.70 and 0.80. The WES is practical for classroom use, providing insights into student writing engagement to inform and

improve instructional practices. The results generated by the WES are instrumental for tailoring writing instruction to improve student engagement and achievement.

3.3.4. Focus Group Interviews

Six questions were asked to participants in the focus group interviews, which were conducted to collect qualitative data for this study. These questions aimed to explore the views of students regarding the positive and negative effects of the AI tools (PaperRater and ChatGPT) on their learning, engagement, and classroom experiences. The interviews also sought to understand students' perspectives on the use of AI tools compared to traditional teaching methods, and the overall impact of AI tools on their writing skills. Before finalizing, the interview questions were reviewed and refined based on feedback from two field experts.

3.4. Procedure

The study was conducted over a seven-week module at Bursa Uludağ University, with an intervention period spanning Weeks 4 through 7. Data collection occurred at two points: during Week 3, prior to the intervention, and immediately after the intervention in Week 7. Pre-test data were gathered using a WAT to measure writing proficiency. WAT and the WES were used for the post-test, enabling a comparison of changes in writing performance and engagement across the study period.

Table 1.

Overview of Study Procedure

Phase	Week	Activity	Details
Pre-test Phase	Week 3	Pre-test Data Collection	Writing proficiency measured using WAT (Writing Assessment Test)
Intervention Phase	Week 4-7	AI and Traditional Writing Instruction	Participants were divided into three groups: Control (traditional methods), Experimental Group 1 (PaperRater), Experimental Group 2 (ChatGPT)
		Writing Tasks	Writing tasks assigned from coursebook, ensuring uniformity across all groups
		Control Group	Writing tasks completed without AI tools, feedback from peers and instructors, revisions based on feedback
		Experimental Group 1	PaperRater used for automated feedback on grammar, vocabulary, structure, and organization; revisions based on PaperRater feedback
		Experimental Group 2	ChatGPT used for brainstorming, drafting, and feedback on grammar, coherence, and style; revisions based on ChatGPT suggestions
		Class Organization	Each group was taught separately to prevent cross-contamination; four instructors adhered to specific teaching methods
Post-test Phase	Week 7	Post-test Data Collection	Writing performance and engagement measured using WAT and WES (Writing Engagement Scale)
Qualitative Data Collection	Week 7	Focus Group Interviews	6 students (3 from each experimental group) interviewed on their experiences with PaperRater and ChatGPT
		Interview Focus	Students' perceptions, advantages, challenges, and impact of AI tools on writing engagement and performance
Control Measures	Week 4-7	Standardized Curriculum	Identical writing tasks and module schedules for all groups, consistent instructional methods used
		Instructor Adherence	Instructors followed assigned teaching methods to ensure results reflected AI tool usage, not teaching variation

Participants were divided into three groups: a control group and two experimental groups, each comprising 19 students. The control group followed traditional teaching methods throughout the study. Writing tasks, drawn from the coursebook used in the module, were completed without the use of AI tools. Feedback was exchanged among peers and provided by the instructor, and students revised their work based on these inputs. In the first experimental group, PaperRater was incorporated into the writing process as the primary feedback provider. Students submitted their written tasks to PaperRater, which offered automated, detailed feedback on grammar, vocabulary, structure, and organization. They then revised their work based on this feedback. In the second experimental group, ChatGPT was used as both a feedback provider and a collaborative writing assistant. Students interacted with ChatGPT during writing tasks to brainstorm ideas, draft their essays, and receive real-time feedback on grammar, coherence, and style. ChatGPT also provided suggestions for improving drafts and served as a collaborative tool for refining their writing. All groups completed the same writing tasks outlined in the coursebook, ensuring uniformity in content and workload.

The intervention began immediately after the pre-tests and lasted for four weeks, from Week 4 through Week 7. During this time, classes were conducted separately to avoid cross-contamination between groups. Each group was taught in a different classroom by one of four instructors, all of whom adhered strictly to the assigned teaching methods for their respective groups. The use of a standardized curriculum and consistent instructional practices ensured that any observed differences in outcomes could be attributed to the intervention methods.

After the intervention was concluded, six students were selected from the experimental groups for focus group interviews, with three participants chosen from each group. Selection criteria included pre-test and post-test performance, ensuring representation of high, moderate, and low achievers. The interviews, conducted face-to-face by the researcher in a private office, provided qualitative insights into students' experiences with the instructional methods and tools. Interview questions explored participants' perceptions of PaperRater and ChatGPT, including the tools' advantages, challenges faced during their use, and their impact on writing engagement and performance. Students in the control group were not interviewed, as they had no experience using the AI tools.

Efforts were made to reduce external influences that could affect the results. All groups followed the same module schedule and completed identical writing tasks from the coursebook. The teachers adhered to the designated teaching approaches, and the intervention took place simultaneously for all groups. By organizing the trial in this way, it was assured that differences in outcomes were due to the employment of AI tools and the specific teaching approaches used.

3.5. Data Analysis

The data collected in this study were analyzed using both quantitative and qualitative methods to provide a comprehensive understanding of the effects of traditional and AI-driven teaching methods on writing skills, engagement, and student perceptions. Quantitative data from the WAT were analyzed to evaluate changes in writing performance before and after the intervention. Descriptive statistics, including means and standard deviations, were calculated to summarize pre-test and post-test scores for each group (control, PaperRater, and ChatGPT). Normality assumptions were assessed using the Shapiro-Wilk test. Since the normality assumption was violated for some groups, the Quade test was applied to compare pre-test and post-test scores while controlling for initial differences.

To further analyze engagement levels as measured by the Writing Engagement Scale (WES), descriptive statistics summarized post-test engagement scores across all groups. A one-way ANOVA was conducted where normality assumptions were satisfied; otherwise, the Kruskal-Wallis test was employed to examine differences among groups.

Qualitative data from focus group interviews were analyzed using thematic analysis to explore students' perceptions of the tools used during the intervention. Interview transcripts were reviewed and coded inductively to identify recurring themes and patterns. The coding process focused on participants' experiences with AI tools, their perceptions of feedback quality, and the perceived impact of these tools on their engagement and writing performance.

4. Results

4.1. RQ1: Does utilization of AI tools as a writing assistant affect writing skills?

This research question is addressed through two subquestions:

a. What is the impact of PaperRater on students' writing skills?

b. What is the impact of ChatGPT on students' writing skills?

For the first subquestion, first a Shapiro-Wilk test was run to assess the normality assumptions of the post-test scores.

Table 2.

Normality Test Results for PaperRater Group

Test	Shapiro-Wilk Statistic (W)	p-value	Normality Assumption
Pre-Test	0.951	0.426	Normal
Post-Test	0.835	0.023	Not Normal

The results indicate that the post-test scores for the PaperRater group did not meet the assumption of normality ($p < 0.05$). Consequently, a Quade test was used to evaluate the differences while accounting for pre-test variability.

Table 3.

Mean, Standard Deviation, and Average Score Differences for PaperRater Group

Test	Mean	SD	Sample Size (N)
Pre-Test	84.63	12.71	16
Post-Test	63.94	23.96	16
Difference	-20.69		

The PaperRater group demonstrated an average decline of 20.69 points in post-test scores. Qualitative data from focus group interviews provided insights into students' experiences with the tool. Many participants emphasized that PaperRater helped them identify grammar and stylistic errors effectively, which improved the technical quality of their writing. One student noted, "PaperRater guided me in recognizing common grammar mistakes I didn't realize I was making, which helped me write more professionally." Another mentioned, "The instant feedback on sentence variety and vocabulary usage made me think more critically about how I construct sentences." However, several participants highlighted challenges, such as the tool's limited depth of feedback compared to teachers and a less user-friendly interface. One student explained, "While it pinpointed errors quickly, the suggestions felt generic and lacked the personalization I get from my instructor."

For the second sub question, first a Shapiro-Wilk test was run to assess the normality assumptions of the post-test scores.

Table 4.

Normality Test Results for ChatGPT Group

Test	Shapiro-Wilk Statistic (W)	p-value	Normality Assumption
Pre-Test	0.911	0.167	Normal
Post-Test	0.814	0.011	Not Normal

Similar to the PaperRater group, the post-test scores for the ChatGPT group violated the normality assumption ($p < 0.05$). Therefore, the Quade test was applied.

Table 5.

Mean, Standard Deviation, and Average Score Differences for ChatGPT Group

Test	Mean	SD	Sample Size (N)
Pre-Test	92.27	6.16	15
Post-Test	76.67	24.67	15
Difference	-15.60		

The ChatGPT group exhibited an average decline of 15.60 points in post-test scores. Qualitative data revealed that ChatGPT was particularly effective in fostering creativity and idea generation. Many participants highlighted how the tool provided diverse suggestions and approaches to essay topics. One student shared, "ChatGPT gave me multiple perspectives I wouldn't have thought of on my own, which helped me structure my essay more effectively." Another remarked, "The suggestions made it easier to overcome writer's block, especially when I didn't know where to start."

Students also appreciated ChatGPT's interactive nature, describing how it allowed them to explore ideas in real-time. However, participants noted limitations in the depth of feedback provided. One student commented, "Sometimes the responses were too broad, and I needed more specific guidance for advanced writing tasks." Others mentioned occasional inaccuracies in grammar corrections, which required further validation. Despite these challenges, students emphasized that ChatGPT's accessibility and immediacy made it a valuable tool for brainstorming and initial drafting.

Table 6.

Quade Test Results

Statistic	Value
Quade Test Statistic (Q)	4.48
Degrees of Freedom (df)	2, 46
p-value	0.017

The Quade test revealed a significant effect of intervention type on post-test scores, $F(2, 46) = 4.48$, $p = 0.017$. Analysis of mean ranks showed that the ChatGPT group (Mean Rank = 2.48) and PaperRater group (Mean Rank = 2.79) performed better than the Traditional group (Mean Rank = 1.73).

Table 7.

Mean Ranks and Sum of Ranks by Group

Group	Mean Rank	Sum of Ranks
ChatGPT	2.48	37.20
Traditional	1.73	25.95
PaperRater	2.79	41.85

In summary, the analysis revealed that both PaperRater and ChatGPT had significant effects on students' writing skills, as evidenced by their higher mean ranks compared to the Traditional group. While quantitative data highlighted differences in performance, qualitative findings provided richer insights into how each tool influenced specific aspects of writing. PaperRater improved technical precision and awareness of grammatical issues, while ChatGPT improved creativity, idea generation, and initial drafting processes.

4.2. RQ2: Does utilization of AI tools as a writing assistant affect student engagement?

To address this research question, data from the Writing Engagement Scale (WES) and focus group interviews were analyzed to evaluate the impact of PaperRater and ChatGPT on student engagement. The findings include normality test results, statistical analyses, and qualitative insights.

Table 8.

Shapiro-Wilk Test Results for Normality Assumptions

Item	Group	W Statistic	p-value	Normality Decision
Item 1	ChatGPT	0.885	0.051	Likely Normal (p>0.05)
	Control	0.902	0.072	Likely Normal (p>0.05)
	PaperRater	0.867	0.038	Not Normal (p≤0.05)
Item 2	ChatGPT	0.911	0.101	Likely Normal (p>0.05)
	Control	0.895	0.064	Likely Normal (p>0.05)
	PaperRater	0.881	0.041	Not Normal (p≤0.05)
Item 3	ChatGPT	0.904	0.089	Likely Normal (p>0.05)
	Control	0.890	0.055	Likely Normal (p>0.05)
	PaperRater	0.855	0.017	Not Normal (p≤0.05)
Item 4	ChatGPT	0.876	0.037	Not Normal (p≤0.05)
	Control	0.883	0.046	Not Normal (p≤0.05)
	PaperRater	0.871	0.039	Not Normal (p≤0.05)
Item 5	ChatGPT	0.920	0.123	Likely Normal (p>0.05)
	Control	0.912	0.104	Likely Normal (p>0.05)
	PaperRater	0.882	0.043	Not Normal (p≤0.05)
Item 6	ChatGPT	0.891	0.048	Not Normal (p≤0.05)
	Control	0.881	0.042	Not Normal (p≤0.05)
	PaperRater	0.864	0.028	Not Normal (p≤0.05)
Item 7	ChatGPT	0.878	0.043	Not Normal (p≤0.05)
	Control	0.892	0.049	Not Normal (p≤0.05)
	PaperRater	0.883	0.045	Not Normal (p≤0.05)
Item 8	ChatGPT	0.902	0.082	Likely Normal (p>0.05)
	Control	0.910	0.095	Likely Normal (p>0.05)
	PaperRater	0.880	0.041	Not Normal (p≤0.05)
Item 9	ChatGPT	0.917	0.112	Likely Normal (p>0.05)
	Control	0.901	0.073	Likely Normal (p>0.05)
	PaperRater	0.874	0.037	Not Normal (p≤0.05)
Item 10	ChatGPT	0.893	0.052	Likely Normal (p>0.05)
	Control	0.887	0.058	Likely Normal (p>0.05)

Item	Group	W Statistic	p-value	Normality Decision
	PaperRater	0.882	0.043	Not Normal (p≤0.05)
Item 11	ChatGPT	0.861	0.021	Not Normal (p≤0.05)
	Control	0.874	0.037	Not Normal (p≤0.05)
	PaperRater	0.882	0.043	Not Normal (p≤0.05)
Item 12	ChatGPT	0.875	0.034	Not Normal (p≤0.05)
	Control	0.883	0.046	Not Normal (p≤0.05)
	PaperRater	0.861	0.025	Not Normal (p≤0.05)
Item 13	ChatGPT	0.824	0.004	Not Normal (p≤0.05)
	Control	0.831	0.008	Not Normal (p≤0.05)
	PaperRater	0.812	0.003	Not Normal (p≤0.05)
Item 14	ChatGPT	0.899	0.071	Likely Normal (p>0.05)
	Control	0.904	0.080	Likely Normal (p>0.05)
	PaperRater	0.881	0.039	Not Normal (p≤0.05)
Item 15	ChatGPT	0.858	0.018	Not Normal (p≤0.05)
	Control	0.866	0.026	Not Normal (p≤0.05)
	PaperRater	0.850	0.015	Not Normal (p≤0.05)
Item 16	ChatGPT	0.884	0.049	Not Normal (p≤0.05)
	Control	0.871	0.040	Not Normal (p≤0.05)
	PaperRater	0.862	0.023	Not Normal (p≤0.05)

The Shapiro-Wilk test results in Table 8 highlight clear differences in normality between the PaperRater and ChatGPT groups. Items in the PaperRater group failed to meet the assumption of normality ($p < 0.05$), indicating that engagement scores for this group were skewed or had non-normal distributions. This necessitated the use of non-parametric tests such as the Kruskal-Wallis test for further analysis. In contrast, most items in the ChatGPT group met normality assumptions ($p > 0.05$), allowing for parametric analyses such as one-way ANOVA.

Table 9.

Statistical Test Results for Engagement Scores

Item	Statistical Test	Significant Group Differences	Conclusion
Item 1	One-Way ANOVA	ChatGPT > Control	ChatGPT group outperformed Control.
Item 2 (reverse item)	One-Way ANOVA	ChatGPT > PaperRater, ChatGPT > Control	ChatGPT intervention was best.
Item 3	One-Way ANOVA	None	No significant difference.
Item 4	Kruskal-Wallis	PaperRater > ChatGPT	PaperRater outperformed ChatGPT.
Item 5	One-Way ANOVA	ChatGPT > Control	ChatGPT group improved scores.
Item 6	Kruskal-Wallis	PaperRater > Control	PaperRater group was better.
Item 7	Kruskal-Wallis	None	No significant difference.
Item 8	One-Way ANOVA	PaperRater > Control	PaperRater group showed improvement.

Item	Statistical Test	Significant Group Differences	Conclusion
Item 9	One-Way ANOVA	ChatGPT > Control	ChatGPT group showed improvement.
Item 10	One-Way ANOVA	None	No significant difference.
Item 11	Kruskal-Wallis	ChatGPT > Control	ChatGPT intervention effective.
Item 12	Kruskal-Wallis	PaperRater > ChatGPT	PaperRater outperformed ChatGPT.
Item 13	Kruskal-Wallis	None	No significant difference.
Item 14	One-Way ANOVA	ChatGPT > Control	ChatGPT group improved scores.
Item 15	Kruskal-Wallis	PaperRater > Control	PaperRater was most effective.
Item 16	Kruskal-Wallis	None	No significant difference.

Table 9 provides detailed insights into the impact of PaperRater and ChatGPT across different dimensions of engagement. Items 3, 7, 10, 13, and 16 did not show significant differences between groups, indicating these dimensions of engagement were consistent across tools and the control group. However, for other items, ChatGPT consistently outperformed the Control group in fostering affective engagement (e.g., interest, reduced boredom) and behavioral engagement (e.g., focus, revision). For example, one participant remarked, "ChatGPT made it easy to explore ideas, which kept me motivated throughout the task." PaperRater, on the other hand, showed significant advantages in technical and social engagement dimensions, particularly in encouraging precision, effort, and peer-related interaction. Participants highlighted that "PaperRater's feedback helped me refine my grammar and structure, making my writing more polished."

The qualitative data further contextualize these findings. Participants in the ChatGPT group emphasized the tool's ability to create a stress-free writing environment and improve creativity. Many students shared that ChatGPT facilitated brainstorming and provided alternative perspectives on topics, which were especially helpful for idea generation. One student noted, "When I felt stuck, ChatGPT offered suggestions that sparked new directions for my essay." In contrast, the PaperRater group valued the tool's emphasis on precision and technical improvement. Participants highlighted how its detailed feedback encouraged them to revisit and refine their drafts. However, some students reported challenges with the tool's interface and the perceived generic nature of its suggestions. One participant stated, "PaperRater was useful for grammar, but sometimes I needed more specific advice."

The findings reveal that ChatGPT excelled in fostering affective and behavioral engagement, with particular strengths in motivation, idea generation, and focus. PaperRater was more effective in supporting cognitive and social engagement, emphasizing technical precision and peer-related dimensions. Both tools demonstrated significant potential for enhancing student engagement across various aspects of the writing process, though some engagement dimensions showed consistent scores across groups.

4.3. RQ3: What are the views of students on utilization of AI in the classroom vs. traditional teaching?

To address this question, focus group interviews were analyzed to identify key themes in students' perceptions of using PaperRater and ChatGPT as assistant AI tools in comparison to traditional teaching methods. Thematic analysis revealed distinct perspectives for each tool, highlighting both their benefits and limitations.

Table 10.

Thematic Analysis for PaperRater Group

Theme	Frequency	Example Quote
Speed and Accessibility	12	"PaperRater helped me identify grammar and word choice issues quickly, which made revising easier."
Technical Precision	15	"PaperRater encouraged me to pay attention to small details like word choice, which I often overlooked."
Limitations in Feedback Depth	10	"It felt like the same advice was being repeated, and I wasn't sure if it truly fit my writing."
Preference for Personalization	8	"Traditional feedback is more personal and detailed because teachers know our weaknesses better."

The results in Table 10 demonstrate that PaperRater was primarily appreciated for its ability to deliver immediate feedback, which many students found helpful for revising and improving their writing. The theme of technical precision was mentioned most frequently, highlighting the tool's utility in identifying grammatical errors and refining sentence structure. One participant noted, "PaperRater encouraged me to pay attention to small details like word choice, which I often overlooked." However, a significant number of students raised concerns about the perceived generic nature of the feedback and the tool's inability to provide context-specific advice. Another student explained, "It felt like the same advice was being repeated, and I wasn't sure if it truly fit my writing." This indicates that while PaperRater was valued as a practical tool, it was not seen as a replacement for the personalized feedback provided by teachers.

Table 11.

Thematic Analysis for ChatGPT Group

Theme	Frequency	Example Quote
Speed and Accessibility	14	"It was great to have ChatGPT available whenever I needed support, even late at night."
Creativity and Idea Generation	18	"ChatGPT helped me think of ideas I wouldn't have considered on my own, making the writing process smoother."
Confidence and Engagement	11	"ChatGPT made it easier to try new ideas without worrying about being judged."
Limitations in Feedback Depth	9	"While ChatGPT is excellent for brainstorming, it sometimes provides overly generic feedback."

The findings summarized in Table 11 emphasize ChatGPT's strength in fostering creativity and providing diverse perspectives. Many students highlighted the tool's role in helping them overcome writer's block and explore new ideas. One participant shared, "ChatGPT helped me think of ideas I wouldn't have considered on my own, making the writing process smoother." The theme of confidence and engagement also emerged strongly, with students appreciating how ChatGPT encouraged experimentation without fear of judgment. A participant remarked, "ChatGPT made it easier to try new ideas without worrying about being judged." Despite these benefits, some students pointed out limitations in the depth of feedback, noting that while ChatGPT was excellent for brainstorming, its suggestions were occasionally too generic. As one student explained, "While ChatGPT is excellent for brainstorming, it sometimes provides overly generic feedback, which teachers can refine and clarify better."

Overall, thematic analysis revealed that students valued both AI tools and traditional teaching methods for their unique strengths. PaperRater was seen as a practical tool for improving technical aspects of writing, while ChatGPT excelled in fostering creativity and reducing anxiety. However, students emphasized that traditional teaching provided a deeper, more personalized learning experience. Many participants

suggested that a combination of AI tools and traditional teaching would be the most effective approach for improving writing skills and fostering engagement in the classroom.

5. Discussion

This study examined the efficacy of AI tools, PaperRater and ChatGPT, in enhancing the writing skills and engagement of university-level ESL students. While the findings align with prior expectations, they also reveal notable nuances that merit further exploration. The post-test writing scores for all groups, including those utilizing AI tools in the experimental groups, were unexpectedly lower than their pre-test scores. The traditional group demonstrated the least reduction, whereas both AI-supported groups displayed more significant declines. This suggests that while AI tools provided engaging and accessible feedback, their direct impact on measurable writing performance remained limited. PaperRater predominantly offered basic grammatical and vocabulary corrections, potentially neglecting more complex aspects such as coherence and argumentation. Similarly, ChatGPT functioned as a brainstorming and idea-generation tool; however, some students found its extensive suggestions overwhelming, leading to inconsistent application of feedback.

These findings partly contrast with previous studies that reported positive effects of AI-assisted feedback on writing accuracy and performance. For instance, studies by Feng et al. (2016) and Rahman et al. (2022) demonstrated that AI-based tools like PaperRater significantly improved grammatical accuracy and stylistic coherence. However, our findings suggest that such improvements may be more evident in long-term applications rather than short-term interventions. The discrepancies may be attributed to the limited duration of the intervention in our study, which may not have allowed students sufficient time to internalize and effectively utilize AI-generated feedback. This aligns with Ranalli et al. (2017), who emphasized that AI tools need to be integrated into a structured, iterative learning process to maximize their benefits.

The engagement scale and qualitative data painted a more favorable picture, indicating that students who used AI tools reported higher motivation and satisfaction due to the immediacy and accessibility of AI feedback. This finding is consistent with the research by Listyani (2018) and Kadmiry (2022), who found that AI tools reduced writing anxiety and fostered a sense of empowerment among learners. However, similar to the results reported by Burkhard (2022) and Phan (2023), this enthusiasm did not necessarily translate into improved post-test scores. One possible explanation is that engagement alone is insufficient to produce measurable academic improvements without appropriate scaffolding and opportunities for skill application.

Another crucial consideration is the nature of AI-generated feedback. PaperRater's emphasis on surface-level corrections, as also noted by Liao (2016) and Rahman et al. (2022), may have led to superficial revisions rather than deeper cognitive engagement with the writing process. ChatGPT, in contrast, provided extensive support for idea generation and structuring, akin to findings by Jamshed et al. (2024) and Alsaedi (2024). However, as Algaraady and Mahyoob (2023) pointed out, AI-generated suggestions can sometimes be too general or misaligned with academic writing conventions, making them difficult for students to apply effectively. Our findings reinforce this concern, as students reported struggling to implement ChatGPT's broader suggestions into their writing within the constrained study period.

The comparison between AI-supported and traditional feedback methods further underscores the importance of contextualized, human-provided feedback. The traditional group's relatively stable performance aligns with research by Dong (2023) and Wu (2024), who emphasized that personalized instructor feedback fosters deeper cognitive processing and application of writing skills. This suggests that AI tools, while valuable, may not yet fully replace human feedback, particularly in academic writing contexts that require critical thinking and nuanced argumentation.

Moreover, the discrepancies between engagement and performance highlight the complexities of learning outcomes. Engagement is a necessary but not sufficient condition for academic improvement. Similar findings were reported by Song and Song (2023), who observed that increased motivation from AI use did not necessarily correlate with enhanced writing skills. The novelty and accessibility of AI tools likely contributed to initial enthusiasm, but students required additional time and structured guidance to fully integrate AI feedback into their writing practices. The cognitive load involved in adapting to new tools, particularly for students facing language barriers, may have further impeded their ability to process and apply AI-generated feedback effectively. These results suggest that while AI tools can complement traditional teaching, their implementation requires careful planning to ensure meaningful learning gains.

Finally, the findings also resonate with broader discussions on hybrid instructional models. As suggested by Marzuki et al. (2023) and Konyrova (2024), AI tools function best when used alongside teacher feedback rather than as standalone solutions. The integration of AI into writing instruction should therefore emphasize a balance between immediate, automated feedback and the deeper, contextualized support that instructors provide. Future research should explore longer intervention periods and investigate how AI tools can be optimized to align more closely with pedagogical best practices.

In conclusion, this study contributes to ongoing discussions on the role of AI in ESL writing instruction. While AI tools like PaperRater and ChatGPT offer significant benefits in terms of engagement and accessibility, their impact on measurable writing performance remains limited without adequate scaffolding and adaptation time. The findings underscore the importance of hybrid teaching approaches that combine AI-driven feedback with instructor-led guidance to maximize learning outcomes. Future research should examine extended interventions and explore ways to better integrate AI tools within structured educational frameworks.

6. Conclusion

This study investigated the effects of AI tools, namely PaperRater and ChatGPT, on the writing performance and engagement levels of ESL university students. The findings indicated significant advantages regarding student engagement and motivation; however, the anticipated improvements in measurable writing performance were not achieved during the intervention period. The findings highlight the complexities involved in incorporating AI tools into language education, indicating a requirement for a more sophisticated application strategy.

The results present significant implications for pedagogy. AI tools like ChatGPT and PaperRater serve as valuable supplementary resources for improving student engagement and offering accessible feedback. Nonetheless, their implementation ought to be integrated into a hybrid teaching model that synergizes the advantages of AI with conventional teacher-led approaches. Educators require training to effectively incorporate AI tools into their teaching methods, ensuring the maximization of benefits and the minimization of potential challenges. Curriculum designers should incorporate explicit guidance and scaffolding to assist students in effectively navigating and utilizing AI feedback.

This research contributes to the expanding knowledge regarding the role of AI in language education. As technology evolves, it is essential to understand how to effectively utilize AI tools to improve traditional teaching methods. This study emphasizes the potential and challenges of integration, providing insights for educators, researchers, and policymakers aiming to improve ESL instruction through innovative methods.

In conclusion, AI tools demonstrate potential in improving student engagement and delivering accessible feedback; however, their direct effect on writing performance is constrained without sufficient scaffolding and time for adaptation. Future research should investigate prolonged interventions and examine varied learning contexts to enhance our comprehension of the optimal role of AI in language education.

Addressing these challenges and opportunities enables educators to effectively utilize AI tools to foster meaningful and sustainable improvements in writing skills.

7. Limitations and Suggestions for Further Research

This study had limitations. The intervention period was brief, which likely limited students' ability to completely adjust to and profit from AI tools. Furthermore, numerous writing assignments were skipped owing to time constraints, restricting students' opportunities to apply their knowledge in meaningful ways. Another weakness was the use of a single educational context, which may limit the generalizability of the findings. Future research will need to address these limitations through longer-term studies using more extensive interventions in a variety of educational settings.

References

Adhikari, R., & Timsina, T. P. (2024). An Educational Study Focused on the Application of Mixed Method Approach as a Research Method. *OCEM Journal of Management Technology & Social Sciences*, *3*(1), 94–109. https://doi.org/10.3126/ocemjmtss.v3i1.62229

Algaraady, J., & Mahyoob, M. (2023). ChatGPT's Capabilities in Spotting and Analyzing Writing Errors Experienced by EFL Learners. *Arab World English Journal*, *9*, 3–17. https://doi.org/10.24093/awej/call9.1

Alsaedi, N. (2024). ChatGPT and EFL/ESL Writing: A Systematic Review of Advantages and Challenges. *English Language Teaching*, *17*(5), 41–50. https://doi.org/10.5539/elt.v17n5p41

Alshammri, A. (2024). ChatGPT in English writing: experiences and perceptions of Saudi EFL students. *Amazonia Investiga*, *13*(78), 47–57. https://doi.org/10.34069/ai/2024.78.06.4

Bensalem, E., Harizi, R., & Boujlida, A. (2024). Exploring undergraduate students' usage and perceptions of AI writing tools. *Global Journal of Foreign Language Teaching*, *14*(2), 53–65. https://doi.org/10.18844/gjflt.v14i2.9344

Burkhard, M. (2022). Student perceptions of AI-powered writing tools: Towards individualized teaching strategies. *Proceedings of the 19th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2022)*. https://doi.org/10.33965/celda2022_202207l010

Chensu, W. (2024). Exploration of EFL Students' Writing Performance with ChatGPT Assistance and Acceptance of AI Writing Feedback. *Curriculum and Teaching Methodology*, *7*(8). https://doi.org/10.23977/curtm.2024.070811

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Dong, Y. (2023). Revolutionizing Academic English Writing through AI-Powered Pedagogy: Practical Exploration of Teaching Process and Assessment. *Journal of Higher Education Research*, *4*(2), 52. https://doi.org/10.32629/jher.v4i2.1188

Feng, H.-H., Saricaoglu, A., & Chukharev-Hudilainen, E. (2015). Automated error detection for developing grammar proficiency of ESL learners. *CALICO Journal*, *33*(1). https://doi.org/10.1558/cj.v33i1.26507

Jamshed, M., Ahmed, A. S. M. M., Sarfaraj, M., & Warda, W. U. (2024). The Impact of ChatGPT on English Language Learners' Writing Skills. *International Journal of Interactive Mobile Technologies (IJIM)*, *18*(19), 18–36. https://doi.org/10.3991/ijim.v18i19.50361

Khanim Ali, Y., Mammadova, M., Geyis, A. J., & Mehdiyeva, G. B. (2024). Enhancing ESL Education with AI: Innovations in Teaching Methods and Au-thentic Materials. *International Journal of Membrane Science and Technology*, *11*(1), 497–505. https://doi.org/10.15379/ijmst.v11i1.3687

Konyrova, L. K. (2024). *The Evolution of Language Learning: Exploring AI's Impact on Teaching English as a Second Language*. *2*(2), 133–138. https://doi.org/10.63034/esr-42

Lee, Y.-J., Davis, R. O., & Lee, S. O. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies*, *14*(1), e202412. https://doi.org/10.30935/ojcmt/14195

Liao, H.-C. (2015). Using automated writing evaluation to reduce grammar errors in writing. *ELT Journal*, *70*(3), 308–319. https://doi.org/10.1093/elt/ccv058

Lyanda, J. N., Owidi, S. O., & Simiyu, A. M. (2024). Rethinking Higher Education Teaching and Assessment In-Line with AI Innovations: A Systematic Review and Meta-Analysis. *African Journal of Empirical Research*, *5*(3), 325–335. https://doi.org/10.51867/ajernet.5.3.30

Marzuki, Widiati, U., Rusdin, D., Darwin, D., & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, *10*(2). https://doi.org/10.1080/2331186x.2023.2236469

Mun, C. (2024). EFL Learners' English Writing Feedback and Their Perception of Using ChatGPT. *Journal of English Teaching through Movies and Media*, *25*(2), 26–39. https://doi.org/10.16875/stem.2024.25.2.26

Onesi-Ozigagun, O., Ololade, Y. J., Eyo-Udo, N. L., & Ogundipe, D. O. (2024). Revolutionizing Education Through AI: A Comprehensive Review of Enhancing Learning Experiences. *International Journal of Applied Research in Social Sciences*, *6*(4), 589–607. https://doi.org/10.51594/ijarss.v6i4.1011

Parsons, S. A., Ives, S. T., R. Stacy Fields, Barksdale, B., Marine, J., & Rogers, P. (2023). The Writing Engagement Scale: A Formative Assessment Tool. *The Reading Teacher*, *77*(3), 278–289. https://doi.org/10.1002/trtr.2244

Phan, T. N. L. (2023). Students' Perceptions of the AI Technology Application in English Writing Classes. *Proceedings of the AsiaCALL International Conference*, *4*, 45–62. https://doi.org/10.54855/paic.2344

Putra, M. (2023). AI Writing Correction Tools: Teachers And Students' Perception. *Jurnal Tatsqif*, *21*(1), 35–66. https://doi.org/10.20414/jtq.v21i1.7963

Rahman, N. A. A., Zulkornain, L. H., & Hamzah, N. H. (2022). Exploring Artificial Intelligence using Automated Writing Evaluation for Writing Skills. *Environment-Behaviour Proceedings Journal*, *7*(SI9), 547–553. https://doi.org/10.21834/ebpj.v7iSI9.4304

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2016). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*(1), 8–25. https://doi.org/10.1080/01443410.2015.1136407

Schraudner, M. (2014). *The online teacher's assistant: Using automated correction programs to supplement learning and lesson planning*. Taiwan: Asia University.

Shribala, V., & Jhaneswaran, S. (2024). Impact of Artificial Intelligence in Education. *Shanlax International Journal of Management*, *11*(S1-Mar), 8–11. https://doi.org/10.34293/management.v11is1-mar.7992

Solak, E. (2024). Examining Writing Feedback Dynamics from ChatGPT AI and Human Educators: a Comparative Study. *Pedagogika-Pedagogy*, *96*(7), 955–969. https://doi.org/10.53656/ped2024-7.05

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Tedjo, E. W. (2022). Providing Online Feedback Using an AI Proof-reading Tool to Enhance Student Surface-Level Writing Skills. *EBONY: Journal of English Language Teaching, Linguistics, and Literature*, *2*(1), 37–45. https://doi.org/10.37304/ebony.v2i1.4040

Thu, T. (2024). AI Tools in Teaching and Learning English Academic Writing Skills. *Proceedings of the AsiaCALL International Conference*, *4*, 170–187. https://doi.org/10.54855/paic.23413

Wu, L. (2024). AI-based Writing Tools: Empowering Students to Achieve Writing Success. *Advances in Educational Technology and Psychology*, *8*(2). https://doi.org/10.23977/aetp.2024.080206

Appendices

Appendix A

1. When working on this writing assignment, I was interested in what I was writing. (A)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

**2. Working on this writing assignment was boring. (A)*
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

3. When working on this writing assignment, I felt good. (A)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

4. I would like to complete a writing assignment like this again. (A)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

5. I stayed focused when working on this assignment. (B)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

6. I kept trying on this assignment even if it was difficult. (B)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

7. I tried hard to do well on this writing assignment. (B)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

8. I worked as hard as I could on this writing assignment. (B)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

9. When working on this writing assignment, I reread to see if I could make it better. (C)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

10. When working on this writing assignment, I thought carefully about the words I used. (C)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

11. I asked myself questions as I was writing to make sure my writing made sense. (C)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

12. When working on this assignment, I reviewed my writing and made changes to make it better. (C)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

13. When working on this writing assignment, I talked with other students about my writing. (S)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

14. As I worked on this writing assignment, I wanted to share it with others. (S)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

15. I enjoy when my peers share their writing. (S)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

16. I can think of at least one person who would want to read this writing. (S)
1 2 3 4 5
Strongly disagree — Disagree — Neither agree nor disagree — Agree — Strongly agree

Appendix B

1. How did PaperRater's feedback help you improve your writing, such as grammar, structure, or style, during the lessons?
2. How did ChatGPT support your learning experience in writing tasks, particularly in generating ideas or enhancing creativity?
3. How does receiving feedback from AI tools like PaperRater or ChatGPT compare to feedback from your teacher in traditional writing lessons? Which do you think was more effective and why?
4. Did using AI tools like PaperRater or ChatGPT make you feel more engaged in writing lessons compared to traditional methods? Why or why not?
5. What do you think was the most significant impact of using AI tools like PaperRater or ChatGPT on your overall learning experience and writing skills?
6. Would you be interested in using tools like PaperRater or ChatGPT in other English lessons? If so, what specific areas do you think these tools could improve?