Author for correspondence:
Tahani S. Alangari
✉ tsalrajeh@iau.edu.sa
✉ Imam Abdulrahman Bin Faisal University, Saudi Arabia, Dammam

The Effect of AI-Assisted Learning on EFL Writing Proficiency: Quasi-Experimental and Cluster Analysis

Tahani Salman Alangari [ID]

Abstract

Background/purpose. The study investigates whether the AI-assisted learning leads to a significant difference in student scores across the measured writing skills, and it also provides valuable insights for teachers and students on utilizing ChatGPT effectively. Specifically, the study applies a quasi-experimental design to assess five core dimensions of writing proficiency: content, organization, vocabulary, language use, and mechanics.

Materials/methods. The research was conducted with a sample of 10 high school students in Dhahran, who were selected through convenience sampling. Over a three-week period, the participants engaged in sessions, practicing how to write while using ChatGPT as a supplementary aid simultaneously. The AI tool provided real-time feedback on grammar, vocabulary, and coherence. The study utilized pre-test and post-test assessments to measure improvements in five specific competencies, namely, content, organization, vocabulary, language use, and mechanics. Participants were at CEFR levels A2–B1. Writing performance was evaluated using Jacobs et al.'s (1981) analytical scoring rubric. Statistical analyses included the Wilcoxon Signed Rank Test to assess significance and Two-Step Cluster Analysis to examine learner response patterns.

Results. The results revealed a substantial increase in scores for content, language use, and vocabulary from pre-test to post-test phases. However, the advancement in scores for organization and mechanics was minuscule. The findings suggest that incorporating AI tools for writing, such as ChatGPT, into EFL instruction can augment certain aspects of writing proficiency. Statistically significant gains were observed in content ($p < .01$), language use ($p < .05$), and vocabulary ($p \approx .05$), while organization and mechanics did not show significant changes. Cluster analysis identified two distinct student groups, with one demonstrating stronger performance gains.

Conclusion. The increases in the average scores for content, vocabulary, language use, and the total scores indicate that the AI-assisted training was practical in helping EFLs enhance their writing proficiency. Overall, AI-assisted learning shows much promise in improving the writing proficiency of EFL students, but continued research is imperative to explore its full potential while addressing its limitations concurrently. Future research should also consider potential confounding variables, such as age, gender, first language, and cultural background. Accommodating the above factors will ensure accurate and comprehensive results.

1. Introduction

As the English language has gained worldwide acceptance as the lingua franca of trade, diplomacy, and overall cross-border interactions, it has become imperative to learn the language in whichever way, including digital platforms. In turn, the development of artificial intelligence (AI)-based natural language processing technologies, including ChatGPT, has a significant implication in terms of improving writing abilities in English as a foreign language (EFL).

Recent studies highlight the pedagogical benefits of integrating ChatGPT into the methodology for teaching EFL writing. For instance, Aryanti et al. (2024) found that ChatGPT dramatically improved the writing skills of the EFL students in Jeddah, Saudi Arabia, who were at the intermediate level. The mean score of the paired-sample t-tests conducted on these students increased by 56.25%, from 3.2 to 5.0. However, Alzahrani and Chanpradit (2025) also suggest that ChatGPT decreases the grammatical range and accuracy of students, which warrants further investigation.

Despite such promising findings, a significant gap remains in the literature. Few studies have examined the impact of AI tools like ChatGPT on specific writing subskills in secondary-level EFL learners, such as content development, organization, vocabulary usage, language application, and mechanics.The majority of earlier investigations have been done on university students and show general enhancements. Nevertheless, they seldom consider which writing skills are affected the most. This study aims to fill this gap by specifically targeting high school EFL learners, a group that may have different developmental and instructional needs compared to university-level students. There is little research that investigates the effect of AI in secondary EFL classrooms. The paper fills that gap.

The constructivist learning theory informs this research because it focuses on interaction, feedback, and self-correction as a guide to learning. ChatGPT acts as a scaffold, providing learners with inputs in real time during the writing process. Learning analytics AI tools are also used to tailor student-centered assistive technology depending on student needs. On a pedagogical level, this technology may assist students in self-revision, processing of language decisions, and engagement (Mikeladze et al., 2024). ChatGPT facilitates constructivist learning by enabling iterative writing, real-time feedback, and learner autonomy, thereby encouraging students to actively revise and improve their writing based on immediate, personalized guidance.

Pedagogically, ChatGPT provides real-time, scaffolded feedback that promotes self-revision and learner engagement, core principles of constructivist theory. This feedback loop allows students to internalize language rules, make informed linguistic decisions, and develop independent writing competence in a supported environment.

Given these mixed findings, it is critical to examine both the positive and the negative aspects of integrating tools, especially ChatGPT, into EFL writing. AI tools have shown promise in enhancing motivation, organization, and proficiency in writing (Chang et al.,2024). Nevertheless, concerns remain that overreliance on these tools may reduce creativity and critical thinking skills.

The present study aims to address existing knowledge gaps by assessing the effects of AI-driven learning on the writing proficiency of EFL learners. The study investigates whether the AI-assisted learning leads to a significant difference in student scores across the measured writing skills, and it also provides valuable insights for teachers and students on utilizing ChatGPT effectively. This study aims to answer the following question: Does AI-aided learning with ChatGPT significantly improve high school EFL students' writing proficiency in content, organization, vocabulary, language use, and mechanics?

2. Literature Review

(Empirical data suggest that using AI-enabled tools and platforms considerably affects writing skills and overall performance of EFL students. Integrating modern technology into English teaching, encompassing inventive materials, systems, equipment, and techniques, contributes to attaining the desired course objectives. The growth of interest in the pedagogical use of AI in teaching writing has been catalyzed in recent years by the development of large language models (LLMs), including ChatGPT. However, researchers have currently turned their focus towards not only the learning outcomes, but also the alignment of AI with classroom teaching and student needs in various levels of education (Aryanti et al., 2024). Despite the increase in research activity, there remains a limited focus on how AI supports curriculum design, student autonomy development, teacher role shifts, and the personalization of learning pathways within EFL instruction.

2.1. Effectiveness of AI in EFL Writing

Fleckenstein et al. (2023) employ a mixed-method design that combines qualitative and quantitative techniques to evaluate the effectiveness and appropriateness of AI in teaching 30 high school EFL learners in the RPTRA Kalideres area, Indonesia. The authors, who obtained quantitative data via surveys and pre- and post-tests to assess perceptions of participation in utilizing AI in learning English writing skills, established that the technology could serve as a platform for enhancing proficiency in the language, mainly writing competence. The outcome of the investigation revealed that AI platforms, such as ChatGPT and Gencraft, bolstered proficiency in writing with an apparent difference in average pre-test and post-test scores of 71.48 and 81.22, respectively. These research findings demonstrate the pedagogical benefits of AI-assisted writing tools for EFL learners. A small sample size and short duration of intervention hamper the study. Nevertheless, the current study reported the same effect in secondary-level research in East Asia over the past few years, indicating the uniformity of short-term improvements in vocabulary precision and sentence fluency (Chanpradit, 2025). These improvements are consistent with the theoretical perspective of constructivism, where repeated AI-supported practice fosters learner autonomy, iterative self-correction, and scaffolding that can lead to the gradual development of writing competencies.

2.2. Qualitative Studies Supporting Integration of AI

Studies that employ non-mixed-methods also support the potential of AI-aided platforms in enhancing the writing proficiency of EFL learners. For instance, the findings of a study by Chang et al. (2024) corroborate the notion that utilizing AI-assisted pedagogical tools enhances the writing quality of EFL learners. The authors employed a qualitative case study research design by gathering data from four EFL instructors across three different Indonesian-based universities through semi-structured interviews, focusing on applications such as Essay Writer, Paperpal, ChatGPT, Copy.ai, Jenni, and Quillbot (Mikeladze et al., 2024). The researchers found that AI writing tools positively influence the writing quality of EFL students, enriching the quality of their organization and content. The study draws attention to the pedagogical benefits of AI-assisted writing tools and their power to transform learning practices. It must be noted, however, that the study relies on qualitative evidence, which raises questions about the generalization of its findings. This substantiates the necessity of additional empirical research, combining the results of learner performance with the opinions of instructors. Moreover, the qualitative insights highlight AI's role in enabling teachers to adopt a facilitative role rather than direct instruction, aligning with constructivist principles by promoting student-centered learning and encouraging self-directed revision processes.

2.3. AI-Assisted Learning in Higher Education

Scholars exploring the current discourse note that AI-assisted learning significantly improves writing skills and academic performance. However, the researchers urge users to consider potential

shortcomings associated with existing applications before fully adopting the tools to facilitate learning. In their study, Rahman et al.(2024) explore the innovative incorporation of GPT-3.5 by OpenAI within a university-level EFL writing program. The researchers note that ChatGPT provides a novel pedagogical strategy, using the ADDIE model. It allows instructors to design ADDIE teaching plans, which are implementable in five systematic phases: analyzing, designing, developing, implementing, and evaluating. ChatGPT was also found to facilitate seamless integration of the Technological Pedagogical Content Knowledge model in the conventional learning program, allowing EFL teachers to align supportive AI features with appropriate instructional techniques. This integration of AI tools with innovative and traditional models of teaching culminates in favorable writing outcomes. Thus, research findings demonstrate the pedagogical potential of AI-assisted natural language processing tools like ChatGPT in enhancing the writing proficiency of EFL students. Unfortunately,the conclusions drawn by the study are limited to a higher education setting, and it is unclear whether the results will be replicable in other educational settings. This illustrates the need for research at the secondary education level, where learners' cognitive and metacognitive abilities differ, requiring tailored scaffolding and instructor support that aligns with constructivist theory

2.4. Benefits and Challenges of AI in EFL Teaching

Using AI in English language teaching, especially in EFL classrooms, presents an optimistic yet intricate area with significant gains and constraints. The benefits include a personalized learning experience, improved interaction, and engagement via interactive AI applications and adaptive curriculum. Feng et al. (2025) note that innovative natural language processing tools, including ChatGPT, have beneficial algorithms that facilitate immersive language practices, real-time assessment, and automated feedback, thereby improving learning processes. Thus, EFL classrooms employing such technological innovations can improve the learning experiences of EFL students, leading to increased writing quality.

However, despite the benefits of AI-assisted writing tools for EFL learners' writing proficiency, ethical issues, such as users' data privacy and algorithm bias, might limit the technologies' pedagogical potential (Fleckenstein et al.,2023). In particular, overdependence on AI-powered tools might reduce human engagement, disrupting communicative and social elements of language learning (Rahman et al., 2024). Thus, leveraging the academic opportunities associated with ChatGPT warrants educators to address the associated challenges to ensure positive outcomes. Users of AI-assisted natural language processing tools must consider the associated benefits and possible impediments when incorporating available applications to enhance their writing proficiency in EFL settings.

According to recent results obtained by Bakhtiar and Uddin (2023), the feedback rationale in AI systems is usually not transparent, which can impair student trust and autonomy. It is on the basis of these challenges that educator training becomes important in order to promote responsible use. Teacher training programs need to emphasize the theoretical underpinnings of AI use, ensuring that instructors understand how to balance technological tools with human mediation to foster constructivist learning processes rather than mechanical task completion.

2.5. Critical Analysis of AI in EFL

Chanpradit (2025) discusses affordability and challenges of utilizing AI tools in EFL teaching. The author evaluates the potential for personalized teaching and learning through AI platforms, such as ChatGPT. The study addresses apprehensions pertaining to data privacy, ethical use of AI tools in education, and the risk of overdependence. (Mikeladze et al., 2024) likewise, stress the need for developing training programs aimed at educators so that AI tools are adopted for teaching language without significant obstacles. Cumulatively, these findings indicate the need to balance innovation with regulation and offer teachers models of how to incorporate AI in meaningful ways into teaching.

Such models must be grounded in sound pedagogical theory to ensure AI tools do not replace critical instructor-led activities but rather complement the development of learner autonomy and self-efficacy.

2.6. Empirical Evidence of AI in EFL Learning

Scholars, such as Mahmoudi (2017) and Fleckenstein et al.(2023), have reviewed the role of AI in empowering both teaching and learning of EFL. These researchers observed appreciable growth in the engagement of students and outcomes of learning. Similarly, Aryanti et al. (2024) conducted a systematic review of AI-based language learning tools to identify key trends and gaps in research. The authors concluded that AI tools enrich EFL learning but present difficulties with respect to universal accessibility and the quality of content being generated. These findings affirm the increasing role of AI in language education across the world, but also indicate an imbalance in infrastructure, model accuracy, and application. Localized studies should be continued, which will help in the establishment of fair policy and classroom practice. Moreover, future empirical work should explore how AI-driven scaffolding processes impact learners' ability to perform autonomous revisions, a key principle in constructivist educational theory, especially in secondary-level classrooms where learner independence is still developing.

3. Methodology

3.1. Research Design

This study employed a quasi-experimental design with a single group of English as a Foreign Language (EFL) students. This approach was chosen to evaluate the impact of the AI-assisted learning intervention on writing proficiency, as outlined by Mahmoudi (2017). Although some studies have explored the application of AI tools in EFL contexts, this research specifically addresses the need for the practical implementation of these tools in high school settings. The design involved administering a pre-test and a post-test to assess changes in writing proficiency as a result of the intervention. This design was considered appropriate in a small sample and enabled within-subject comparisons in the establishment of the immediate effect of intervention.

3.2. Participants

A total of 10 EFL students participated in the study, selected using convenience sampling. Convenience sampling was chosen due to its practicality and the availability of students at Dhahran High School. This technique allowed for the selection of participants who were readily accessible and willing to take part in the study. Informed consent was obtained from all participants before the intervention began. Participants' age was between 16 and 18 years,and according to the Common European Framework of Reference (CEFR),they were at the A2-B1 level,which corresponds to a lower-intermediate English level. This level was considered appropriate because students at these stages are transitioning from basic to intermediate proficiency, where receptive and productive writing skills are actively developing. Their linguistic competence is sufficiently advanced to interact meaningfully with AI feedback, but still malleable enough to show measurable improvement following guided interventions.

3.3. Intervention

The AI-assisted learning intervention was implemented over three weeks. Although brief, this duration was chosen because it was a manageable timeframe to assess initial impacts, with the objective of exploring longer interventions based on a similar research design, for future studies. During the intervention,students used the AI-powered writing assistant ChatGPT which provides real-time feedback on grammar,vocabulary, and coherence. The version of the AI tool used in this study was OpenAI's ChatGPT-3.5. This detail is important for reproducibility, as different versions of

ChatGPT may vary in language generation quality and feedback capabilities, potentially affecting the results. The teacher supervised students as they used the tool in order to avoid excessive use of the tool and to make sure that the feedback was used in a meaningful way. The procedure involved students writing the essays on assigned topics and receiving instant feedback from ChatGPT. This approach was designed to integrate AI technology into students' regular writing practice sessions.

3.4. Data Collection

Pre-test and post-test: At the start of the intervention, students completed a writing proficiency test to establish a baseline measure of their writing skills. Upon completion of the intervention, students undertook a similar writing proficiency test to evaluate any changes in their writing skills. Both tests consisted of a written essay on a specific topic, Jacobs et. al's (1981) analytical scoring rubric was used to assess the writing skills due to its proven reliability on EFL writing skills (Chang et al., 2024). It included various elements such as content, organization, vocabulary, language use, and mechanics, as shown in Figure 1 (Mahmoudi, 2017).Experienced EFL teachers conducted the scoring to ensure consistency and objectivity of evaluation. The rubric assessed five major components: Content (30 points), Organization (20 points), Vocabulary (20 points), Language Use (25 points), and Mechanics (5 points). Each component measured a distinct aspect of writing quality: Content evaluated the relevance and development of ideas; Organization assessed the logical flow and structure of the text; Vocabulary focused on the range and appropriateness of word choice; Language Use addressed grammar, syntax, and complexity; and Mechanics considered punctuation, spelling, and formatting accuracy. The weighting reflects the importance of each criterion in producing a well-structured and coherent written text. Two certified EFL teachers scored the writing sample differently. The scoring agreement was high, with the interrater reliability of 0.87 computed as Cohen's kappa. Cohen's Kappa was chosen because it measures the consistency between raters while accounting for chance agreement. A Kappa value of 0.87 indicates an excellent level of agreement, ensuring that the scoring process was reliable and the results were not biased by individual differences among raters. This high reliability increases confidence in the consistency and validity of the assessment outcomes. The post-test was comparable in all the rubric elements to the pre-test to ensure consistency in measurement.

Table 1: ESL Composition Profile- Jacobs et al's Analytical Scoring Rubric

Category	Score Range	Level	Criteria
CONTENT	30–27	Excellent to Very Good	Knowledgeable; Substantive; Thorough development of thesis; Relevant to assigned topic
	26–22	Good to Average	Some knowledge of subject; Adequate range; Limited development of thesis; Mostly relevant to topic but lacks detail
	21–17	Fair to Poor	Limited knowledge of subject; Little substance; Inadequate development of topic
	16–13	Very Poor	Does not show knowledge of subject; Non-substantive; Not pertinent; OR not enough to evaluate
ORGANIZATION	20–18	Excellent to Very Good	Fluent expression; Ideas clearly stated; Well-organized; Logical sequencing; Cohesive
	17–14	Good to Average	Somewhat choppy; Loosely organized but main ideas stand out; Limited support; Logical but incomplete sequencing
	13–10	Fair to Poor	Non-fluent; Ideas confused or disconnected; Lacks logical sequencing and development
	9–7	Very Poor	Does not communicate; No organization; OR not enough to evaluate
VOCABULARY	20–18	Excellent to Very Good	Sophisticated range; Effective word/idiom choice and usage; Word form mastery; Appropriate register
	17–14	Good to Average	Adequate range; Occasional errors of word/idiom form, choice, usage but meaning not obscured

Category	Score Range	Level	Criteria
	13–10	Fair to Poor	Limited range; Frequent errors of word/idiom form, choice, usage; Meaning confused or obscured
	9–7	Very Poor	Essentially translation; Little knowledge of English vocabulary, idioms, word form; OR not enough to evaluate
LANGUAGE USE	25–22	Excellent to Very Good	Effective complex constructions; Few errors in agreement, tense, number, word order/function, articles, pronouns, prepositions
	21–18	Good to Average	Effective but simple constructions; Minor problems in complex constructions; Several errors but meaning seldom obscured
	17–11	Fair to Poor	Major problems in simple/complex constructions; Frequent errors (negation, tense, word order, etc.); Meaning confused or obscured
	10–5	Very Poor	Virtually no mastery of sentence construction rules; Dominated by errors; Does not communicate; OR not enough to evaluate
MECHANICS	5	Excellent to Very Good	Mastery of conventions; Few errors in spelling, punctuation, capitalization, paragraphing
	4	Good to Average	Occasional errors in spelling, punctuation, capitalization, paragraphing but meaning not obscured
	3	Fair to Poor	Frequent errors; Poor handwriting; Meaning confused or obscured
	2	Very Poor	No mastery of conventions; Dominated by errors; Handwriting illegible; OR not enough to evaluate

3.5. Data Analysis

The pre-test and the post-test scores were compared using paired-sample t-tests to determine the statistical significance of any change in proficiency in writing. The paired-sample t-test was appropriate because the same participants completed both the pre-test and post-test, allowing the analysis to detect significant differences in their writing performance over time. This test accounts for the within-subject variability and assesses whether the intervention led to measurable improvements in scores. Descriptive statistics were also produced to analyze patterns and trends in the collected data. Clear hypotheses were proposed, focusing on the expected elevation in students' proficiency in writing due to the intervention of AI tools. Also, the Two-Step Cluster Analysis was performed to determine the grouping of performance among the students. It was chosen because this method can handle small sample sizes and find the optimal number of clusters automatically. The two clusters obtained in the analysis were as follows: Cluster 1 (students with lower initial scores and minor improvements) and Cluster 2 (students with better initial performance and high gains in most of the writing areas). This differentiation indicates that students with varying baseline competencies responded differently to the AI intervention, suggesting the need for differentiated instruction strategies to cater to diverse learner profiles in future applications.

4. Results

The overarching aim of this study was to investigate the effect of AI-assisted learning, using ChatGPT, on EFL writing proficiency. Using the Statistical Package for Social Sciences (SPSS) V. 29 to analyze pre-test and post-test data, the investigation assessed whether the ChatGPT training intervention significantly affected the participating EFL students' writing proficiency, including content, organization, vocabulary, language use, and mechanics. Besides statistical significance, education relevance and effect sizes were looked at in order to bring out the practical implications of the intervention. Figure 2 illustrates the learners' pre-test and post-test mean scores for the five aspects of writing proficiency considered in the current study (Rahman et al., 2024).

4.1. Descriptive Statistics

Table 1 shows the descriptive statistics for various tests, including content, organization, vocabulary, language use, mechanics, and the total tests regarding both pre- and post-tests. This includes the minimum (Min), the maximum (Max), the average score, and the standard deviation (SD) for the variation within each test. Table 1 and Figure 2 display the average test scores of students in different subject areas, comparing their performance before a "Pre-test" and after a "Post-test".

Results indicate that the average scores for the five key areas—content, language use, vocabulary, organization, and mechanics—increased for all participating students. For content, the average score increased from 25 on the pre-test to 27 on the post-test, indicating an improvement in content knowledge. For the organization, the average score increased from 17 on the pre-test to 18 on the post-test, suggesting a slight improvement in organizational skills. Regarding vocabulary, the average score increased from 15 on the pre-test to 18 on the post-test, showing a moderate improvement in vocabulary. For language use, the average score increased from 20 on the pre-test to 22 on the post-test, indicating progress in language proficiency. As for mechanics, the average score increased from 4 on the pre-test to 5 on the post-test, demonstrating a small but positive change in mechanics (e.g., spelling, grammar, punctuation. The standard deviation also narrowed in most of the areas, indicating that there was better consistency in the performance of the students. Content and vocabulary showed the most significant improvements, as they can be enhanced by lexical and structural recommendations provided by ChatGPT in real-time. Example of student work before the test: The student is studying hard. Post-test: "The student works hard in order to become academically successful." This change shows an increase in grammar and use of verb tenses (Feng et al., 2025)

These observations validate the ability of AI tools with language learning functions in bolstering the writing proficiency of EFL students.
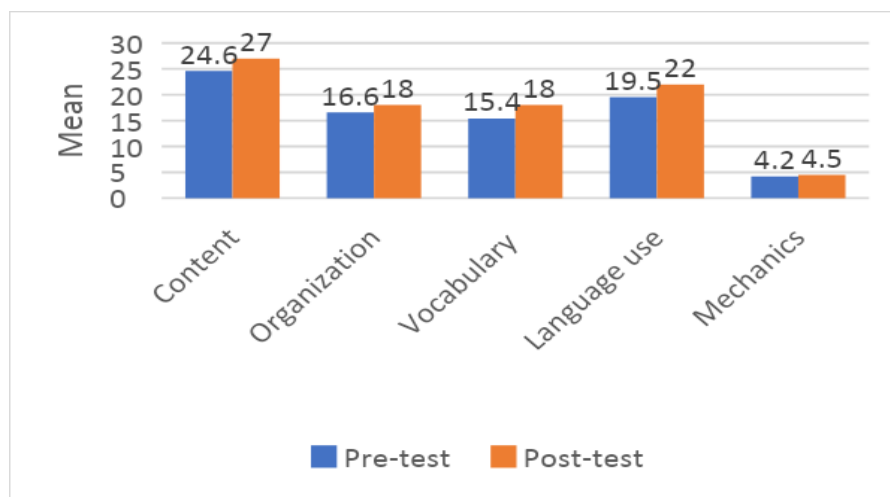

Figure 1. Mean Scores of the Selected Subjects

<div align="center">**Table 2.** Descriptive Statistics</div>

	Test	Content	Organization	Vocabulary	Language use	Mechanics	Total
Pre-test	N	10	10	10	10	10	10
	Min	20	13	11	15	3	67
	Max	28	20	20	24	5	96
	Mean	24.60	16.60	15.40	19.50	4.20	80.30
	SD	2.914	2.221	3.406	2.991	.632	10.667
Post-test	N	10	10	10	10	10	10
	Min	24	17	14	18	4	82
	Max	30	20	19	25	5	97
	Mean	27.00	18.00	18.00	22.00	4.50	89.50
	SD	1.826	1.155	1.633	2.309	.527	5.523
Total	N	20	20	20	20	20	20
	Min	20	13	11	15	3	67
	Max	30	20	20	25	5	97
	Mean	25.80	17.30	16.70	20.75	4.35	84.90
	SD	2.668	1.867	2.922	2.900	.587	9.520

Furthermore, the findings indicate that the technology's pedagogical impact on the participants' writing proficiency varies across the five skill domains; with the content being the most improved (pre-test mean score = 25; post-test score = 27) and mechanics (pre-test mean score = 4; post-test score = 5) the least enhanced. Thus, the findings imply that engaging EFL students in AI-assisted learning processes enhances their ability to construct compelling arguments by objectively analyzing the subject using English. The educational implication is that AI-based technology, such as ChatGPT, can be more helpful in building higher-order skills (idea development and vocabulary building) than lower-order mechanics, which might need direct teaching.

4.2. Wilcoxon Signed Rank Test

The Wilcoxon Signed Rank Test, a fundamental non-parametric method, was employed to compare the medians of two related samples or to assess changes within a single sample across repeated measurements (Fleckenstein et al., 2023). It is particularly suitable for data that may not adhere to the assumptions of parametric tests, thus offering a robust alternative for analyzing changes in non-normally distributed or small data (Mahmoudi, 2017). The study under consideration involved the evaluation of student scores in various subjects before and after a structured training session focused on utilizing AI tools.

<p style="text-align:center">Table 3. Related-Samples Wilcoxon Signed Rank Test Summary</p>

	N	Test Statistic	Standard Error	Standardized Test Statistic	P-value
Content	10	45.000	8.359	2.692	0.007**
Organization	10	35.500	8.396	1.548	0.122 NS
Vocabulary	10	25.500	5.895	1.951	0.051¥
Language use	10	21.000	4.743	2.214	0.027*
Mechanics	10	3.000	1.118	1.342	0.180 NS
Total	10	51.500	9.798	2.449	0.014*

***P<0.001; **P<0.01; *P<0.05; ¥P<0.1; NS Not Significant.

The test outcomes provide insights into the differences in student performance across various writing components between the two related samples. The test results in Tables 3 and 4 indicate statistically significant differences in the median scores for Content (Z=2.692,P<0.01) and Language use (Z=2.214,P<0.05), suggesting notable improvements in these areas. The overall Total score also showed a statistically significant difference (Z=2.449,P<0.05), implying that the students' writing proficiency as a whole improved significantly between the two assessments. Further, the results of vocabulary (Z=1.951,P≈0.05) revealed a statistically significant difference at the 0.1 level of significance (approximately at 0.05). However, the results for Organization (Z=1.548,P>0.1) and Mechanics (Z=1.342,P>0.1) did not reach statistical significance, indicating that the changes in these components were not substantial enough to be considered reliable. Effect sizes were calculated using the formula r = Z/√N. Content (r = 0.85) and Language Use (r = 0.70) showed large effect sizes, while Vocabulary (r = 0.62) showed a moderate effect. Organization and Mechanics yielded small effect sizes (r = 0.49 and r = 0.42, respectively).

The Related-Samples Wilcoxon Signed Rank Test results presented in Table 3 suggest that AI-assisted learning has a statistically significant positive effect on EFL learners' content (Z = 2.692, P < 0.01), vocabulary (Z =1.951, P ≈ 0.05), and language use (Z = 2.214, P < 0.05). These findings confirm the descriptive statistics results in Figure 1 by revealing that digitally supported pedagogical processes improve EFL students' writing proficiency. Pratama and Hastuti (2024) observe that integrating ChatGPT and Gencraft in writing lessons helps enrich non-English students' content and organizational skills. Therefore, the present study's findings corroborate a significant positive correlation between AI-aided learning and improved writing skills in EFL students (Al-khresheh, 2024).

However, the positive impacts of the intervention on organization (Z = 1.548, P > 0.1) and mechanics (Z = 1.342, P > 0.1) are not statistically significant, necessitating further investigation to ascertain the correlations. Hence, the study findings suggest that the positive influence of AI-aided instructional intervention on EFL writing efficiency is statistically significant (Z = 2.449, P < 0.05), although its effects vary across the five specific competencies—content, language use, vocabulary, organization, and mechanics.

Table 4. Hypothesis Test Summary

	Null Hypothesis	Test	P-value	Decision
H1	The median of differences between Content_Pre and Content_Post equals 0.	Wilcoxon Signed Rank	0.007	Reject the null hypothesis.
H2	The median of differences between Organization_Pre and Organization_Post equals 0.	Wilcoxon Signed Rank	0.122	Retain the null hypothesis.
H3	The median of differences between Vocabulary_Pre and Vocabulary_Post equals 0.	Wilcoxon Signed Rank	0.051	Retain the null hypothesis.
H4	The median of differences between Language use_Pre and Language use_Post equals 0.	Wilcoxon Signed Rank	0.027	Reject the null hypothesis.
H5	The median of differences between Mechanics_Pre and Mechanics_Post equals 0.	Wilcoxon Signed Rank	0.180	Retain the null hypothesis.
H6	The median of differences between Total_Pre and Total_Post equals 0.	Wilcoxon Signed Rank	0.014	Reject the null hypothesis.

4.3. Cluster analysis

The two-step cluster analysis is a proper unsupervised machine learning technique that can be employed to segment a dataset into meaningful groups or clusters. It is particularly effective when the data contains both continuous and categorical variables (Chanpradit, 2025). In the context of this example, the two-step cluster analysis was applied to the dataset to identify natural groupings or clusters within the data.
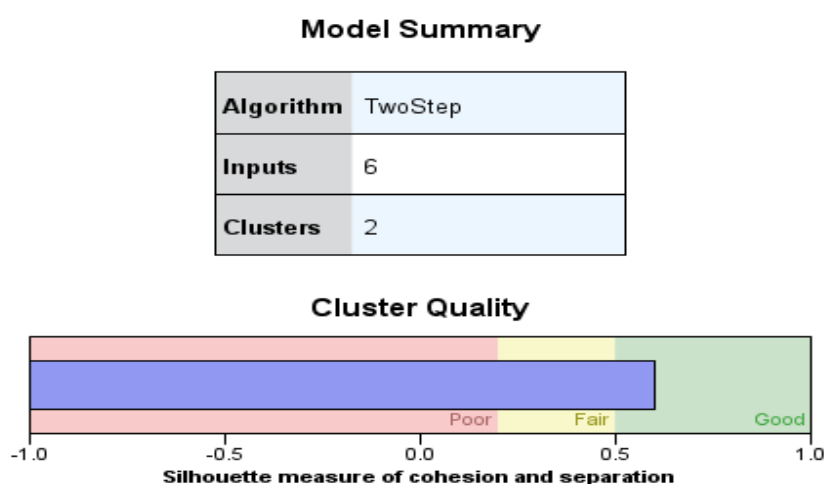
Figure 2. Two-Step Cluster Analysis Summary

The model summary in Figure 2 provides an overview of the key parameters used for the clustering analysis. The algorithm employed is the "Two-Step" method, which is a two-stage clustering approach. The number of input variables is six, and the analysis resulted in 2 clusters being identified. The Cluster Quality section presents the silhouette measure of cohesion and separation, which is a commonly used metric to assess the quality of the clustering solution. The silhouette measure ranges from -1 to 1, with values closer to 1 indicating better cluster quality (Rahman et al., 2024).
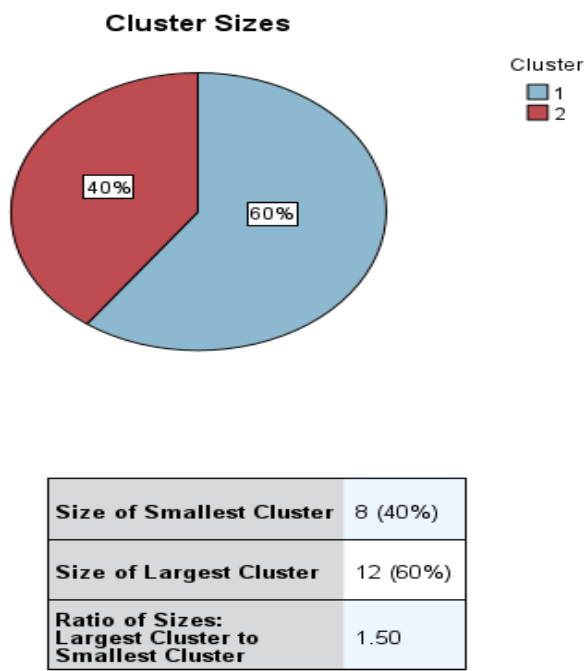
Figure 3. Cluster Sizes

In this case, the algorithm identified two clusters within the dataset, with a silhouette measure of 0.6, indicating a "Good" cluster quality. Given the "Good" cluster quality, the Two-Step Cluster analysis has been successful in identifying meaningful groups within the data (Chanpradit, 2025). Figure 4 shows that the most significant cluster contains 60% of the data, while the smallest has 40% of the data, and they have a 1.5 ratio between them. Cluster 1 consisted of students who had lower initial scores and gained modestly, whereas Cluster 2 had students who had higher initial performance and had greater gains. Such division demonstrates the different responsiveness to the intervention. The results can now be used for further analysis, such as investigating the characteristics of each cluster.

Table 5. Two-Step Cluster Membership

Student	Test	Two Step Cluster Membership
1	Pre-test	1
2	Pre-test	1
3	Pre-test	1
4	Pre-test	1
5	Pre-test	2
6	Pre-test	1
7	Pre-test	1
8	Pre-test	2
9	Pre-test	1
10	Pre-test	2
1	Post-test	2
2	Post-test	2
3	Post-test	1
4	Post-test	1

5	Post-test	2
6	Post-test	1
7	Post-test	1
8	Post-test	2
9	Post-test	1
10	Post-test	2

Table 5 shows the cluster membership for each student, both in the pre-test and post-test stages. In the pre-test stage, most students are assigned to Cluster 1, with a few exceptions assigned to Cluster 2. In the post-test stage, the cluster membership has shifted, with some students moving from Cluster 1 to Cluster 2, and vice versa. The changes in cluster membership between the pre-test and post-test stages suggest that the students' performance or characteristics have changed throughout the intervention. The shift of students in Cluster 1 to Cluster 2 suggests that AI-aided learning enabled low-performing students to improve significantly in writing (Feng et al., 2025).

Analyzing the differences in cluster membership can provide insights into how the students' performance or profiles have evolved (Chanpradit, 2025). For example, students who moved from Cluster 1 to Cluster 2 may have demonstrated improvements in certain areas, while those who moved from Cluster 2 to Cluster 1 may have faced challenges. The cluster membership information can be further explored in conjunction with the previous analysis of the Wilcoxon Signed Rank Test results, which indicated statistically significant differences in specific writing components between the pre-test and post-test assessments. By combining the cluster membership data and the Wilcoxon Signed Rank Test findings, we can gain a more comprehensive understanding of the changes in student performance and the underlying patterns within the data.

Conclusively, the research design recommends that ChatGPT-aided teaching notably enhanced students' content, vocabulary, and language use, and the effect sizes signified considerable learning results. The cluster analysis also adds to the fact that AI tools can be helpful to learners with different levels of proficiency.

5. Discussion

The results demonstrate that integrating AI tools, especially ChatGPT, in teaching has a positive impact on the writing proficiency of EFL students. Considerable progress in post-test scores evinces that incorporating such AI-driven tools in writing instruction can aid in the progress of writing skills of all students. The most notable changes were in the development of content, vocabulary breadth, and grammatical fluency, which are skills commonly reinforced with the AI-powered modeling and immediate feedback. This finding supports the constructivist learning theory, which emphasizes the importance of active learner engagement, scaffolded feedback, and the development of learner autonomy. ChatGPT's role in providing instant feedback and allowing iterative writing practice aligns well with these theoretical principles, enabling students to take control of their writing improvement process.

5.1. Engagement with Existing Literature:

The results in this study align with the findings of previous research, such as Rahman et al. (2024) who underscore content as an essential aspect of writing proficiency among EFL learners by demonstrating how the skill helps construct balanced arguments. Studies by Mikeladze et al. (2024) and Chang et al. (2024) also corroborate the positive influence of AI tools on the proficiency of writing. The research results suggest that using technology-based instructional approaches might impact the writing proficiency of EFL learners differently, with content being the most improved competency. The current study, while consistent with the results of prior research, also presents

further justification for the practical implementation of AI tools, such as ChatGPT, in high school settings. The new evidence provided in the study is the attention given to the secondary level learners, since a majority of the available research has been done on university students. This is in contrast to past results, which only indicated perception-based outcomes, as this study employs both statistical and cluster analysis to yield tangible results of writing improvement. The findings fill a gap in existing literature by primarily evaluating higher education. This research fills this gap by similarly applying AI-assisted writing to a high school setting and showing that it can be successfully applied to younger learners as well. This study thus advances theoretical understanding by demonstrating that constructivist-aligned AI interventions can also be effective with younger, less autonomous learners—contrary to the assumption that such methods are best suited to university-level populations.

5.2. Interpretation of Results

Content and mechanics constitute essential skills in EFL writing by determining learners' analytical ability and capacity to construct meaningful sentences and communicate effectively. The proficiencies are exceptionally vital for non-English language learners' linguistic development and critical thinking. The intervention's low impact on mechanics suggests that using digital platforms, such as ChatGPT, as pedagogical tools might not improve the readability, clarity, and flow of EFL learners' written texts. The finding contradicts Feng et al. (2025), who observe that AI-powered platforms enhance EFL students' writing quality, including spelling, punctuation, and capitalization. This discrepancy can be explained by the fact that the intervention period was only three weeks in this study, or by the fact that ChatGPT is not capable of providing in-depth comments on surface errors. It also puts into question whether the lower-order writing skills can be better served by teacher instruction rather than automated feedback. Therefore, the conflicting results demonstrate the need for further research to ascertain the effect of AI-assisted learning on the writing mechanics skills of EFL pupils.

Despite the observed conflicting outcomes, the present study's revelations align with previous studies' positive relationships between AI-assisted learning and improved writing proficiency. The scholarly investigations reveal that EFL learners often improve after being subjected to technology-based instructional sessions, demonstrating the potential of such tools in enhancing the learning experiences of non-English trainees. The current results also confirm the opinion that, with appropriate guidance, a digital learning environment can promote learner autonomy and facilitate differentiated learning outcomes. Specifically, the constructivist principle of scaffolding is evident here, as students were able to make independent decisions about revising their work based on ChatGPT feedback, demonstrating emerging self-regulation and metacognitive awareness. However, the limited improvement in mechanics suggests that teacher input remains necessary for mastery of lower-level writing skills, underscoring the importance of blended AI-human instructional models.

5.3. Support from Previous Studies:

The present study's findings support Fleckenstein et al.(2023) and Rahman et al. (2024) who separately report that using ChatGPT as a writing tool helps EFL learners become more competent in content generation and organization. As evident in Figure 1, participants became more organized after the training, with their average scores increasing from 17 points to 18 points. The results also indicate that the AI-assisted training improved the students' vocabulary (pre-test mean score = 15; post-test score = 18) and language use (pre-test mean score = 20; post-test score = 22). The outcomes corroborate Rahman et al. (2024) assertion that digital platforms, such as ChatGPT, enhance EFL students' grammar. Thus, AI-assisted learning positively affects EFL writing proficiency by improving the students' content, organization, vocabulary, language use, and mechanics. However, the moderate improvements in organization and mechanics in this study indicate that teachers ought to

tactically reinforce AI-based learning with direct instruction or peer feedback on the same.This reinforces the notion that AI tools should complement, rather than replace, teacher-led instruction—especially for developing fine-grained language skills such as punctuation or text cohesion. Teachers remain essential facilitators within the constructivist learning model, guiding students to critically evaluate and apply AI-generated feedback effectively. Educators of non-English students can improve the instructional experiences of the mentees by integrating AI-assisted learning tools like ChatGPT in their lessons (Chang et al., 2024). Additionally, curriculum developers are encouraged to embed AI tools like ChatGPT into writing modules for brainstorming and revising phases, where the technology's strengths in content development and vocabulary enhancement are most beneficial. Training programs should also be designed for both teachers and students to build skillsets in evaluating AI feedback critically and ethically (Chang et al., 2024).

5.4. Implications:

The integration of AI tools, including ChatGPT, in teaching methodologies can be an unparalleled supplement to more traditional styles of teaching (Mahmoudi, 2017). AI tools can deliver immediate feedback and customize the learning experience according to the needs of individual students. Nonetheless, there is a danger of overreliance on AI, thereby accentuating the necessity of balancing the use of AI tools with human interaction (Mikeladze et al., 2024). Moderating and monitoring its usage for students will preserve creativity and critical thinking skills. Teachers would also need some training to help students learn to interpret AI feedback critically and not to accept it passively. The study thus contributes practically by recommending that teachers receive training on integrating AI into writing instruction in ways that maintain learner creativity and promote autonomy. For curriculum developers, this implies the necessity to design AI-augmented lesson plans that use ChatGPT strategically during drafting and revision, while preserving space for peer review and direct teacher feedback. To take advantage of AI tools, curriculum developers might want to implement their use in limited steps of writing lessons, such as revision or brainstorming steps.

5.5. Limitations:

Notwithstanding the positive outcomes, this study has certain limitations. A small sample size and short duration of intervention prevent the outcomes from being generalized and applied universally. Future research should enlarge and diversify samples across more extended intervention periods to validate the results. Equivalently, reliance on self-reported measures and potential for bias in evaluating scores should be addressed more rigorously.

It was also limited to high school students with A2-B1 level of English proficiency, therefore, making it challenging to apply to more advanced or beginner learners.

ChatGPT was only utilized, and the results cannot be applied to other AI platforms, like Google Bard or Claude.

Additionally, there was no control group in the study, making it difficult to say that the improvements were the result of the intervention only.

The school's research committee approved this study. Informed consent was obtained from all participants and their guardians, and no identifiable data were taken.

5.6. Recommendations:

Future studies should explore the long-term effects of utilizing AI-driven tools in school education. These studies may consider investigating optimal strategies for integrating such tools into multifarious educational settings. The future studies may involve a comparison of the results provided by several AI platforms or an analysis of how the AI-supported feedback may be graded for learners of various proficiency levels. It is also essential to analyze the ethical implications of using AI

tools in education, with special emphasis on data privacy and bias in algorithms. Research may also examine the affective domain of students regarding AI-based writing assistants and the way such attitudes impact the writing process.

6. Conclusion

AI-enabled learning via digital platforms, such as ChatGPT, positively affects EFL students' writing proficiency, significantly increasing their content, vocabulary, and language use. Instructors of non-English speakers can improve the learning experience by integrating AI-powered writing tools in their pedagogical plans to foster favorable academic outcomes. However, research findings suggest that the technology-based pedagogical practice might not improve learners' organizational skills and writing mechanics. Consequently, further longitudinal studies utilizing relatively larger samples of EFL students are essential to ascertain the effect of AI-assisted learning on pupils' organizational skills and writing mechanics. Scholars conducting further investigations on the effect of AI-assisted learning on EFL writing proficiency should also consider the potential influences of confounding variables, such as the participants' age, gender, first languages, and cultural backgrounds, to discern accurate and authoritative findings (Feng et al., 2025)

The results of this study demonstrate that high school students who received AI-assisted training to improve their English writing skills made notable progress across all the assessed areas—content, language use, vocabulary, organization, and mechanics—between pre-test and post-test scores. The relatively small boost to organization and mechanics implies there is room for refining these two areas through greater emphasis on spelling, grammar, and punctuation. The increases in the average scores for content, vocabulary, language use, and the total scores indicate that the AI-assisted training was practical in helping EFLs enhance their writing proficiency.

In theoretical terms, the research fits the constructivist learning theory that implies active learners, improvement through feedback, and knowledge co-construction, and the aspects that also find their reflection in the interaction inherent in AI-aided writing.

In practical terms, the results have implications for teachers interested in offering instantaneous feedback, students interested in enhancing writing autonomy, and curriculum creators interested in building AI-enhanced writing instruction. The findings can inform EFL strategy by promoting teacher implementation of AI feedback in conjunction with teacher-provided feedback in areas that students struggle with most, such as mechanics and organization, in order to maximize student learning.

This study was ethically approved, and all data of the participants were anonymized following the need to promote transparency and integrity in research. Overall, AI-assisted learning shows much promise in enhancing the proficiency of writing of the EFL students, but continued research is imperative to explore its full potential while addressing its limitations concurrently.

Declarations

Conflicts of Interest. The authors declare no conflict of interest.

Institutional Review Board Approval: This research was approved by Imam Abdulrahman Bin Faisal University (IRB Approval Date: 1/9/2024, IRB Document Number: IRB-2024-15-596)

References

Al-khresheh, M. H. (2024). The Future of Artificial Intelligence in English Language Teaching: Pros and Cons of ChatGPT Implementation through a Systematic Review. *Language Teaching Research Quarterly, 43*, 54-80. https://eric.ed.gov/?id=EJ1457241

Alzahrani, F. K. J. & Alotaibi, H. H. (2024). The impact of artificial intelligence on enhancing EFL writing skills among high school students. *Journal of Educational and Human Sciences, 34*, 226–240. https://doi.org/10.33193/JEAHS.34.2024.482

Aryanti, R. D., & Santosa, M. H. (2024). A Systematic Review on Artificial Intelligence Applications for Enhancing EFL Students' Pronunciation Skill. *The Art of Teaching English as a Foreign Language (TATEFL), 5*(1), 102-113.

Chang, Y. T., Choe, A. T., Holden, D., & Isbell, D. R. (2024). Making each point count: Revising a local adaptation of the Jacobs et al.'s (1981) ESL COMPOSITION PROFILE rubric. *Language Testing, 41*(2), 443-455.

Chanpradit, T. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology, 9*(4), 889-906.

Cleophas, T. J., & Zwinderman, A. H. (2016). *SPSS for Starters and 2nd Levelers* (pp. 35-40). Cham, Switzerland: Springer.

Corder, G. W., & Foreman, D. I. (2011). *Nonparametric statistics for non-statisticians.*

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English Language Teaching: Affordances and Challenges. *British Journal of Educational Technology, 55*(6), 2503–2529. https://doi.org/10.1111/bjet.13460

Dugosija, T. (2024). Benefits and challenges of artificial intelligence in English language teaching. *Knowledge International Journal, 62*(2), 275–280. https://ikm.mk/ojs/index.php/kij/article/view/6590

Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in Education: Opportunities, Challenges, and Ethical Considerations. Magna Scientia Advanced Research and Reviews, 10(2), 006–013.

Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2023). *Artificial Intelligence and English Language Teaching: Preparing for the Future.* British Council.

Feng, Haiying, Kexin Li, and Lawrence Jun Zhang. "What does AI bring to second language writing? A systematic review (2014-2024)." (2025). https://scholarspace.manoa.hawaii.edu/items/c86f4480-2a2b-45b5-80ea-576d1fb990da

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics.* Sage Publications.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence, 6*, 1162454. https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1162454/full

George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference.* Routledge.

Haswell, R. (2005). Researching Teacher Evaluation of Second Language Writing. In T. Silva &P. K. Matsuda (Eds*.), Second Language Writing Research: Perspectives on the Process of Knowledge.* Construction (pp.105–120). London: Lawrence Erlbaum Associates.

Jiang, R. (2022). How Does Artificial Intelligence Empower EFL Teaching and Learning Nowadays? A Review on Artificial Intelligence in the EFL Context. *Frontiers in Psychology, 13,* 1049401.

King, R. S. (2015). *Cluster analysis and data mining: An introduction.* Mercury Learning and Information.

Kushmar Lesia, V., Vornachev, A. O., Korobova, I. O., & Kaida, N. O. (2022). Artificial Intelligence in Language Learning: What Are We Afraid Of. *Arab World English Journal.*

Mahmoudi, A. (2017). Effect of planning on Iranian intermediate EFL learners' mastery of writing skill. *Theory and Practice in Language Studies, 7*(3), 219. https://core.ac.uk/download/pdf/266995161.pdf

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), 1–18. https://doi.org/10.1080/2331186X.2023.2236469

Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2013). *Performing data analysis using IBM SPSS.* John Wiley & Sons.

Mikeladze, T., & Meijer, P. C. (2024). Exploring the Application of Artificial Intelligence in Foreign Language Education within School Settings: Systematic Literature Review. *Journal of Education in Black Sea Region, 10*(1). https://jebs.ibsu.edu.ge/jms/index.php/jebs/article/view/328

Pratama, M. D. & Hastuti, D. P. (2024). The use of artificial intelligence to improve EFL students' writing skills. *English Learning Innovation, 5*(1), 13–25. https://doi.org/10.22219/englie.v5i1.30212

Rahman, A., Raj, A., Tomy, P., & Hameed, M. S. (2024). A comprehensive bibliometric and content analysis of artificial intelligence in language learning: Tracing between the years 2017 and 2023. *Artificial Intelligence Review, 57*(4), 107.

Setyowati, L., Sukmawan, S., & El-Sulukiyyah , A. (2020). Exploring the Use of ESL Composition Profile for College Writing in the Indonesian Context. *International Journal of Language Education, 4*(2), 171-18. https://doi.org/10.26858/ijole.v4i2.13662

Shivakumar, A., Shukla, S., Vasoya, M., Kasrani, I. M., & Pei, Y. (2019). AI-Enabled Language Speaking Coaching for Dual Language Learners. *IADIS International Journal on WWW/Internet, 17*(1).

Son, J.-B., Ružić, N. K., & Philpott, A. (2023). Artificial Intelligence Technologies and Applications for Language Learning and Teaching. *Journal of China Computer-Assisted Language Learning,* 0.

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*, 1–14. https://doi.org/10.3389/fpsyg.2023.1260843

Tkaczynski, A. (2017). Segmentation using two-step cluster analysis. *Segmentation in social marketing: Process, methods and application,* 109-125.

Tseng, Y., & Lin, Y. (2024). Enhancing English as a foreign language (EFL) learners' writing with ChatGPT: A university-level course design. *The Electronic Journal of e-Learning, 22*(2), 78–97. https://doi.org/10.34190/ejel.21.5.3329

Widiati, U., Rusdin, 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), 1–18. https://doi.org/10.1080/2331186X.2023.2236469

Woo, J. H., & Choi, H. (2021). *Systematic Review for Ai-Based Language Learning Tools.* arXiv Preprint arXiv:2111.0445

About the Contributor(s)

Tahani Salman Alangari, Department of Curriculum and Instruction, Collage of Education, Imam Abdulrahman Bin Faisal University, P.O. Box. 1982, Dammam, Saudi Arabia, Saudi Arabia, Dammam Email: tsalrajeh@iau.edu.sa