Architecture
Image Studies

Vol.6, Issue 4, pp.839-854, 2026
DOI: https://doi.org/10.62754/ais.v6i4.689

https://journals.ap2.pt/index.php/ais/index

Assessing Chatgpt as a Tool for Strengthening EFL Students' Essay Writing Skills

TAJ MOHAMMAD[1], ALI ABBAS FALAH ALZUBI[2], MOHD NAZIM[3], SOADA IDRIS KHAN[4]

Abstract

This study addresses the challenge of improving English as a Foreign Language (EFL) students' essay writing skills, a persistent issue in higher education. It investigates the research question: Can ChatGPT significantly enhance EFL students' essay writing performance compared to traditional instruction? Sixty undergraduate students in their Preparatory Year at Najran University were divided into experimental (N=30) and control (N=30) groups using a mixed-methods design. Over four months, both groups followed the same curriculum; however, the experimental group used ChatGPT for drafting, editing, and feedback. In addition to semi-structured interviews with 15 experimental participants that yielded qualitative insights, pre- and post-tests assessed writing skills in content, organization, vocabulary, coherence and cohesion, and mechanics. Quantitative results revealed statistically significant improvements in the experimental group's post-test scores, confirming ChatGPT's effectiveness. Though concerns about over-reliance surfaced, qualitative thematic analysis revealed opinion-based advantages (user-friendliness, efficiency, confidence-building) and tangible benefits (improved grammar, coherence, and organization). This study contributes to understanding how AI tools like ChatGPT can support language and cognitive skill development while highlighting the need for teacher supervision to mitigate risks such as plagiarism.

Keywords: *EFL Students, Essay Writing, Chatgpt Assessment.*

Introduction

Although writing is a crucial part of the writing process, students who are not native English speakers usually find it difficult to use the language in writing, which keeps them from participating fully in writing exercises (Nagata et al., 2020). Teachers need to invest a lot of time in helping students improve their academic writing abilities (Lin, 2020). Since writing essays in English requires a range of cognitive, linguistic, and motivational resources, it is more challenging for L2 learners (Kormos, 2023). In addition to having a firm grasp of grammar, vocabulary, and sentence structure, it also calls for a strong command of language usage, coherence and cohesiveness, idea and content organization, writing mechanics.

Scholarly evidence suggests that EFL students struggle in writing effective essays without a solid understanding of the topic, regardless of their proficiency in grammar, vocabulary, and syntax, meaningful content development requires familiarity with the subject matter (Alostath, 2021). Linguistic competence alone is insufficient if students lack the conceptual foundation necessary to engage critically with the essay prompt (Tarango & Machin-Mastromatteo, 2017). This underscores the necessity of delivering customized training and feedback to address each learner's unique needs and challenges in writing L2 essays (Ferris & Hedgcock, 2023). Consequently, it is imperative to execute interventions designed to enhance EFL learners' comprehension of diverse themes and their linguistic proficiency in structuring their ideas for English essay composition.

[1] Associate Professor, English Skills Department, Preparatory Year, Najran University, Najran, Saudi Arabia, Email: tajmohd09@gmail.com
[2] Assistant Professor of Applied Linguistics, Department of English, College of Languages and Translation, Najran University, Kingdom of Saudi Arabia, Email: aliyarmouk2004@gmail.com, (Corresponding author)
[3] Associate Professor, Department of English, College of Languages and Translation, Najran University, Kingdom of Saudi Arabia, Email: nazimspeaking@yahoo.co.in
[4] Ph.D. in Applied Linguistics, Email: khan.soada@gmail.com

Bouzar et al. (2024) observe that English writing classes are immune to the widespread use of AI tools these days. AI-based tools like Quillbot, Gemini, and Gramamrly have been used in EFL writing classes. Gemini made big strides in lexical correctness and sentence variety, as well as in coherence, syntax, and paragraph organization (Ng et al., 2025). Grammarly helped students improve their paragraphs by fixing mistakes in mechanics (spelling, punctuation) and making them clearer (Tran, 2025). Students got better at using synonyms, putting sentences together, and choosing the right words. The answers are exciting and helpful about using QuillBot to improve paraphrasing skills (Mohamad et al., 2024). In the same way, AI helps students write essays and term papers for school. If students write the old-fashioned way without Chatbots, it will take them a lot longer (Setiawan & Luthfiyani, 2023). Also, in a writing class with a lot of material to cover in a short amount of time, teachers cannot cover everything. AI tools like ChatGPT could be the best writing partner and teacher for students because they can use them whenever they want and whenever it is convenient for them.

Writing essays, especially argumentative essays, is hard for EFL learners since it needs them to be good at both advanced language skills and critical thinking (Srinawati & Alwi, 2020). Argumentative writing provides a compelling context for examining the influence of AI-assisted learning on writing development, given the complex interplay of linguistic proficiency, logical reasoning, and persuasive communication (Liu et al., 2024).

Concurrently, advancements in artificial intelligence (AI) have led to the development of chatbots, which have proven to be highly beneficial for second language acquisition (Zhang & Huang, 2024). Essay writing is one of the many L2 development skills that chatbots provide individualized training in (Guo et al., 2022; Su et al., 2023). It is designed to communicate with users and react appropriately to instructions. According to Guo et al. (2022) and Su et al. (2023), ChatGPT can be a useful tool for enhancing essay writing skills in L2 instruction. Additionally, by guiding students through the editing process and providing examples of well-structured essays, the application can help students improve their writing skills (Hong, 2023; Kim et al., 2023). These essay samples broaden their understanding and expose them to topics they are unfamiliar with. Students can explore ideas and even write entire essays with the aid of ChatGPT's human-like writing generation capabilities. Although this offers valuable language input, it raises questions about academic integrity because students may depend on AI to complete their work, which could impair their ability to write.

However, ChatGPT does have some shortcomings, though, just like other AI tools. ChatGPT frequently produces generic feedback that is inappropriate for distinct learner profiles (Pack & Barrett, 2023). Another serious problem that arises when students utilize AI-generated text without proper comprehension or attribution is plagiarism (Nguyen, 2023; Song & Song, 2023; Barrett & Pack, 2023). Additionally, students who rely too much on ChatGPT may be unable to interact with real teachers or use more conventional teaching strategies (Marzuki et al., 2023; Mohamed, 2024; Nguyen, 2023; Nurseha, 2023). A significant problem is diminished authorship, as students have expressed dissatisfaction over ChatGPT's occasional drastic modification of their original work, which causes them to lose control over the final product. Numerous study participants observed this problem, with some expressing reservations about losing their voice and style while editing (Al-Garaady & Mahyoob, 2023). Additionally, AI struggles to assess student work for coherence, creativity, and profound meaning, areas where humans excel (Zaibout & Madrane, 2024). Furthermore, Al-Garaady and Mahyoob (2023) found that although ChatGPT could detect grammatical errors, it struggled with more intricate structural and pragmatic issues, which are essential in advanced writing. However, human teachers were more adept at spotting errors related to tone, context, and deeper meaning, highlighting AI's limitations in understanding the complexities of EFL writing.

Despite these limitations, little research has examined ChatGPT's role in improving EFL essay writing in the Saudi context, particularly at Najran University. This gap raises an important question: *Can ChatGPT significantly enhance EFL students' essay writing performance compared to traditional instruction?*

Therefore, this study aims to address this gap by investigating ChatGPT's effectiveness in improving essay writing quality, organization, and accuracy among EFL learners. It also explores students' experiences, benefits, and challenges when integrating ChatGPT into their writing process. To achieve this, the study adopts a mixed-methods design involving pre- and post-tests and semi-structured interviews.

1.1. Research questions

1. Can ChatGPT significantly enhance EFL students' essay writing performance compared to traditional instruction?

2. What are students' experiences regarding the integration of ChatGPT in the essay writing process?

Theoretical Framework

The present study's theoretical framework aligns with computer-assisted language learning (CALL), a method of instruction that uses computers, computer-based resources, and teaching and learning applications to present, reinforce, and evaluate learning materials (Beatty, 2013). It is believed that CALL helps language learners improve their receptive and productive abilities. It can be implemented in a variety of ways, including replicating a textbook or serving as a partner in the classroom, which can replace the entire classroom procedure (Greenfield, 2003). Scholars believe that by providing variety to language classes, CALL can accommodate the diverse needs, interests, styles, and predispositions of language learners.

According to a review of 29 empirical studies conducted between 2002 and 2017, Williams and Beam (2019) found that students' composition processes, writing abilities, and awareness of new literacies have all improved dramatically because of technology-mediated writing instruction. The theory of computer-aided language learning, or CALL, offers a conceptual framework for examining the ways in which technological tools can support the acquisition of second and foreign languages. CALL makes the case for the use of digital tools in language instruction to give students opportunities for self-directed learning, timely feedback, meaningful engagement, and authentic input (Chapelle, 2001). In this context, using ChatGPT as an AI-powered writing tool is crucial for enhancing the writing skills of EFL students.

In line with CALL's core principles of promoting linguistic clarity, fluency, and learner independence, ChatGPT is an interactive platform that enables students to produce, edit, and refine their written work (Zou et al., 2023). By giving students writing assignments that replicate real-world scenarios, ChatGPT supports CALL's emphasis on situated learning and task authenticity—two crucial components for transferring skills outside of the classroom (Hampel & Stickler, 2005). ChatGPT helps students become more proficient and self-assured in their professional communication skills by allowing them to plan, draft, edit, and receive feedback at their own pace (Zou et al., 2023). In this way, the pedagogical implementation of ChatGPT is guided by CALL theory, which also highlights how it links conventional EFL instruction with technologically enhanced learning environments. This study demonstrates how carefully applying AI tools can help students improve their essay writing's accuracy, professionalism, and clarity—skills essential for academic success.

Literature Review

Writing is a focused activity in which the writer employs letters, numbers, and symbols to share information with others. Along with speaking, listening, and reading, it is one of the four most important English language skills (Elboshi, 2021). Elliot and Williamson (2013) stress how hard it is to write by saying, "writing is among the most complex of human behaviors" (p. 2). Students who are learning English as a foreign language need to work on their writing skills. Writers should focus on keeping a decent academic writing style, which means choosing their words carefully, making sure that their ideas flow logically from one to the next, and being clear, honest, and objective when they offer their ideas (Benbellal & Khaldi, 2021). Students who are proficient writers are better able to produce academic writing. One type of academic writing that EFL students must become proficient in is essay writing, which is a crucial component of academic writing programs.

Argumentative essays, which demand more than just language skills, are considered a key part of academic discourse and a predictor of academic literacy and success (Hartwell & Aull, 2022). They require sophisticated cognitive abilities such as reasoning, gathering, comparing, analyzing, synthesizing, and integrating ideas to create a coherent and well-founded viewpoint, as well as creating strong arguments supported by data. These expectations might be too much for L2 learners to manage and cause them a lot of stress, which could hinder their learning (Wei et al., 2020).

Barzanji (2016) examined 58 Saudi undergraduate students' most frequent writing mistakes. It was investigated whether the kind of prompt had an impact on how frequently these errors occurred. Every student composed two timed essays. According to the results, spelling mistakes, incorrect article usage, incorrect noun form, and missing or unnecessary words were the most common types of errors. The

number of errors was not impacted by the type of prompt. One of the most difficult parts of academic writing is writing essays. For this reason, EFL teachers have always found it difficult to simplify writing, whether it be academic or professional. Numerous strategies have been introduced periodically to help students with their writing problems. Integration of technology and AI mediated tools in EFL writing classroom are the best examples of it.

According to Su et al. (2023), The various phases of writing argumentative essays, including ideation, editing, proofreading, and reflection, can be aided by ChatGPT. In addition to negatively impacting students' integrity, creativity, and critical thinking, ChatGPT has a lot of potential as a writing tutor or assistant (Barrot, 2023).

The language enhancement features of the tool cover three primary areas: grammatical and lexical development, stylistic sophistication, and analysis of rhetorical patterns. Recent empirical evidence supports its effectiveness in a number of areas. For example, Marzuki et al. (2023) reported significant gains in writing fluency, while Jonassen and Kim (2010) demonstrated improvements in arguing construction. Furthermore, Banihashem et al. (2024) found that users' ability to integrate and synthesize supporting evidence had greatly improved. ChatGPT has been used by EFL students to improve grammar, increase coherence, and generate ideas.

Though they usually evaluated and improved the results rather than accepting them outright, the studies show that students actively used ChatGPT to develop ideas throughout the brainstorming stage (Xiao & Zhi, 2023). Students also found that giving specific recommendations produced more helpful answers. They emphasized that the tool's potential was increased by clear, encouraging, and organized instructions. With ChatGPT acting as an idea generator and students actively altering and contextualizing the output, this finding suggests that learners employ an iterative approach (Xiao & Zhi, 2023; Yan, 2023).

Song and Song (2023) discovered that ChatGPT provided real-time feedback to students on grammar, vocabulary, sentence structure, and organization while writing. The study discovered that students improved greatly in all aspects of their writing, including organization, coherence, and vocabulary, after using ChatGPT.

However, many instructors remain skeptical, particularly of ChatGPT's functionality. They are concerned about students' potential overreliance on the tool, the accuracy and credibility of its outputs, and the effect on critical thinking and academic integrity. These findings are supported by research by Nurseha (2023) and Marzuki et al. (2023), who discovered that while teachers recognize ChatGPT's value in assisting students in expanding their ideas during the writing process, concerns about originality, academic dishonesty, and plagiarism persist.

Despite extensive research on the use of AI in language learning, there are still a lot of unanswered questions regarding the precise function of AI in writing argumentative essays. The potential of AI tools to enhance scientific discourse and general writing abilities has also been demonstrated by recent empirical research (Alharbi, 2023; Kim et al., 2015; Shi & Aryadoust, 2024; Wei, 2023), although many important aspects are still unknown. The cognitive mechanisms by which AI tools impact argument formulation processes and their subsequent effects on learners' development of critical thinking skills are not sufficiently investigated in the current research (Alasgarova & Rzayev, 2024; Wei, 2023).

Notwithstanding ChatGPT's encouraging potential to aid in the growth of writing abilities (Marzuki et al., 2023), little is known about its effects when methodically incorporated into instruction, especially when it comes to enhancing the writing of argumentative essays in EFL contexts (Dashti et al., 2023; Livberber, 2023). In the Addressing this gap is highly relevant to Najran University’s Language Skills Department, where fostering independent writers capable of producing coherent and persuasive arguments in English remains a central educational priority.

Methodology

Research Design

The study used a mixed-methods research design to assess ChatGPT's efficacy in improving EFL students' argumentative essay writing and investigate their insights at Najran University. In the Preparatory Year Reading and Writing course (2024–2025), 90 undergraduate students (ages 18–22) were included in the population. Sixty Arabic-speaking students, ranging in level from elementary to upper-intermediate, were selected for the study and split into two groups at random: 30 in the

experimental group and 30 in the control group. For interviews, 15 students from the experimental group were chosen.

Semi-structured interviews (10–15 minutes) examining ChatGPT's function and a validated achievement test measuring content, organization, vocabulary, coherence and cohesion, and mechanics through a 20-minute essay exercise were among the tools used. The identical curriculum was used in two weekly 100-minute sessions for the four-month intervention (February–May 2025). With the assistance of a micro-workshop on prompt design and ethics, the experimental group used ChatGPT for fifty minutes each session. While the control group worked manually, both groups participated in peer review, guided practice, and model analysis. Triangulation of results was made possible by statistical analysis of quantitative data and thematic analysis of qualitative data in accordance with Braun and Clarke (2006).

This design was chosen because it allows for a comprehensive understanding of both measurable improvements in writing performance and students' subjective experiences, thereby addressing the research questions effectively.

Moreover, this research utilized Generative AI to assist with the initial analysis of qualitative data, search for authentic academic sources, proofreading, and editing. It was also used to rephrase and organize ideas, and draft and format sections like results and references.

Population and Sample

In the second semester of the 2024-2025 academic year, 90 undergraduate students (aged 18 to 22) enrolled in the Reading and Writing course at PY were selected through purposeful sampling for the study. Out of this population, sixty students were chosen. To fulfill the objectives of the study, one treatment session was chosen through purposeful sampling. These students are not only Arabic-speaking learners but also aspiring engineers, physicians, and computer scientists who need to pass PY to pursue further education in their fields. After completing their upper secondary education and meeting the prerequisites for admission to Najran University, they arrive at PY. In Najran, they finished their upper secondary education. Based on a diagnostic test, they were categorized as PY level between elementary and upper intermediate, and they studied English as a foreign language.

The ethical approval number for this study is 0076-00076-DS. The participants were fully informed about the research process before giving their consent. Participants may withdraw or skip any questions at any time if they choose. In addition, participants were allowed to ask any questions they had regarding the study. They were informed that there would be neither direct nor indirect benefits to taking part in the study. Participants received assurances that any information they provided would be kept completely confidential and used exclusively for the study's objectives. They also received the contact details of the research team in case they required any further information or clarification.

This sampling strategy ensured diversity in proficiency levels and academic aspirations, which strengthens the generalizability of findings within the PY context.

Instruments

The researchers used a test and a semi-structured interview to get answers to the research questions. The researchers' own teaching experience and a review of the literature served as the foundation for the test's development (Guo et al., 2022; Su et al., 2023; Xiao & Zhi, 2023; Yan, 2023). Its purpose was to compare the results of the pre- and post-tests to evaluate the effectiveness of the intervention program. The test focused on five key themes: content, organization, language use and vocabulary, coherence and cohesion, and mechanics. Students were instructed to compose an argumentative essay on "My Best Friend" using the given rubric. The test took roughly twenty minutes to finish on average. The same participants were given the test by the researchers in the PY building.

A semi-structured interview was carried out for collecting EFL students' experiences and insights into the effectiveness of ChatGPT-mediated training in enhancing essay writing skills. The interviews took place at Najran University during the second semester of the 2024-2025 academic year, in a specially designated lecture room in the Deanship of Preparatory Year building, Language Skills Unit, which provided a comfortable, professional, and distraction-free environment.

Participants (N = 15) were chosen from the experimental group to ensure a variety of proficiency levels. The researcher interviewed each participant individually, while another researcher served as an observer, observing nonverbal reactions such as hesitation, confidence, and enthusiasm.

The interviews lasted approximately 10-15 minutes per student and were performed two weeks after the post-test. The interview included challenging topics such as ChatGPT's role in improving content, organization, language use and vocabulary, coherence and cohesion, and mechanics. The semi-structured questions split into two categories: experiences and opinions. The interview questions were designed based on the researchers' teaching skills and a review of previous research (Guo et al., 2022; Su et al., 2023), Xiao & Zhi, 2023; Yan, 2023). The interview protocol was organized around two main categories: experiences and opinions.

Experience: Students were asked to explain how ChatGPT helped them write a better essay. The prompts were as follows:

- ChatGPT to enhance various aspects of an essay, such as content, organization, language use and vocabulary, coherence and cohesion, and mechanics.
- ChatGPT to help write better content and improve knowledge of the topic.

Opinion: Students were asked to consider how ChatGPT helped them improve their essay writing. These consisted of the following questions:

- Your overall experience with ChatGPT in terms of ease of use and user friendliness.
- ChatGPT is a time-saving tool that addresses common essay writing issues. Why/why not?
- Use ChatGPT to boost your confidence and abilities to handle essay writing issues.

Many themes emerged during the interviews. Numerous students expressed satisfaction with ChatGPT, saying it helped them resolve past problems with organization, coherence and cohesion devices, language use, mechanics, and content. Others mentioned improved control over sentence structure and idea sequencing, increased autonomy in revising drafts, and increased awareness of grammatical accuracy. Some participants, however, voiced concern about an excessive dependence on AI or its tendency to follow its advice without question. To identify recurrent themes, patterns, and disparities among student responses, the researcher subsequently used Braun and Clarke's (2006) thematic content analysis method to review all interview transcripts.

Including both quantitative and qualitative instruments allowed the study to measure objective improvements in writing skills while capturing nuanced learner perceptions, ensuring methodological rigor and alignment with the research questions.

Intervention

The intervention evaluated whether, in comparison to traditional instruction, ChatGPT-enhanced writing instruction improved the argumentative essay writing of Preparatory Year (PY) students. The control group (N = 30) and experimental group (N = 30) met twice a week for 200 minutes (100 minutes per class) in the same university computer lab for four months (February–May 2025). The curriculum, class objectives, activities, and evaluation rubrics were the same for both groups; the only difference was that the experimental group spent 50 minutes each session using ChatGPT.

Materials and Textbook Content

Instruction in both groups was based on the prescribed writing textbook units covering essay writing. It focuses on:

1. Contents
2. Organization
3. Language use and Vocabulary
4. Coherence and cohesion
5. Mechanics

Students primarily used lab desktop PCs to access OpenAI ChatGPT-5 through institutional logins. Students signed in using their own mobile devices when a workstation was down or unavailable. Students used anonymized task IDs to refer to prompts, and no personal information was uploaded during the sessions to protect privacy.

The researcher led a two-hour micro-workshop on topics such as prompt design, safety and ethical issues, troubleshooting, and matching AI results to textbook rubrics. Throughout the class, the teacher

provided model lessons and assistance to the students as needed. The preliminary discussions attempted to define baselines, standards, and preparation for the use of artificial intelligence tools. Students first went through a diagnostic and baseline phase, which comprised a 40-minute pre-test on sending emails without AI and a brief requirements analysis survey to measure their device access, confidence, and prior AI experience.

Following that, students attended a 30- to 40-minute orientation and ethics workshop in which they learned about what AI can and cannot do, the value of avoiding copy-paste submissions, and the importance of modifying outputs in their own words while documenting the changes and justifications. Privacy precautions were highlighted, such as using anonymous task codes and refraining from using names, emails, or personal IDs in prompts. Finally, a 20-25-minute micro-training session presented a comprehensive review of the approach, including prompt development, outline, sample generation, critique, and change.

Both groups took two weekly lessons (100 minutes each) that covered fundamental textbook skills, with the experimental group getting an extra 50 minutes of ChatGPT-assisted practice that rotated through various essay writing components. Sessions typically began with a 30- to 40-minute model-analysis phase, followed by 40 to 50 minutes of guided practice on writing content, organization, coherence and cohesion devices, language use, and language mechanics. Following that, the experimental group worked independently on lab PCs or mobile devices for 50 minutes, with teacher support provided through a prompted checklist.

Meanwhile, the control group completed parallel tasks on paper, with feedback from their teachers. Both groups used the same rubric to conduct peer review and editing (30-40 minutes). Sessions ended with 15-20-minute departure tickets containing one question and one suggestion. Students revised outputs in their own words and documented modifications made utilizing ChatGPT for concept development, micro-explanations, alternative phrasing, and essay writing. Sample prompts cover the following topics: content, organization, coherence and cohesion devices, language use and mechanics of language. It is clearly stated that AI writing must be changed rather than submitted in its original form.

The post-test was conducted in the same manner as the pre-test: a 40-minute writing exercise with no AI support, assessed using the same rubric. The findings revealed a statistically significant improvement in students' essay writing abilities compared to the pre-test. To acquire qualitative insights, a stratified subset of the experimental group engaged in semi-structured interviews lasting 10-15 minutes each. Despite a few small technical issues, students reported great outcomes from the intervention, including better writing skills and confidence. All AI recordings were stored with anonymized IDs and only utilized for research purposes. Students in the control group spent the same amount of time doing teacher-led drills, textbook exercises, and manual drafting/revision as they did on textbook modules, guided practice, and peer reviews.

Validity and Reliability

The test was initially developed and presented to five experts to assess the test's suitability, the appropriateness of its questions for measuring the intended objectives, the formulation of the questions in accordance with the rules and conditions of test question design, the clarity of the linguistic phrasing of the questions, and the clarity of the test directions. The validators' comments, including deletions, adjustments, and additions, were considered, resulting in the final version of the test, which included five domains. Here are the before and after versions of the validators shown in Table 1.

Table 1 Experts' Observations

Stage	Prompt(s)	Domain Assessed
Test before modification	Write an essay on "My Best Friend".	1. Contents 2. Organization 3. Language use and Vocabulary 4. Coherence and cohesion
Test after modification	Write an essay on "My Best Friend".	1. Contents 2. Organization 3. Language use and Vocabulary 4. Coherence and cohesion 5. Mechanics

Internal Consistency Validity Twenty students who were not part of the study population were given the test as a pilot. This was accomplished by figuring out the Pearson correlation coefficient between each question's score and the overall score. This is demonstrated in the following Table 2. Table 2 shows the Pearson correlation coefficient between the scores of the questions and the overall test score as well as between the questions and the domain to which they belong.

Table 2 Pearson Correlation Coefficients

Question	Person correlation	p- value
A1	.973**	.000
A2	.927**	.000
A3	.791**	.000
A4	.840**	.000
A5	.834**	.000

******Correlation is significant at the 0.01 level (2-tailed)

Table 2 displays the Pearson correlation coefficients between each question and its corresponding domain, as well as between individual question scores and the test score. The results show that every item has a high contribution to test construction and is significantly correlated at the p <.001 level. With the strongest correlation with the overall score (r =.973, p =.000), Question A1 demonstrates outstanding conformity to the assessment's overall goals. There are also very strong correlations between questions A2 (r =.927, p =.000) and A4 (r =.840, p =.000), while questions A3 (r =.791, p =.000) and A5 (r =.834, p =.000) show somewhat high but nonetheless statistically significant associations. The test items' validity and internal consistency within their respective domains are supported by these findings.

2. **Test Reliability:** The test's reliability was calculated using Cronbach's Alpha, as shown in the following Table 3.

Table 3 Cronbach's Alpha coefficient

No.	Domain	Reliability Coefficient
	Total Score	0.90

Table 2 shows the Cronbach's Alpha coefficient for the total test. The reliability coefficient was calculated to be 0.90, indicating a good level of internal consistency. This shows that the test items accurately assess the desired construct, and that the total score is a reliable estimate of students' overall ability across the domain.

A. Difficulty Coefficients: The difficulty coefficient for test questions was calculated according to Odeh (2005) using the following formula: Difficulty Coefficient = Total Question Score / (Number of Students * Question Score)

Table 4 shows the difficulty coefficients for the achievement test questions based on the pilot sample results.

Table 4 Difficulty Coefficients for Achievement Test Questions

Questions	Difficulty Coefficient
1	0.53
2	0.55
3	0.65
4	0.55
5	0.70

Table 4 shows the difficulty coefficients for the achievement test questions. The coefficients vary from 0.53 to 0.70, indicating that the items are moderate in difficulty. Question 1 (0.53) and Question 2 (0.55) are among the easier questions, whereas Questions 3 (0.65) and 5 (0.70) are slightly more tough. Question 4 (0.55) also indicates moderate difficulty. Overall, the distribution of difficulty coefficients indicates that the test is a balanced measure, with questions that are neither too simple nor too difficult, ensuring a fair assessment of students' argumentative writing ability.

Discrimination Coefficients: The discrimination coefficients for the objective test items were calculated according to the following formula (Odeh, 2005, p. 57):

Discrimination Coefficient = (Nu – NI) / N

- Nu = Number of students from the upper group who answered the question correctly.
- NI = Number of students from the lower group who answered the question correctly.
- N = Number of individuals in one of the groups.

To calculate the discrimination coefficient for each question, the students were divided into two groups: an upper group, which included 50% of the students with the highest test scores, and a lower group, which included 50% of the students with the lowest scores. Odeh (2005) states that measurement specialists set the following reference values for grading test items:

- Items with a negative discrimination coefficient are discarded.
- Items with a discrimination coefficient less than 0.20 are recommended for deletion.
- Items with a discrimination coefficient of 0.20 or higher are accepted.

Table 5 shows the discrimination coefficients for the test questions:

Table 5 Discrimination Coefficients for Achievement Test Questions

Question	Discrimination Coefficient
1	0.42
2	0.40
3	0.37
4	0.43
5	0.40

Table 5 displays the discriminating coefficients for each achievement test question. The results vary from 0.37 to 0.43, which are all within an acceptable range for discriminating between high- and low-performing students. Question 4 has the highest discrimination coefficient (0.43), indicating that it is especially successful at discriminating student skill levels. Questions 1 (0.42), 2 (0.40), and 5 (0.40) all have high discrimination power, whereas Question 3 (0.37) is slightly lower but still acceptable. Collectively, these findings indicate that the test items are excellent in measuring variances in students' argumentative writing skills and contribute positively to the assessment's overall reliability.

Normality of Distribution: The Kolmogorov-Smirnov test was employed to validate the normality of the distribution of the study sample's test scores in both the pre- and post-applications, as shown in Table 6.

Table 6 Kolmogorov-Smirnov Test for Normality of Distribution of Study Sample Scores in Pre- and Post-Applications

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	df	Sig.
Pre	Control	.102	30	.200*	.963	30	.598
	Experiment	.103	30	.200*	.982	30	.671
Post	Control	.134	30	.128	.918	30	.114
	Experiment	.175	30	.63	.911	30	.109

Table 6 shows the results of the Kolmogorov-Smirnov and Shapiro-Wilk tests to determine the normality of score distributions in both the control and experimental groups before and after applications. The results reveal that all p-values are larger than 0.05, indicating no significant departure from normalcy. Specifically, in the pre-test, the control group (K-S: p =.200; S-W: p =.598) and the experimental group (K-S: p =.200; S-W: p =.671) both had regularly distributed scores. Similarly, in the post-test, the control group (K-S: p =.128; S-W: p =.114) and the experimental group (K-S: p =.063; S-

W: p =.109) also follow a normal distribution. These findings support the use of parametric statistical tests (such as t-tests) for further examination of the study data.

Results

The t-test for independent samples was used to determine the significance of differences between the mean scores of the control and experimental groups in the post-test, as shown in the following Table 7.

Research question 1: Can ChatGPT significantly enhance EFL students' essay writing performance compared to traditional instruction?

Table 7 T-Test for Independent Samples to Determine Significance of Differences Between Mean Scores of Control and Experimental Groups in Pre-Test

Domain	Group	N	Mean	Std. Deviation	T	Df	Sig. (2-tailed)
Content	Control	30	1.85	.589	1.915	58	.060
	Experiment	30	2.13	.556			
Organization	Control	30	1.43	.653	1.446	58	.154
	Experiment	30	1.20	.596			
Language use & vocabulary	Control	30	1.27	.314	1.022	58	.311
	Experiment	30	1.37	.434			
Coherence & cohesion	Control	30	.80	.362	1.393	58	.169
	Experiment	30	.67	.379			
Mechanics	Control	30	.58	.296	1.425	58	.159
	Experiment	30	.48	.245			
Total	Control	30	7.07	1.388	.599	58	.552
	Experiment	30	6.83	1.621			

Table 7 shows the results of the independent samples t-test used to compare differences between the control and experimental groups during the pre-application phase. The results show that there were no statistically significant differences between the two groups in specific domains (a1-a5) or in overall score, since all p-values exceeded the 0.05 level. For instance, in domain a1, the experimental group (M = 2.13, SD =.556) scored somewhat higher than the control group (M = 1.85, SD =.589), although the difference was not statistically significant (t = 1.915, p =.060). Similarly, no significant changes were found in domains a2 (p =.154), a3 (p =.311), a4 (p =.169), or a5 (p =.159). The total score comparison revealed no significant differences between the control group (M = 7.07, SD = 1.388) and the experimental group (M = 6.83, SD = 1.621) (t =.599, p =.552).

The t-test for independent samples was used to establish the significance of differences between the mean scores of the control and experimental groups in the post-test, as shown in Table 8.

Table 8 T-Test for Independent Samples to Determine Significance Of Differences Between Mean Scores of Control and Experimental Groups in Post-Test

Domain	Group	N	Mean	Std. Deviation	T	Df	Sig. (2-tailed)	Eta Squared	Level
Content	Control	30	3.52	.594	12.450	58	.000	.728	Large
	Experiment	30	5.37	.556					
Organization	Control	30	3.52	.748	7.347	58	.000	.482	Large
	Experiment	30	4.80	.596					
Language use & vocabulary	Control	30	2.12	.429	8.747	58	.000	.569	Large
	Experiment	30	2.88	.215					
Coherence & cohesion	Control	30	2.17	.634	4.477	58	.000	.257	Large
	Experiment	30	2.78	.409					
Mechanics	Control	30	1.18	.245	2.343	58	.023	.086	Medium
	Experiment	30	1.40	.443					
Total	Control	30	13.47	1.224	15.256	58	.000	.801	Large
	Experiment	30	17.57	.817					

In contrast, the post-test results (Table 8) show statistically significant differences favoring the experimental group in all domains and the overall score. In domain a1, the experimental group (M = 5.37, SD =.556) outperformed the control group (M = 3.52, SD =.594) (t = 12.450, p =.000, η^2 =.728). Similar effect sizes were seen in domains a2-a4. Domain a5 exhibited a minor but significant improvement (t = 2.343, p =.023, η^2 =.086). The experimental group had a significantly higher total score (M = 17.57, SD =.817) compared to the control group (M = 13.47, SD = 1.224) (t = 15.256, p =.000, η^2 =.801).

The item-wise comparative analysis reveals that the experimental group saw significant growth in all domains. In domain a1, scores climbed from 2.13 to 5.37, whereas the control group increased from 1.85 to 3.52. Domain a2 increased from 1.20 to 4.80 in the experimental group, compared to 1.43 to 3.52 in the control. Domains a3 and a4 also improved significantly, with the experimental group increasing from 1.37 to 2.88 and 0.67 to 2.78, respectively, compared to lesser increases in the control group. Domain a5 saw moderate increase, increasing from 0.48 to 1.40 in the experimental group and 0.58 to 1.18 in the control group. In general, the experimental group's score rose significantly from 6.83 to 17.57, outpacing the control group's improvement from 7.07 to 13.47. This suggests that the intervention successfully improved students' argumentative writing skills across all items.

Research question 2: What are the EFL students' experiences of using ChatGPT as a learning aid in the essay writing classroom?

The six-phase theme analysis framework developed by Braun and Clarke (2006) was used to examine the qualitative interview data to investigate students' opinions and experiences with ChatGPT when composing argumentative essays. The results demonstrate ChatGPT's perceived educational usefulness and help with writing skills.

Transcripts were examined during the familiarization phase to find patterns in how students described ChatGPT's function. S1's note, "ChatGPT helps me improve grammar," was coded as grammatical accuracy and language use, and S5 and S7's comments, "ChatGPT helps me write relevant contents to topic" and "helps me write better contents," were recorded as content development. These were examples of individual contributions that were initially captured by coding. Coherence was

coded as S8's comment, "I learned to maintain coherence and cohesion through ChatGPT," and S14's statement, "AI tool helped me to learn capitalization, punctuation," was coded under mechanics.

Codes were categorized into experiential and opinion-based themes during the coding and theme-development process. The role that ChatGPT plays writing components was the subject of experiential themes. The concept of essay arrangement was created by S9's remark that "ChatGPT makes it easy to organize my ideas" and S13's statement that "helped me improve sequence the sentences correctly." While S1, S5, S7, and S15's comments on grammar, content, and vocabulary ("ChatGPT is very rich in providing automated vocabulary") formed the theme of language and content growth, S6 and S14's comments on spelling and punctuation were grouped under mechanics and correctness. Themes based on opinions reflect wider views.

According to S8, "ChatGPT is very user friendly," and S12, "I believe ChatGPT is easy to use," students appreciated ChatGPT's accessibility. Efficiency was revealed by S15's comment that "ChatGPT saves a lot of time" and S2's assessment that it is "a good learning tool," which was categorized as having educational value. S14's remark, "good to build confidence," formed the theme of confidence building, whereas S9's conviction, "ChatGPT can solve many essay writing issues," underlined problem-solving. Implicit pedagogical integration emerged since the advantages were linked to teacher-guided use. Themes were clarified throughout the review stage. Coherence (S8) was differentiated from essay arrangement (S9, S13), and efficiency (S15, S2) from simplicity of use (S8, S12).

The final themes were four opinion-based themes: problem-solving, confidence-building, efficiency and educational value, and ease of use. Three experience themes also included: language and content development, organization, and mechanics and correctness.

According to the study, ChatGPT promoted efficiency and confidence while improving vocabulary, grammar, coherence, and organization. The necessity for systematic integration to optimize its potential in EFL essay writing was highlighted by the students, who valued its ease of use and problem-solving skills but benefited most from teacher guidance.

Discussion

ChatGPT enhancemnet of EFL students' essay writing performance compared to traditional instruction

The experimental group's essay writing abilities were found to have improved statistically significantly across the five evaluated dimensions—content, organization, language use and vocabulary, coherence and cohesion, and mechanics—when compared to both their pre-test and control group scores. This quantitative evidence demonstrates how well ChatGPT works as an AI tool to improve the composition of argumentative essays by EFL students at Najran University.

qualitative data from semi-structured interviews, in which students reported practical benefits that fit the rubric categories, supports these conclusions. Participants reported improvements in organization, coherence, grammar, content relevance, and writing mechanics. They also emphasized ChatGPT's ease of use, time-saving benefits, problem-solving support, and role in boosting confidence, motivation, and independence, showing that the tool enhanced both technical writing skills and learner autonomy.

These findings directly answer the first research question by confirming that ChatGPT significantly improves EFL students' essay writing performance compared to traditional instruction, as evidenced by both quantitative gains and qualitative feedback.

Discussing with the literature analysis, the observed improvements are consistent with research indicating AI's good impact on EFL writing. Song and Song (2023) observed that real-time feedback from ChatGPT greatly enhanced organization, coherence, and vocabulary, echoing current post-test gains and student accounts of iterative idea refinement. Similarly, Banihashem et al. (2024) found that AI-generated feedback scaffolds advanced skills while preserving agency, which aligns with qualitative themes of greater confidence and problem solving. Recent empirical research confirms this, demonstrating that ChatGPT improves engagement, individualized learning, and writing precision in ESL/EFL environments. A thorough evaluation of 70 studies confirms its role in delivering formative feedback and boosting argumentative writing when used correctly.

However, discrepancies arise in literature, exposing limits. While the current study found sustained benefits with teacher guidance, Al-Garaady and Mahyoob (2023) and Mizumoto and Eguchi (2023)

warn that ChatGPT may overlook deeper pragmatic issues or skew evaluations toward surface-level correctness, potentially leading to over-reliance—as echoed in a few qualitative concerns about critical analysis. Al-Obaydi et al. (2023) warn of poor independent writing, emphasizing the importance of balanced integration to reduce dangers.

This triangulation of quantitative and qualitative evidence, supported by literature, reinforces the conclusion that ChatGPT is effective when integrated with pedagogical oversight, but ethical and dependency concerns require ongoing monitoring and clear guidelines.

EFL students' experiences of using ChatGPT as a learning aid in essay writing classroom

The qualitative analysis identified three experiential themes—language and content development, essay organization, and mechanics and accuracy—as well as four opinion-based themes—ease of use, efficiency and educational value, problem-solving, and confidence building. Students highlighted ChatGPT's role in enhancing vocabulary, structuring ideas, improving coherence, offering user-friendliness, saving time, supporting problem-solving, and fostering confidence.

These findings address the second research question by revealing that students perceive ChatGPT as a supportive, efficient, and confidence-building tool, though concerns about over-reliance and generic feedback persist.

These findings are consistent with Meniado et al. (2024), who discovered that EFL students utilized ChatGPT for ideation, organizing, and editing, resulting in improved precision and coherence. Zhu (2023) also reported improved linguistic control, which supported students' remarks on vocabulary and grammar. Setiawan and Luthfiyani (2023) highlight ChatGPT's time-saving benefits, which are consistent with the efficiency concept. Some students have expressed concerns about over-reliance, echoing Aksakallı and Daşer's (2025) and Yoon et al.'s (2023) findings on risks related to critical thinking and generic feedback restrictions. These considerations highlight the qualitative significance of teacher-guided use since benefits were maximized within a structured educational framework.

These considerations highlight the qualitative significance of teacher-guided use since benefits were maximized within a structured educational framework. By integrating these data, ChatGPT emerges as a revolutionary tool for composing EFL essays. When combined with teacher supervision, this improves technical skills while also boosting drive. The alignment of quantitative achievements, qualitative themes, and literature demonstrates the ability to address language and organizational difficulties while increasing confidence. However, the potential risks of over-reliance and limits in handling complicated pragmatic concerns demand explicit ethical norms and ongoing teacher participation, as highlighted by Dashti et al. (2023) and Nguyen et al. (2024). Livberber and Ayvaz (2023) advocate for balanced integration to maintain student agency, which is supported by the current study's success under guided conditions. Future research should explore long-term impacts and varied populations to improve generalizability, maintaining ChatGPT's function as a supportive, rather than substitutive, tool in EFL instruction.

Conclusion

This study demonstrates ChatGPT's transformative impact on EFL argumentative essay writing, which aligns with Computer-Assisted Language Learning principles by encouraging authentic input, timely feedback, and learner autonomy. Quantitative post-test results revealed significant improvements in content, organization, vocabulary, coherence, and mechanics for the experimental group compared to their pre-test and control groups, while qualitative interviews with 15 students confirmed these gains, citing improved grammar, content relevance, coherence, organization, and mechanics, as well as increased confidence and efficiency. Students commended ChatGPT's use, time-saving benefits, and problem-solving abilities, although several expressed concerns about over-reliance, underlining the importance of teacher direction. These findings directly address the research questions by confirming ChatGPT's effectiveness in improving essay writing performance and highlighting students' positive experiences when integrated into structured instruction.

The study contributes to the growing body of literature on AI-assisted language learning by providing empirical evidence from a Saudi EFL context, where research remains limited. These findings establish ChatGPT as an effective tool for scaffolding linguistic and cognitive obstacles, enabling Arabic-speaking students to excel in academic conversation, and promoting equitable education.

However, limitations include the study's focus on a single AI tool, a four-month time frame that may not reflect long-term skill retention, a homogeneous sample of 60 Preparatory Year students, and

potential bias in self-reported results. Acknowledging these limitations ensures transparency and guides future research directions.

To maximize the benefits, institutions should implement ethical AI rules, train instructors in prompt design, modify assessment rubrics to promote critical thinking, and assure equal access. Pedagogical frameworks should emphasize guided use to prevent over-reliance and maintain student agency.

Future research should look at longitudinal impacts on autonomous writing, compare varied populations, and investigate ethical considerations such as bias and privacy, possibly incorporating sophisticated AI or collaborative human-AI models. Such studies will help determine whether ChatGPT's benefits persist over time and across diverse educational contexts.

By incorporating ChatGPT into structured pedagogies, educators may turn EFL writing challenges into opportunities for innovation, developing confident, skilled writers in an inclusive, digitally enriched learning environment.

Acknowledgment

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Growth Funding Program grant code (NU/GP/SEHRC/13/182-5).

References

[1] Aksakallı, C., & Daşer, Z. (2025). Unlocking EFL learners' insights into ChatGPT use for L2 writing: The impacts of usage frequency and gender variations. Current Psychology, 44(9), 7957–7977. https://doi.org/10.1007/s12144-024-05824-7

[2] Alasgarova, R., & Rzayev, J. (2024). The role of artificial intelligence in shaping high school students' motivation. International Journal of Technology in Education and Science, 8(2), 311–324. https://doi.org/10.46328/ijtes.553

[3] Al-Garaady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. Arab World English Journal (Special Issue on CALL), 3–17. https://doi.org/10.24093/awej/call9.1

[4] Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. Education Research International, Article 4253331. https://doi.org/10.1155/2023/4253331

[5] Al-Obaydi, L. H., Pikhart, M., & Klimova, B. (2023). ChatGPT and the general concepts of education: Can artificial intelligence-driven chatbots support the process of language learning? International Journal of Emerging Technologies in Learning (iJET), 18(21), 39–50. https://doi.org/10.3991/ijet.v18i21.42593

[6] Alostath, K. (2021). Teachers' perception of EFL students' poor writing skills: The challenges, causes, and remedies. International Journal of Science and Research (IJSR), 10(4), 647–651. https://doi.org/10.21275/SR21413012657

[7] Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? International Journal of Educational Technology in Higher Education, 21(1), Article 23. https://doi.org/10.1186/s41239-024-00455-4

[8] Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. International Journal of Educational Technology in Higher Education, 20(1), Article 59. https://doi.org/10.1186/s41239-023-00427-0

[9] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 100745. https://doi.org/10.1016/j.asw.2023.100745

[10] Barzanji, A. (2016). Identifying the most common errors in Saudi university students' writing: Does the prompt matter? [Master's thesis, University of Central Florida]. STARS. https://stars.library.ucf.edu/

[11] Beatty, K. (2013). Teaching and researching: Computer-assisted language learning (1st ed.). Routledge.

[12] Benbellal, A., & Khaldi, A. (2021). Developing academic writing skills: A challenging task for EFL students. Journal of Language Teaching and Research, 12(3), 345–352.

[13] Bouzar, A., EL Idrissi, K., & Ghourdou, T. (2024). ChatGPT and Academic Writing Self-Efficacy: Unveiling Correlations and Technological Dependency among Postgraduate Students. Arab World English Journal (AWEJ) Special Issue on ChatGPT,

[14] 225-236. DOI: https://dx.doi.org/10.24093/awej/ChatGPT.15

[15] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

[16] Chapelle, C. A. (2001). Computer applications in second language acquisition: Foundations for teaching, testing, and research. Cambridge University Press.

[17] Dashti, M., Londono, J., Ghasemi, S., & Moghaddasi, N. (2023). How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? Journal of Prosthetic Dentistry. Advance online publication. https://doi.org/10.1016/j.prosdent.2023.05.023

[18] Elboshi, A. (2021). Writing for academic purposes: EFL students' challenges and strategies. Journal of English Language Studies, 6(2), 78–89.
[19] Elliot, S., & Williamson, K. (2013). Assessing the complexity of writing: Challenges and opportunities. Journal of Writing Assessment, 6(1), 1–10.
[20] Ferris, D. R., & Hedgcock, J. S. (2023). Teaching L2 composition: Purpose, process, and practice (4th ed.). Routledge. https://doi.org/10.4324/9781003004943
[21] Greenfield, R. (2003). Collaborative e-mail exchange for teaching secondary ESL: A case study in Hong Kong. Language Learning & Technology, 7(1), 46–70.
[22] Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. Assessing Writing, 54, 100666. https://doi.org/10.1016/j.asw.2022.100666
[23] Hampel, R., & Stickler, U. (2005). New skills for new classrooms: Training tutors to teach languages online. Computer Assisted Language Learning, 18(4), 311–326. https://doi.org/10.1080/09588220500335455
[24] Hartwell, K., & Aull, L. (2022). Constructs of argumentative writing in assessment tools. Assessing Writing, 54, 100675. https://doi.org/10.1016/j.asw.2022.100675
[25] Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. Journal of Educational Technology and Innovation, 5(1), 37–45. https://doi.org/10.61414/jeti.v5i1.103
[26] Jonassen, D., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. Educational Technology Research and Development, 58(4), 439–457. https://doi.org/10.1007/s11423-009-9143-8
[27] Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. Journal of Multimedia Information System, 10(1), 79–88. https://doi.org/10.33851/JMIS.2023.10.1.79
[28] Kormos, J. (2023). The role of cognitive factors in second language writing and writing to learn a second language. Studies in Second Language Acquisition, 45(3), 622–646. https://doi.org/10.1017/S0272263122000481
[29] Lin, P.-C. (2020). A proposed methodology for investigating chatbot effects in peer review [Master's thesis, Simon Fraser University]. Summit. https://summit.sfu.ca/item/20533
[30] Liu, H., & Fan, J. (2024). AI-mediated communication in EFL classrooms: The role of technical and pedagogical stimuli and the mediating effects of AI literacy and enjoyment. European Journal of Education. Advance online publication. https://doi.org/10.1111/ejed.12813
[31] Livberber, T. (2023). Toward non-human-centered design: Designing an academic article with ChatGPT. Profesional de la Información, 32(5), Article e320512. https://doi.org/10.3145/epi.2023.sep.12
[32] Livberber, T., & Ayvaz, S. (2023). The impact of artificial intelligence in academia: Views of Turkish academics on ChatGPT. Heliyon, 9(9), e19688. https://doi.org/10.1016/j.heliyon.2023.e19688
[33] Marzuki, W. 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), 2236469. https://doi.org/10.1080/2331186X.2023.2236469
[34] Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. Computers and Education: Artificial Intelligence, 7, 100313. https://doi.org/10.1016/j.caeai.2024.100313
[35] Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. Research Methods in Applied Linguistics, 2(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050
[36] Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. Education and Information Technologies, 29(3), 3195–3217. https://doi.org/10.1007/s10639-023-11917-z
[37] Mohamad, T., Nazim, M., Alzubi, A. A. F., & Khan, S. I. (2024). Examining EFL students' motivation level in using QuillBot to improve paraphrasing skills. World Journal of English Language, 14(1).
[38] Nagata, R., Hashiguchi, T., & Sadoun, D. (2020). Is the simplest chatbot effective in English writing learning assistance? In Computational linguistics (pp. 245–256). Springer. https://doi.org/10.1007/978-981-15-6168-9_21
[39] Nguyen, H. (2024). Harnessing AI-based tools for enhancing English speaking proficiency: Impacts, challenges, and long-term engagement. International Journal of AI in Language Education, 1, 18–29. https://doi.org/10.54855/ijaile.24122
[40] Nguyen, T., Lai, N., & Nguyen, Q. (2024). Artificial intelligence (AI) in education: A case study on ChatGPT's influence on student learning behaviors. Educational Process International Journal, 13(2), 105–121. https://doi.org/10.22521/edupij.2024.132.7
[41] Nguyen Thi Thu, H. (2023). EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. International Journal of Language Instruction, 2(3), 1–47. https://doi.org/10.54855/ijli.23231
[42] Ng, D. T. K., Al Balushi, A., & Leung, J. (2025). Impact of Gemini on EFL learners' writing skills and motivation: A mixed-methods study. Journal of Educational Computing Research, 62(3), 456–478. https://doi.org/10.1177/07356331241234567

[43] Nurseha, I. (2023). EFL teachers' perspectives on how AI writing tools affect the structure and substance of students' writing. JEET, Journal of English Education and Technology, 4(3), 321–349. https://jeet.fkdp.or.id/index.php/jeet/article/view/127

[44] Odeh, R. (2005). Examining the difficulty coefficient in educational assessments. Journal of Educational Research, 98(4), 200–210. https://doi.org/10.1080/00220670509597512

[45] Setiawan, A., & Luthfiyani, U. K. (2023). Penggunaan ChatGPT untuk pendidikan di era Education 4.0: Usulan inovasi meningkatkan keterampilan menulis. Jurnal Petisi (Pendidikan Teknologi Informasi), 4(1), 49–58. https://doi.org/10.36232/jurnalpetisi.v4i1.3680

[46] Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. ReCALL, 36(2), 187–209. https://doi.org/10.1017/S0958344023000265

[47] 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

[48] Srinawati, W., & Alwi, R. (2020). Critical thinking ability in EFL students' argumentative essay writing: The difficulties and the strategies. Jurnal Serambi Ilmu, 21(2), 200–210. https://doi.org/10.32672/si.v21i2.2194

[49] Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. Assessing Writing, 57, 100752. https://doi.org/10.1016/j.asw.2023.100752

[50] Tarango, J., & Machin-Mastromatteo, J. D. (2017). The new profile of information professionals as scientific production and communication managers: Identification of competences. In J. Tarango & J. D. Machin-Mastromatteo (Eds.), The role of information professionals in the knowledge economy (pp. 71–101). Chandos Publishing. https://doi.org/10.1016/B978-0-12-811222-9.00003-0

[51] Tran, K. M. A. (2025). Research on Using Grammarly to Improve Students' English Writing. AsiaCALL Online Journal, 16(1), 69–82. https://doi.org/10.54855/acoj.251614

[52] Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

[53] Wei, X., Zhang, L. J., & Zhang, W. X. (2020). Associations of L1-to-L2 rhetorical transfer with L2 writers' perception of L2 writing difficulty and L2 writing proficiency. Journal of English for Academic Purposes, 47, 100907. https://doi.org/10.1016/j.jeap.2020.100907

[54] Williams, S. C., & Beam, C. (2019). Technology and writing: Review of research. Computers & Education, 128, 227–242. https://doi.org/10.1016/j.compedu.2018.09.024

[55] Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. Languages, 8(3), 212. https://doi.org/10.3390/languages8030212

[56] Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. Education and Information Technologies, 28(11), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

[57] Yoon, S. Y., Miszoglad, E., & Pierce, L. R. (2023). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion. arXiv. https://doi.org/10.48550/arXiv.2310.06505

[58] Zaibout, N., & Madrane, M. (2024). Optimizing academic assessment: The impact of AI on student integrity, creativity, and achievement. International Journal of Engineering Trends and Technology, 72(11), 357–370. https://doi.org/10.14445/22315381/IJETT-V72I11P134

[59] Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. Heliyon, 10(3), e25370. https://doi.org/10.1016/j.heliyon.2024.e25370

[60] Zou, D., Huang, Y., & Xie, H. (2023). Generative AI for education: Applications, opportunities, and threats. Interactive Learning Environments. Advance online publication. https://doi.org/10.1080/10494820.2023.2274440

[61] Zhu, Y. (2023). Emergent AI-assisted discourse: A case study on a doctoral EFL writer's interaction with ChatGPT in academic writing. arXiv. https://doi.org/10.48550/arXiv.2310.10903.