Assessing the Impact of ChatGPT on EFL Students' Writing Productivity and Proficiency

Mona A. Alwasidi

College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Khaloud S. Al-Khalifah

College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

Abstract—With an emphasis on its effects on students' writing productivity and proficiency, this study investigates the effects of incorporating ChatGPT, an AI-based tool, into English as a Foreign Language (EFL) course. The research was conducted with 52 Saudi female university students and employed a pretest-posttest design to evaluate changes in key writing skills, including topic sentence formation, organization, use of supporting details, mechanics, and word count. Results revealed noticeable improvements in students' writing proficiency, particularly in generating effective topic sentences and providing related details, while the impact on mechanics and organization was less pronounced. The findings highlight ChatGPT's potential as a helpful tool for beginning EFL learners, assisting with idea generation, reducing redundancy, and fostering writing clarity. However, the study underscores the need for judicious application to avoid overreliance and ensure students develop essential writing skills independently. These insights contribute to advancing AI integration in EFL contexts and align with Saudi Vision 2030's emphasis on technological innovation in learning environments.

Index Terms—ChatGPT, AI, writing productivity, writing proficiency

I. INTRODUCTION

The fluency and mastery of English as a Foreign Language (EFL) have emerged to be important, especially for people hoping to advance their studies in foreign countries or pursue a highly competitive and professional career. Writing is one of the necessary components to become proficient in any language. Therefore, imparting writing skills to EFL learners has been highly dependent on classroom sessions using the appropriate methods and tools. But there is a change for the better with the advancement of AI technologies in educational institutions assisting in this Fourth Industrial Revolution. An example of such is ChatGPT, which is a language model built by OpenAI. It has the capabilities of text generation, dialoguing, and providing instant feedback, thus making it a great addition to the current ways of teaching languages.

In situations where native speakers are difficult to find or where there is no scope for individualized teaching, such as that experienced by many EFL students, ChatGPT serves as a constant conversational partner practicing and interacting with students in a way that wouldn't have been possible before. In this research paper, we consider the question of the effectiveness of ChatGPT usage as a means to increase the productivity and accuracy of writing in EFL learning contexts. By considering cases of EFL students interacting with this AI tool, we aim to make sense of how technology is changing the development and learning of languages in the current era.

There has been much research on the impact of technology in general and on the development of writing skills in particular on language acquisition. Ahmadi (2018) cites studies showing this correlation, saying that the effective application of new technologies enhances learners' language learning skills. Specifically, Zhao (2023) has maintained that the quality of English compositions is improved by digital writing devices. Also, tools like ChatGPT, which are incorporated with AI technologies, can also enhance the learning of languages. It is claimed that ChatGPT possesses such traits as "generating human-like responses to user requests" (Nguyen-Trung et al., 2023).

However, although ChatGPT is gaining popularity in EFL writing classes, students' writing productivity and proficiency are not broadly covered in the literature. Moreover, some studies suggested that some positive aspects of implementing ChatGPT may include facilitating students' idea generation and improving the efficiency of procedures in the writing tasks (Akpur & Enes, 2024). Other studies expressed doubts regarding the reliability and authenticity of the information provided by AI (Oh et al., 2023) and its limitations in detecting complex writing errors (Nguyen, 2023; Algaraady & Mahyoob, 2023).

This gap in the existing body of knowledge warrants an examination that takes into consideration the EFL writing instruction with ChatGPT that is encouraged for beginners and relatively inexperienced people who haven't sufficiently shaped their own writing style as yet. All of the aforementioned studies have reported the general outcomes, as most of the studies conducted were qualitative, using either a self-report, interviews, or observation. There is a lack of experimental research that seeks to analyze the use of ChatGPT among its enthusiasts and non-enthusiasts in terms of output, such as essays, words per minute rate, and the number of mistakes. This study thereby creates an opportunity to conduct further research regarding this gap. In this regard, the study focuses on this problem and confirms or refutes the

vector of high-level quantitative indicators of writing when using ChatGPT for writing, as well as when writing drafts or essay outlines without using AI. This research will also help improve the practice of integrating AI writing tools into the learning and writing process of beginner students as it fills in the gap regarding the empirical data on the impact of these tools on the students' writing skills development.

II. REVIEW OF LITERATURE

In the process of mastery over any foreign language, the improvement of writing skills has been one of the most significant and, at the same time, complex tasks for EFL learners. Attribution to this process includes a low range of language proficiency, such as vocabulary, choosing appropriate words, grammar, organization, and comprehension of style and structure (Sun et al., 2024). Additionally, anxiety concerning writing is another factor that can affect EFL learners in their written work (Sari et al., 2024). Syntax complexity stands out as another invisible barrier in writing quality for EFL students (Wang et al., 2023). Models such as collaborative writing, flipped learning models, peer assessment, as well as some integrated new technologies, like AI applications, wikis, and computer-aided argument mapping are said to have a positive impact on writing among EFL learners (Liu et al., 2022; Al-Rashidi et al., 2022; Yesuf & Anshu, 2022; Dai et al., 2023; Alsmari, 2022). Targeting these factors and difficulties would enable educators to augment EFL students in their writing performance and effectiveness.

Technology in EFL learning has been incorporated to improve writing skills on various levels. Starting from basic components like word processors to sophisticated AI systems, different technologies have been developed. Lately, it has been reported that AI has some positive impact on EFL learning as it offers learners a range of opportunities to learn through creating their writing and speaking (Lira-Gonzales et al., 2023). Also, software such as automated writing assessment software, feedback software, machine translation software, and automatic writing software have also been reported to assist in improving learners' written academic language in classrooms (Roe et al., 2023). In this regard, the use of AI tools may be beneficial since they allow for individualized feedback, tailored instruction, and active engagement that may help alleviate the majority of the problems encountered by EFL students.

As a result of AI's evolution, several instruments and programs have been developed that can help students who study EFL increase writing productivity and proficiency. One of these technologies is ChatGPT – an artificial intelligence created by an American organization named Open AI (Khan et al., 2023). ChatGPT proved to predict the next word within the chain of the words so that even a bunch of content could be channeled into complete and grammatically sensible pages (Cox & Tzoc, 2023). ChatGPT has been proven to help students by providing them with appropriate feedback on their writing (Qadir, 2023). Any student can benefit from the aid of AI writing tools for proofing, editing, language check, and content generation, which enable opportunities to enhance their writing skills and generate quality content (Mudawy, 2024).

When it comes to incorporating AI tools within writing instruction, Escalante (2023) focused on how students' writing performance would improve, if AI-based feedback was employed over the university EFL students who received feedback from ChatGPT with those who received feedback from human tutors. The findings suggested that feedback given to students using AI made a positive contribution to their writing. Furthermore, Almelhes (2023) pointed out that chatbots driven by artificial intelligence are effective in situations where there aren't enough qualified human tutors to be able to teach the second language sufficiently.

Arab postgraduate students in India have reported benefits associated with academic tasks or writing tasks in the span of language through the use of ChatGPT (Kovačević, 2023). Moreover, the tool has been observed to be motivating for students' learning, especially the younger users, which shows its great possible effect on their education (Mohammed et al., 2023). All these findings suggest the clear improvement that ChatGPT can bring in enhancing the writing productivity and proficiency of EFL students and have also maintained the need to encourage the students to perfect their prompts and be accustomed to the tool for maximum output.

What separates ChatGPT from other conversational machines is that it is specifically tailored to sound more passive in order to enhance accessibility for the users. Because of this focus on dialogue, it is also able to deal with more conversationally oriented prompts more soundly (De Angelis et al., 2023). In addition, ChatGPT represents a new generation of text generators that are not only sophisticated but also imaginative, which allows it to be used in different contexts, such as in teaching (Zhang, 2023). There are studies that show the possibility of ChatGPT being utilized by EFL students in performing better in their writing tasks due to providing timely feedback that encourages students to revise their work and increase their writing quality as well as efficiency (Boudouaia et al., 2024).

There have been several studies that incorporated ChatGPT for educational purposes, mainly in writing involving ChatGPT. Alneyadi's (2023) quantitative study records substantial gains in the academic performance and creative potential of students when ChatGPT is employed. On the other hand, Roman et al. (2023), Abouammoh et al. (2023), and Mohammed et al. (2023) offer qualitative data that allows them to present the value of ChatGPT for time management, depth of guidance, and the success of students in writing-related educational tasks. While there are good indications that ChatGPT will be effective in assisting learners as well as teachers, other aspects such as critical thinking, ethics, and comprehension of complex ideas center the need to use ChatGPT judiciously in education. Since AI tools are relatively new in the domain of the EFL classroom by offering a unique combination of opportunities and challenges, further studies are needed to enhance understanding of these interventions including AI tools.

It is clear that a knowledge gap exists in terms of quantitative research on writing production and editing within the EFL context. Moreover, the effects of these tools on learners of different levels of English have been researched, but only to a certain degree. This research aims to adopt a quantitative approach in assessing the effectiveness of ChatGPT in writing tasks in an EFL context. This study aims to fill these gaps by employing a quantitative approach to assess the impact of ChatGPT on EFL students' writing productivity and proficiency. By comparing writing tasks completed with and without the help of ChatGPT and analyzing the quality and quantity of the outputs, this research will provide practical evidence on the tool's effectiveness. The research participants will be students from an EFL college who will be able to evaluate the effectiveness of the different functions of ChatGPT, and in doing so, they will enhance their understanding to develop an AI framework that meets the educational standards and requirements of the EFL environment. Accordingly, this research will provide empirical insights into the effectiveness of ChatGPT, helping educators and developers improve AI tools for varied learning needs in EFL contexts.

III. METHODOLOGY

A. Research Design and Questions

The study uses a one-group pretest-posttest design to track how integrating ChatGPT affects EFL students' writing productivity and proficiency. So, quantitative measures were used to answer the following research questions
  - Does the integration of ChatGPT affect students' writing productivity?
  - How can ChatGPT, as an AI-assistant tool, improve Saudi EFL learners' writing proficiency?

B. Participants

Fifty-two EFL students participated in this study. Their IELTS scores range from band 4 to band 5.5. Therefore, they can be described as independent, limited, and moderate language users. The study was conducted during writing classes in an English Language and Literature Department at a Saudi university. The course mode of instruction was a traditional one where students had three contact hours a week with their instructor. All participants were female learners studying academic writing for the first time during the first year of their bachelor's degree.

C. Instrument

Pre/Post Test:

A definition paragraph topic that requires students to define "A good friend" was administered in the pre/post-test. Tests were specific to students' composition of their first drafts. The writing topic of the pre/post-test was reviewed by two professors in the applied linguistics major to ensure learners' acquaintance with ideas, vocabulary, and transitions. So, defining "a good friend" in the pre/post-test has been approved as an appropriate tool that can measure what it is supposed to measure.

The definition paragraph rubric was provided by RCampus.com, a website that allows teachers to create their own rubrics or choose from pre-existing ones to help with student work evaluation and feedback. The rubric also enabled students to know the performance expectations and have the ability to understand their grades in a meaningful way. Its validity and reliability have been established through expert review and pilot testing. According to the rubric, students' writing performance was graded based on four criteria: topic sentence, details, organization, and mechanisms; and each criterion was scaled from 2 pts (unsatisfactory) to 5 pts (exemplary).

To avoid the pre-test effect on the post-test, a time interval between tests is essential. Concerning estimating reliability, Brown et al. (2007) suggested that students should not have remembered how they answered the first time if an exam were to be repeated somewhere between three and six weeks. Accordingly, nine weeks of the intervention program in this study would be sufficient for learning, practice, and for students to forget nearly all of the ideas and sentences of their pretest.

Students used pen and paper to handwrite both tests in the classroom, where mobiles and computers were not allowed. Therefore, the originality of students' written paragraphs was guaranteed, and no plagiarism was anticipated. To assure the marks awarded are impartial, consistent and reliable, a second internal writing instructor double-marked 20% of the pre and post written paragraphs using the same assessment rubric. The second marking process is completed independently where the second reviewer determined the mark without knowing the grade awarded by the first marker. Results showed high agreement between the two raters' final holistic score by up to 92%.

D. Procedures

The study started at the beginning of the second semester of the academic year 2024 and ended by the end of the semester. So, it continued for about three months. The first week was dedicated to explaining the general format and rules of academic writing, examining modal paragraphs, and having some exercises about writing mechanics. To reinforce students' understanding of the academic writing rules, they were also directed to work individually and collaboratively on a number of activities in the Learning Management System. In the second week, the elements, structure, and purpose of the definition paragraph were explained, and modal samples were illustrated and analyzed. In the same week, students started conducting their pretest. In the following week, students were introduced to ChatGPT and required to upload it for the purpose of using it before writing any in-class assignment. Learners were given about 30 minutes to seek the

assistance of the chatbot to generate ideas for their writing, organize their thoughts, or even warm up before writing. Then, students were allowed an hour to hand-write their paragraphs. Using the same procedure, students completed writing three definition paragraphs, two opinion paragraphs, and two procedure paragraphs. After 9 weeks had elapsed, the same pretest topic was administered in the post-test.

E. Data Analysis

To assess the influence of ChatGPT on students' writing productivity and proficiency, a paired sample t-test is used to compare their scores before and after the treatment. By analyzing the pre and post-test scores, the study aims to quantify the changes in students' writing performance and determine the effectiveness of integrating ChatGPT into the writing process.

F. Ethical Considerations

Before conducting the study, the Institutional Review Board approval (IRB) was gained from the deanship of scientific research at the university. Participants' rights were respected by adhering to the ethical principles. All participants were informed about the study's purpose, the voluntary nature of their participation, and their right to withdraw at any time without penalty. Consent forms were obtained before the commencement of the study. Privacy and confidentiality of participants' data were rigorously maintained throughout the research process. The study was conducted honestly, and findings were reported truthfully without inventing or falsifying any data.

IV. RESULTS AND DISCUSSION

Results:

To answer the research questions, standard deviations and a paired sample t-test of students' results in both tests were calculated. Students' writing proficiency and productivity were tested and compared based on topic sentence, details, organization, mechanics, and word count. Each level of performance was described and represented by a descriptive label (unsatisfactory, satisfactory, good, exemplary) and a numerical value (2, 3, 4, 5).

A. Topic Sentence

Students were expected to write a clear topic statement that creates interest, concisely defines the theme, and provides a roadmap for the reader to guess the details that will follow. An assessment of the pre topic sentence showed that about 71% of the performance was either unsatisfactory or satisfactory but lacking the unique defining characteristics of the topic. However, 25% of the students wrote good topic sentences that were well-developed but not very creative. Only 4% of the pre-topic sentences were exemplary, which skillfully denotes the uniqueness of the term.

TABLE 1
TOPIC SENTENCE SCORES IN THE PRE- AND POST-TEST

	Topic Sentence															
Pre	Score	2 (unsatisfactory)				3 (satisfactory)				4 (good)				5 (exemplary)			
	N/ 52	17				20				13				2			
	%	32.7%				38.5%				25%				3.8%			
Post	Score	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
	N/ ..	3	5	5	4	0	9	5	6	0	2	5	6	0	0	1	1
	%	17.7%	29.4%	29.4%	23.5%	0%	45%	25%	30%	0%	15.4%	38.5%	46.1%	0%	0%	50%	50%

Evaluation of the topic sentence in the posttest recorded considerable developed scores. Specifically, Table 1 shows that about 31% of the topic sentences were one step developed from one level to the following one, and 29% of them were even two and three levels better. About 35% of the topic sentence scores remained consistent in the posttest, noting that most of them were satisfactory and good in the pretest. Only 5% of the topic sentence scores were less in the posttest, asserting that it was a one level decline (Table 1).

TABLE 2
MEANS, STANDARD DEVIATION, AND PAIRED SAMPLE T-TEST RESULTS OF PRE AND POST TOPIC SENTENCE

	Test	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Topic Sentence	pre	52	3.00	.863	-5.983*	51	.000
	post	52	3.90	.975			

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

Generally, Table 2 results showed statistically significant differences at (= 0.05) between the means of pre- and post-topic sentence scores (t=-5.983, P=0.000) in favor of the post-test.

B. Details/Examples

Students were expected to provide concrete and enough details and examples that support their topic sentences. These details were expected to be different including facts, anecdotes, quotes, counterarguments or any other examples that

would elaborate on the topic sentence. However, it wouldn't be enough for students to state some examples, but they need to elaborate on these examples to show how they relate to the main idea. Therefore, any inclusion of tangential information would affect the assessment of this criterion.

The supporting sentences for most of the students in the pretest were either unsatisfactory (25 students) or satisfactory but not specific or not enough to adequately define the topic (22 students). On the other hand, four students wrote good concrete examples in the pretest, and only one student provided exemplary supporting details.

TABLE 3
DETAILS/EXAMPLES SCORES IN THE PRE- AND POST-TEST

		Details/Examples															
Pre	Score	2 (unsatisfactory)				3 (satisfactory)				4 (good)				5 (exemplary)			
	N/ 52	25				22				4				1			
	%	48.1%				42.3%				7.7%				1.9%			
Post	Score	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
	N/ ..	4	15	3	3	3	8	3	8	0	0	1	3	0	0	0	1
	%	16%	60%	12%	12%	13.6%	36.4%	13.6%	36.4%	0%	0%	25%	75%	0%	0%	0%	100%

In the posttest, results also documented significant improvement in developing good examples. Out of the 25 students who got 2 (unsatisfactory) in the pretest, 15 students developed their supporting details and achieved 3 (satisfactory), 3 of them got 4 (good), and 3 students exceeded expectations to achieve 5 (exemplary). When considering those who wrote satisfactory details in the pretest, 50% of them improved this writing aspect to achieve good and exemplary scores. Moreover, 3 out of 4 students developed their supporting details from good to exemplary level (Table 3).

TABLE 4
MEANS, STANDARD DEVIATION, AND PAIRED SAMPLE T-TEST RESULTS OF PRE AND POST DETAILS/EXAMPLES SCORES

	Test	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Details/Examples	pre	52	2.63	.715	-6.945*	51	.000
	post	52	3.58	1.054			

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

Table 4 reveals a statistically significant difference at (α= 0.05) between the means of pre- and post-details/examples scores (t=-6.945, P=0.000) in favor of the post-test.

C. Organization

An exemplary paragraph organization utilizes transitional phrases that smoothly link the details and examples to the topic sentence in a clear and coherent manner. So, different logical patterns, such as the use of chronological order and order of importance, raise the writing level to high standards. Moreover, paragraph unity is assumed to be achieved through the use of grammatical and lexical elements that binds individual sentences together.

The pretest results in the following table show that writing organization was challenging for many students. About half of the students struggled to organize their thoughts, transition between ideas, or maintain the same organizational framework throughout the paragraph. The 36.5% of the students who had satisfactory paragraph organization wrote the details in discernible order but not enough to clearly define the topic in a concrete way. 11.5% of the writing organization was good, although not very engaging, and 1.9% was exemplary.

TABLE 5
ORGANIZATION SCORES IN THE PRE- AND POST-TEST

		Organization															
Pre	Score	2 (unsatisfactory)				3 (satisfactory)				4 (good)				5 (exemplary)			
	N/ 52	26				19				6				1			
	%	50%				36.53%				11.53%				1.9%			
Post	Score	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
	N/ ..	8	12	5	1	1	5	10	3	0	1	3	2	0	0	0	1
	%	30.8%	46.2%	19.2%	3.8%	5.2%	26.3%	52.6%	15.9%	0%	16.7%	50%	33.3%	0%	0%	0%	100%

However, students' writing organization has improved through little instruction and feedback, with lots of ChatGPT practice. 46.2% of the unsatisfactory organization in the pretest has improved to a satisfactory level, 19.2% to a good level, and 3.8% to an excellent level. In addition, more than half of those whose organization was assessed as satisfactory in the pretest have also developed the organization to reach a good level and 15.9% of them to an exemplary level. However, out of the six students whose writing organization was good in the pretest, only 2 students developed this writing aspect to attain an exemplary level. The one well-organized paragraph in the pretest maintained the same score in the post-test. Although percentages in Table 5 highlight a significant improvement in the general writing organization, a considerable number of students did not improve this aspect, especially those good ones.

TABLE 6
MEANS, STANDARD DEVIATION, AND PAIRED SAMPLE T-TEST RESULTS OF PRE AND POST ORGANIZATION SCORES

	Test	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Organization	pre	52	2.65	.764	-7.125*	51	.000
	post	52	3.46	.959			

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

Table 6 shows there are statistically significant differences at (α= 0.05) between the means of pre and post organization scores (t=-7.125, P=0.000), in favor of the posttest. When compared to topic sentence and details criteria, organization was the less improved writing aspect.

D. Mechanics

The technical aspects of students' writing were also examined before and after using ChatGPT. So, excellent writing should be free from spelling, grammar, punctuation, and even paper-formatting errors. Evaluation of the writing mechanics also involves the use of a combination of simple, compound, and complex sentences. Although simple sentences might be effective and safe haven for beginners, yet they make the writing style appear overly straightforward and lacking variety. Thus, using some sophisticated sentences would enable writers to communicate more complex thoughts and show a more substantial mastery of language.

Table 7 shows that 38.5% of the students had frequent and distracting spelling, capitalization, punctuation, and grammar mistakes in the pretest. 48.1% of the writing mechanics were satisfactory as few spelling and grammar errors were detected and caused little meaning distraction. Six students communicated the topic with effective writing mechanics in the pre-test but did not reach the perfection. Only one student wrote a perfect paragraph with accurate spelling, proper punctuation, correct formatting, and sound grammar.

TABLE 7
MECHANICS SCORES IN THE PRE- AND POST-TEST

Mechanics																	
Pre	Score	2 (unsatisfactory)				3 (satisfactory)				4 (good)				5 (exemplary)			
	N/ 52	20				25				6				1			
	%	38.5%				48.1%				11.5%				1.9%			
Post	Score	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
	N/ 52	12	5	3	0	4	14	6	1	0	3	2	1	0	0	0	1
	100 %	60%	25%	15%	0%	16%	56%	24%	4%	0%	50%	33.3%	16.7%	0%	0%	0%	100%

In the posttest, results also documented significant improvement in formulating words with better spelling and sentences with more efficient structure. Of the 20 unsatisfactory results, 8 students improved their writing mechanics to attain satisfactory and good results. A small number of the students who obtained satisfactory scores in the pre-test have improved to good and exemplary levels [7/25]. Out of the six students whose writing mechanics were good in the pretest, only one student developed this aspect to achieve an exemplary score. From these results, it can be observed that the effect of the treatment was getting less as the writing mechanics level improved in the pre-test.

TABLE 8
MEANS, STANDARD DEVIATION, AND PAIRED SAMPLE T-TEST RESULTS OF PRE AND POST MECHANICS SCORES

	Test	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Mechanics	**pre**	52	2.77	.731	-2.615*	51	.012
	post	52	3.10	.955			

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

Table 8 shows statistically significant differences at (α= 0.05) between the means of pre- and post-mechanical scores (t=-2.615, P=0.012) in favor of the post-test. However, from the above tables, it can be observed that the mechanics criterion was the least influenced by the experiment. 58% of the students whose writing mechanics were below the good level did not develop their grammar and spelling after using ChatGPT.

E. Word Count

As beginners, students were instructed to write short paragraphs with a title, a topic sentence, three supporting sentences that relate to the controlling idea, a detail that elaborates on each supporting sentence, and a concluding sentence. Using concise, academic language is also mandatory, where they should avoid redundancy and any unnecessary filler words to keep their paragraphs tight and impactful.

	Test	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Word Count	Pre	52	119.27	41.911	1.051*	51	.298
	Post	52	110.35	32.301			

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

Table 9 shows no statistically significant differences at (= 0.05) between the means of pre- and post-word count (t=-1.051, P=0.298). The mean word count in the pre-test was 119 words, and 110 words in the post-test. So, students wrote paragraphs that were nearly the same length before and after using ChatGPT.

Discussion:

This research evaluated how ChatGPT affected writing productivity and writing proficiency of the Saudi EFL students on five parameters: topic sentence construction, details/examples, organization, mechanics, and word count. Several remarkable patterns emerge about the effect ChatGPT has on the performance of students' writings. The first research question assessed the outcome of ChatGPT on the writing productivity of students. Generating a certain amount of text within a particular time limit was often an important measure of writing output and, hence, productivity.

A word count used to be considered one of the main elements in writing tests. For teachers, setting a word count range can enable students to communicate their ideas effectively and succinctly. It also guarantees that all students have the same opportunity to demonstrate their knowledge and writing skills. For them, standardized word count would also improve students' time management skills: by planning the time they need to allocate for each element, students can manage their time efficiently.

On the other hand, the majority of EFL learners believe that more words entail a clear and expressive piece of writing. However, this is not necessarily the case. Clarity, coherence, and relevance are more important factors in determining a paragraph's quality than its word count. When comparing the word count in the pre and post-test, results showed insignificant difference, yet students provided piece of writings with better quality in the post-test.

The pretest is characterized by extensive use of filler words that add little to the sentence's meaning, affecting the writing formal tone and reflecting the writer's internal monologue. Such focus on quantity creates sentences with obscured meaning that distract readers and make the writing unclear. In the posttest, students' writing becomes more concise, and the overuse of unnecessary words is eliminated.

Redundancy is another aspect of students' pretest writing. Most of the students repeated ideas, words, and phrases unnecessarily for the sake of reaching a specific word count. This redundancy obscures the main idea and reduces clarity in a way that bore and disengage readers. In the post-test, students became more mindful of redundancy and conveyed their thoughts more directly. This improvement proves that ChatGPT affects the quality of students' writing production in a way that fewer redundant ideas and filler words are used. However, the more chatting students have with the chatbot, the more knowledge and writing skills they acquire. So, writing quality rather than quantity will be their primary concern. This improvement in productivity aligns with findings by Mohammed et al. (2023), who found that Arab postgraduate students reported significant improvements in their academic writing productivity when using ChatGPT. Similarly, Abouammoh et al. (2023) noted that participants experienced significant timesaving and effort-saving benefits.

One of the study's main findings is that the lower the writing quality is in the pretest, the more notable improvement is in the posttest. However, these improvements involve all the tested writing aspects, and most of them are one level further. The quantitative improvements in our study are parallel to those found by Alneyadi (2023), who reported significant enhancements in post-test scores when using ChatGPT in writing. This suggests that ChatGPT can be more helpful to learners with low writing level than those with average and high levels.

The second research question investigated how ChatGPT improved the students' writing proficiency across four other measures: construction of the topic sentence, supporting sentences and details/examples, organization, and mechanics. Most of the topic sentence analysis revealed an increase in mean scores, indicating the results' significance after incorporating ChatGPT into the writing sessions. The increase in mean scores from 3.00 to 3.90 (t=-5.983, p<.001) indicates that students benefited substantially from ChatGPT assistance in developing their topic sentences. Improving on this aspect was the inability of students to perform at an acceptable level, which decreased from 32.7% to 17.7%, which means that ChatGPT was quite beneficial for those students who could not even perform basic topic sentence construction.

The observed improvements can be attributed to several key factors. ChatGPT's ability to provide immediate, specific feedback on sentence structure and clarity, as noted by Boudouaia et al. (2024), likely played a crucial role in helping students improve their topic sentences. Furthermore, ChatGPT gave learners different effective models of topic sentences because of the limited chances of coming across real examples which EFL students are said to face (Rahmatunisa, 2014). Apart from the decrease in writing anxiety when using AI assistance, a factor pointed out by Sari et al. (2024) could also have assisted students in improving their opening statements.

Some overall analyses brought about even more significant improvements concerning students' abilities to give more detail and relevance to examples in order to substantiate their main ideas. Statistical analysis indicated a considerable gain, with the mean scores for this construct going up from 2.63 in the pre-test to 3.58 in the post-test. This improvement has been more evident in the students who had problems in the first instances of sufficiently supporting their ideas with

adequate material during an argument. The fact that supporting details were better written can be attributed to the understanding of ChatGPT on how to present the main ideas in various ways, which is one of the problems experienced in EFL writing where students are unable to provide the right number of examples and elaboration (Sun et al., 2024).

One of the main noted problems in the pretest is that most students do not adhere to the paragraph organization principles. For example, students' supporting details are most often detached from the tone and direction proposed by the topic sentence. In addition, within the supporting sentences, some ideas do not flow clearly and logically. Most of the paragraphs are also not cohesive, lacking the grammatical and lexical linking within sentences.

However, improvements in the writing organization after using ChatGPT cannot outweigh improvements in other writing aspects. This can be attributed to the fact that organization and its elements of coherence and cohesion is an advanced step of writing skills. In the posttest, a considerable number of students' paragraphs make sense as a whole and readers can follow the writer's thoughts. Moreover, paragraph unity is sometimes achieved as a good number of the sentences relate to the controlling idea proposed in the topic sentence, and fewer irrelevant sentences are observed. Some students also varied the sentence structure, which made the paragraph more engaging and easier to read. These notable improvements suggest that writing prompts generated by ChatGPT encourage students to formulate their paragraphs logically, helping them practice organizational skills.

Compared to other writing aspects, spelling and grammar are the least writing aspects to be affected by ChatGPT. However, expectations for spelling and grammar in student writing can vary based on factors like proficiency level and the language learning settings. As beginners, students are not expected to improve their writing mechanics highly due to some exposure to proper grammar and spelling input. It would be sufficient if students' repertoire of writing mechanics is improved to the extent that they develop one level further.

However, early studies on using AI for improving EFL students' writing focused on implementing ChatGPT to proofread students' writing. Noy and Zhang (2023) found that AI tools like ChatGPT can effectively detect surface-level errors in EFL learners' writing, improving their grammatical accuracy and syntax. In the present study, students did not use ChatGPT to check their work or to receive feedback to identify their errors but to particularly enhance their repertoire of FL grammar and spelling before writing. Therefore, access to other resources that can influence the writing mechanics was prevented while writing. To assess the impact of language exposure provided by ChatGPT, students were not allowed to have access to writing assistant materials such as dictionaries, style guides, and spell-checking tools while writing their first drafts. In addition, students were notified about their grammar and spelling mistakes during the experiment but did not receive any instruction or explicit corrective feedback on their spelling and grammar.

Results of this study verifies the tool's effectiveness in improving students' general writing mechanics. This can be reflected in mimicking AI structural models and in their recognition of their common grammatical errors such as misuse of articles, improper word order, and subject-verb agreement. Results suggest that EFL students' struggle with writing correctly spelled words and well-formed sentences can be eliminated using ChatGPT. They also provide evidence that sentence structures and spelling can improve as students are exposed to additional AI templates.

V. CONCLUSION

Findings and implications:

Since ChatGPT assists students with idea generation, outlines, organization of thoughts, and grammar and spelling checks, it is a great timesaving tool for them. Although research on AI tools for language learning is promising, concerns about heavy reliance on them must be raised. Language learners, especially beginners, are at a stage where they need to learn and practice specific basic skills and not only make use of the app as a central medium of information and feedback.

The research indicates that ChatGPT and other artificial tools can, in the appropriate context, be sources of enhancement of writing performance for students. This research, from an education perspective of teaching ESL students, will help reduce the gaps in pedagogical practices and curriculum planning. By assessing the effects of ChatGPT use on the scores of student writing through a standardized rubric, the current study identifies prospects, challenges, and constraints of this technology in EFL writing instruction. This information will assist EFL teachers in knowing the correct and inappropriate contexts in which ChatGPT and other similar applications can be used to improve students' performance. Also, it could provide teachers with valuable recommendations with respect to best practices for the appropriate use of AI-based tools in EFL classrooms.

To conclude, the results of this research can be beneficial since they are in accordance with the educational objectives specified by the Saudi Vision 2030. This study supports Saudi Arabia's digital advancement in education by showcasing ChatGPT as an aid to students' writing skills development. The results are quite satisfactory in providing some ways in which teaching and learning of the English language can be improved, which is necessary in order to prepare Saudi students with the required skills to pursue further education and employment in international environments. There is also a growing body of evidence on the educational impact of AI tools, which this research seeks to address in the context of language learning domains. Such studies are particularly useful for Saudi Arabia's transition towards a knowledge-based economy since those studies provide innovative practices integrating education and technology. The resultant findings are promising as they support the integration of AI tools such as ChatGPT as means contributing towards the attainment of the Kingdom's goals on educational excellence.

Limitations:

This study does come with its share of limitations that need to be acknowledged. First, a quasi-experimental research design with an assignment of a control group can be performed for more accurate manipulation of the independent variable and to account for the possibility of generalizability. Another primary concern is the relatively small number of participants involved, which might limit how broadly our findings can be applied to the general EFL student population. Also, the study was conducted over a relatively short period, which may not fully capture long-term effects and improvements in language proficiency. Another point to consider is that there was variability in how much each participant engaged with ChatGPT, which could have influenced their individual outcomes. This variability is an important factor when considering the consistency and reliability of our results.

FUNDING

REFERENCES

[1] Abouammoh, N., Alhasan, K., Raina, R., Malki, K., Aljamaan, F., Tamimi, I., Muaygil, R., Wahabi, H., Jamal, A., Al-Tawfiq, J., Al-Eyadhy, A., Soliman, M., & Temsah, M.-H. (2023). *Exploring Perceptions and Experiences of ChatGPT in Medical Education: A Qualitative Study Among Medical College Faculty and Students in Saudi Arabi*. https://doi.org/10.1101/2023.07.13.23292624

[2] Ahmadi, D. M. R. (2018). The Use of Technology in English Language Learning: A Literature Review. *International Journal of Research in English Education*, 3(2), 115–125. https://doi.org/10.29252/ijree.3.2.115

[3] Al-Garaady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journals, Special Issue on CALL*, (9).

[4] Akpur, A., & Enes, K. (2024). Evaluating ChatGPT's Role in Assessing Turkey's Medical Tourism Trends. *Jordan Medical Journal*, 58(4).

[5] Almelhes, S. A. (2023). A Review of Artificial Intelligence Adoption in Second-Language Learning. *Theory and Practice in Language Studies*, 13(5), Article 5. https://doi.org/10.17507/tpls.1305.21

[6] Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology*, 15(4), ep448. https://doi.org/10.30935/cedtech/13417

[7] Al-Rashidi, A. H., Asif, M., Vanani, M. G., & Aberash, A. (2022). Learner-oriented assessment (LOA) practice: The comparative study of self-assessment, peer assessment, and teacher assessment on EFL learners' writing complicity, accuracy, and fluency (CAF), speaking CAF, and attitude. *Language Testing in Asia*, 12(1), 59. https://doi.org/10.1186/s40468-022-00209-x

[8] Alsmari, N. A. (2022). Integrating Computer-aided Argument Mapping into EFL Learners' Argumentative Writing: Evidence from Saudi Arabia. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(10), Article 10. https://doi.org/10.14569/IJACSA.2022.0131013

[9] Boudouaia, A., Mouas, S., & Kouider, B. (2024). A Study on ChatGPT-4 as an Innovative Approach to Enhancing English as a Foreign Language Writing Learning. *Journal of Educational Computing Research*, 07356331241247465. https://doi.org/10.1177/07356331241247465

[10] Brown, G. T. L., Irving, S. E., & Keegan, P. J. (2007). *An Introduction to Educational Assessment, Measurement and Evaluation: Improving the Quality of Teacher-Based Assessment*. Pearson.

[11] Cox, C., & Tzoc, E. (2023). *ChatGPT: Implications for academic libraries | Cox | College & Research Libraries News*. https://doi.org/10.5860/crln.84.3.99

[12] Dai, J., Wang, L., & He, Y. (2023). Exploring the effect of wiki-based writing instruction on writing skills and writing self-efficacy of Chinese English-as-a-foreign language learners. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.1069832

[13] De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Frontiers in Public Health*, 11. https://doi.org/10.3389/fpubh.2023.1166120

[14] Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. https://doi.org/10.1186/s41239-023-00425-2

[15] Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. (2023). ChatGPT - Reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences*, 39(2), Article 2. https://doi.org/10.12669/pjms.39.2.7653

[16] Kovačević, D. (2023). Use of ChatGPT in ESP Teaching Process. *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, 1–5. https://doi.org/10.1109/INFOTEH57020.2023.10094133

[17] Lira-Gonzales, M.-L., Wa Chao Chao, K., David, J., Lira-Gonzales, M.-L., Wa Chao Chao, K., & David, J. (2023). Technologies in English as a Second/Foreign Language Writing Classes. *Folios*, 58, 171–187. https://doi.org/10.17227/folios.58-16435

[18] Liu, G.-Z., Rahimi, M., & Fathi, J. (2022). Flipping writing metacognitive strategies and writing skills in an English as a foreign language collaborative writing context: A mixed-methods study. *Journal of Computer Assisted Learning*, 38(6), 1730–1751. https://doi.org/10.1111/jcal.12707

[19] Mahmud, F. A. (2023). Investigating EFL Students' Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool. *Journal of Language Teaching and Research*, 14(5), Article 5. https://doi.org/10.17507/jltr.1405.28

[20] Mohammed, A. A. Q., Al-ghazali, A., & Alqohfa, K. A. S. (2023). Exploring ChatGPT Uses in Higher Studies: A Case Study of Arab Postgraduates in India. *Journal of English Studies in Arabia Felix*, 2(2), Article 2. https://doi.org/10.56540/jesaf.v2i2.55

[21] Moqbel, M. S. S., & Al-Kadi, A. M. T. (2023). Foreign Language Learning Assessment in the Age of ChatGPT: A Theoretical Account. *Journal of English Studies in Arabia Felix*, *2*(1), 71–84. https://doi.org/10.56540/jesaf.v2i1.62

[22] Mudawy, A. M. A. (2024). *Investigating EFL Faculty Members' Perceptions of Integrating Artificial Intelligence Applications to Improve the Research Writing Process: A Case Study at Majmaah University*. OSF. https://doi.org/10.31235/osf.io/k4ub8

[23] Nguyen Thi Thu, H. (2023). EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. Nguyen, TTH (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.

[24] Nguyen-Trung, K., Saeri, A., & Kaufman, S. (2023). *Applying ChatGPT and AI-powered tools to accelerate evidence reviews*. https://doi.org/10.31219/osf.io/pcrqf

[25] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

[26] Oh, N., Choi, G. S., & Lee, W. Y. (2023). ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Annals of Surgical Treatment and Research*, *104*(5), 269-273.

[27] Qadir, J. (2023). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. *2023 IEEE Global Engineering Education Conference (EDUCON)*, 1–9. https://doi.org/10.1109/EDUCON54358.2023.10125121

[28] Rahmatunisa, W. (2014). Problems Faced by Indonesian EFL Learners in Writing Argumentative Essay. *English Review: Journal of English Education*, *3*(1), 41–49.

[29] Roe, J., Renandya, W., & Jacobs, G. (2023). A Review of AI-Powered Writing Tools and Their Implications for Academic Integrity in the Language Classroom. *Journal of English and Applied Linguistics*, *2*(1). https://doi.org/10.59588/2961-3094.1035

[30] Roman, A., Al-Sharif, L., & Gharyani, M. A. (2023). The expanding role of ChatGPT (Chat-Generative Pre-Trained Transformer) in neurosurgery: a systematic review of literature and conceptual framework. *Cureus*, *15*(8).

[31] Sari, Y. P., Quthny, A. Y. A., & Hamdani, B. (2024). "I Get Blank When I Start Writing" Analysis of EFL Student's Anxiety in Writing Descriptive Text. *Journey: Journal of English Language and Pedagogy*, *7*(1), Article 1. https://doi.org/10.33503/journey.v7i1.4016

[32] Sun, J., Motevalli, S., & Chan, N. N. (2024). Exploring Writing Anxiety during Writing Process: An Analysis of Perceptions in Chinese English as a Foreign Language (EFL) Learners. *Qualitative Research in Education*, 1–16. https://doi.org/10.17583/qre.12938

[33] Wang, W., Duan, M., & Zhang, H. (2023). Corpus-based Development of Syntactic Complexity in EFL Writing. *SHS Web of Conferences*, *152*, 04001. https://doi.org/10.1051/shsconf/202315204001

[34] Yesuf, M. Y., & Anshu, A. H. (2022). The Impact of the Use of Collaborative Writing on Attitude of EFL Students towards Learning Writing Skills. *International Journal of Education and Literacy Studies*, *10*(3), 113–120. https://doi.org/10.7575/aiac.ijels.v.10n.3p.113

[35] Zhang, B. (2023). ChatGPT, an Opportunity to Understand More About Language Models. *Medical Reference Services Quarterly*, *42*(2), 194–201. https://doi.org/10.1080/02763869.2023.2194149

[36] Zhao, X. (2023). Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal*, *54*(3), 890–894. https://doi.org/10.1177/00336882221094089

Mona A. Alwasidi born in May 25, 1985, in Riyadh, Saudi Arabia, earned a PhD in Applied Linguistics from Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, in 2020. She completed her MA in Applied Linguistics from the same institution in 2011 and her BA in English Language and Literature from IMSIU's College of Languages and Translation in 2008.

She has been a faculty member at the College of Languages and Translation at Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, since 2011. Dr. Alwasidi also served as an external reviewer for the BA English language program at princess Norah University in Riyadh. She also collaborated with King Abdullah Institute for Translation and Arabization and reviewed number of translated books. Her fields of interest are language acquisition, language teaching and learning, and sociolinguistic studies. ORCID ID: 0009-0005-0611-1195

Khaloud S. Al-Khalifah, born in March 3, 1985, in Riyadh, Saudi Arabia, earned a PhD in Applied Linguistics from Imam Muhammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, in 2020. She completed her MA in Applied Linguistics from the same institution in 2012 and her BA in English Language and Literature from IMSIU's College of Languages and Translation in 2008.

She has been a faculty member at the College of Languages and Translation at Imam Muhammad Ibn Saud Islamic University (IMSIU), located in Riyadh, Saudi Arabia, since 2009. She has also translated articles for the Information Technology page of Al-Riyadh newspaper. Dr. Al-Khalifah has published extensively, including works like The Effect of Arabic Language on Reading English for Arab EFL Learners: An Eye-Tracking Study (2011), Error Analysis of Pretrained Language Models in English-to-Arabic Machine Translation (2024), and Halwasa: Quantify and Analyze Hallucinations in Large Language Models: Arabic as a Case Study (2024). Her research focuses on EFL reading processes, language acquisition, language learning and teaching, eye-tracking studies, and machine translation models. ORCID ID: 0009-0009-1194-8909