



OPEN ACCESS

EDITED BY

Thomas Hartung,
Johns Hopkins University, United States

REVIEWED BY

Mohammad Najib Jaffar,
Islamic Science University of Malaysia,
Malaysia
Kee-Man Chuah,
University of Malaysia Sarawak, Malaysia

*CORRESPONDENCE

Reza Kafipour
✉ rezakafipour@gmail.com
Farhad Pakdel
✉ ffpakdel@yahoo.com

RECEIVED 01 July 2024

ACCEPTED 25 November 2024

PUBLISHED 11 December 2024

CITATION

Shahsavari Z, Kafipour R, Khojasteh L and Pakdel F (2024) Is artificial intelligence for everyone? Analyzing the role of ChatGPT as a writing assistant for medical students. *Front. Educ.* 9:1457744. doi: 10.3389/feduc.2024.1457744

COPYRIGHT

© 2024 Shahsavari, Kafipour, Khojasteh and Pakdel. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Is artificial intelligence for everyone? Analyzing the role of ChatGPT as a writing assistant for medical students

Zahra Shahsavari, Reza Kafipour*, Laleh Khojasteh and Farhad Pakdel*

Department of English Language, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

This study explores the potential impact of ChatGPT on the academic writing skills development of medical students enrolled in a compulsory 3-unit writing course at a medical university. The research focuses on two primary objectives, which are formulated as two research questions: Firstly, does the use of ChatGPT enhance medical students' English academic writing skills compared to conventional writing training? Secondly, how does the use of ChatGPT impact on different components of academic writing? A longitudinal intervention design was employed with 83 participants from two writing classes in the experimental and control groups. The findings demonstrated ChatGPT's significant impact on enhancing medical students' English academic writing skills, with large effect sizes. ChatGPT enhanced students' writing skills, especially content, organization, vocabulary, and mechanics in the experimental group, while its impact on language use is limited. AI tools like ChatGPT can be valuable in assisting with certain aspects of writing, but they should not be considered a one-size-fits-all solution for enhancing writing skills. The result of the study can be beneficial for educators, particularly those interested in teaching writing.

KEYWORDS

AI-assisted learning, ChatGPT, academic writing, medical students, EFL

1 Introduction

Recently, Artificial Intelligence (AI) plays a crucial role in various fields, and education is no exception (Dempere et al., 2023). AI has an effective role to improve personalization, engagement, and efficiency in language learning by providing its new methods (Alneyadi et al., 2023). In fact, authentic and interactive materials made available by AI technology allow language learners not only to interact with the target language in novel ways but also to develop their language skills more effectively and meaningfully (Alneyadi et al., 2023). Some researchers who applied AI in education believe that using AI enables students from diverse backgrounds to obtain high quality educational opportunities (e.g., Madasamy et al., 2022). Gningue et al. (2022) mention that AI has a potential to offer suitable learning materials, suggest areas to improve students' knowledge, and adjust the difficulty of learning. This functionality guarantees that each student gets the essential support to achieve his maximum learning potential by using AI tools. Harry (2023) emphasizes that a major advantage of AI in education is the ability to provide personalized learning that allows each student to learn at his own speed in a meaningful way. Nonetheless, it is crucial to tackle challenges concerning privacy, security, trust, costs, ethics, and bias.

Recently, different AI tools and platforms have been used in language learning; for example, Akyuz (2020) used Intelligent Tutoring Systems (ITS) to personalize instruction and provide feedback based on the individual needs and learning pace of students. Adaptive Learning Technologies were also employed to facilitate learning by adjusting the content and assessments based on real-time analytics of the student performance (Capuano and Caballé, 2020). Other researchers like Tomiak et al. (2020) used Assessment and Feedback Tools as an effective AI tool to automate grading and provide instant feedback on assignments. These tools help them not only in saving time but also in giving faster feedback to students. A Plagiarism Detection Tools like Turnitin was applied by other researchers to check the originality of the manuscripts and maintain academic integrity in their work (e.g., Jiffriya et al., 2021). In another study, Learning Management Systems (LMS) with AI features were applied to enhance the learning experience by offering personalized learning paths, using predictive analytics to evaluate student performance, and automating grading systems (Cavus et al., 2021). Vijayakumar (2024) used Chatbots and Virtual Assistants to assist students with questions, offer administrative support, and even help with tutoring; Chatbots can be used by educators to organize their thoughts, provide feedback on their work, code writing, and summarize the research literature (Hutson, 2022). Another invaluable AI tool which has attracted significant attention from educational scholars and practitioners across various fields is ChatGPT (Xiao and Zhi, 2023); following its release by OpenAI in November 2022, the potential of a Large Language Model (LLM) trained to mimic the statistical patterns of language in an enormous database of human-generated text combined from text in books, articles, and websites across a wide range of domains (Stokel-Walker, 2023). The purpose of this study is to discuss the prospective use of ChatGPT in writing programs at higher education levels and the future of teaching writing skills in classrooms.

1.1 The use of ChatGPT tool in education

ChatGPT can assist scientists with material organization, draft creation, and proofreading, making it a valuable tool in research and publishing (Lo, 2023). Since its release, however, ChatGPT has caused a heated debate in academia, especially about extended writing forms (e.g., essays, project reports, etc.). Many raised further queries about the purpose of using AI in education, how it is being used, at what levels (e.g., individual, collective, or transnational), where and by whom, and finally, how it works within the educational system (Imran and Almusharraf, 2023).

1.2 The role of ChatGPT in writing

In terms of assisting in writing, the introduction of ChatGPT has led to increased efficiency in generating written content, allowing students and educators to save time and focus on other aspects of their work (Lund and Wang, 2023; Yan, 2023; Pakdel et al., 2024). Kasneci et al. (2023) and Taucharungroj (2023) suggest that ChatGPT can aid idea generation by suggesting topics, themes, and perspectives that students may not have considered. This idea supports Elhossiny et al., 2022 theory of integrating psychology, thinking style, and technology.

Also, Al-Jarf (2010) highlights that one of the constant challenges ESL/EFL students face in writing is the problem of idea generation, as they often struggle to generate original and relevant content. Lametti (2022) argues that using ChatGPT can solve the above problem since ChatGPT would not kill the college essay because it would not replace 'flesh and blood authors' and that teachers and learners should enjoy working with this new technology and take chatting with ChatGPT as fun (p.3). Additionally, many researchers mention the ability of ChatGPT to translate text from one language to another, benefiting students who write in a non-native language by ensuring accuracy and grammatical correctness (e.g., Lametti, 2022; Lund and Wang, 2023; Stock, 2023). Stacey (2022) emphasizes that ChatGPT's access to vast information results in more accurate and consistent content. The latest versions of ChatGPT, according to Rasul et al. (2023) and Suaverdez and Suaverdez (2023), have the most striking ability to produce a human-like performance for various academic and professional tasks such as generating longer essays and more creative writings. In this perspective, Geher (2023), and McMurtrie (2022) discuss how ChatGPT enables improved collaboration among students and educators through simultaneous project work, proofreading, and editing capabilities, leading to enhanced writing quality and reduced errors. Research conducted by Elhossiny et al., 2022 reveal that undergraduate students improved their writing quality by incorporating ChatGPT's suggestions into argument structuring and evidence support. This integration of ChatGPT has enhanced learners' understanding of effective writing techniques, improving their writing proficiency. Numerous studies have demonstrated the capabilities of ChatGPT in generating various academic documents, including abstracts, research papers, dissertations, and essays across diverse subjects (to name as a few, Aljanabi et al., 2023; Ariyaratne et al., 2023; Gao et al., 2023). Furthermore, using ChatGPT in conjunction with DaVinci-003 has resulted high-quality essays in physics, achieving top grades in the UK higher education system (Yeadon et al., 2023).

While ChatGPT seems to have emerged as both an innovative and revolutionary tool for language education, concerns have also appeared regarding the potential risks associated with its inappropriate use such as fairness, copyright infringement, and breaches of academic integrity (Kasneci et al., 2023). According to Ahmed and Roche (2021), students using English as an Additional Language (EAL) in their studies sometimes unintentionally breach academic integrity rules when they summarize, paraphrase, and synthesize incorrectly. These rules are founded on cultural notions of text authorship and ownership (Pennycook, 1996). In university written assessments, criticisms have been attributed to EAL students' lack of awareness and command of standard academic English (e.g., Flowerdew and Li, 2007; Nejad et al., 2019). In other studies, criticisms have also been raised regarding the accuracy and reliability of the information generated by ChatGPT, as it has been reported that the data it generates is a combination of both true and entirely fabricated information (Alkaissi and McFarlane, 2023). Experiments by Bašić et al. (2023) found that ChatGPT did not significantly enhance essay grades, with issues of text authenticity and contextual understanding affecting the outcomes. The lack of human-like intuition and adaptability in chatbot responses hindered language acquisition and writing proficiency progress, particularly in ESL students (Shumanov and Johnson, 2021). Bašić et al. (2023) also conducted a study on using ChatGPT-3.5 as a writing assistance tool for students. She examined the essay-writing performances of students with and without

ChatGPT as an assistance tool. The control group (traditional essay-writing) and experimental group (ChatGPT-assisted essay-writing) received an average grade of C, with no significant differences between the two groups regarding essay scores. While the experimental group had slightly higher text unauthenticity and more potential AI-generated texts, overall essay similarity was low across the sample. The results suggest that ChatGPT did not improve essay quality, writing speed, or text authenticity. The study indicates that the effectiveness of ChatGPT-assisted writing may depend on the user's previous knowledge and skills, potentially causing confusion in inexperienced users and resulting in poorer essay performance. Similarly, Farrokhnia et al. (2023) highlight concern regarding ChatGPT's limited 'comprehension' of topics, especially when paired with students' limited knowledge, potentially resulting in unreliable outcomes. In another study, Fyfe (2023) found that students expressed concerns about being unable to identify the sources of generated text, leading to distractions during writing tasks. To address this issue, AI tools should be used as supplementary resources for writing improvement rather than complete replacements. Other studies have suggested that students often struggle with constructing coherent arguments in their writing, whether in general writing tasks or specifically in argumentative essays, so using ChatGPT may further complicate this issue (Banihashem et al., 2023; Farrokhnia et al., 2023; Ranjbaran et al., 2023).

Additional studies on students' perceptions of ChatGPT are also warranted (Yan, 2023; Zou and Huang, 2024). Imran and Almusharraf (2023) report that humanities and social sciences-related journals had the least number of related documents on ChatGPT and its role in writing tasks, while Vincent (2023) report that most literature related to ChatGPT and writing themes were published in medical journals. He reported that scholars in the USA, UK, and Australia produced more articles on the role of ChatGPT as a writing assistant. These findings highlight the need for further research on improving chatbots' contextual grasp, personalized feedback, and adaptability to cater to ESL and EFL learners' unique needs. To address some of the existing research gaps, the current study applies a longitudinal intervention design to investigate the effects of ChatGPT on L2 medical students' academic writing skills development over time. Specifically, this study is guided by the following two research questions (RQs):

RQ1: Does the use of ChatGPT enhance medical students' English academic writing skills compared to conventional writing training?

RQ2: How does the use of ChatGPT impact on different components of academic writing?

2 Methodology

2.1 Design and method

The present study investigated the effects of ChatGPT on the development of L2 medical students' academic writing skills. An experimental research design was used to compare the experimental and control groups with traditional writing instruction.

In this study, the theoretical framework is drawn from Elhossiny et al., 2022 theory, which contains three parts: psychology, thinking style, and technology. In this regard, the conceptual framework shows

how students psychologically view artificial intelligence as well as the potential impact of using AI tools like ChatGPT on their writing. From a thinking style perspective, the framework shows the potential advantages and weaknesses of incorporating ChatGPT into students' writing. A technological perspective presents how technology such as ChatGPT can affect their learning (see Figure 1).

2.2 Participants

The target population in this study was all medical students enrolled for the compulsory 3-unit academic writing course at Shiraz University of Medical Sciences (SUMS). After the participants provided their written informed consent to participate in this study, 83 students were randomly separated into the experimental group (n = 42/ two writing classes) and the control group (n = 41/ two writing classes). We just assessed the writing of 60 students since 9 students withdrew from the course and 14 others failed to submit all of their assignments for evaluation. In addition, the homogeneity tests were run to ensure students' writing abilities were equivalent in the pre-test. To rate medical students' academic writing performance before treatment, each writing sample in the pre-test was scored using Jacobs' et al. (1981) rubrics. Then, the pre-test scores of control and treatment groups were tested through ANCOVA. No statistically significant difference was found between the writing components and the overall writing scores. It is worth noting that the scores showed the writing proficiency level of the learners as intermediate. It showed that both groups were homogenous in writing before treatment (for more information on the rubrics, please see section 2.6.)

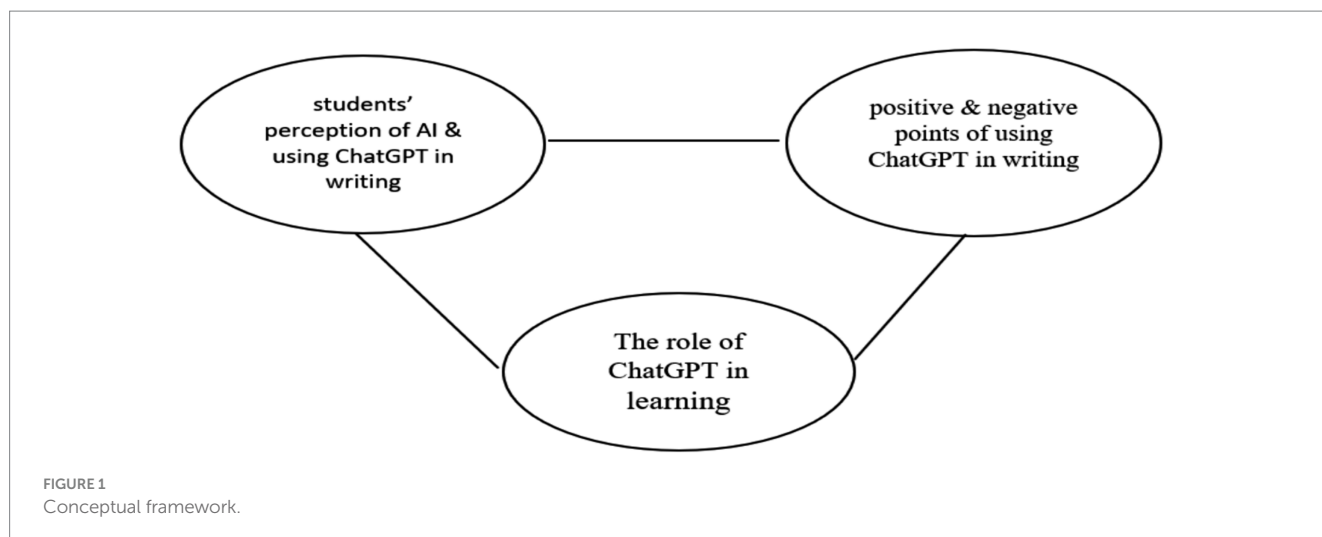
The writing instructors in both the experimental and control groups had obtained their Ph.D. (Doctor of Philosophy) degrees in English language education. They had taught English writing for more than 15 years.

2.3 Research context

The research was carried out at a medical university located in Iran. It should be noted that the university did not officially endorse the incorporation of ChatGPT in teaching, and many academics were unaware of the potential benefits offered by ChatGPT before the data collection process. However, a few writing instructors expressed concerns about plagiarism and the originality of the essays submitted by medical students as part of their course requirements. Notably, from the first author's observations, it was evident that most students had prior experience using ChatGPT. Consequently, as writing instructors for the compulsory three-unit course aimed at junior medical students, we conducted an experimental study to determine whether providing training in ChatGPT usage and covering the regular writing syllabus would enhance their writing skills.

The study was carried out within a 17-week English writing course which consisted of two sessions a week. It aimed to enhance medical students' English writing performance in various genres such as argumentative, cause-and-effect, problem-solution, and analytical essays (the genres pertinent to their final exam).

All assignment topics were the same in both groups and related to health since the students were pursuing a medical major. The writing instructors for the experimental group (2 parallel classes) incorporated



ChatGPT activities, while the instructor for the control group (2 parallel classes) applied the conventional writing approach.

2.4 Procedure

The process-oriented approach was implemented to teach writing in the control group. The writing instructors emphasized grammar accuracy, vocabulary breadth, and the model texts in the coursebook. The students were asked to write one draft each week, and the instructors gave written feedback, focusing on the lexical and grammatical aspects.

The same process-oriented approach was implemented in the experimental group to teach writing. However, in these classes, the students were trained about the usage of ChatGPT and the ways with which instructions should be provided for ChatGPT to get the maximum results. In this group, the students were asked to write their writing assignments in the class under the observation of the writing instructor; after that, they typed their homework at home and inserted it into the ChatGPT dialog box with the following instruction, "Proofread and edit the text by providing reasons for every single edition." Then, the students were instructed to submit a before-and-after snapshot of their previous work - the original version and the one edited and proofread by ChatGPT. Furthermore, the students were mandated to provide a Persian or an English explanation for each ChatGPT's suggestions (see [Supplementary Appendix A](#)). The writing instructors needed to check the students' comprehension of ChatGPT feedback which ensured them that the students were not only receiving the feedback but also understanding it. The students' explanations for ChatGPT's suggestions allowed the instructors to critically evaluate and apply the feedback provided. They believed this procedure may reinforce the learning outcomes of using ChatGPT as a writing tool and encourage the students to engage with feedback to improve their writing skills actively. Additionally, checking the students' comprehension of the feedback helped to identify any misunderstandings or challenges that the students may be facing, allowing the instructors to provide additional support or clarification as needed.

To alleviate external factors affecting learners' writing performance, biweekly pre-class teaching training sessions were

organized with the two writing instructors for the experimental group to guide them in implementing the same teaching procedures. As the writing instructor for the other two writing classes (control group) was the same, she was told to follow the syllabus and act as she was routinely doing in her writing classes.

2.5 Data collection

The data was collected within a 17-week English writing course. At the pre-test stage, a writing test, with a health-related topic, was arranged by the Department of English at this university, and the students had to finish the writing test in the classroom. Writing samples were collected to obtain baseline data from the experimental and the control group. Then, 17-week writing instructions were carried out in the experimental group. At the post-test stage, another writing test was conducted immediately after the interventions.

In both stages, medical students were required to write an essay with around 200 words within 30 min under exam conditions. Three experienced English writing instructors were consulted to assess the comparability and feasibility of the two writing topics.

2.6 Grading rubric for written assignments

To rate medical students' academic writing performance, each writing sample in the pre-test and post-test was scored using [Jacobs et al.'s \(1981\)](#) rubric. The rubric comprises five differentially weighted scales: content, organization, language use, vocabulary, and mechanics. The content is assessed through some descriptors such as knowledgeable, substantive, and thorough development of the thesis and relevant to the assigned topic. Organization is tested in terms of fluency of expression, clarity in the statement of ideas, support, organization of ideas, sequence, and the development of ideas. Vocabulary is examined in terms of the sophisticated range, effective word choice, word form mastery, and appropriate register. Language use is concerned with the use of effective complex construction, agreement, tense, number, and word order. Mechanics dealt with the attention to the use of spelling, punctuation, capitalization, and paragraphing. The scores allocated to each trait are as follows:

Content = 25, Organization = 25, Language use = 25, Vocabulary = 15, and Mechanics = 10. The total mark is 100 points.

The main reason for choosing the rubric is that [Jacobs et al. \(1981\)](#) developed this comprehensive rubric through a collaborative effort to evaluate writing quality. They reviewed existing literature on writing assessment, interviewed teachers and students, analyzed writing samples, and tested various rubric designs to determine which factors were the most relevant and important for evaluating writing. Second, although the rubric is old, it is the most commonly one that carried out in various educational settings by focusing on its validity and reliability in evaluating students' writing (e.g., [Ghanbari et al., 2012](#); [Setyowati et al., 2020](#)).

To grade students' writing, two researchers, ZSH and FP were given a set of essays to score individually. They were asked to compare their scores to see if they had given the same marks for the same compositions. If their scores differed, they were asked to discuss and explain why they had given a particular score. This helped them agree on a standard for scoring, after which they were allowed to evaluate the rest of the essays independently. Finally, the inter-rater reliability of their scoring was measured (87%), which showed that the scoring was highly reliable.

2.7 Data analysis

To address the research questions, the authors initially assessed the normality of the scores using the Kolmogorov–Smirnov test. This test yielded non-significant results, leading to the employment of parametric tests for this study. An independent sample *t*-test, a paired sample *t*-test, and an ANCOVA analysis were conducted to answer the first research question “Does the use of ChatGPT enhance medical students' English academic writing skills compared to conventional writing training?” A paired sample *t*-test was employed to reply the second research question “How does the use of ChatGPT impact on different components of academic writing?”

3 Results

Research question 1: “Does the use of ChatGPT enhance medical students' English academic writing skills compared to conventional writing training?” As an independent sample *t*-test illustrated in [Table 1](#), there was a significant difference between students' post-test scores in both groups. The mean differences indicated that the experimental group ($M = 89.20$) surpassed the control group ($M = 79.50$) in overall writing performance. The conclusion can be drawn that the application of ChatGPT, when juxtaposed with traditional writing instruction, resulted in a more pronounced enhancement in the writing skills of the learners.

Although the experimental group preformed significantly better than control group in their post-test, a paired sample *t*-test of pre-test and post-test scores in the control group indicated a significant ($p < 0.05$) improvement in the students' writing in the control group (see [Table 2](#)).

As detailed in [Table 2](#), the Standard Deviation (SD) for pre-test scores in the control group was 12.47, while for the post-test it dropped to 5.50. The initially high SD in the pre-test reflects a broader variability in student performance at the start of the course, suggesting that some students entered with stronger foundational skills than others. The subsequent reduction in SD by the post-test implies a convergence of performance levels within the group. This narrowing

TABLE 1 The comparison of post-test scores between two groups.

Group	N	Mean	SD	<i>t</i>	df	Sig.
Control	30	79.50	5.50	6.63	58	0.001
Experimental	30	89.20	5.82			

TABLE 2 The comparison of the students' writing scores in the control group.

Test	N	Mean	SD	<i>t</i>	df	Sig.
Pre-test	30	68.33	12.47	4.73	29	0.001
Post-test	30	79.50	5.50			

TABLE 3 The comparison of the students' writing scores in the experimental group.

Test	N	Mean	SD	<i>t</i>	df	Sig.
Pre-test	30	82.07	6.14	4.90	29	0.001
Post-test	30	89.20	5.82			

of the performance gap could indicate that the instructional methods employed throughout the course were effective in reducing variability, helping to standardize skill levels among students. In particular, the control group's alignment by the end of the course suggests that even without the targeted intervention, the general curriculum may have contributed to more uniform learning outcomes.

Regarding the experimental group, the learners exhibited superior performance in the post-test ($M = 89.20$) compared to the pre-test ($M = 82.07$). A paired sample *t*-test revealed this difference to be significant ($p < 0.05$), leading to the conclusion that the utilization of ChatGPT has indeed enhanced medical students' English academic writing skills. (see [Table 3](#) and [Figure 2](#)).

Moreover, to conduct a thorough comparison, an ANCOVA analysis was conducted between the control and experimental groups for each writing component. Partial eta squared (η_p^2) and Cohen's *d* were used for measuring the effect sizes for ANCOVA. The interpretation of the effect size was based on [Cohen \(1992\)](#) classification that η_p^2 values of 0.01, 0.06, 0.14; *d* values of 0.20, 0.50, and 0.80, and *r* values of 0.10, 0.30, and 0.50 were considered small, medium, and large, respectively.

As illustrated in [Table 4](#), the experimental group outperformed the control group in all writing components. However, the difference in the language component was insignificant ($p > 0.05$). Therefore, it can be concluded that ChatGPT improved the content, organization, vocabulary, and mechanics more significantly than traditional writing instruction. The effect sizes of the writing components further corroborate this finding. Regarding the effect size, the language use component exhibited a small effect size, while the content component demonstrated a large effect size. The organization, vocabulary, and mechanics components are followed with medium effect sizes. This suggests that while ChatGPT can significantly enhance certain aspects of writing, its impact on language use component is less pronounced, followed by mechanics, vocabulary, organization, and content components. This nuanced understanding of the tool's effectiveness can guide its application in educational settings.

To reply the second research question “How does the use of ChatGPT impact on different components of academic writing?” a paired sample *t*-test was employed. As revealed in [Table 5](#), there was

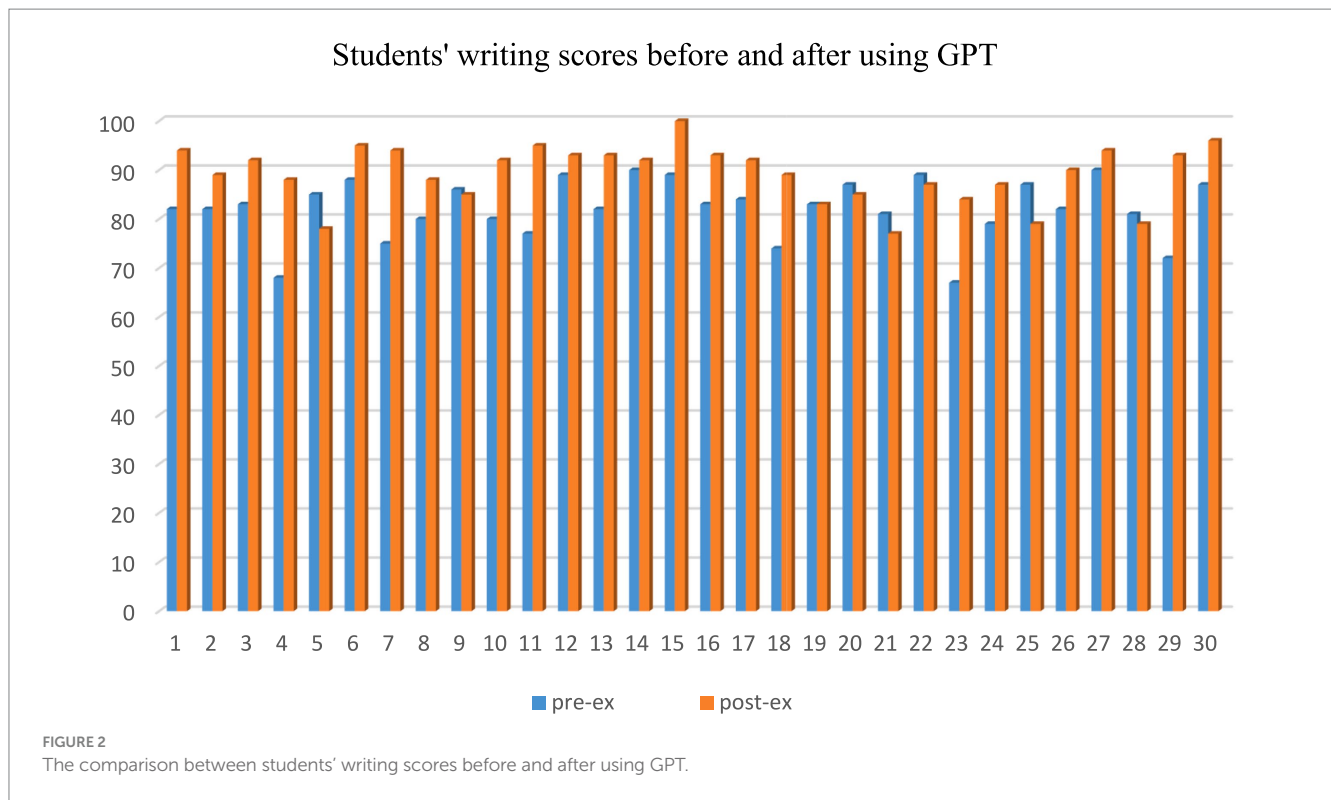


TABLE 4 ANCOVA for writing component scores between two groups in the post-test.

Writing component	Group	N	Mean	SD	df	Mean sq.	F	Sig.	Eta Squared
Content	Control	30	23.03	2.54	1	149.92	14.54	0.00	0.85
	Experimental	30	27.80	2.92					
Organization	Control	30	17.00	0.98	1	29.60	16.90	0.00	0.78
	Experimental	30	19.07	1.20					
Vocabulary	Control	30	16.00	1.53	1	18.69	10.49	0.00	0.71
	Experimental	30	17.27	1.05					
Language	Control	30	19.53	2.21	1	13.69	2.99	0.09	0.28
Use	Experimental	30	20.73	1.96					
Mechanics	Control	30	3.93	0.64	1	2.34	5.04	0.03	0.69
	Experimental	30	4.33	0.71					

a significant difference in the content, organization, and mechanics components when using ChatGPT. However, no significant difference ($p > 0.05$) was observed in the vocabulary and language use components. This suggests that while ChatGPT can significantly enhance certain aspects of writing, its impact on vocabulary and language use component is less pronounced, followed by content, organization, and mechanics components.

4 Discussion

Regarding the initial research question, the findings of this study demonstrate that employing ChatGPT has a substantial influence on enhancing medical students' English academic writing abilities. The study observed a statistically significant improvement in performance

from the pre-test to the post-test, with a large effect size. These results align with previous research emphasizing the advantages of artificial intelligence tools in language learning and improving writing. For instance, Al-Raimi et al., 2024 study reveal that automated writing evaluation tools improved students' writing skills and overall performance. Similarly, Dong, 2023 research indicate that students who utilized AI-powered writing tools exhibited significant writing proficiency and accuracy advancements. However, it is essential to acknowledge that while numerous studies have reported positive perceptions and beliefs regarding AI's impact on writing, not all have measured actual improvements in students' work. For this reason, we excluded studies (e.g., Bašić et al., 2023; Khalifa and Albadawy, 2024; Ginting et al., 2023) from our discussion, as they did not assess the actual impact of AI on students' writing performance. Our findings are consistent with Elhossiny et al., 2022 theoretical framework for using

TABLE 5 Comparison between writing component scores before and after using GPT.

Writing component	Test	N	Mean	SD	t	df	Sig.
Content	Pre-test	30	23.97	3.22	5.99	29	0.01
	Post-test	30	27.80	2.92			
Organization	Pre-test	30	17.57	1.63	4.22	29	0.01
	Post-test	30	19.07	1.20			
Vocabulary	Pre-test	30	16.73	1.41	1.59	29	0.12
	Post-test	30	17.27	1.05			
Language	Pre-test	30	19.90	1.70	1.73	29	0.09
	Post-test	30	20.73	1.96			
Mechanics	Pre-test	30	3.90	0.71	2.53	29	0.01
	Post-test	30	4.33	0.71			

AI writing assistants in medical education. It suggests that by using personalized feedback and support, AI tools such as ChatGPT can greatly enhance students' academic writing skills. Also, the observed improvement in medical students' writing abilities supports Elhossiny's argument that AI can be an effective supplementary resource in developing critical writing competencies. However, our study points out the limitations of AI's abilities, which echoes Elhossiny's caution about over-reliance on AI without adequate human oversight.

Contrary to the findings of our study, previous research has highlighted certain limitations in the effectiveness of AI tools in improving writing skills. Some of these differences can be attributed to the specific AI tools and methods used in the studies, as well as the varying contexts and populations involved. For instance, Link et al. (2022) demonstrate that AI tools can be beneficial in offering feedback on grammar and syntax but may not be as successful in enhancing higher-level writing abilities like critical thinking and argumentation. Bašić et al. (2023) conducted experiments showing that ChatGPT did not significantly enhance essay grades, with issues like text authenticity and understanding of context impacting the results. This finding could be due to the limitations of ChatGPT in addressing certain writing aspects.

Moreover, Shumanov and Johnson (2021) and Lee et al. (2020) find that a lack of human-like intuition in chatbot responses can hinder language acquisition and writing progress, particularly for ESL students. This could be because human interaction and feedback are crucial for language learners, especially those with limited proficiency in the target language. Farrokhnia et al. (2023) express concerns about ChatGPT's limited grasp of topics, which could lead to unreliable results especially when students have a restricted knowledge base. This might be due to the AI model's limitations in understanding complex or domain-specific content. Fyfe (2023) argues that students struggled to identify the sources of AI-generated text, causing disruptions during writing tasks. This could be because AI-generated content might not always be completely distinguishable from human-written text, leading to confusion and potential plagiarism concerns. Banihashem et al. (2023), and Ranjbaran et al. (2023) highlight that students often struggle with constructing coherent arguments in their writing, suggesting that the use of ChatGPT could further complicate this issue. This might be due to the AI model's inability to fully grasp the nuances and complexities of human thought processes, which seem essential for crafting well-structured arguments.

The findings of the second research question show the influence of ChatGPT on different components of academic writing. This result is consistent with previous research that has shown the potential of AI tools in enhancing language skills (Alkhaldeh and Khasawneh, 2023). The use of ChatGPT in this study resulted in a more substantial enhancement in content, organization, and mechanics components of writing, indicating its effectiveness in improving various aspects of academic writing skills. The observed enhancement in content and organization can be attributed to ChatGPT's ability to provide suggestions for structure, as supported by previous studies (Zheng and Zhang, 2019). By offering alternative ideas and organizational strategies, ChatGPT assists students in developing more well-structured essays.

But comparing the students' writing in the experimental group showed a significant difference in the content, organization, and mechanics components by using ChatGPT, while no significant difference was observed in using the vocabulary. This finding is contrary to the findings of Lund and Wang (2023), who observed that AI-based writing tools can expand students' vocabulary usage. They also believe that improving vocabulary and writing fluency can be credited to ChatGPT's capability to suggest synonyms and alternative wordings. As a result, students can express their ideas more effectively and diversify their writing style.

Likewise, it is notable that the language use component did not significantly improve when using ChatGPT compared to traditional writing instruction. This finding is supported by previous literature highlighting the challenge of AI tools in addressing higher-order language skills such as coherence and cohesion. While ChatGPT may excel in generating text based on input prompts, it may struggle with the nuances of language use and cohesion, which are crucial in academic writing (Zheng and Zhang, 2019). This suggests that a combination of AI tools and traditional writing instruction (hybrid) may be more effective in addressing all components of the students' writing skills.

Last but not least, our finding is contrary to some studies which have suggested that AI tools may not significantly improve essay grades or writing quality, with factors like text authenticity, understanding of context, and a lack of human-like intuition in chatbot responses (Farrokhnia et al., 2023). Moreover, the contradictory findings emphasize the importance of further research to better understand the limitations and potential improvements in AI-assisted writing support (Banihashem et al., 2023; Ranjbaran et al., 2023).

5 Implications of the study

The findings of this study indicate that ChatGPT can play a valuable role in supporting medical students' writing tasks. The results can help students and other members, such as professors, administrators, and researchers enhance the quality of their writing. Likewise, it assists educators in generating ideas, drafting, editing, revising any piece of writing, and developing critical writing skills by using ChatGPT which supports the integration of psychology, thinking style, and technology noted by Elhossiny et al. (2022).

6 Limitations of the study

Like any study, this study had some limitations. First, this study relied on convenience sampling that potentially compromising the

findings' generalizability. This limitation arises from participants selected for ease of access rather than representativeness. Hence, replicating the study based on random sampling would strengthen the results of this study in the future. Another limitation was the number of essays practiced in writing classes. The outcome would be different if the students were to practice more essays. Also, the students applied Chat GPT to their argumentative, cause-and-effect, problem-solution, and analytical essays. Further research could be undertaken to examine the role of Chat GPT in different types of writing, such as journalistic writing, article writing, thesis writing, or creative writing. Moreover, the dual role of the instructor-researcher introduces potential bias. Perhaps the instructors' familiarity with the participants might influence the data collection and interpretation processes, which may unintentionally lead to subjectivity that skews the results. This dual role may create power dynamics that affect participants' responses by impacting the study's internal validity. To strengthen future research, separating the instructors and researchers' roles is recommended to enhance the research objectivity and generalizability. Moreover, while AI tools have demonstrated potential in providing feedback on grammar, syntax, and text structure, they may not be as effective in improving higher-level writing abilities such as critical thinking and argumentation. This highlights the need for continuous human intervention and guidance in developing these advanced writing skills. Also, future studies could be conducted to address the identified challenges such as improving the AI models' understanding of context, domain-specific knowledge, and human-like intuition. In this study, no significant differences were found in writing components related to language use. Further research on this topic is recommended. Moreover, additional research is required to investigate how different AI tools can be combined or integrated with traditional teaching methods to optimize the students' effectiveness in language learning and writing improvement. By taking the suggested ways, we can better understand the true potential of AI in supporting students' writing development and tailor our approaches accordingly.

7 Conclusion

This study shows that using ChatGPT significantly enhanced medical students' English academic writing skills. It enhanced students' writing skills, especially content, organization, vocabulary, and mechanics, while its impact on language use was limited. The results advocated a supportive role for applying AI and reinforced Elhossiny's framework while acknowledging its limitations in educational settings. We found that AI tools like ChatGPT can be valuable in assisting with certain aspects of writing, but they should not be considered a one-size-fits-all solution for enhancing writing skills. Hence, combining AI support and human guidance may yield the most effective results in improving students' writing abilities.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study has been approved by the Ethics Committee of Shiraz University of Medical Sciences under the ethics code IR.SUMS.REC.1402.616. Written informed consent from the participants was obtained to participate in this study.

Author contributions

ZS: Data curation, Methodology, Validation, Writing – review & editing. RK: Conceptualization, Data curation, Writing – review & editing. LK: Formal analysis, Project administration, Resources, Supervision, Writing – original draft. FP: Investigation, Software, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work has received financial support from the Shiraz University of Medical Sciences Research Council through grant number 29649.

Acknowledgments

We would like to sincerely thank the Shiraz University of Medical Sciences for granting us permission to conduct this study and express our appreciation to all medical students who participated in this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2024.1457744/full#supplementary-material>

References

- Ahmed, S. T., and Roche, T. (2021). Making the connection: examining the relationship between undergraduate students' digital literacy and academic success in an English medium instruction (EMI) university. *Educ. Inf. Technol.* 26, 4601–4620. doi: 10.1007/s10639-021-10443-0
- Akyuz, Y. (2020). Effects of intelligent tutoring systems (ITS) on personalized learning (PL). *Creat. Educ.* 11, 953–978. doi: 10.4236/ce.2020.116069
- Aljanabi, M., Ghazi, M., Ali, A. H., and Abed, S. A. (2023). ChatGpt: open possibilities. *IJCSM* 4, 62–64. doi: 10.52866/20ijcsm.2023.01.01.0018
- Al-Jarf, R. (2010). Teaching medical terminology with mind-mapping software. *Lang. Lit.* 29, 1–10.
- Alkaiissi, H., and McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus* 15:e35179. doi: 10.7759/cureus.35179
- Alkhalwaleh, M. A., and Khasawneh, M. A. S. (2023). Language learning tools and improvements in speaking and writing skills. *J. Southwest Jiaotong Univ.* 58, 428–444. doi: 10.35741/issn.0258-2724.58.5.33
- Alneyadi, S., Wardat, Y., Alshannag, Q., and Abu-Al-Aish, A. (2023). The effect of using smart e-learning app on the academic achievement of eighth-grade students. *Education* 19:em2248. doi: 10.29333/ejmste/13067
- Al-Raimi, M., Al-Yafaie, Y., Al-Maashani, S., and Mudhsh, B. A. (2024). Utilizing artificial intelligence tools for improving writing skills: exploring Omani EFL learners' perspectives. *Forum Ling. Stud.* 6:1177. doi: 10.59400/fls.v6i2.1177
- Ariyaratne, S., Iyengar, K. P., Nischal, N., Chitti Babu, N., and Botchu, R. (2023). A comparison of ChatGPT-generated articles with human-written articles. *Skeletal Radiol.* 52, 1755–1758. doi: 10.1007/s00256-023-04340-5
- Banihashem, S. K., Noroozi, O., den Brok, P., Biemans, H. J. A., and Taghizadeh Kerman, N. (2023). Modeling teachers' and students' attitudes, emotions, and perceptions in blended education: towards post-pandemic education. *Int. J. Manag. Educ.* 21:100803. doi: 10.1016/j.ijme.2023.100803
- Bašić, Ž., Banovac, A., Kružić, I., and Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Human. Soci. Sci. Commun.* 10, 1–5. doi: 10.1057/s41599-023-02269-7
- Capuano, N., and Caballé, S. (2020). Adaptive learning technologies. *AI Mag.* 41, 96–98. doi: 10.1609/aimag.v41i2.5317
- Cavus, N., Mohammed, Y. B., and Yakubu, M. N. (2021). Determinants of learning management systems during COVID-19 pandemic for sustainable education. *Sustain. For.* 13, 1–23. doi: 10.3390/su13095189
- Cohen, J. (1992). Statistical power analysis. *Curr Direc Psychol Sci.* 1, 98–101.
- Dempere, J., Modugu, K., Hesham, A., and Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Front. Educ.* 8:1206936. doi: 10.3389/feduc.2023.1206936
- Dong, Y. (2023). Revolutionizing academic English writing through AI-powered pedagogy: practical exploration of teaching process and assessment. *J. High. Educ. Res.* 4:52. doi: 10.32629/jher.v4i2.1188
- Elhossiny, M., Eladly, R., and Saber, A. (2022). The integration of psychology and artificial intelligence in e-learning systems to guide the learning path according to the learner's style and thinking. *Int. J. Adv. Appl. Sci.* 9, 162–169. doi: 10.21833/ijaas.2022.12.020
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A. (2023). A SWOT analysis of ChatGPT: implications for educational practice and research. *Innov. Educ. Technol. Int.* 6, 1–15. doi: 10.1080/14703297.2023.2195846
- Flowerdew, J., and Li, Y. (2007). Language re-use among Chinese apprentice scientists writing for publication. *Appl. Linguis.* 28, 440–465. doi: 10.1093/applin/amm031
- Fyfe, P. (2023). How to cheat on your final paper: assigning AI for student writing. *AI & Soc.* 38, 1395–1405. doi: 10.1007/s00146-022-01397-z
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., et al. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digit. Med.* 6:75. doi: 10.1038/s41746-023-00819-6
- Geher, G. (2023). ChatGPT, artificial intelligence, and the future of writing. *Psychol. Today.*
- Ghanbari, B., Barati, H., and Moinsadeh, A. (2012). Rating scales revisited: EFL writing assessment context of Iran under scrutiny. *Lang. Test. Asia.* 2, 83–100. doi: 10.1186/2229-0443-2-1-83
- Ginting, P., Batubara, H. M., and Hasnah, Y. (2023). Artificial intelligence powered writing tools as adaptable aids for academic writing: insight from EFL college learners in writing final project. *Int. J. Multidiscip. Res. Anal.* 6, 4640–4650. doi: 10.47191/ijmra/v6-i10-15
- Gningue, S. M., Peach, R., Jarrah, A. M., and Wardat, Y. (2022). The relationship between teacher leadership and school climate: findings from a teacher-leadership project. *Educ. Sci.* 12:749. doi: 10.3390/educsci12110749
- Harry, A. (2023). Role of AI in education. *Injury* 2, 260–268. doi: 10.58631/injury.v2i3.52
- Hutson, M. (2022). Could AI help you to write your next paper? *Nature* 611, 192–193. doi: 10.1038/d41586-022-03479-w
- Imran, M., and Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: a systematic review of the literature. *Contemp. Educ. Technol.* 15:ep464. doi: 10.30935/cedtech/13605
- Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., and Hughey, J. (1981). Writing rubrics: development of a comprehensive writing rubric for assessing writing quality in educational settings. *Educ. Assess.* 3, 135–152.
- Jiffriya, M. A., Jahan, M. A., and Ragel, R. (2021). Plagiarism detection tools and techniques: a comprehensive survey. *J. Sci. FAS-SEUSL* 2, 47–64.
- Kasneji, E., Seřler, K., Kūchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* 103:102274. doi: 10.1016/j.lindif.2023.102274
- Khalifa, M., and Albadawy, M. (2024). Using artificial intelligence in academic writing and research: an essential productivity tool. *Comp. Methods Programs Biomed. Update* 5:100145. doi: 10.1016/j.cmpbup.2024.100145
- Lametti, D. (2022). AI could be great for college essays. *Slate.* (Accessed March 1, 2023).
- Lee, J. H., Yang, H., Shin, D., and Kim, S. (2020). Chatbots. *ELT J.* 74, 338–344. doi: 10.1093/elt/ccaa035
- Link, S., Mehrzad, M., and Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Comput. Assist. Lang. Learn.* 35, 605–634. doi: 10.1080/09588221.2020.1743323
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Educ. Sci.* 13:410. doi: 10.3390/educsci13040410
- Lund, B. D., and Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? Library Hi Tech News. Available at: <https://ssrn.com/abstract=4333415> (Accessed August 2, 2024).
- Madasamy, S. K., Raja, V., al-bonsrulah, H. A. Z., and al-Bahrani, M. (2022). Design, development and multi-disciplinary investigations of aerodynamic, structural, energy and exergy factors on 1 kW horizontal-axis wind turbine. *Int. J. Low Carbon Technol.* 17, 1292–1318. doi: 10.1093/ijlct/ctac091
- McMurtrie, B. (2022). AI and the future of undergraduate writing. *Chronicles High. Educ.* 15, 1–14. Available at: https://www.chronicle.com/article/ai-and-the-future-of-undergraduate-writing?cid=gen_sign_in
- Nejad, A. M., Pakdel, F., and Khansir, A. A. (2019). Interaction between language testing research and classroom testing practice. *Educ. Process* 8, 59–71. doi: 10.22521/edupij.2019.81.4
- Pakdel, F., Nouri Nezhad, S., and Shahsavari, Z. (2024). Shiraz medical students' attitudes toward using artificial intelligence in writing. *Sadra Med. Sci. J.* 12, 95–102.
- Pennycook, A. (1996). Borrowing others' words: text, ownership, memory, and plagiarism. *TESOL Q.* 30, 201–230. doi: 10.2307/3588141
- Ranjbaran, F., Babae, M., Akhteh Khaneh, P., Gohari, M., Daneshvar Ghorbani, B., Taghizadeh Kerman, N., et al. (2023). Students' argumentation performance in online learning environments: bridging culture and gender. *IJTE* 6, 434–454. doi: 10.46328/ijte.460
- Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., et al. (2023). The role of ChatGPT in higher education: benefits, challenges, and future research directions. *J. Appl. Learn. Teach.* 6, 41–56. doi: 10.37074/jalt.2023.6.1.29
- Setyowati, L., Sukmawan, S., and El-Sulukiyah, A. A. (2020). Exploring the use of ESL composition profile for college writing in the Indonesian context. *Int. J. Lang. Educ.* 4, 171–182. doi: 10.26858/ijole.v4i2.13662
- Shumanov, M., and Johnson, L. (2021). Making conversations with chatbots more personalized. *Comput. Hum. Behav.* 117:106627. doi: 10.1016/j.chb.2020.106627
- Stacey, S. (2022). Cheating on your college essay with ChatGPT. *Bus. Insid.*
- Stock, L. (2023). ChatGPT is changing education, AI experts say—but how? Available at: <https://www.dw.com/en/chatgpt-is-changing-education-ai-experts-say-but-how/a-64454752> (Accessed August 26, 2024).
- Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 613, 620–621. doi: 10.1038/d41586-023-00107-z
- Suaverdez, J. B., and Suaverdez, U. V. (2023). Chatbots impact on academic writing. *Global J. Busin. Integral. Sec.* 2.
- Taecharungroj, V. (2023). "What can ChatGPT do?" analyzing early reactions to the innovative AI Chatbot on twitter. *BDCC* 7:35. doi: 10.3390/bdcc7010035
- Tomiak, A., Braund, H., Egan, R., Dalgarno, N., Emack, J., Reid, M.-A., et al. (2020). Exploring how the new trustworthy professional activity assessment tools affect the quality of feedback given to medical oncology residents. *J. Cancer Educ.* 35, 165–177. doi: 10.1007/s13187-018-1456-z
- Vijayakumar, M. (2024). A study on chatbots and virtual assistants in customer engagement: a review. *Int. J. Eng. Manag. Res.* 14, 204–208. doi: 10.5281/zenodo.10791697
- Vincent, J. L. (2023). How artificial intelligence will affect the future of medical publishing. *Crit. Care* 27:271. doi: 10.1186/s13054-023-04511-9
- Xiao, Y., and Zhi, Y. (2023). "An exploratory study of EFL learners' use of ChatGPT for language learning tasks: experience and perceptions" in *Languages*, vol. 8 (MDPI: AG. Retrieved from). doi: 10.3390/languages8030212

Yan, D. (2023). How ChatGPT's automatic text generation impact on learners in a L2 writing practicum: an exploratory investigation. *Hall Open Sci.* doi: 10.35542/osf.io/s4nfz

Yeadon, W., Inyang, O.-O., Mizouri, A., Peach, A., and Testrow, C. P. (2023). The death of the short-form physics essay in the coming AI revolution. *Phys. Educ.* 58, 1–13. doi: 10.1088/1361-6552/acb2c2

Zheng, Z., and Zhang, G. (2019). Practical research of pre-service teachers' TPACK development based on design-based. *China Edu. Technol.* 389, 86–94.

Zou, M., and Huang, L. (2024). The impact of ChatGPT on L2 writing and expected responses: voice from doctoral students. *Educ. Inf. Technol.* 29, 13201–13219. doi: 10.1007/s10639-023-12397-x