

Exploring Chat GPT Potential in Flipped Classrooms for Improving Lexico-grammatical Features in Algerian EFL Learners' Essay Writing

Amina Abde lhadi¹

¹Ibn Khaldoun University, English Department, LICL Laboratory (Algeria).

Email Author: amina.abdelhadi@univ-tiaret.dz

Received: 04/2023

Published: 08/2023

Abstract

This paper attempts to investigate the potential benefits of integrating Chat GPT in flipped classrooms to enhance lexical and grammatical features in Algerian university EFL students' essay writing. The study opts for a quasi-experimental approach, employing an independent samples t-test on a randomly selected group of third year EFL students. The objective is to determine whether the proper use of Chat GPT during pre-writing and while-writing essay stages leads to a significant difference in lexico-grammatical aspects between the pretest and posttest assessments. The findings of the unpaired samples t-test demonstrate a statistically significant difference in the means between the pretest and posttest, supporting the alternative hypothesis, with a preset significance level ($\alpha = 0.05$). The data obtained in the SPSS outputs indicate that the appropriate and ethical incorporation of Chat GPT by teachers in flipped classrooms can offer EFL students the opportunity to enhance their essay writing skills by improving their lexico-grammatical competence. This study ends with recommendations and suggests further research directions to explore the implications of using AI-based tools in language learning and teaching environments.

Keywords: Artificial intelligence, Chat GPT, flipped learning, essay writing, lexico-grammatical competence.

Introduction

Chat GPT, an artificial intelligence-based language model developed by Open AI in 2022, and its capabilities in generating human-like texts have sparked intrigue among educators (Kumar, 2023; Yu et al., 2023), particularly those involved in teaching academic writing (Yan, 2023). As a teacher of academic writing, my initial exposure to this language model across various paragraphs, essay styles, article formats, reports, and summaries left me disconcerted. The prospect of assessing my students' work became a challenge, as I grappled with the potential of AI tools that entirely craft their written assignments. Additionally, I found myself pondering the strategies necessary to captivate their attention and discourage overreliance on this powerful language model. In essence, I felt both in competition with and overcome by this chatbot.

This raised a fundamental question: can ChatGPT genuinely function as an effective academic writing assistant? Chat GPT (The Generative Pre-Trained Transformer 3.5) revealed its myriad of potential applications (Rudolph et al., 2023) yet also acknowledged certain limitations (Gill & Kaur, 2023; Zhu et al., 2023). These limitations include the production of biased or inaccurate results, difficulties in controlling the tone or style of generated content, and its limitations in tasks requiring common-sense reasoning, such as understanding and resolving logical contradictions. Nevertheless, despite these drawbacks, the inevitability of using Chat GPT for authorship calls for a thoughtful exploration of how to maximize its benefits effectively and ethically instead of simply banning students from using it as a writing assistant (Su et al., 2023). Against this backdrop, the present study endeavors to shed light on the potential impact of integrating Chat GPT into the flipped classroom approach for teaching writing essay skills, more precisely the lexico-grammatical competence, and ultimately contributing valuable insights to the field of English language education. This study aims to address the following main research question:

1. Does incorporating Chat GPT into flipped classroom approaches lead to lexico-grammatical features improvement in EFL students' essay writing?

To this end, the following hypotheses have been put forth as anticipated answers to the research question:

1.1. Null Hypothesis (H₀): If teachers receive assistance from Chat GPT in the flipped classroom approach, it will NOT enhance the lexico-grammatical features in EFL students' essay writing.

1.2. Alternative Hypothesis (H₁): If teachers receive assistance from Chat GPT in the flipped teaching of writing, it will enhance the lexico-grammatical features in EFL students' essay writing.

2. Embracing Chat GPT in My Classroom

Research on the integration of Chat-GPT and the use of its generated outputs for educational purposes is still in its nascent stages. Recent studies have focused on the utilisation of Chat GPT as a powerful tool for improving programming skills (Beth, 2022; Rahmane and Watanobe, 2023; Taecharungroj, 2023), overcoming writer's block, enhancing brainstorming, and generating human like texts that are challenging to distinguish from human authored texts (van Dis et al., 2023) and consequently, restrictions have been imposed on its usage in some educational institutions (Kahlan, 2023). However, similar studies also find it imperative to acknowledge its limitations, particularly the presence of a hallucination effect, limited capabilities and security concerns that render it unsuitable for certain fields (Deng & Lin, 2022; Shen et al., 2023).

Recently, Chat GPT and academic writing enhancement has been a subject of interest, illustrating the extensive support it provides for various tasks such as outlining, content revision, proofreading, and post-writing reflection (Su et al., 2023). Similarly, Yan (2023) corroborates these findings while illuminating the concerns of L2 learners regarding the issues of plagiarism, academic dishonesty, integrity, and educational inequity. As the existing educational policies regarding AI plagiarism and the ethical use of AI tools for authorship are still ambiguous, Dergaa et al. (2023) highlight the need to engage in thorough discussions concerning the potential benefits, threats, and limitations of AI tools, placing a strong emphasis on human intelligence, creativity, skepticism, and critical thinking as pivotal components of the educational process.

Quite similar studies by Javaid et al., (2023) and Ward et al. (2023) have raised thought-provoking questions about the incorporation of ChatGPT and similar AI models in education. These studies have placed a spotlight on Chat GPT as a cutting-edge tool with the potential to revolutionize teaching and learning methods. While it can serve as both a companion to teachers and an assistant to students, the crucial need to equip both teachers and students with the necessary skills for its judicious use is duly emphasized. Despite the scarcity of research on integrating AI tools like Chat GPT into flipped learning (Lo & Hew, 2023) leaving uncertainties about whether its adoption will indeed lead to an effective shift in educational practices due to ethical and privacy concerns (Baskara, 2023), teachers may consider implementing the flipped classroom strategy, allowing students to complete certain activities without relying solely on AI tools.

3. Methodology

To ascertain the soundness of the set hypotheses of the present study, we opt for a quasi-experimental approach involving an independent samples t-test. Quasi-experimental design in terms of t-test for independent groups, as a type of inferential statistics, is used to determine if there is a significant difference between the pre-and post-test mean scores of forty (40) third year student, in the Department of English Language at Ibn Khaldoun University of Tiaret (Algeria), who were split equally into experimental (20) and control (20) groups. The data in the target sample is obtained randomly using a probability sampling; yet later, the students assigned as an experimental group were asked whether they have the necessary skills, digital tools, and access to Internet for the adoption of the treatment. The phases and the procedures attempted for conducting the quasi-experiment are as follows:

1. *The first stage (measuring the dependent variable)* was through performing a pretest to assess both unrelated groups' essay writing performance in terms of lexico-grammatical competence. The pre-test was about producing an argumentative essay with no more than 300 words about a

specific topic which is ‘Artificial intelligence assistance in essay production’ and which had to be completed in class under surveillance.

2. **The second stage** was through exposing the experimental group to a treatment which is the use of Chat GPT in pre-writing and while-writing stages (drafting) of their essays and which involved six sessions with an average of two sessions per week.

2.1. Teaching the control group: the control group is taught using the traditional ‘face to face’ learning environment, without AI assistance, and following the process-based approach to writing. The process approach is based on the notion of recursiveness (Raimes, 2010) which means teachers intervene in students’ writing during the pre-writing, while writing and post writing stages.

2.2. Teaching the experimental group: the experimental group is also taught using the process-based approach, yet the pre- and while writing stages are achieved with the assistance of Chat GPT in a flipped learning environment. Please refer to table (1) below for a detailed description of the adopted approach used along with the objectives designed for teaching the experimental group:

Writing Stages	Flipped Learning Environment	Experimental Group Instructions	Objectives (Revised Taxonomy)
Pre-writing	Pre-class	Ask ChatGPT with effective prompts to brainstorm the topic given.	Generating relevant ideas and taking notes
	In class	Review the content generated by ChatGPT	Evaluating student’s work prior to class.
	Diagnostic Assessment		
Whilst writing (Drafting)	Pre-class	-Use ChatGPT to generate and regenerate sections (introduction, arguments, and conclusion) about the topic then review them.	Selecting and Linking ideas
	In class	Edit and proofread the sections generated to outline the essay	Composing the initial draft
		Proving peer feedback after exchanging papers.	Discussing the success or failure of the essay’s structures. Editing and proofreading
Post-writing	Formative Assessment		

	Post-class	In class, write the final draft of the argumentative essay to be handed back	Creating appropriate written productions
Summative Assessment			

- The post-test was *the last stage* that occurs after the treatment. It was performed fundamentally to check the effectiveness of the manipulated independent variable (the proper use of Chat GPT) on the dependent variable (students' English lexico-grammatical competence).

To ensure that both tests are appropriate and accurately scored, some factors are considered, including mainly how and what is assessed in EFL students' essay writing. The format of both tests includes producing argumentative essays about Artificial intelligence assistance in essay production. The researcher (as a teacher) assesses the sample's post-test following the same procedures used in the pre-test, opting for an analytic assessment to evaluate separately three main aspects of lexico-grammatical competence including: (1) grammar, (2) vocabulary and (3) collocation. The table below clearly shows these elements with a Likert scale that provides five rating points to measure quality:

Elements	P	S	G	VG	E
1. Grammar					
2. Vocabulary					
3. Collocation					

1. P= Poor; 2.S= Sufficient; 3. G= Good; 4. VG= Very good; 5. E= Excellent

Table1. An analytic Assessment of Lexico-grammatical competence aspects

5. Results

In this section, the global findings of the study at hand are presented and analyzed using SPSS version 26. The focus is on the assessment aspects of the participants' essay writing, specifically their lexico-grammatical competence before and after receiving the experimental treatment. Subsequently, the results of each test are compared for conducting the null hypothesis testing.

5.1. Pre-test Results

Initially, the pre-test results are displayed to assess the target sample' lexico-grammatical competence in essay writing before receiving the treatment.

4.1.1. Overall Pre-test written performance

Groups	Grammar		Vocabulary		Collocation	
	Experimental	Control	Experimental	Control	Experimental	Control
Mean	2.10	2.15	1.60	1.80	1.55	1.35
Mode	2	2	2	2	1	1
SD	,96	,93	,68	,63	,75	,58
Minimum	1	1	1	1	1	1
Maximum	4	4	3	3	3	3

SPSS Output 1. Overall Pre-test written Performance

The SPSS output 1 provides inferential statistics organized into different columns, including 'Mean,' 'Mode,' 'Standard Deviation (SD) and the 'Minimum & Maximum values'. The analysis focuses on the assessment aspects of the participants' academic essay writing, specifically examining the variations in lexico-grammatical competence before and after the experimental treatment. Before the treatment, both groups exhibited the lowest mean scores in the 'collocation' aspect, with values of

1.55 and 1.35, respectively. Conversely, the highest mean scores were observed in the 'grammar' aspect, with values of 2.10 and 2.15. For the 'vocabulary' aspect, the means were 1.60 and 1.80 for the respective groups. The standard deviation (SD) as a measure of dispersion was found to be similar across all elements, indicating low variability. These results suggest that students' performance levels were relatively consistent within each aspect.

Moreover, the output reveals that, prior to the treatment, both groups consistently rated '1' (representing poor) as the minimum score for all evaluation criteria. Only the 'grammar' aspect had the maximum rating point of '4' (representing very good) at 20%. For the other aspects, 'vocabulary' and 'collocation' the maximum rating was '3,' indicating 'good' quality. Notably, the maximum rating of '5' was not attained by any participants in either group.

4.2. Post-test Results

The post test results are displayed to identify whether the performance of the selected sample has developed owing to the experimental treatment.

Groups	Grammar		Vocabulary		Collocation	
	Experimental	Control	Experimental	Control	Experimental	Control
Mean	3.30	2.40	3.20	2.35	2.70	1.60
Mode	3	2	3	2	3	2
SD	,73	,82	,58	,63	,57	,51
Minimum	2	1	2	1	2	1
Maximum	5	4	4	4	4	3

SPSS Output 2. Overall Post-test Performance

The analysis of the SPSS output 2 reveals that both groups of participants achieved higher mean scores in the assessment aspects compared to the pretest results. The standard deviation (SD) remains low, indicating that despite the improvement in students' essay writing performance, there is still no significant variability. Notably, the Experimental group shows a noteworthy increase in both the minimum and maximum values. The lowest rating point of all criteria is '2,' denoting 'sufficient' quality. Conversely, six post-test takers in the Experimental group achieved the highest grade of '5,' representing an 'excellent' measure, specifically in the assessment aspect of 'grammar.' Furthermore, the column labeled 'Mode' indicates that the rating point '1' (representing 'poor') has a frequency of zero, implying that no participants received this lowest rating after the treatment.

4.3. Unpaired Sample T-Test: Comparing Posttest Means and Hypothesis Testing

Methods	Aspects	Mean	SD	Sig. (P-value)
Post-test Means	Use of ChatGPT in flipped Classroom Model	3	3.06	.32
	'Face to face' classroom with no AI assistance'	3	2.11	.44
Means difference			0.95	0.046

SPSS Output 3. Unpaired samples T-Test

In the SPSS Output 3, the column labeled 'Means difference' displays the difference between the posttest means (0.95). The standard deviation is observed to be low and close to zero (0.32/0.44) due to the clustering of all three assessment elements' rating points around the mean value. This indicates that the gathered data is homogeneous, with no significant variability across the five-point rating scale of the criteria. Specifically, the posttest mean for the experimental group is

3.06, indicating that the rating points are clustered around the measure 'good' given the low standard deviation. On the other hand, the posttest mean for the control group is 2.11, with rating points clustered around the quality measure 'sufficient.' Though these findings suggest a consistent and narrow distribution of performance levels within each group based on the respective assessment elements, the experimental group's essay writing performance is progressed more than the control group.

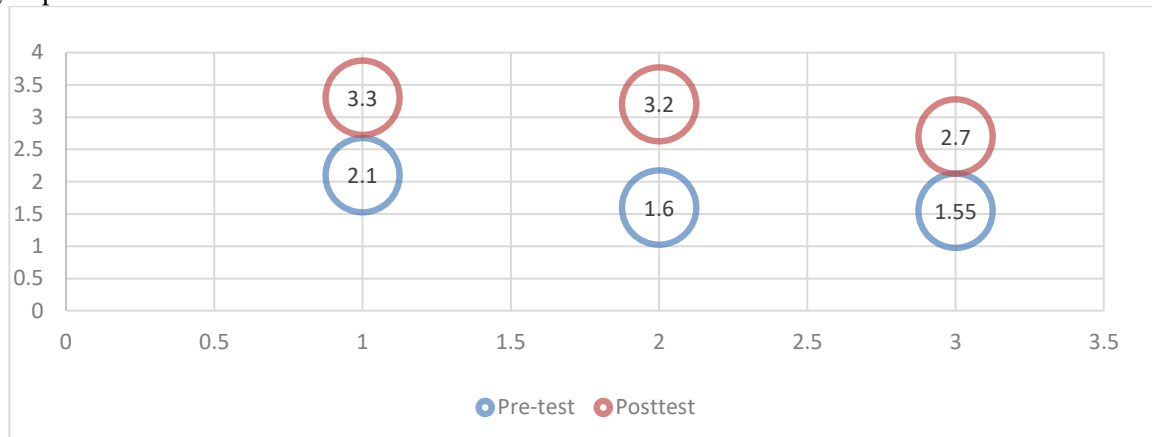


Figure 1. Comparing the Experimental Group's assessment aspects Means

Upon comparing the experimental group's pretest and posttest means for each assessment aspect, as depicted in Figure 1, it becomes apparent that while the difference in the means may not be exceptionally significant, it is still noteworthy across almost all three aspects. The magnitude of the slight difference in means, ranked from large to small, indicates the significance of improvement in each aspect: Vocabulary (-1.6), Grammar (-1.2), and Collocation (-1.15).

4.3.2. Hypothesis Testing

Hypothesis testing may be based on three main criteria: (1) the nature of the hypothesis, (2) the preset significance level (α), and (3) the p-value (represented as sig), obtained from the analysis. In this study, the hypothesis is one-tailed, and the researcher has set α at 0.05 (5% the results may be due to chance). The observed level of significance (sig) is indicated in the SPSS output 3 as 0.0046. Following the decision rule, since the p-value (0.0046) is less than the α level (0.05), the null hypothesis (H_0) is rejected. This indicates that there is sufficient evidence to conclude that the appropriate integration of ChatGPT in the Flipped Classroom Model for teaching essay writing contributes to the enhancement of EFL students' lexico-grammatical competence. Although there is a noticeable increase in the highest rating point '5' and no occurrence of the lowest rating point '1', the difference in means discussed earlier suggests that the enhancement in lexico-grammatical competence for the experimental group with higher post-test scores is not highly remarkable.

.5. Synthesis of the Main Finding

The present study aims to investigate the impact of incorporating Chat GPT into the flipped classroom model for enhancing EFL students' academic writing. The findings from the analysis indicate a significant improvement in the academic writing skill of the target sample after receiving the treatment. This improvement is evident across three aspects: grammar, vocabulary, and collocation, confirming the alternative hypothesis. Statistically, the unpaired samples t-test reveals a reliable difference between the mean scores of the writing posttests for the experimental group ($M=3.06$, $SD=0.32$) and the control group ($M=2.11$, $SD=0.44$) with a p-value of 0.0046, which is less than the significance level ($\alpha=0.05$).

The notable increase, though not highly remarkable, in the posttest means for the experimental group is attributed to the purposeful use of ChatGPT as the independent variable, confirming its effectiveness in enhancing the dependent variable, students' lexico-grammatical features in essay writing. In the posttest results, the experimental group demonstrates improvement from 'sufficient' to 'excellent' ratings, with no instances of 'poor' ratings.

The data obtained reveal that the order of the slight difference in means for each aspect is highest in 'vocabulary' (-1.6), followed by 'grammar' (-1.2), and then 'collocation' (-1.15). This

substantial increase in all aspects provide support for the positive effects of the treatment. Consistent with other research in various contexts (Sok& Heng, 2023; Lawan et al., 2023; Hess, 2023), the incorporation of ChatGPT proves to be an effective in a flipped instructional method for pre-writing and while writing (drafting) stages, ensuring that students are truly engaging in meaningful learning experiences rather than merely outsourcing tasks to AI and bringing post-writing stage into the classroom under close observation. These results contribute to the growing body of evidence that incorporating ChatGPT in classrooms can pose challenges. However, with the right skills and adherence to ethical standards, its integration may have the potential to be effective.

6. Conclusion

Teaching writing, known as a frequently assessed productive skill, can be time-consuming, especially when using the process-based approach in large classes. In this context, the integration of AI based tools, particularly in a flipped classroom model, can offer valuable assistance. Students can benefit from the help provided by AI based tools with the availability of instructors to guide them, addressing doubts, exercising judgment, and promoting critical thinking when evaluating the content generated by AI. This symbiotic relationship between students, teachers, and AI has the potential to revolutionize the teaching and learning process.

To maximize the effectiveness of this approach, teachers must be well-prepared and skilled in using AI tools during flipped class sessions, where students interact without direct AI writing assistance. In this regard, familiarizing students with AI writing applications and equipping them with the necessary skills to create effective prompts and critically evaluate AI-generated results may enhance their learning outcomes.

Though the findings of the present study offer compelling evidence that integrating AI-based tools in language learning and teaching, particularly for essay writing skill, holds promise in transforming learning approaches, it is worth mentioning to recognize certain limitations. The research was conducted with a small sample size and within a limited period, suggesting that the results may not be applicable to all research contexts. Besides, this study does not cover all aspects of AI writing assistance and further research may be necessary to explore these aspects thoroughly.

References

- Abdulmalik Ahmad Lawan, Basheer Ridwan Muhammad, Ahmad Muhammad Tahir, Kamaluddeen Ibrahim Yarima, Abubakar Zakari, Aliyu Hassan Abdullahi II, Adamu Hussaini, Hafizu Ibrahim Kademi, Abdullahi Aliyu Danlami, Mustapha Abdulkadir Sani, Alhassan Bala, Safiya Lawan. (2023). Modified flipped learning as an approach to mitigate the adverse effects of generative artificial intelligence on education. *Education Journal*, 12(4), 136-143. DOI: 10.11648/j.edu.20231204.14
- Baskara, F. R. (2023). Chatbots and flipped learning: enhancing student engagement and learning outcomes through personalised support and collaboration. *IJORER: International Journal of Recent Educational Research*, 4(2), 223-238. DOI: 10.46245/ijorer.v4i2.331
- Beth, M. (2022, December 13). AI, and the Future of Undergraduate Writing. The chronicle of higher education. <https://www.chronicle.com/article/ai-and-the-future-of-undergraduate-writing> (accessed on 20 April 2023)
- Deng, J., & Lin, Y. (2022). The benefits and challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems*, 2(2), 80-83. DOI: 10.54097/fcis.v2i2.4465
- Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of sport*, 40(2), 615–622. DOI: 10.5114/biol sport.2023.125623
- Gill, S. S., & Kaur, R. (2023). ChatGPT: Vision and challenges. *Internet of Things and Cyber-Physical Systems*, 3(n/a), 262–271. DOI: 10.1016/j.iotcps.2023.05.004
- Hess, R. (2023, January 26). Will ChatGPT Unflip the Classroom? January 2023. Education week. <https://www.edweek.org/technology/opinion-will-chatgpt-unflip-the-classroom/2023/01>. (Accessed on 18 June 2023).
- Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *Bench Council Transactions on Benchmarks, Standards and Evaluations*, 3(2), 100-115. DOI: 10.1016/j.tbench.2023.100115

- Kalhan, R. (2023, January 3). ChatGPT Banned from New York City Public Schools' Devices and Networks. News. <https://www.nbcnews.com/tech/tech-news/new-york-city-public-schools-ban-chatgpt-devices-networks-rcna64446> (Accessed on 12 April 2023).
- Kumar, A. H. (2023). Analysis of ChatGPT Tool to Assess the Potential of its Utility for Academic Writing in Biomedical Domain. *Biology, Engineering, Medicine and Science Reports*, 9(1), 24–30. DOI: 10.5530/bems.9.1.5
- Lo, C. K., & Hew, K. F. (2023). A review of integrating AI-based chatbots into flipped learning: new possibilities and challenges. In *Frontiers in Education* 8(1175715), 1-7. DOI: 10.3389/educ.2023.1175715
- Rahman, M.M.; Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Appl. Sci.* 13(9), 57-83. DOI : 10.3390/app13095783
- Raimes, A. (2010). *Pocket keys for writers*. 3rd ed. Boston: Wadsworth Cengage Learning.
- Rudolph, S., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1),1-22.DOI: 10.37074/jalt.2023.6.1.9
- Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology*, 307(2), e230163. DOI: 10.1148/radiol.230163
- Sok, Sarin and Heng, Kimkong, ChatGPT for Education and Research: A Review of Benefits and Risks (March 6, 2023), 1-12. Available at SSRN: <https://ssrn.com/abstract=4378735> or <http://dx.doi.org/10.2139/ssrn.4378735>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57(100752), 1-10. DOI: 10.1016/j.asw.2023.100752
- Taecharungroj, V. “What can ChatGPT do?” Analyzing early reactions to the innovative AI Chatbot on Twitter. *Big Data Cogn. Comput*, 7 (35), 1-10. <https://doi.org/10.3390/bdcc7010035>
- van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224–226. DOI: 10.1038/d41586-023-00288-7
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7), 1-18. DOI: 10.29333/ejmste/13272
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*. DOI: 10.1007/s10639-023-11742-4
- Yu, P., Chen, J., Feng, X., & Xia, Z. (2023). CHEAT: A Large-scale Dataset for Detecting ChatGPT-Written Abstracts. *Preprint (n/a)*. <https://arxiv.org/abs/2304.12008>
- Zhu, Y., Han, D., Chen, S., Zeng, F., & Wang, C. (2023). How can ChatGPT benefit pharmacy: A Case Report on Review Writing. *Preprints*, (n/a), 2-7. DOI: 10.20944/preprints202302.0324.v1