

Exploring the Efficacy of ChatGPT in Language Teaching

Nguyen Thi Hoang Bau^{1*}, Tran Thi Dieu Hien¹

¹ University of Foreign Languages, the University of Danang, Da Nang, Vietnam

*Corresponding author's email: nthbau@ufl.udn.vn

*  <https://orcid.org/0000-0001-5988-2130>

 <https://doi.org/10.54855/acoj.2314210>

©Copyright (c) 2023 Nguyen Thi Hoang Bau, Tran Thi Dieu Hien

Abstract

This study examines the potential of using Artificial Intelligence (AI), specifically ChatGPT, for language teaching. As AI technology continues to advance and be integrated into education, there is growing interest in its role in improving language learning and evaluation. To explore the efficacy of ChatGPT, it was assigned to evaluate ten writing essays from advanced English students, providing detailed feedback and assigning grades based on the established criteria. Then, the results were viewed by an experienced senior teacher who specializes in teaching advanced English writing. The study employs a qualitative approach, with an in-depth interview with the teacher to explore her perspective on the reliability and effectiveness of ChatGPT's evaluation. The findings show a significant similarity between the grades assigned by the teacher and ChatGPT for the set of writing essays. The results from interviewing the teachers showed that ChatGPT is a useful tool for reducing teachers' workload but can only replace human teachers partially. This study supports AI as a potential complement to human teaching and underscores the importance of human expertise for comprehensive feedback, suggesting AI-human collaboration for improved writing instruction and evaluation.

Keywords: Artificial Intelligence, ChatGPT, language assessment, English writing assessment, language teaching.

Introduction

The landscape of education has undergone substantial change in recent years due to the continuous march of technological innovation, with artificial intelligence (AI) emerging as a major disruptor. In fact, the use of AI in language education has drawn a lot of attention.

Language teaching and evaluation approaches have undergone a paradigm shift as a result of the use of AI-powered models (Schmidt & Thomas, 2022). Besides, AI-driven chatbots are supplanting or supplementing the traditional classroom approach, which relies on human teachers and standardized assessments (Rudolph et al., 2023). This paradigm change offers opportunities and difficulties that require careful examination (Rebolledo et al., 2023). Thus, it is crucial to carefully assess AI's function in language education as educators and learners increasingly embrace it.

As the paper authors, we aim to explore the many facets of AI-driven chatbot-ChatGPT in language teaching through its automated evaluation in this study. Our investigation will thoroughly examine ChatGPT's capabilities in evaluating college-level writing skills that serve

the purpose of language teaching. Yet, there are several limitations to the use of AI in language assessment. We will also critically evaluate any potential drawbacks in using AI for assessing writing skills.

Literature review

A chatbot is a software designed to simulate human communication by responding to user input. ChatGPT, a San Francisco-based company OpenAI creation, was made available for free public testing on November 30, 2022. (Opara, Theresa, & Tolorunleke, 2023). OpenAI states that the ChatGPT model uses reinforcement learning from human feedback (RLHF), following methods similar to InstructGPT, albeit with slight differences in data collection. As a paradigm change in educational practice, AI-driven chatbots like ChatGPT have been employed for language teaching and evaluation. Research has found that using AI-driven chatbots such as ChatGPT to practice continuously and receive rapid feedback can enhance language ability (Orsi et al., 2020). With the assistance of AI, language learners become more confident in using their target language to communicate (Hoang et al., 2023). In addition, the utilization of AI in language instruction, exemplified by ChatGPT and other AI models, has clearly demonstrated its effectiveness in enhancing English writing proficiency. Recent research has shown the value of AI-driven systems in enhancing the grammar, syntax, vocabulary use, and coherence of writing (Rahman et al., 2022). As a cutting-edge language model, ChatGPT can offer learners detailed insights into their writing abilities and places for progress through real-time feedback on written compositions (Bishop, 2023). Additionally, the assessment of grammar, vocabulary, pronunciation, and conversational skills is now easier, thanks to AI systems' real-time assessment and feedback capabilities (Rusmiyanto et al., 2023). AI can be used to evaluate creativity and originality in writing assignments, offering helpful ideas to improve composition quality (Abdullayeva & Musayeva, 2023). The scalability and effectiveness of AI-powered assessment make it a beneficial tool for teachers entrusted with assessing a lot of student writing (Zheqian et al., 2019). Due to students' access to rapid feedback, which encourages iterative writing improvements, this technology-driven method in which AI is employed also fosters learners' autonomy and self-regulated learning (Beck & Levine, 2023). AI-powered models are also found to assist educators in gaining insights into various aspects of their teaching, such as the linguistic demands of exercises, the specific characteristics of individual learners, their errors and deficiencies, strengths and weaknesses, as well as affective attributes and more (Schmidt & Thomas, 2022). Though AI's assistance in language learning and assessing is outstanding, the human element is an essential quality of a teacher since the teacher serves as a mentor, resource, and compass for students as they mature and develop morally and ethically (Manuel et al., 2023).

The field of AI-based writing assessment still needs help, such as the requirement for strong models that can comprehend subtleties and context, as well as resolving issues with privacy and fairness in automated grading (Cardon et al., 2023). Despite these difficulties, the developing discipline of AI-driven writing assessment shows promise for assisting English writing instruction.

Research Questions

The study aimed to provide answers to the following research questions.

1. How does ChatGPT evaluate students' essays in advanced English classes at the University of Foreign Language Studies (UFLs)?
2. To what extent can ChatGPT effectively assess writing skills from the perspective of a UFLs lecturer?

Methods

Pedagogical Setting & Participants

UFL students are required to attend advanced English classes to enhance their language proficiency in reading, writing, speaking, and listening. Graduation eligibility hinges on passing the Vietnamese Standardized Test of English Proficiency (Vstep), a standardized language test.

We have chosen to focus on writing skills to evaluate ChatGPT's effectiveness in supporting language teaching. In our study, ten essays on a specified topic, authored by senior ESP students, were randomly selected for assessment using an AI-driven chatbot. This assessment is based on Vstep's criteria for writing, which include Task Fulfillment, Organization, Vocabulary, and Grammar. To meet the Task Fulfillment criteria, essays had to be a minimum of 250 words, and the ideas had to be relevant to the assigned genre and topic. Organization requires clear, logical, and coherent essay structures, with the use of complex sentences and minimal grammar errors to fulfill Grammar requirements.

A qualitative method of semi-structured interviews is employed for the study. An experienced UFL lecturer who has more than 10 years of experience in teaching and assessing language skills was invited to join the study, and she was tasked with viewing the assessments generated by ChatGPT and providing her perspectives. Additionally, the interviews with the lecturer revealed valuable insights into teaching language skills, particularly in the context of writing.

Data collection & analysis

Ten essays with the topic "*Some people think that what children watch influences their behavior, while others believe the amount of time they spend on television influences their behavior the most. Discuss both views and give your opinion*" were initially evaluated by ChatGPT using the Vstep writing criteria for task 2. The topic that belongs to the genre of Opinion /Argumentative was thoughtfully chosen to align with the standardized English tests' format. Students were assigned the task of composing essays as part of their homework, which they were required to type on their computers. They were expected to adhere to specific criteria, including Task Fulfillment, Organization, Vocabulary, and Grammar.

Scoring follows the system of 10/10 for each criterion, and the overall score is the sum of each scoring component. The scoring system of 10/10 is in accordance with the score bands utilized at the University of Foreign Language Studies. The primary objective of this scoring system is to aid in teaching and provide teachers with a basis for assessing students' academic performance. Specifically, scores in the range of 8.5 to 10 correspond to an A, 7 to 8.4

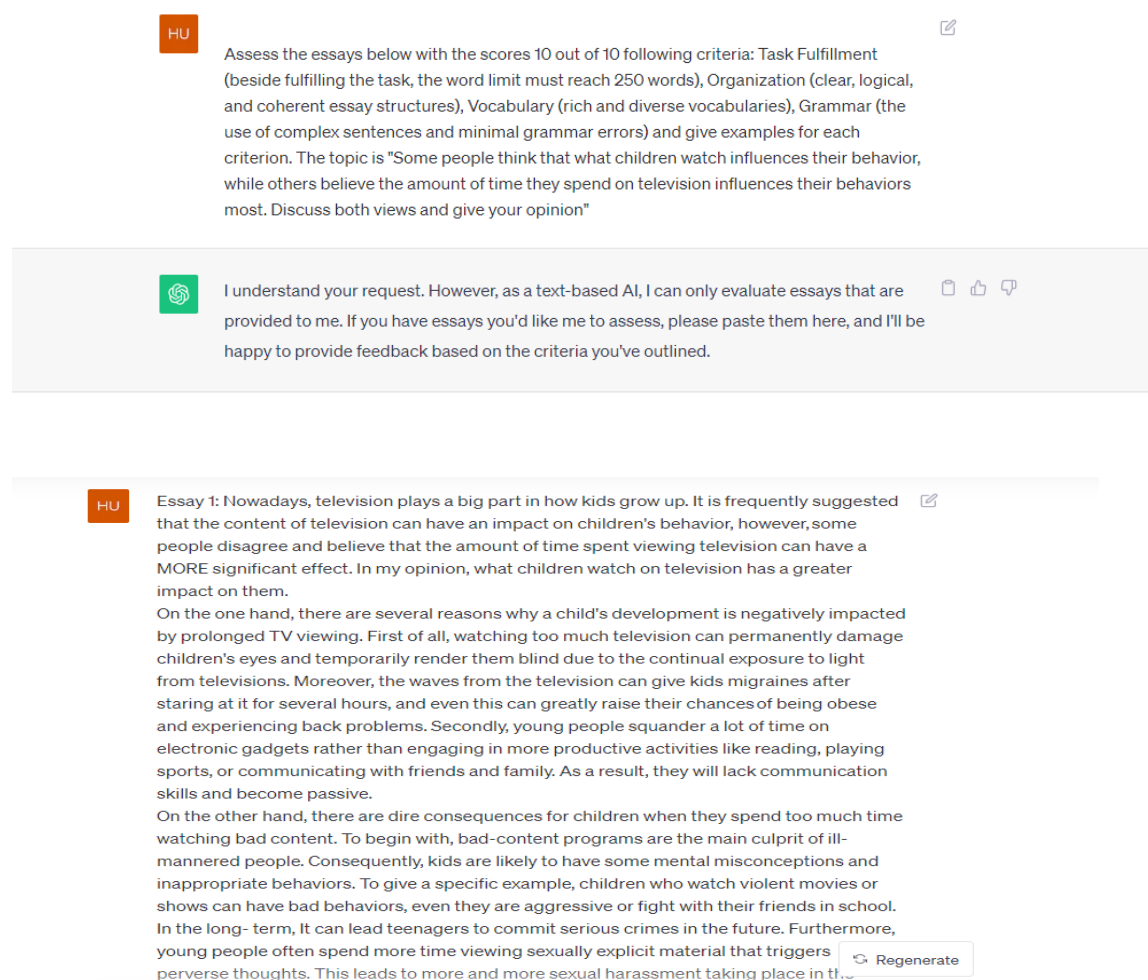
correspond to a B, 5.5 to 6.9 correspond to a C, and 4 to 5.4 correspond to a D. ChatGPT was instructed to score and generate improved versions of these essays based on the originals written by the students. Subsequently, the lecturer reviewed both the initial assessments and the improved texts.

Data collection also involved a semi-structured interview conducted in spoken Vietnamese and later translated into English. Given the shared mother tongue between the authors and the participant, interviews were conducted in Vietnamese to create a comfortable and confident environment for the participant to share her insights. The collected data was then subjected to thematic analysis related to teaching and assessing writing skills with AI support, with notable ideas highlighted for emphasis.

The implementation of ChatGPT evaluating essays.

First, the authors typed the instructions for ChatGPT's evaluation as follows.

1. ChatGPT was required to evaluate the essays based on Vstep criteria of writing and the topic given.



The screenshot displays a chat interface with a user prompt and an AI response. The user prompt, marked with an 'HU' icon, asks for an evaluation of essays based on Vstep criteria (Task Fulfillment, Organization, Vocabulary, Grammar) and provides a topic: "Some people think that what children watch influences their behavior, while others believe the amount of time they spend on television influences their behaviors most. Discuss both views and give your opinion". The AI response, marked with the OpenAI logo, states: "I understand your request. However, as a text-based AI, I can only evaluate essays that are provided to me. If you have essays you'd like me to assess, please paste them here, and I'll be happy to provide feedback based on the criteria you've outlined." Below this, the user's essay is shown, starting with "Essay 1: Nowadays, television plays a big part in how kids grow up. It is frequently suggested that the content of television can have an impact on children's behavior, however, some people disagree and believe that the amount of time spent viewing television can have a MORE significant effect. In my opinion, what children watch on television has a greater impact on them." The essay continues with two paragraphs discussing the negative impacts of television on children's development and the consequences of watching bad content. A "Regenerate" button is visible at the bottom right of the text area.

ChatGPT generated its evaluation based on the essay given.

Let's evaluate the provided essay using the criteria you mentioned:

Task Fulfillment: 4/10 (The essay falls significantly short of the 250-word requirement, reaching only 236 words.)

Organization: 5/10 (The essay lacks clear organization, as it jumps between points about the effects of television and its content without a smooth transition. The introduction and conclusion could be improved for better coherence.)

Vocabulary: 4/10 (While the essay uses some descriptive words, there is room for improvement. The vocabulary could be more diverse and precise.)

Grammar: 5/10 (The essay contains numerous grammatical errors and awkward phrasing, such as "even this can greatly raise their chances of being obese." These issues affect the overall readability.)

Examples for Each Criterion: Regenerate

2. Then, ChatGPT was required to make improvements to the essay given.

give me an improved version of the essay 1 given above

Certainly, here's an improved version of the essay you provided with better organization, vocabulary, grammar, and meeting the 250-word requirement:

Television has become a substantial part of children's lives, and its influence on their behavior is a topic of ongoing debate. Some argue that what children watch on television significantly impacts their behavior, while others believe that the time they spend in front of the screen is the primary factor. In this essay, I will explore both viewpoints and provide my perspective on the matter.

On one hand, proponents of the "what children watch" argument have valid points. The content children are exposed to can indeed shape their attitudes and behaviors. For instance, children who regularly watch violent or aggressive characters may imitate such behavior, potentially leading to issues with aggression or desensitization to violence.

On the other hand, the "screen time" argument also holds weight. Excessive time spent

Regenerate

It is highly noticeable that ChatGPT may produce varying scores and outcomes in response to any changes in the instructions provided. Hence, it is advisable to prioritize precision and thoroughness in ChatGPT's instructions by maintaining uniformity across all assessed essays. This approach ensures a standardized and equitable evaluation process, fostering a clearer understanding of the system's performance across the essays. Employing consistent instructions for all essays not only streamlines the assessment procedure but also contributes to a more reliable and fair outcome, facilitating both ChatGPT's evaluation and subsequent review by the lecturer.

Results/Findings and discussion

ChatGPT's evaluation of students' essays

Explicit guidelines for ChatGPT's evaluation were provided for each criterion. Additionally, essays were expected to showcase rich and diverse vocabularies.

ChatGPT was also instructed to provide detailed explanations for scoring and errors during the scoring process. Moreover, improved versions of the original essays were generated based on the student's initial submissions. The results of ChatGPT's scoring are presented in Table 1.

Table 1. *ChatGPT's scoring results*

No of essays	Task fulfillment	Component scores			Overall score
		Organization	Vocabulary	Grammar	
E1	7	5	6	7	6
E2	6	4	5	6	5
E3	4	4	3	5	4
E4	8	7	7	8	7
E5	5	4	5	6	5
E6	4	3	5	4	4
E7	7	7	7	6	6.5
E8	8	7	8	7	7.5
E9	8	7	8	7	7.5
E10	9	8	8	9	8.5

Observing the data in the table, we can see that 20% of students received a grade of C, 40% received a grade of D, 30% earned a B, and a mere 1% managed to attain an A. Clearly, it is evident that more than half of the students did not perform strongly in the writing task. After reviewing the assessment, the EFL lecturer expressed full agreement with the scores and was particularly impressed by ChatGPT's comprehensive comments and explanations for each essay.

The participant states the essays were meticulously scored in accordance with the criteria and is amazed by the extensive academic vocabulary and complex grammatical structures suggested by ChatGPT, surpassing her expectations. Examples of ChatGPT's suggestions are provided in Table 2.

Table 2. *Suggestions made by ChatGPT*

No	Original texts	Suggested texts
1	spend too much time watching TV	<i>excessive time watching television</i>
2	...the consequence of children watching inappropriate materials would be deleterious. Because they can not distinguish between the right or the wrong contents...	... <i>that exposure to inappropriate content can have deleterious consequences for children is their inability to distinguish between right and wrong...</i>
3	...children would suffer from physical disease...	... <i>children may suffer from physical ailments...</i>
4	It makes them more likely to imitate bad behaviors as well as have negative thoughts.	<i>This exposure can easily lead children to imitate these negative behaviors and develop negative thoughts.</i>
5	...the long time spent sitting in front of the screen can lead to lack of outdoor activities.	... <i>prolonged hours spent in front of the screen can lead to a lack of engagement in outdoor activities, resulting in a sedentary lifestyle.</i>
6	...the young could become inactive and easily sensitive to others' behaviors	... <i>children may become inactive and overly sensitive to the actions of others, leading them to exhibit rude and impolite behavior.</i>
7	...counter-educational programs are the main culprit of ill-mannered people.	<i>Counter-educational programs, for example, can be a major contributor to ill-mannered behavior.</i>
8	Excessive exposure to TV is likely to be addictive to children.	<i>Excessive exposure to TV can lead to dependency, causing children to neglect their responsibilities at school and become less inclined to engage in physical activities.</i>
9	...confining themselves in their room to watch TV causes them to lose the capacity to interact with people, resulting in a lack of physical communication skills.	... <i>isolating themselves in their rooms to watch TV can hinder their ability to interact with others, resulting in a deficiency in communication skills.</i>
10	They may be exposed to violent and sexual content, thus forming rude and aggressive behavior.	<i>Exposure to violent and sexual content can have a detrimental impact on children's behavior. Such content may contribute to the formation of rude and aggressive tendencies.</i>

In Table 2, ChatGPT undeniably showcased its capacity to offer a diverse vocabulary and maintain impeccable grammar, qualities that greatly benefit students in their writing learning. This aligns with the results of Schmidt-Fajlik's (2023) study, which suggests that ChatGPT is the most beneficial and superior solution for assisting Japanese English Language Learners (ELLs) in enhancing their grammar proficiency.

Furthermore, the participant emphasizes that the improved texts are highly beneficial to her students as they align with the original essays but are in improved versions with better organization, task fulfillment, grammar, and vocabulary. As a lecturer, she is constantly burdened with assessing her students' writing skills and guiding them on error identification and improvement, which demands significant time and effort.

However, she points out a few drawbacks in ChatGPT's assessment. Firstly, ChatGPT primarily focuses on identifying grammatical and spelling mistakes in essays, but it tends to repeat comments on sentence structures across different essays. For instance, ChatGPT comments on Essay 1, stating, "the phrase '*the waves from the television can give kids migraines*' should be revised to '*the waves from the television can give kids migraines*.'" However, a similar comment appears in Essay 2, "*Attention should be paid to subject-verb agreement, as in 'excessive screen time would deprive teenagers,' which should be 'excessive screen time would deprive teenagers.'*"

Another concern is the quality of feedback provided by ChatGPT. It offers generic comments on essays that should ideally be more specific for each one. For instance, for Task Fulfillment, which scores more than 6, it provides a repetitive comment like, "*The writing effectively addresses the given topic by discussing both views on the influence of television on children's behavior and provides the author's opinion. The writer presents arguments supporting both perspectives and concludes by acknowledging that both factors play a role. Therefore, the task is fulfilled.*" Conversely, some comments are overly general across different essays, such as "*The writer presents arguments for each perspective in separate paragraphs. However, the conclusion could be more concise and reaffirm the main points made throughout the essay.*"

In such cases, the importance of language teachers becomes evident, as they can evaluate the accuracy of factual claims and the depth of understanding of specific topics or provide context-specific suggestions for improvement. Otherwise, teachers need to check AI's trustworthiness in language teaching (Mark Bedoya Ulla, William F. Perales, & Stephenie Ong Busbus, 2023).

Lecturer's perspectives towards the use of AI tool (ChatGPT)

Benefits

The participant agrees with ChatGPT's consistent overall assessment based on different writing criteria set by the lecturer. Within a click, it gives a full and detailed description of students' errors, especially in lexical use, structures, and syntax (Rahman et al., 2022), which helps save a great deal of time and effort in detecting errors and giving feedback to the students. Within 1 minute for an essay's grading and comment, it costs just one hour and 40 minutes for 100 essays in comparison with 45 minutes spent by the lecturer for one essay. Respectively, it is 45 times faster than human capacity with high accuracy. In this aspect, it helps reduce the high workload for lecturers, setting them free time for other tasks and contributing to job satisfaction. This is in line with the ideas from Azevedo, Orsi Koch Delgad, Sebastian, and Silva (2020) and Zhejiang, Miao, and Man (2019), who claim that rapid feedback can help save time to enhance language ability and supports lecturers with massive writing assessment.

ChatGPT is a natural language processor that is installed with standardized English. Therefore, the comments look native-like like this: "*Overall, while the text covers the main points of the topic, it could benefit from better organization, more balanced arguments, and improved clarity in vocabulary and grammar*". "*We aren't English speakers, so we can't be proficient writers. It's undeniable that ChatGPT supplies students with more standard English than us.*" said the participant.

The suggestions for improvement upon the students' original writing work prove to be of great efficiency since they reflect the fundamental objectives of teaching and assessment. It offers and fosters helpful ideas and well-structured development of ideas for composition quality, which is in line with the research of Abdullayeva and Musayeva (2023). Nevertheless, the participant finds this job *“requires considerable physical effort and high level of English proficiency while with ChatGPT, it spares no effort at all.”*

Challenges

For a record, the participant describes some technical problems that ChatGPT often makes due to its machinery nature. First, ChatGPT provides suggested essays with unbalanced length and viewpoints. Based on the ideas from an original piece of work, it only focuses on improvement without concerning the logical length between paragraphs, making some look irrelevantly long and others too short. The arguments prove the imbalance systematically. This phenomenon puts these suggested essays on the verge of losing their effect with reliability and persuasiveness, which are of little merit for students in identifying the suitable layout of the essay.

“As a kind of tool, ChatGPT gives plenty of comments based on the given writing criteria. However, these comments are largely generalized, not specified in certain areas like grammar and vocabulary”. Examples regarding this are presented in ChatGPT's evaluation of students' essays.

The participant stresses the fact that these improved comments on vocabulary and grammar don't count on students' language levels, namely intermediate or advanced, to make suggestions. This is considered to be one of the most important factors that support students in developing their skills at the appropriate language competencies. As a result, teaching objectives still need to be fully met. This opinion sheds far different light from research by Cardon, Fleischmann, Aritz, Logemann, and Heidewald (2023), who mention the subtleties and context matter that ChatGPT can't assist with.

The role of ChatGPT and lecturers in teaching and assessment

“If ChatGPT is implemented in writing assessment, would you like to try it? Definitely yes, I would really love to and delighted to experience such a new useful thing.” says the participant. She mentions that with this new tool in classrooms, the role of lecturers definitely changes. Lecturers will take advantage of ChatGPT's benefits in grading, correcting mistakes, and recommending improved writing essays for students; despite the fact that a large amount of work it can do, it can never substitute for the lecturer's role. Literally, ChatGPT is considered stereotyped and mechanical, i.e., it merely provides similar and inflexible forms of feedback for all essays. Lecturers, on the other hand, are competent to review the drawbacks and carefully examine contextual factors like students' language competency, their targets and age group, the learning environment, and course objectives to make the utmost use of teaching and assessment. In the final analysis, ChatGPT will be a processor of information, lecturers, reviewers, and adapters of writing pieces of work. Manuel, Giuseppe, Salvatore, and Mario (2023) look into AI roles from the same angle, which states that lecturers always play the role of mentors for students' best development.

Recommendations

This research should be recommended to other lecturers at UFLS with this supporting tool, which is practical as a lecturer's helper in dealing with writing task evaluation. We want to propose some recommendations. Concerning the university stakeholders, it is imperative to heighten awareness regarding the utilization of AI in writing evaluations. This can be achieved through workshops, conferences, and meetings aimed at both lecturers and students. Furthermore, ChatGPT should be implemented in writing skills and other aspects of language competencies. More importantly, lecturers are encouraged to incorporate ChatGPT into their lessons, providing improved feedback to students through a balanced combination of AI tools and human collaboration.

Conclusion

In conclusion, while AI-driven technologies like ChatGPT clearly have a significant role in improving language instruction and evaluation, they should be seen as supplements rather than a replacement for human language teachers. This conclusion is supported by a study conducted by Opara, Theresa, and Tolorunleke in 2023, underscoring that while AI can enhance education, it should not be employed as a substitute for it. Indeed, AI can significantly alleviate the workload of human teachers, streamlining the evaluation process and saving valuable time. This assistance in overall assessment contributes to more efficient and effective teaching, benefiting both educators and students. Furthermore, the most effective approach lies in synergy, where AI complements educators by offering proficient grammar and style checks, enriching vocabulary, and automating assessments. Meanwhile, human teachers play a vital role in double-checking AI evaluations, nurturing a deep comprehension of language that transcends mere syntax and semantics. This combined effort ensures a holistic and comprehensive approach to language education. Indeed, teachers can give students access to a more thorough and productive environment for language learning by fusing the advantages of AI with the knowledge of language educators. This study contains potential limitations. Firstly, the minimal sample size leads to unintentional bias in findings and discussions. The fact that the interview with one participant reflects a viewpoint from only one angle may need to be revised. Secondly, the literature review provides a narrow scope of research with restricted access to existing literature related to AI since AI has appeared as a new term in recent years.

Further studies are necessary to gain more insights into the role of human teachers in language teaching and evaluation in light of AI as a learning tool when students can easily assess the tool as teachers do.

References

- Abdullayeva, M., & Musayeva, Z. M. (2023, April). The Impact of ChatGPT on Student's Writing Skills: An Exploration of Ai-Assisted Writing Tools. In *International Conference of Education, Research and Innovation* (Vol. 1, No. 4, pp. 61-66).
- Beck, S. W., & Levine, S. R. (2023). Backtalk: ChatGPT: A powerful technology tool for

- writing instruction. *Phi Delta Kappan*, 105(1), 66-67.
- Bishop, L. (2023). A computer wrote this paper: What ChatGPT means for education, research, and writing. *Research, and Writing (January 26, 2023)*.
- Cardon, Peter & Fleischmann, A. Carolin & Aritz, Jolanta & Logemann, Minna & Heidewald, Jeanette. (2023). The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age. *Business and Professional Communication Quarterly*. <https://doi.org/10.1177/23294906231176517>.
- Hoang, N. T., Duong, N.H., & Le, D. H. (2023). Exploring Chatbot AI in improving vocational students' English pronunciation. *AsiaCALL Online Journal*, 14(2), 140–155. <https://doi.org/10.54855/acoj.231429>.
- Manuel, G., Giuseppe, C., Salvatore, P., & Mario, A (2023). Do we still need teachers? Navigating the paradigm shift of the teacher's role in the AI era. *Frontiers in Education*. <https://doi.org/10.3389/educ.2023.1161777>.
- Mark Bedoya Ulla, William F. Perales & Stephenie Ong Busbus (2023) 'To generate or stop generating response': Exploring EFL teachers' perspectives on ChatGPT in English language teaching in Thailand. *Learning: Research and Practice*. <https://doi.org/10.1080/23735082.2023.2257252>
- Opara, Emmanuel & Theresa, Adalikwu & Tolorunleke, Caroline. (2023). ChatGPT for Teaching, Learning and Research: Prospects and Challenges. 5. 33-40. <https://org.doi/10.36348/gajhss.2023.v05i02.001>.
- Orsi Koch Delgado, Heloísa & Azevedo, Aline & Sebastiany, Maria & Silva, Asafe. (2020). Artificial intelligence adaptive learning tools: The teaching of English in focus. *BELT - Brazilian English Language Teaching Journal*. <https://doi.org/10.15448/2178-3640.2020.2.38749>
- Schmidt-Fajlik, R. (2023). ChatGPT as a Grammar Checker for Japanese English Language Learners: A Comparison with Grammarly and ProWritingAid. *AsiaCALL Online Journal*, 14(1), 105–119. <https://doi.org/10.54855/acoj.231417>
- Schmidt, Torben & Strasser, Thomas. (2022). Artificial Intelligence in Foreign Language Learning and Teaching: A CALL for Intelligent Practice. 33. 165-184.
- Rahman, N. A. A., Zulkornain, L. H., & Hamzah, N. H. (2022). Exploring Artificial Intelligence using Automated Writing Evaluation for Writing Skills. *Environment-Behaviour Proceedings Journal*, 7(SI9), 547-553.
- Rebolledo Font de la Vall, Roxana & Gonzalez Araya, Fabian. (2023). Exploring the Benefits and Challenges of AI-Language Learning Tools. *International Journal of Social Sciences and Humanities Invention*. <https://doi.org/10.18535/ijsshi/v10i01.02>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1).
- Rusmiyanto, R., Huriati, N., Fitriani, N., Tyas, N., Rofi'i, A., & Sari, M. (2023). The Role Of

Artificial Intelligence (AI) In Developing English Language Learner's Communication Skills. *Journal on Education*, 6(1), 750- 757. <https://doi.org/10.31004/joe.v6i1.2990>

Zheqian, Su & Miao, Liu & Man, Jiang. (2019). Artificial Intelligence Promotes the Evolution of English Writing Evaluation Model. IOP Conference Series: Materials Science and Engineering. <https://doi.org/10.1088/1757-899X/646/1/012029>

Biodata

Nguyen Thi Hoang Bau holds an M.S in Tourism Management from Chinese Culture University, Taiwan (2015) and an M.A in Applied Linguistics from the University of Melbourne, Australia (2018). She is a lecturer at the Faculty of English for Specific Purposes, University of Foreign Languages Studies, University of Danang, where she obtained her Bachelor of Business English in 2012. She is an accomplished author of local and international papers and has guided students in research. Her research interests span tourism and linguistics.

Tran Thi Dieu Hien hold an M.A of Applied Linguistics from Curtin University, Australia in 2010. She received a bachelor's degree in English linguistics and literature from University of Social Sciences and Humanities, Ho Chi Minh City. She's currently a lecturer at the Faculty of English for Specific Purposes, University of Foreign Languages Studies, the University of Da Nang. Her work focuses mainly in Business English and English for Tourism. Her research interests cover several aspects across language teaching and assessment.