The Application of ChatGPT in Language Test Design – The What and How

Nguyen Thi Phuong Thao[1*]

University of Languages and International Studies, Vietnam National University, Hanoi, Vietnam
*Corresponding author's email: thaontp.ulis@vnu.edu.vn
https://orcid.org/0000-0001-8444-8053
https://doi.org/10.54855/paic.2348

ABSTRACT

Keywords: artificial intelligence, ChatGPT, language test, test design	ChatGPT, an artificial intelligence-based chatbot, has recently gone viral among users thanks to its ability to create human-like interactive experiences. It is now a trend to make use of this artificial intelligence (AI) tool in different fields of language education, including language testing and assessment. With a concern related to the potential and challenges of applying ChatGPT in designing language tests, this study has investigated the current use of the chatbot among language teachers at different schools and institutions. Participants are expected to share their experiences utilizing the chatbot to assess their learners. Data from a survey and in-depth interviews will reveal initial findings to answer questions related to the popularity of this chatbot among language teachers and its application in language test design. Based on the result discussion, the presentation concludes with some implications for language teachers to make the most of their future use of ChatGPT in the field of language testing.

Introduction

ChatGPT – an artificial intelligence-based chatbot, has reached one million users only five days after its official launch (Mirati, 2022). This software application has promptly gone viral due to its capabilities to engage in human-like conversation with a huge database created from existing reference sources. As users interact with the system via a question-response form, users can ask about whatever issue they are concerned about and then receive a result within seconds. Therefore, there is no doubt that using ChatGPT is of interest to various groups of people in different fields in general and educators and learners in particular. In the field of language education, ChatGPT has marked its appearance in the classroom context when it can provide authentic language use by initiating real-life conversations and also act as a tool for a personalized learning environment (Hong, 2023). For teachers, ChatGPT is reported to support them in multiple aspects related to planning lessons, assessing students' writing papers, and

generating tasks and scenarios (Rudolph et al., 2023). However, as the chatbot has been recently launched, there are not many existing studies with empirical evidence about the application of ChatGPT in current teaching and learning activities. Available studies seem to start from a review of the benefits, opportunities, and challenges of using the software application. More research on the use of this chatbot in different educational contexts should be implemented to evaluate its effectiveness in real-life situations. As a result, this study is conducted with the expectation of contributing to the research field of ChatGPT application in a certain aspect known as language test design. The researcher aims to investigate how the chatbot supports language teachers in developing different kinds of tests and whether it can become an effective educational tool for teachers to take advantage of in their teaching context.

Literature review

What ChatGPT is

Being officially launched in November 2022, ChatGPT is considered a "state-of-the-art chatbot" (Hong, 2023). It is "a general-purpose conversation chatbot" and "designed to generate human-like text based on a given prompt or conversation and has the ability to engage in natural, open-ended conversations on a wide range of topics" (Zhai, 2023). It helps compare existing data to provide the most likely relevant responses based on users' questions. Thus, Hong (2023) makes a precise summary that ChatGPT is a text-generating search engine that may not suggest up-to-date information but excels at mimicking human interactions and filtering out irrelevant information while processing its responses.

Opportunities for using ChatGPT in language education

There is no doubt to mention two main subjects: teachers and learners. A number of studies have reported the benefits of chatbots to teaching and learning activities. On the students' side, ChatGPT acts as a tool to personalize learning and support a personal learning environment. For example, the application can act as a personal language tutor by creating topics for discussion or writing prompts to practice (TESOL International Association, 2023). Also, like using other search engines, students can look for whatever information they wish by giving feasible instructions to the chatbot. According to Kasneci et al. (2023) and Thai (2023), students can enhance their engagement and follow personalized learning with this application with precise answer provision. Interestingly, some studies have investigated the chatbot's ability to generate answers in formal examinations, and the results show that it can reach passing scores (Choi et al. (2023), Dao et al. (2023)). Furthermore, based on the capability to summarize information, ask follow-up questions, and clarify, this chatbot can provide authentic language use for students pursuing language study (Hong, 2023). Consequently, Nisar and Aslam (2023) conclude that ChatGPT can be utilized as a prompt reference and self-studying instrument for learners. There exists an argument that the application may limit students' creativity as almost everything they need is available with the tool. However, according to Zhai (2022), this language model chatbot can encourage the use of creativity and critical thinking skills among users via AI-involved learning tasks solving real-world problems. The application of ChatGPT in enhancing language learners' critical digital literacies is also investigated in Tran & Tran

(2023).

For teachers, as soon as the chatbot has been promoted, Moore et al. (2022) show that it can help evaluate the quality of students' generated answers. In this sense, ChatGPT acts as a tool to support assessment activities. Later on, through a review of two journal articles and eight preprints on ChatGPT and higher education, Rudolph et al. (2023) summarize that this software application is useful in marking students' writing papers, creating lesson plans, offering suggestions, generating tasks, and initiating scenarios. ChatGPT does play some roles in language assessment. To clarify, the chatbot can help create assessment tasks based on users' instructions. As a text generation tool, the chatbot also produces texts that may be used to compose test items (Tate et al., 2023). Additionally, as Rudolph et al. (2023) discuss, the application supports instructors in assessing their students' writing and then reducing their workload.

Challenges of using ChatGPT in language education

Despite many opportunities ChatGPT may open for both educators and learners, the chatbot also has some negative impacts on teaching and learning activities, as reported by some studies. Yeadon et al. (2022) claim that ChatGPT may become a threat to the credibility of short-form essays as an assessment method. The borderline between plagiarism using the chatbot and using it as a reference seems not to be clear enough for learners. Teachers and policymakers are also aware of the negative impacts of AI tools in this aspect (Nguyen, 2023). This is also what Cotton et al. (2023) are concerned about AI-powered writing assistants. As students using AI may lack creativity and autonomy, it is crucial for teachers to design reliable assessment tasks to evaluate and assess learners' levels while not stopping them from using such tools. The potential negative impacts of the application on students' critical thinking skills are also reported by Thai (2023). Le et al. (2023) also report that the answers ChatGPT generated were inconsistent and may depend on multiple factors. It means that users should not totally rely on the chatbot's responses. For questions that require a higher level of cognitive thinking, like other AI tools, the chatbot also has difficulty giving accurate responses, according to Dao et al. (2023).

From the brief overview of ChatGPT and some existing studies on the chatbot, it can be seen that more research should be implemented to investigate its application in the field of language testing and assessment. What has been reported in recent studies globally and in Vietnam focuses more on how ChatGPT supports teaching and learning activities. As a result, this current study aims to report the use of ChatGPT in language assessment in general and language test design in particular. The gap to be filled in the literature review should be what aspects of test writing involve this software application and how effective the tool is from language teachers' perspectives.

Research Questions

To fulfill the purpose of the study, three following research questions should be answered:

1. To what extent is ChatGPT used among language teachers?

2. What aspects of language test design can ChatGPT support teachers?

3. How effective is ChatGPT in designing language tests as perceived by teachers?

Methods

Pedagogical Setting & Participants

The research involves the participation of 70 language teachers from different universities and schools in Vietnam. The majority of participants are from universities in the northern region, and the rest are scattered in the central and southern parts. 82.8% of those participants are teaching English at a number of institutions, 8,6% are teachers majoring in Japanese education, while the rest of 8.6% are lecturers and teachers of other languages, including Russian, French, Chinese, German, and Korean. As the data collection is based on convenience sampling, since the researcher specializes in English testing and assessment, it is plausible to see such a gap between English and other languages. Of those respondents, about three-quarters of them have more than 10 years of teaching experience, with 85.7% being female and 14.3% being male. Regarding qualifications, less than one-fifth of the responses are from graduates, while the biggest group includes teachers with Master's degrees at 61.4%, and lecturers with a Ph.D. title comprise 22.9% of the surveyed. That more than half of the participants have fulfilled their Master's program reflects the current situation of teacher's qualifications at Vietnamese language institutions.

Regarding the participants for in-depth data, five educators were invited as they shared in the survey that they had been using ChatGPT regularly and were willing to be asked individually. Four teachers who specialize in English language teaching are from the same university. One educator used to work as a full-time English lecturer but is now a strategy team leader of an education technology company. They have from twelve to fifteen years of working in the field of English language education.

Design of the Study

In order to answer three research questions, the study would employ a mixed method approach of both quantitative and qualitative approaches. According to Cresswell (2016), mixed methods are "an approach to research in which the investigator collects, analyzes, and interprets both quantitative and qualitative data, integrates or combines the two approaches in various ways, and frames the study within a specific type of design or procedure" (p.4). In the scope of this study, a questionnaire and semi-structured interviews were selected to collect data using the mixed method approach. They are expected to collect both at-large and in-depth responses from the participants. The interviews would help fill the questionnaire's gap with more insights.

Data collection & analysis

First, using convenience sampling, an online survey designed on Google Forms was sent to language teachers and lecturers. The questionnaire includes two parts; part 1 aims to collect the participants' relevant background information, while part 2 acts as the main part of compiling responses reflecting participants' use of ChatGPT in language test design. Regarding the application of ChatGPT, the survey focuses on answering two questions: what and how. To be more specific, what aspects of language test design can ChatGPT be involved in, and how effective is it in real-life use? Survey questions were built upon the analysis of findings and discussion in previous studies by Hong (2023) and Rudolph et al. (2023), and based on the researcher's knowledge and hands-on experience of ChatGPT as well. The data is then put into

Excel for analysis. Later on, qualitative data is collected via semi-structured interviews of five participants who completed the survey. The recordings were transcribed and translated into English. The transcript of each participant was coded with numbers, namely T1 to T5, and used for the study purpose only. Relevant themes emerging from the transcription data were individually analyzed and later integrated to provide findings for the study. The interview questions targeted retrieving further information about teachers' application of ChatGPT in test design activities.

Findings and Discussion

Research question 1: To what extent is ChatGPT used among language teachers?

With the purpose of investigating whether ChatGPT is widely used among language teachers, two questions were formed in the questionnaire. Figure 1 presents the extent to which teachers know about the chatbot, while Figure 2 shows its application level among participants.

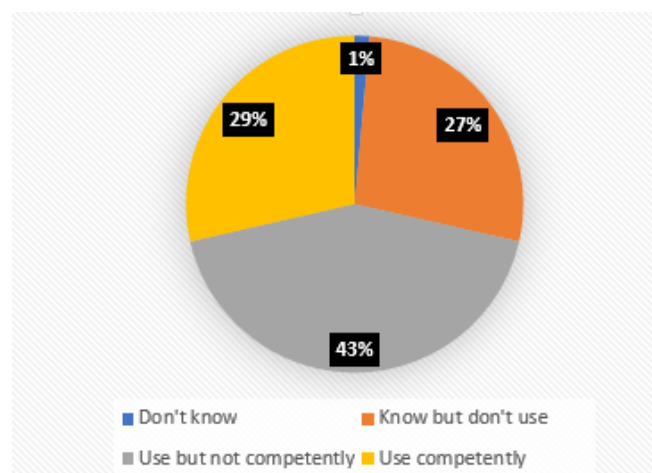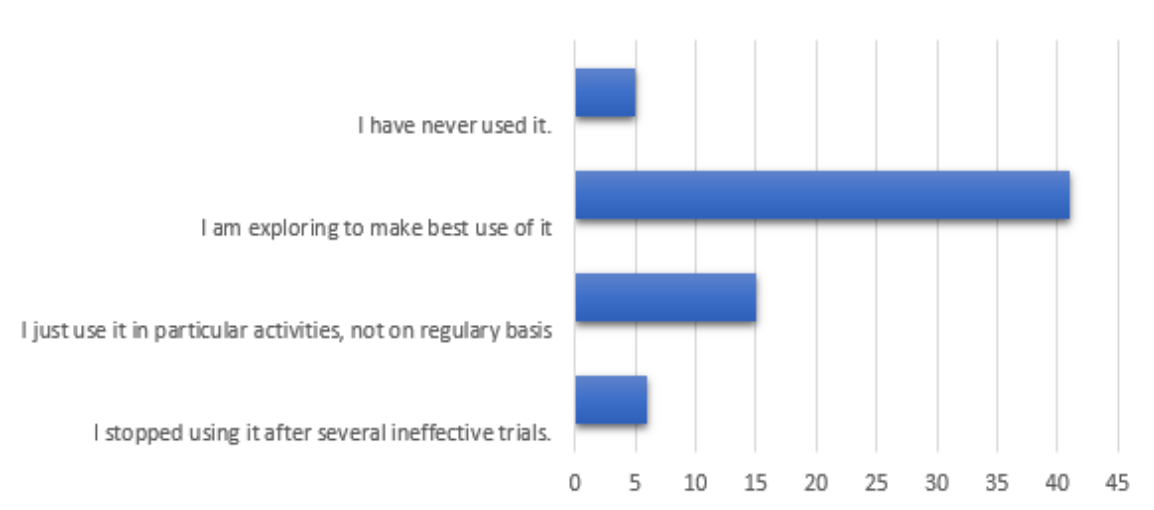Figure 1. Levels of knowing about ChatGPT among language teachers (N=70)

Figure 2: Levels of using ChatGPT among language teachers (N=70)

As can be seen from both figures, most of the respondents know about the chatbot, and only 1% of the participants do not know about it. In Figure 1, nearly one-third of them claimed that they had never used it despite knowing the software application. This is reflected in Figure 2, when 11 participants revealed that they had never used or stopped using it after several ineffective trials. The majority in Figure 1 (72%) responded that they utilized ChatGPT, with more people admitting that they were not very competent in using it. This result can be seen in Figure 2, regarding 60 responses in total. It seems that many of them are on the way to exploring the chatbot to take advantage of it, compared to only 15 respondents applying GPT-3 for particular activities. From the two figures, the surveyed group shows an increasing use of ChatGPT. However, the proportion of competent and regular users is still limited.

Research question 2: What aspects of language test design can ChatGPT support teachers?

As the study focuses on issues related to designing language tests, the questionnaire raised a question that includes some typical features in the field. They are text generation or adaptation and test item composition.

Figure 3: Aspects of language test design that teachers apply ChatGPT (N=70)
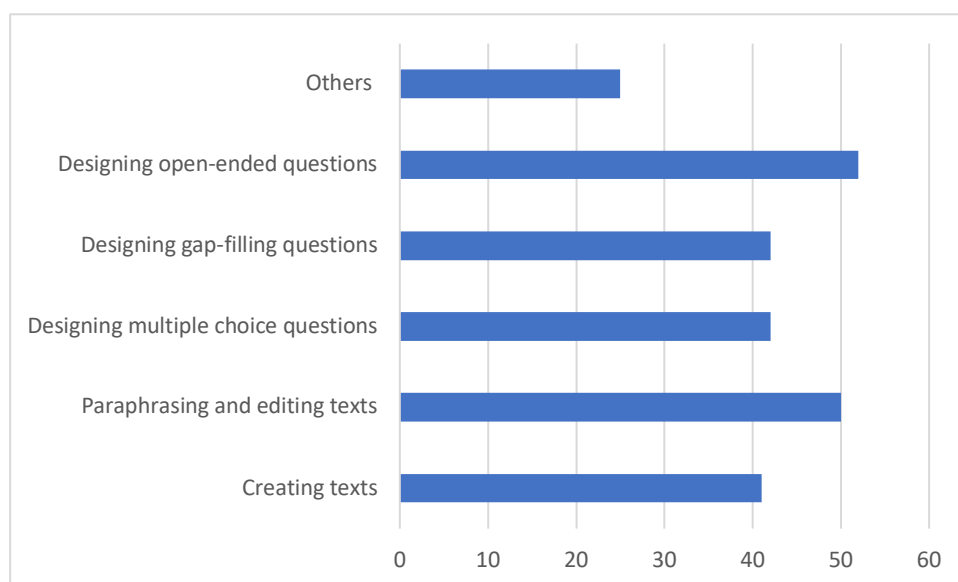
Regarding text generation, more than half of surveyed teachers reported using ChatGPT to create input texts for their own tests. As the chatbot is naturally a text-generation tool, it is understandable to see that many teachers use this function. The application can generate an input text of either Reading or Listening skills based on the user's request. Two out of five interviewed teachers responded that they tried this function by providing detailed guidance for GPT. T1 shared what she did in a reading test:

> "ChatGPT can create a reading passage, and everything is based on your instructions. For example, you need to explain in detail the range of word numbers or how many words you wish to have in your text, what text level is like B1 or B2, what topic it is about, and so on. The more descriptions you provide, the more appropriate text you will have."

Having the opinion in common with T1, T2 clarified what T1 described as "put into the test specification to depict what features the text should possess." The information that T1 shared on word number, text level, topic or domain actually belongs to the test specification's requirement of input text. If teachers want to stick to the test specification, they should break it down and provide detailed questions for GPT (T2). The use of ChatGPT for text generation is similarly reported in some previous studies such as TESOL International Association (2023), Hong (2023), Rudolph et al. (2023), and Tate et al. (2023).

Regarding text adaptation, a significant proportion of about 70% of participants claimed to use ChatGPT in paraphrasing and editing texts. Not surprisingly, all five interviewed teachers reported their frequent use of this function in their test design. According to T3, as ChatGPT can provide authentic language, teachers should take advantage of this benefit to refine their input text before designing questions. To make it more specific, T4 added:

> "If you want to edit your text, you can ask ChatGPT for help by making requests such as simplifying the text to make it easier for level B1 or adding more complex structures to increase its difficulty level. Moreover, ChatGPT helps you refine what you have written as it is a huge database of information, and the language it generates is native-like, so you can make your composition sound much more natural."

As can be inferred from interviewees' sharing, the chatbot is an applicable tool for revising and editing texts by paraphrasing them based on users' requests. So far, the function of adapting texts actually derives from the item writer's need, and this helps to extend the use of this chatbot in the field of language assessment present in the literature review.

In terms of designing test items, the survey asked about three main test questions, including multiple-choice, gap-fill, and open-ended. More than half of all participants have used the software application to write such test question types, especially more than 70% of them designed open-ended questions. These figures show a notable role of ChatGPT in supporting teachers in composing test items, which is often considered a time-consuming task. This finding also reflects what Rudolph et al. (2023) summarize in their study about the capacity of software applications to generate tasks. With clear instructions to require what type of question users want the chatbot to create, they will receive relevant responses. Interviewees had similar reflections on this advantage of the application. T2 and T5 both claimed that they had used it on a regular basis to create gap-fill questions or cloze-text ones. The value of this application in creating separate multiple-choice items for vocabulary-grammar tasks is also appreciated by T1 and T5. Interestingly, T4 had some sharing on the application of ChatGPT in creating different question types for a reading passage:

> "I often put the text in the system and ask whether the chatbot can create multiple choice questions and short-answer questions on sub-skills such as reading for the main idea, specific details, or vocabulary. As a user, it is very important to provide one by one instruction. You should break up your instructions like the first one is to focus on specific details, which go with multiple choice questions, and which go with short-answer questions. Then, the next request is about creating a multiple-

choice question for the main idea."

Another type of question is mentioned in this sharing, known as a short-answer one. T4 added True/False/Not Given questions to the list, and T1 also shared her experience in using the chatbot to create matching and summarizing questions. This may explain the reasons why more than one-third of participants added "others" to their list. Others may refer to other questions to be designed in this sense. Additionally, T4 added a different usage of ChatGPT, which can be categorized as "others". It refers to asking the chatbot to refine the instructions of her test tasks. She considered the features of authentic language use in this case. T2 raised another idea of breaking up the test specification into keywords or short sentences to require the chatbot to create relevant test tasks. In this sense, ChatGPT is considered a bridge connecting test specifications to test tasks, then the test itself.

From the quantitative and qualitative data analysis, it is worth noticing that ChatGPT has been utilized in two main aspects of test design investigated in this study. The first one is working with input texts, covering both generating new texts and editing existing ones. The second factor is item composition with a number of suggested question types. The findings from interviews also reveal additional aspects of test design that ChatGPT may contribute to, such as creating test tasks based on an available test specification. The applications of the chatbot related to text and question generation have been presented in Rudolph et al. (2023)

Research question 3: How effective is ChatGPT in designing language tests as perceived by teachers?

From surveyed aspects of test design in the second research question, the questionnaire continued investigating teachers' perspectives on how effective that chatbot is in such features of test design. A Likert-scale question is designed to ask for participants' opinions on their evaluation in case they have used ChatGPT. The results are displayed in Figure 4 below.

Figure 4: Teachers' perspectives on the effectiveness of ChatGPT in test design
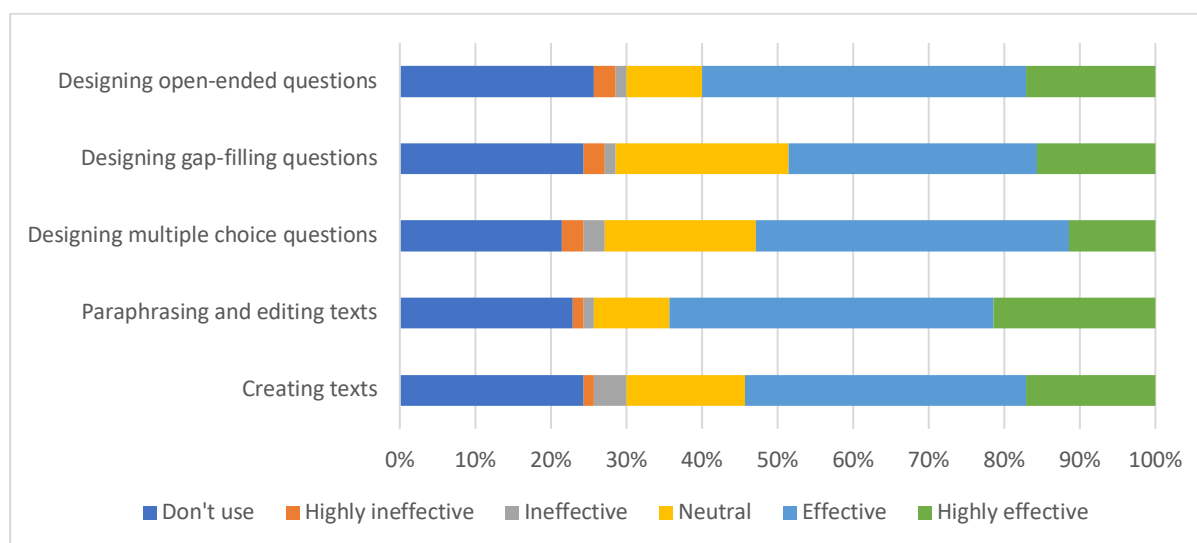
Apart from about one-fifth to a quarter of participants reported not using ChatGPT, the rest expressed various evaluations of the chatbot's advantages on test design. The evaluation is based on their hands-on experience of using the application.

In relation to how potent the language model in text generation and adaptation is, participants showed their confidence in using ChatGPT to perform these actions when more than half of them agreed that it is effective and highly effective. As can be seen from the figure, respondents were more appreciative of the chatbot's ability to edit and paraphrase texts, with 64% choosing effective and highly effective, compared with 54% in the same categories of text generation. This preference is also clearly disclosed in all five teachers' interview data. They appreciated what was revised and refined by ChatGPT when they put a text and gave specific requests. However, T1 and T5 both shared the opinion that it is very common to ask the chatbot to revise the text repeatedly by editing the instructions until they are satisfied with the result. It may be why the proportion of its considerable effectiveness is not significant enough. T1 even admitted that after her check with specialized text analysis tools, the readability and vocabulary level requirement cannot be fulfilled. Hence, a question is raised about whether it takes more time and effort for teachers to use the chatbot rather than polish texts by themselves. However, the interview responses advocated the chatbot as teachers think it is still time-consuming. The more teachers use it, the better their requests will be to navigate it (T4).

Discussing the application of ChatGPT in designing questions, all three investigated question types were reported to be effective and highly effective by nearly 50% to 60% of participants. In this group, the gap-filling questions received the lowest percentage (48%), while open-ended ones gained the highest (60%). Therefore, the proportion for "neutral" opinions also accounted for about one-fifth of respondents in multiple-choice and open-ended questions. This reflects considerable support from surveyed teachers for the function of item design. However, controversies may arise due to different perspectives.

Regarding multiple-choice questions, in-depth data have brought multi-directional viewpoints. T4 and T5 are in favor of requiring ChatGPT to compose such items as T4 shared.

> "As I am teaching a test preparation course, I often use the chatbot to design multiple-choice questions based on a reliable text. I am quite satisfied with the result as distractors meet the requirement of being broad or narrow, being incorrect in detail, etc. Normally, I have to revise the questions at least twice to edit the wording or reduce the word number to follow the description in the test specification. However, I am happy with the questions it creates."

On the other hand, T1, T2, and T3 do not highly value the role of ChatGPT in this situation. T1 thought that distractors were not strong enough to discriminate test takers at different levels. Some distractors seem to be easily eliminated because they are obviously wrong. This is not the art of writing multiple-choice questions (T1). T3 saw this type of question as more suitable for specific details than other sub-skills in a reading comprehension task. The draft needs refining a lot before it is ready to be used. As an educational technology leader, T2 shared a concerning idea:

> "What ChatGPT creates is not very consistent as the chatbot cannot make decisions, and the results change over time. It designs better individual multiple-choice questions in a grammar and vocabulary task than in a reading comprehension exercise, which requires a link between questions and a high cognitive load. Also,

I think it should be applied to classroom tests, not high-stakes ones requiring reliability and consistency."

Some noticeable findings have been drawn from the interviews for other types of questions. Sharing a similar point of view of more than half of the surveyed teachers, interviewees appreciated the quality of gap-fill and open-ended questions. Having been provided enough information from the test specification, ChatGPT can generate pretty good gaps in sentence completion or cloze text (T2). It is also easy to create open-ended questions with the appropriate request for the application (T1). Again, they emphasized the significance of giving the right request. The interviewees also mentioned other types of questions that cannot be well-designed by the chatbot. According to T1, a short-answer question is an instance in which the chatbot has difficulty in limiting the number of words to produce in the answer. She also tried to design a matching task, but the result was not good because ChatGPT was unable to generate pictures. T2 added error recognition as a question type that challenges the language model. The discussion, once again, is about the limitation of ChatGPT in designing a variety of task types. This fact is understandable, as artificial intelligence has its own pros and cons and cannot replace humans in certain aspects.

Conclusion

The study investigated the use of ChatGPT among language teachers at different institutions, mostly in the north of Vietnam. The data analysis from the questionnaire and semi-structured interviews has brought noteworthy findings for the field.

Firstly, the majority of participants claimed they used the chatbot in their teaching and assessment activities. Many of them use it daily and try to make the best use of it. The small percentage of teachers who have never used or stopped using the chatbot after several trials also raises concerns for future resolutions.

More importantly, there is no doubt to infer from the findings that ChatGPT has become increasingly popular in language test design. Its notable application can be clearly seen in two major aspects: input text generation/adaptation and item writing. Such application usage is based on its nature as a text-generation tool. Therefore, it can create something new or revise what is provided with appropriate instructions and requests. The evaluation of its effectiveness shows that text adaptation is preferable to text generation as it is more reliable and consistent. With respect to question design, ChatGPT is reported to be effective in certain types, such as gap-filling, open-ended, and multiple-choice for individual questions. Its effectiveness in designing multiple-choice items in reading comprehension tasks raises some controversies from in-depth interviews. Its limitation in creating matching or error recognition exercises should also be taken into consideration.

It can be concluded from the analysis that ChatGPT has its own pros and cons, like other AI tools, and it is not a one-for-all application. The findings of this study also strengthen theoretical and empirical evidence from existing studies like Hong (2023), Rudolph et al. (2023), and Tate et al. (2023). It is recommended that users edit their instructions many times to navigate the chatbot to follow their requests and meet their requirements. If users are patient enough and sharpen their queries, they will achieve the desired results from the chatbot. Also, another

implication drawn from the study is that teachers should not rely on ChatGPT. It will support teachers in the way they want; however, it cannot replace them. In test design, teachers still control the process of requesting, selecting, revising, and refining. ChatGPT is expected to play the role of a facilitator to reduce their workload.

As the number of participants is not high, this study has its own limitation in generalizing its findings for a bigger group of the population. The research also limits the aspects of language test design that it approached in the questionnaire. Further studies with a greater number of participants and more investigated areas are recommended. Also, a comparison between the perspectives of English teachers and teachers of other languages should be conducted to see if any gap exists. A future study on the application of ChatGPT into language testing and assessment is highly recommended as well.

Acknowledgment

References

Cotton, D.R., Cotton, P.A., & Shipway, J.R. (2023). Chatting and cheating. Ensuring academic integrity in the era of ChatGPT. https://doi.org/10.1080/14703297.2023.2190148

Choi, J.H., Hickman, K.E., & Monahan, A.B. (2023). ChatGPT goes to law school. https://dx.doi.org/10.2139/ssrn.4335905

Dao, X.Q., Le, N.B., Phan, X.D, & Ngo, B.B. (2023). Can ChatGPT pass the Vietnamese National High School Graduation Examination? http://dx.doi.org/10.48550/arXiv.2306.09170

Hong, W.S.H, (2023). The impact of ChatGPT on foreign language teaching and learning: opportunities in education and research. *Journal of Education and Technology and Innovation, 5*(1), 37-45.

E.Kasneci *et al*. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*. 103. https://doi.org/10.1016/j.lindif.2023.102274.

Le, A.V., Bui, T.D., Le, Q.Q., & Vu, V.L. (2023). Khả năng thực hiện bài kiểm tra định kì môn Toán và môn Ngữ văn cấp Trung học của công cụ ChatGPT: Kết quả nghiên cứu và một số khuyến nghị ban đầu. *Tạp chí Khoa học và Giáo dục Việt Nam, 19*(2), 1-10.

Moore, N.S., & MacArthur, C.A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research, 8*(1), 149-175.

Nguyen, Q. H. (2023). AI and Plagiarism: Opinion from Teachers, Administrators and Policymakers. *Proceedings of the AsiaCALL International Conference*, 4, 75–85. https://doi.org/10.54855/paic.2346.

Nisar, S., & Aslam, M.S. (2023). Is ChatGPT a good tool for T&CM students in studying pharmacology? Available at http://dx.doi.org/10.2139/ssrn.4324310

Raford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Retrieved from

 http://www.cs.ubc.ca/amuham01/LING530/papers/radford2018improving.pdf

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of the traditional assessment in higher education. *Journal of Applied Learning & Teaching*, *6*(1), 242-263. Available at https://journals.sfu.ca/jalt/index.php/jalt/index

Tate, T., Doroudi, S., Ritchie, D., & Xu, Y. (2023). Educational research and AI-generated writing: Confronting the coming tsunami. Available at https://doi.org/10.35542/osf.io/4mec3.

TESOL International association. (2023). AI for Language Learning: ChatGPT and the Future of ELT. Available at http://blog.tesol.org/ai-for-language-learning-chatgpt-and-the-future-of-elt/

Thai, T.C.T. (2023). Thái độ và kì vọng của sinh viên sư phạm tiếng Anh đối với ChatGPT: Nghiên cứu tại trường Đại học Sư phạm Hà Nội. *Tạp chí Giáo dục, 23*(10), 51-61.

Tran, T.N., & Tran, H.P. (2023). Exploring the role of ChatGPT in developing critical digital literacies in language learning: a qualitative study. *Proceedings of the AsiaCALL International Conference*, *4,* 45-62. https://doi.org/10.54855/paic.2344.

Yeadon, W., Inyang, O.O., Mizouri, A., Peach, A., & Testrow, C. (2022). The death of the short-form physics essay in the coming AI revolution. *Physics Education, 58*(3). https://iopscience.iop.org/article/10.1088/1361-6552/acc5cf/pdf

Zhai, X. (2022). ChatGPT user experience: Implications for education. Available at http://dx.doi.org/10.2139/ssrn.4312418.

Biodata

Nguyen Thi Phuong Thao (Thao Nguyen) is a lecturer and researcher at VNU University of Languages and International Studies. She earned a Master's degree in English Language Teaching from the University of Southampton, UK. She has participated in research projects on test development and examiner training. Her research interests include testing and assessment, teacher education, and professional development.