Evaluating HUFLIT Lecturers' Perspectives on ChatGPT's Capabilities in Designing English Testing and Assessment

Le Thi Thu Huong[1*]

[1] Ho Chi Minh University of Foreign Languages and Information Technology, Vietnam

[*]Corresponding author's email: huongltt@huflit.edu.vn

[*] https://orcid.org/0009-0003-1074-7589

ABSTRACT

Despite its widespread adoption in education, the potential of ChatGPT in English language teaching and learning, especially in testing and assessment design, still needs to be studied. This paper, therefore, seeks to investigate lecturers' perspectives on the capabilities of ChatGPT in designing English testing and assessment with the participation of thirty lecturers of the Faculty of Foreign Languages at HUFLIT. Data from the questionnaire and semi-structured interview suggest that ChatGPT is a promising tool for designing English testing and assessment owing to its versatility in drafting quicker, more diverse, and more detailed types of questions, resulting in enhanced time efficiency and reduced efforts in test design. The greatest areas of concern, however, lie in ChatGPT's inability to generate questions that match the specific learning outcomes and students' linguistic abilities. The study contributes to providing lecturers with in-depth insights into the potentialities and challenges of ChatGPT in testing and assessment design, assisting them in employing this AI tool more critically and effectively.

Introduction

Evaluation of students' academic progress has always been an indispensable part of the teaching and learning process, through which teachers can gain in-depth insights into such aspects as the quality of their lessons, the varying degrees of knowledge absorption among students, and the appropriateness of the current teaching methodologies. Teachers are also more likely to assess the real challenges facing their students, through which they can seek effective measures to tackle these prevailing issues. On students' side, evaluative activities provide them with invaluable opportunities to reflect on their strengths and weaknesses, the efficacy of the learning strategies they are adopting, the gaps in their knowledge, as well as the potential steps they might take to fill these gaps and improve their academic performance.

A commonly held belief is that testing and assessment are powerful and effective measures of students' achievements. Testing and assessment tools can be employed to provide "systematic feedback for both teachers and learners" (Beikmahdavi, 2016, p. 130) and are a means of checking whether there is a match between the set learning outcomes and students' actual performance; therefore, designing a test or an evaluative activity often requires long hours of sustained efforts and critical thinking and must ensure a variety of objective criteria including "practicality, validity, and reliability" (Beikmahdavi, 2016, p. 130). Traditionally, teachers who were supposed to design tests and different forms of assessment used to take on the responsibility manually, which involved selecting the tested knowledge areas or skills, writing down question types and exercises on paper, generating answers, and proofreading the tests to double-check typical errors. Along with technological advances, the drafting of tests and assessment tasks has become less complicated when teachers are able to compose tests and assessment forms on their computers or laptops and make full use of free checkers online for error rectification, which assists them in saving large amounts of time and energy.

The incorporation of Artificial Intelligence (AI) tools into language teaching and learning activities has brought a new breath of air into educational settings, lifting learners out of the boredom of being confined in classroom walls while simultaneously taking them to a virtual world filled with interactive games and scaffolding learning activities that facilitate the process of knowledge acquisition and skills enhancement with greater ease. In terms of evaluation, AI tools such as ChatGPT and Gemini, to name but a few have made radical transformations to not only the way students are tested but also the job of designing tests and exams carried out by teachers. Instead of devoting hours to composing questions for a test or an exam, teachers, especially those working at tertiary institutions, find it much more straightforward to embark on testing and assessment design with the aid of ChatGPT.

Studies on ChatGPT's potential for devising testing and assessment are available (Almasre, 2024; Anderson et al., 2001; Kolade et al., 2024; Rasul et al., 2023; Rudolph et al., 2023). However, research conducted on the same topic in language teaching and learning is still limited, and studies regarding the utilization of ChatGPT to design English tests and other forms of assessment are also scarce in the context of Vietnamese education. There is an urgent need for more studies to investigate the capabilities of ChatGPT in facilitating lecturers' task of designing tests and relevant forms of assessment.

Literature review

An overview of ChatGPT

The release of ChatGPT by OpenAI in 2022 has entailed a host of opportunities in various major fields of life while simultaneously posing inevitable challenges to its users. Being "built on top of OpenAI's GPT-3 family of large language models and is fine-tuned with both supervised and reinforcement learning techniques" (Rudolph et al., 2023, p. 345). ChatGPT allows engagement in conversations resembling those of humans and produces texts that are challenging to distinguishable from human-generated ones (Roumeliotis & Tselikas, 2023). Utilizing ChatGPT gives users a sense of two-way interaction since it can respond to any

prompts of various forms, such as requests or questions, in a quick manner, and the responses can be meticulous and generally accurate (Gonsalves, 2023; Roumeliotis & Tselikas, 2023; Rudolph et al., 2023). It is ChatGPT's vast storehouse of linguistic knowledge and its competence in pattern identification and analysis that make it likely to undertake a diversity of tasks of varying degrees of complexity required by users (Almasre, 2024; Farazouli et al., 2023), ranging from creating quizzes, answering questions, to planning activities and translating texts as well. In other realms, ChatGPT can demonstrate its capabilities in the production of content, code, and pieces of writing (Almasre, 2024). It can be, therefore, said that ChatGPT has proved its enormous capacity for generating human-like conversations and texts of diverse purposes.

Merits of ChatGPT in education

One of the most outstanding benefits of employing ChatGPT in teaching and learning is this AI tool can provide learning experiences that are "tailored", "customized", "personalized", or "individualized" (Amin, 2023, pp. 1-2; Baskara & Mukarto, 2023, p. 349; Rasul et al., 2023, pp. 4, 6) which, according to Amin (2023), often involves handling issues relevant to each individual student's needs and paces of learning. He also added that this kind of personalized learning proved to be an obstacle to conventional classrooms where teachers could not respond to every student's needs as effectively as ChatGPT. In language learning, personalized learning is of crucial importance as it helps to facilitate the process of acquiring knowledge and skills among learners of different abilities and preferences. What is more, ChatGPT has the potential to assist lecturers in integrating "real-time language practice" (Amin, 2023, p. 1) into their lesson or teaching activities, meaning that the increasing exposure to real-life situations affords students the opportunity to boost their linguistic skills since they can practice applying the learned speaking patterns or lexical items to their daily conversations and in-class activities like debates or role-plays, thereby gaining a thorough understanding of the language they strive to master. From this, skills such as problem-solving and critical thinking are also greatly fostered. Another interesting feature of ChatGPT is the provision of prompt feedback and "round-the-clock access to knowledge" (Graefen & Fazal, 2024; Rasul et al., 2023). Indeed, feedback forms a solid base of precious information that can inform students of their strengths, weaknesses, and methods of improving their existing shortcomings. Since human teachers are not always approachable, gaining feedback from AI assistants such as ChatGPT seems more practical. Ready access to knowledge at all times of the day benefits language learners in an enormous way. By means of making question prompts relevant to the areas of knowledge they wish to delve into, learners could acquire an understanding of certain issues across various domains with greater ease. Last but not least, ChatGPT's ability to expose students to authentic sources of language could allow students to be engaged in meaningful practice that further harnessed their linguistic proficiency and comprehension skills, along with arousing their motivation for learning (Baskara & Mukarto, 2023).

Downsides of ChatGPT in education

In spite of being known for its tremendous benefits in education, ChatGPT also encountered a host of criticisms. One of these is the provision of information or content that is regarded as "incorrect or misleading", "biased or inappropriate" (Borji, 2023; Sallam, 2023, as cited in

Baskara & Mukarto, 2023, p. 347). In their research findings, Baskara & Mukarto (2023) also pointed out that ChatGPT was lacking in "human nuance" which involved an awareness of cultures and language use in an appropriate manner. This is understandable since ChatGPT has been primarily trained on language models to produce texts that bear a resemblance to human ones. Meanwhile, the mastery of language is not confined to the proficiency of linguistic skills, but it demands learners to grasp an adequate understanding of the culture of the targeted language and communicate their ideas with real emotions. Another concern raised among educators regarding the adoption of ChatGPT is that it can lead students to commit plagiarism as students are overreliant on the chatbot, utilizing it to gain better academic achievements while sacrificing their own writing skills. Additionally, the findings drawn from the study undertaken by Pham and Le (2024) indicate that excessive dependence on ChatGPT could undermine the competence of thinking creatively and critically among students. Finally, ChatGPT may encounter difficulties concerning the provision of precise responses when it comes to a particular domain (Abas et al., 2023) due to the lack of inadequate training in that domain, which could detrimentally influence students' academic performance.

Testing and assessment

Testing and assessment methods serve as pivotal tools that "gauge students' knowledge, skills, and progress" (Amin, 2023, p. 2), providing invaluable sources of information for betterment in achieving outcomes in both teaching and learning. The dual purpose of assessment described in the words of Kolade et al. (2024, p. 2) is "facilitation of learning on the one hand, and certification of achievement, on the other." This twofold aim can be fulfilled provided that there are concerted efforts between teachers in employing methodologies and assessment tools that are ideally suited to students' abilities and learners who willingly embrace the teaching and evaluation methods.

Testing and assessment in English language teaching and learning

In the educational domain, especially English language teaching and learning, students' academic performance has been evaluated in a diversity of ways. Traditionally, teacher-based assessment methods served as the principal ones for monitoring students' progress in particular English subjects, with summative assessment methods being adopted to measure what students have achieved for a particular duration of time by means of tests or examinations (Almed et al., 2019; Trumbull & Lash, 2013). Summative assessment featuring a pen-and-paper test format has been favored by numerous teachers owing to their acquaintance with this format and its simplicity (Fadilah et al., 2023). Brown (2004) (as cited in Almed et al., 2019) highlighted the strength of summative assessment, which primarily revolves around the evaluation of students' accomplishment of course objectives, and pointed out its limitation in providing orientation for their future enhancement in learning. It is this limitation that has aroused doubts about the efficacy of summative assessment in assisting students to make academic improvements (Mansory, 2020), leading to the emergence of formative assessment methods that can be implemented on a more regular basis, involving greater interaction, and encompassing a mixed array of tests and evaluative activities, all of which do not aim to form ultimate judgments on students' performance over a specific time span but direct students' learning based on individual learning needs and garner feedback serving as guidelines for facilitating both students'

academic studies as well as teachers' professional development (Trumbull & Lash, 2013; Widiastuti & Saukah, 2017). According to these researchers, the major distinction between these two types of assessment lies in the fact that the former solely focuses on the collection of final results or products, resulting in the absence of feedback for improvement suggestions, whereas the implementation of the latter is an on-going process, measuring students' abilities through a wider range of tasks and instructing students on how to make progress based on feedback gathered from their performance in each activity.

Formative assessment in English language classrooms usually includes such activities as open discussion, task-based learning, self and peer assessment (Black & Jones, 2006), homework, projects, journal writing, role-plays, assessment through portfolios (Chandio & Jafferi, 2015), performance-based assessments, games, and teachers' observations (Moqbel & Al-Kadi, 2023), and so on. These various activities are designed to foster students' linguistic skills and competencies rather than place an emphasis on grammar and memorization abilities (Chandio & Jafferi, 2015).

The potential of ChatGPT in tests and assessment design

Designing exam questions and tests used to be an arduous and time-consuming task. Since the advent of ChatGPT, however, much of teachers' job of devising tests and evaluative activities has been eased in terms of grading and feedback offering owing to two benefits, which are accurate grading automation and detailed comments (Almasre, 2024). Almasre (2024) pointed out in his research when assessing 12,100 essays from the ETS Corpus of Non-Native Written English (TOEFL11) based on the criteria specified in the IELTS Writing Task 2 band score descriptors by means of ChatGPT that the scores of the compositions were perceived to achieve a good degree of reliability and precision. Meanwhile, the feedback provided by ChatGPT was more specific and legible compared to that offered by teachers. ChatGPT has gained recognition for its capability in drafting assessment questions, especially in initial idea generation for assessment design, creation of multiple choice questions, production of case studies drafts, or other forms of assessment for further edition (Rasul et al., 2023). Apart from this, ChatGPT is known for its ability to re-evaluate tasks, focusing on developing students' higher-order thinking, such as creativity and critical thinking (Nieminen et al., 2022, as cited in Rasul et al., 2023) and "the potential to serve as a means of generating different scenarios for students to work together to solve problems and achieve goals" (Rudolph et al., 2023). Moreover, ChatGPT-generated scenarios bear significant resemblance to those of real-life situations (Kolade et al., 2024). This means that ChatGPT is capable of evaluating students' abilities and skills, reflected through its versatility in generating a mixed variety of tasks to assess learning outcomes, and the scenarios generated by the chatbot can be beneficial to language learners in their training of communication, problem-solving, and teamwork skills. Anderson et al. (2001) suggested ways to leverage ChatGPT's capabilities on making lesson plans, scaffolding writing, and drafting assessment activities, highlighting that the tool can be trained to cope with ideas in the 'lower order thinking' category in accordance with Bloom's taxonomy, and the responses generated by ChatGPT should be critically evaluated to assess its competence in handling questions that require higher order thinking skills.

The potential of ChatGPT in English tests and assessment design

Since its arrival, ChatGPT has been perceived as a promising tool for English language teaching and learning as well as assessment design. The results obtained from the study conducted by Zaiarna et al. (2024) suggest that ChatGPT is particularly beneficial to test and assessment design across language aspects, particularly in grammar, writing, and vocabulary. In addition, a study undertaken by Koraishi (2023) reveals that ChatGPT is capable of generating reading passages that can cater to individual preferences, learning needs, and levels of proficiency in accordance with the predetermined topics and contexts. Koraishi (2023) also added its potential to design corresponding kinds of comprehension questions coupled with its capacity to adjust the questions based on rephrased prompts can also help teachers utilize the ChatGPT-generated content but still ensure its suitability for different classes. Designing tests, quizzes, and activities for vocabulary consolidation in diverse formats has become much simpler and more time-efficient for teachers as well with the aid of ChatGPT (Koraishi, 2023). From this researcher's findings, ChatGPT demonstrates its capabilities not only for producing varied question types, placement tests, and standardized tests but also for providing evaluation of students' linguistic levels, thorough, constructive feedback, and recommendations for corrections on the basis of set criteria. In a similar fashion, the results of the study conducted by Nguyen (2023) on the topic of applying ChatGPT in language test design show that this AI assistant tool is adept at generating texts not only for reading but also for listening comprehension provided that teachers give detailed instructions on the characteristics of the outcome they wish to achieve and highlight the adaptation ability of ChatGPT in text refinement to assist teachers to enhance or lower the textual level of difficulty. Moqbel & Al-Kadi (2023) suggested in their research that the employment of ChatGPT could act as a tool for acquiring information related to students' strengths and weaknesses and learning strategies so as to decide on the most appropriate assessment forms. Likewise, Amin (2023) also pointed out the utilization of ChatGPT to provide informative analysis of students' responses to various assessment tasks so that areas of improvement could be identified for enhancing learning outcomes.

The drawbacks of utilizing ChatGPT in tests and assessment design

A limitation concerning the use of ChatGPT in assessment in the findings of Rudolph et al. (2023) is this AI-assisted tool fails to illustrate mathematical problems with diagrams. Likewise, in his research on how to draft multiple-choice questions that challenged ChatGPT, Gonsalves (2023) suggested that teachers should take advantage of questions that require ChatGPT to handle content relevant to figures, images, or charts since it is unable to read visually presented information. This means that if teachers wish to devise exam questions concerning the use of images or diagrams, it may prove a major obstacle to the chatbot. He added that user prompts also had a decisive effect on ChatGPT's competence in handling the task. If a prompt shows a high degree of complexity, it could overlook some parts of the prompt and tend to produce inaccurate answers. Rudolph et al. (2023) found out that for essay writing, the chatbot was likely to produce a composition with an amazing speed; however, the content was regarded as lacking in "breath and depth," which was even deteriorated by its failure to include in-text citations and references. It was also pointed out by Rudolph et al. (2023) that the chatbot was incapable of generating content requiring higher-order thinking, such as critical and analytical

thinking skills. Finally, a shortage of critical thinking originating from the ready availability of information coupled with the simple process of text creation (Sullivan et al., 2023; Tlili et al., 2023, as cited in Kostka & Toncelli, 2023), and the imprecise information produced by ChatGPT (Kostka & Toncelli, 2023) demanded a great emphasis on critical thinking and evaluation abilities (Heaven, 2023) among lecturers.

Research Questions

The purpose of this study is to evaluate the capabilities of ChatGPT in assisting HUFLIT lecturers of English in designing tests and other types of assessment. The research aims to seek the answers to the following questions:

1. What are ChatGPT's potential capabilities in the design of tests and assessments from the standpoint of English lecturers at HUFLIT?
2. What are the challenges HUFLIT's lecturers of English may encounter regarding ChatGPT's competence in tests and assessment design?
3. What can lecturers of English at HUFLIT do to enhance ChatGPT's capabilities in tests and assessment design?

Methods

Pedagogical Setting & Participants

The participating subjects of the study involved thirty lecturers of English of the Faculty of Foreign Languages at Ho Chi Minh University of Foreign Languages and Information Technology (HUFLIT). These lecturers, whose age ranges from 27 to just under 45, possess a diversity of years of teaching experience in English-related subjects, ranging from English skills to English theory and specializations such as English phonetics and phonology, morphology, syntax, semantics, English for business purposes, and English language teaching. The proportion of female lecturers is significantly higher than that of males, and the participants have had experience of varying degrees in adopting ChatGPT to design English tests and other types of evaluative assessment for a range of classes of different English competence levels (primarily from A2 to C1).

Design of the Study

The research paper employed a mixed-methods approach, with the main instruments consisting of a questionnaire of various question types and a semi-structured interview. The rationale for adopting the questionnaire is its great convenience, quickness, and ease of implementation (Kuphanga, 2024), while the semi-structure interview allows the examination as well as the gathering of diverse perspectives and attitudes of participants, thereby providing in-depth insights into the research topic (Jain, 2021), which in turn contributes to the enhancement of the reliability and validity of the findings.

The questionnaire was designed based on a conceptual framework with the adoption of the convenience sampling method. The factors contributing to ChatGPT's competence in devising tests and assessments are placed under one group called independent variables, while the dependent variable is ChatGPT's potential and challenges in test and assessment design. Below

is the illustration of the conceptual framework for framing the questionnaire:

Figure 1. The conceptual framework illustrating the potential capabilities and challenges of employing ChatGPT in test and assessment design
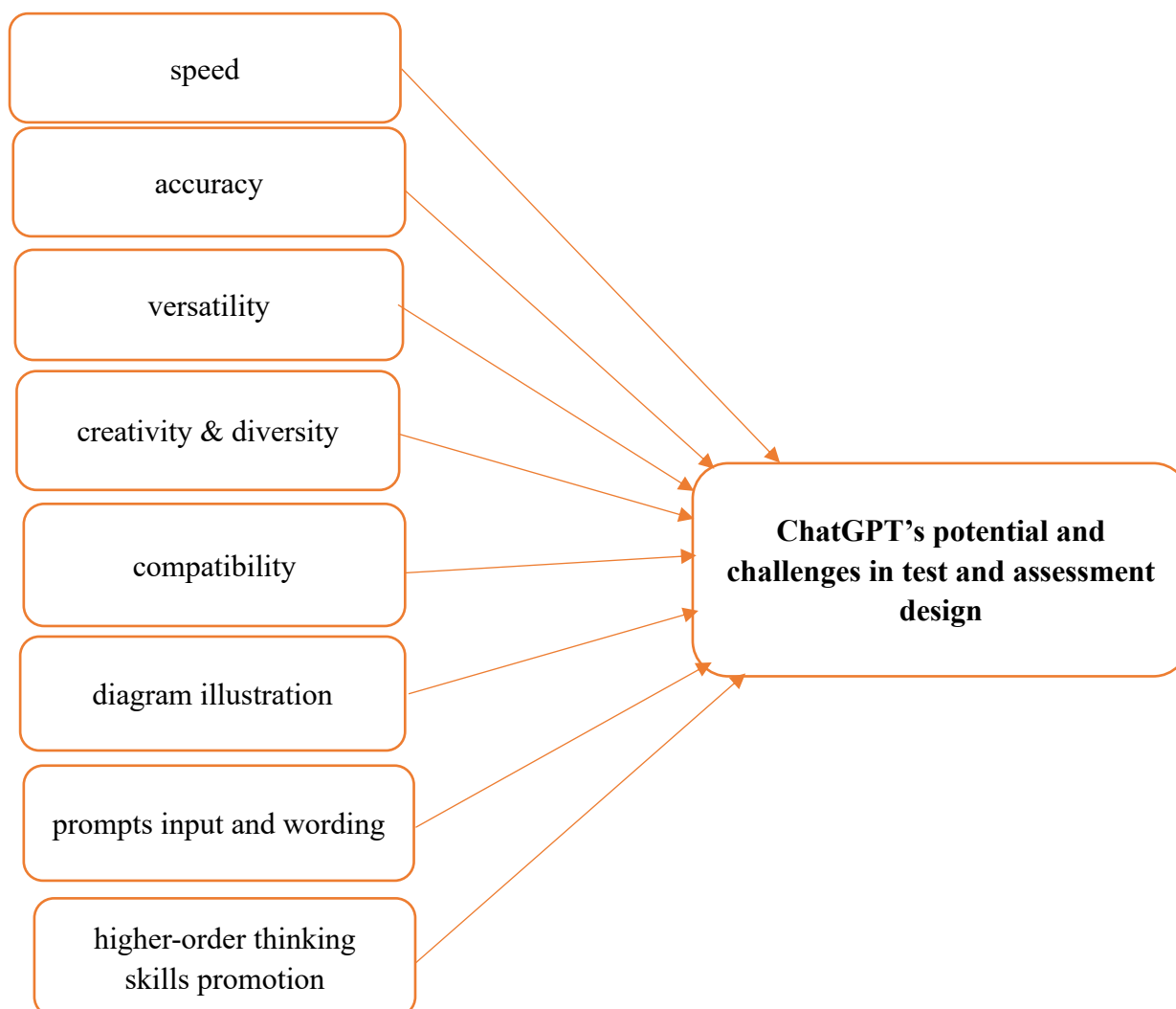
The vast majority of the questionnaire items were designed based on the findings and discussion of Rudolph et al.'s (2023) and Gonsalves' (2023) studies, while the conceptual framework was adapted from the conceptual framework named "The benefits and challenges of ChatGPT – an integrated framework" suggested by Rasul et al. (2023). Several questions with further details were added based on the researcher's experience while utilizing ChatGPT as a test and assessment design tool, but their content still centered around the main concepts in the above-mentioned framework.

The questionnaire encompassed twenty-one questions revolving around four major sections pertinent to the use of ChatGPT in tests and assessments, the benefits of employing ChatGPT in designing English tests and assessments, the challenges concerning the adoption of ChatGPT in English tests and assessment design, as well as lecturers' recommendations for enhancing ChatGPT's capabilities in the drafting of English tests and forms of assessment. Apart from multiple choice and 5-point Likert scale questions, open-ended questions were also included in

the questionnaire to investigate the participants' points of view on the employment of this AI-assisted tool on their tests and assessment designing process. All of the questions were composed on Google Forms, whose link was quickly generated and then distributed to the lecturers via their Zalo accounts.

The semi-structured interview was composed of three fundamental questions and a number of follow-up questions on the basis of the research questions formulated in the previous section, which was conducted on Microsoft Teams with the participation of seven randomly chosen lecturers. Convenience sampling was employed in the construction of the interview so as to select the interview samples which could be easily approachable and eligible for the research study.

Data collection & analysis

The distribution of the questionnaire commenced in the last week of May 2024 and lasted two weeks. The link to the questionnaire was generated from Google Forms and administered through Zalo. All of the survey participants' answers were automatically saved in the tab entitled 'Responses' on Google Forms, which greatly facilitated access to the data later on. Prior to the conduct of the interview, approval was obtained from the lecturers who were invited to respond to the questions on the topic of evaluating ChatGPT's abilities in designing English tests and assessments. Each interview session took place in approximately 15 minutes, and the answers were recorded and transcribed for analysis.

The data garnered from the questionnaire had undergone a meticulous process of analysis before being presented in the form of different figures, such as diagrams and tables, with the support of Google Forms. Thematic analysis was utilized to identify themes prevailing in the interviewees' responses, which enables the provision of detailed insights into the participants' attitudes towards the adoption of ChatGPT in devising English tests and assessments.

Findings and discussion

Research Question 1: What are the potential capabilities of ChatGPT in the design of tests and assessments from the standpoint of lecturers of English at HUFLIT?

### *Results of the questionnaires*

Section 1 of the questionnaire deals with the use of ChatGPT in tests and assessment design.

Figure 2. The types of questions ChatGPT was often required to design
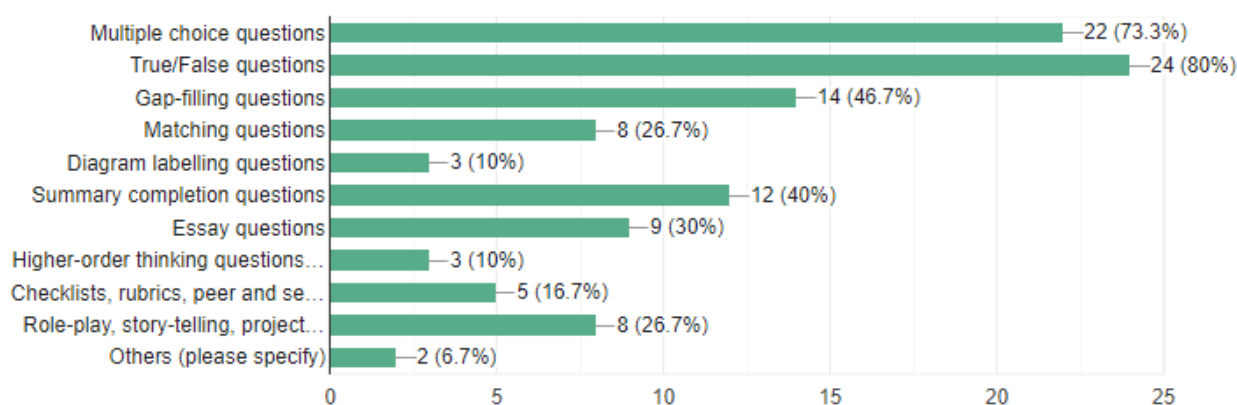
As can be seen from Figure 2, the two most prevalent types of questions ChatGPT was required to design were True/False (80%) and multiple choice questions (73.3%). This was followed by gap-filling, summary completion, and essay questions, making up 46.7%, 40%, and 30% respectively. Matching questions,, as well as role-play, story-telling, and project activities,, equaled in terms of popularity, accounting for 26.7% each, which was more common than such activities as checklists, rubrics, peer, and self-assessment. Diagram labelling and higher-order thinking questions (evaluation, critical thinking, or problem-solving questions, etc.) were the least preferred kinds of questions ChatGPT was requested to devise.

Figure 3. Rating of ChatGPT's capabilities of designing English tests and assessment
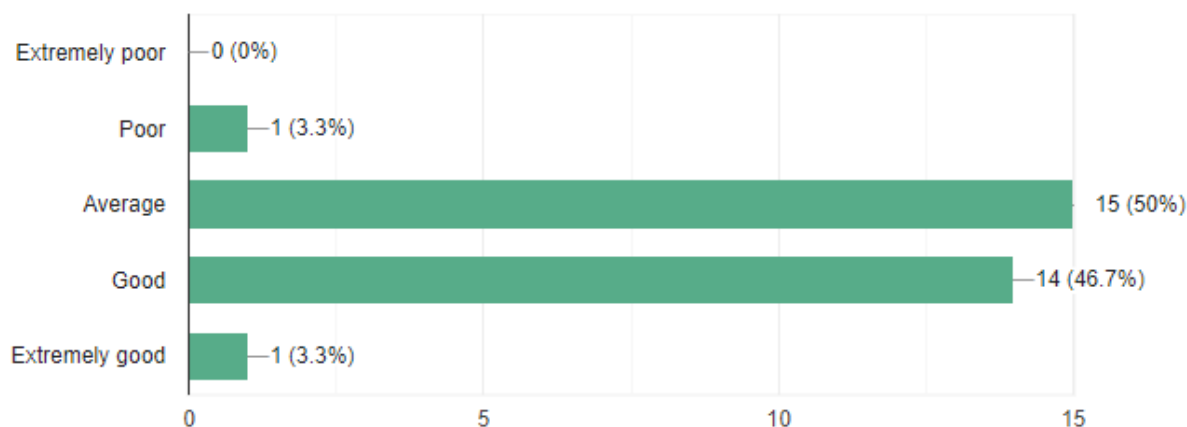
According to Figure 3, the percentages of participants who ranked ChatGPT's competence in English tests and assessment design as good or average were roughly similar, at around half for each, with only one lecturer having perceived the tool to be an excellent tool, whereas one participant considered it to be poor in drafting tests and exam questions.

Section 2 is centered around the lecturers' perspectives on the benefits of employing ChatGPT in designing English tests and assessments.

Table 1. The benefits of employing ChatGPT in designing English tests and assessment

Statements	*Strongly agree*	*Agree*	*Neutral*	*Disagree*	*Strongly disagree*
ChatGPT can create questions that match lecturers' prompts.	*3.3%*	*60%*	33.3%	0%	3.3%
ChatGPT can devise questions more quickly than lecturers can.	*23.3%*	*46.7%*	30%	0%	0%
ChatGPT-produced questions are more accurate than human-generated questions.	0%	36.7%	43.3%	20%	0%
The questions generated by ChatGPT are more interesting than those designed by lecturers.	0%	26.7%	46.7%	23.3%	3.3%
ChatGPT can create a wider range of questions.	*20%*	*50%*	16.7%	10%	3.3%
ChatGPT-generated questions can cover a wide variety of theoretical and application aspects.	*3.3%*	*40%*	40%	16.7%	0%
ChatGPT can draft questions with very detailed instructions on how to implement the activities in the classroom.	*6.7%*	*43.3%*	36.7%	13.3%	0%
ChatGPT can also design questions to evaluate students' higher-order thinking skills.	3.3%	30%	43.3%	23.3%	0%
Utilizing ChatGPT to design tests and assessments can save lecturers large amounts of time and energy.	*36.7%*	*46.7%*	16.6%	0%	0%
Lecturers can learn from the structured format of the questions generated by ChatGPT, from which their competence in designing tests and assessments could be developed.	*16.6%*	*56.7%*	26.7%	0%	0%

A closer look at Table 1 reveals that the most outstanding strength of ChatGPT, from the viewpoint of lecturers of English at HUFLIT, was ChatGPT's potential to save them large amounts of time and energy in drafting questions (83.4%), the opportunities for these lecturers to nurture their competence in designing English tests and assessment through the consultation of structured formats of the questions generated by the chatbot (73.3%), the tool's capabilities to produce a wider range of questions as well as in a quicker manner compared to those devised by human teachers (70% each), and it's ability to create questions that match lecturers' prompts (63.3%). ChatGPT was also highly evaluated for its drafting of questions with very detailed

instructions on implementing the classroom activities. The percentage of lecturers expressing their approval of ChatGPT's competency in covering a wide variety of theoretical and application aspects comprised just under half. Those who agreed with the statement that the tool could generate more precise questions than human-produced questions and evaluate students' higher-order thinking skills accounted for similar proportions, at approximately a third. The figure for lecturers who regarded questions designed by ChatGPT as more interesting than those devised by human teachers was in the minority, at slightly more than a quarter.

The percentages of participants who neither approved nor disapproved of the various capabilities of ChatGPT made up the lowest in terms of time and effort saving (16.6%) and the highest when it comes to accuracy, evaluation of students' higher-order thinking skills, and degree of interest in the questions generated by ChatGPT. The two latter aspects were also the benefits that were the most disagreed by the survey respondents, with each accounting for almost one in four.

Research Question 2: What are the challenges HUFLIT's lecturers of English may encounter regarding ChatGPT's competence in tests and assessment design?

Section 3 consists of a number of statements in which the participants had to express their opinions about the obstacles they encountered when utilizing ChatGPT in drafting exam questions and evaluative activities.

Table 2 highlights the challenges the lecturers in charge of various English-related subjects at HUFLIT experienced while adopting ChatGPT in their tests and assessment design process. The greatest concern shared among the lecturers was ChatGPT's inability to devise questions fulfilling their expectations without providing specific and appropriate prompts, with a massive 80% exhibiting their approval of this. There was also broad agreement among the surveyed participants on the imprecision of the questions created by the chatbot (around two-thirds), with the tendency of not matching students' linguistic competency not far behind, at 63.3%, followed by the limitation of the tool in devising questions to evaluate students' higher-order thinking skills since the questions produced this way proved to be a challenge for learners to have a full grasp of them. Just over half of the respondents were in agreement with ChatGPT's failure to provide picture cards or images when it comes to picture-related activities. As regards questions that are concerned with lower-order thinking skills, a significant percentage considered the questions dealing with this aspect to be rather repetitive and lacking in creativity.

Table 2. The challenges concerning the adoption of ChatGPT in designing English tests and assessment

Statements	*Strongly agree*	*Agree*	*Neutral*	*Disagree*	*Strongly disagree*
Without specific and appropriate prompts, ChatGPT cannot generate questions up to lecturers' expectations.	*__30%__*	*__50%__*	10%	10%	0%
ChatGPT-produced questions to evaluate students' lower-order thinking skills are quite repetitive and lacking in creativity.	3.3%	43.3%	43.3%	10%	0%
Some corresponding answers to questions provided by ChatGPT are not precise.	*__3.3%__*	*__63.3%__*	30%	3.3%	0%
ChatGPT-produced questions to evaluate students' higher-order thinking skills are sometimes challenging for students to fully understand.	**3.3%**	**56.7%**	36.7%	0%	3.3%
ChatGPT-produced questions do not always match students' language levels.	*__3.3%__*	*__60%__*	23.3%	13.3%	0%
ChatGPT-produced questions cannot provide picture cards or images when it comes to picture-related activities.	13.3%	40%	40%	6.7%	0%

The next open-ended question presents other problems associated with the employment of ChatGPT in devising English tests and assessments. Ideas varied across lecturers; however, there were some concerns related to the input of prompts to ask ChatGPT to generate exam or test questions. Many of them emphasized the significance of writing appropriate, specific, and concise prompts that could convey the lecturers' demands in order to enhance the reliability of ChatGPT-produced questions as well as the need for adjustment and insertion of detailed prompts a number of times to create expected questions:

> *"Giving appropriate prompts so that ChatGPT can generate questions that match my expectations."*

> *"The test designer has to input very detailed instructions many times in order to get what he or she wants."*

> *"Without enough specific and concise prompts, ChatGPT will not be supportive and reliable for teachers to design tests."*

Three lecturers pointed out a limitation of ChatGPT in designing tests and assessments concerning the accuracy of answers provided by the chatbot and the level of difficulty of the questions that may not be well-suited to students:

"Some questions are too simple but some are too complex for students to figure out. Sometimes, it also provides wrong answers."

"Sometimes the assessment is not precise."

"ChatGPT cannot provide exact test and assessment for mix-classes."

Another concern raised was the suitability of ChatGPT-generated questions to match students' levels and learning outcomes:

"In my opinion, student level is another problem associated with the adoption of ChatGPT in devising tests and assessment."

"unsuitable level for all learners, not accurately access specific lesson goals"

"Whether the test items designed by ChatGPT match learning outcomes"

One lecturer expressed skepticism about the "reliability and validity" of the test devised by ChatGPT, while another affirmed the tool's inability to design a complete test but pointed out its potential to provide perfect ideas for test design:

"ChatGPT couldn't fully support us in designing the complete tests but its products can be our ideal suggestions when we don't have time."

Two participants suggested that lecturers should spend time checking or revising ChatGPT-produced questions to guarantee their precision and appropriateness to students' linguistic levels, whereas one surveyed respondent attributed his or her problem to a shortage of "technical support" from the faculty in the utilization of ChatGPT for testing design purposes:

*"It takes time to review, sometimes revise some questions because they are not suitable for the content."*

"We have to check both ChatGPT's questions and the answers to make sure that could be used for all students."

Research Question 3: What can lecturers of English at HUFLIT do to enhance ChatGPT's capabilities in tests and assessment design?

The final section revolves around the HUFLIT lecturers' recommendations for enhancing ChatGPT's capabilities in the design of English tests and assessments.

It is apparent from Figure 3 that the most agreed method of enhancing ChatGPT's capabilities in devising English tests and assessment activities was the provision of as specific and exact prompts as possible (93.3%). Adjusting the content of ChatGPT-produced questions to suit students' English language competency was also seen as an effective approach, with a large majority of lecturers opting for it. Another method was to rephrase prompts in many different ways, which could contribute to the training of ChatGPT's versatility in accommodating users' requests. The remaining strategies that could also boost the chatbot's abilities involved inputting challenging questions and prompts at different periods of time, with respective figures being 56.7% and 40%.

Figure 3. Methods of Enhancing ChatGPT's capabilities in English tests and assessment design
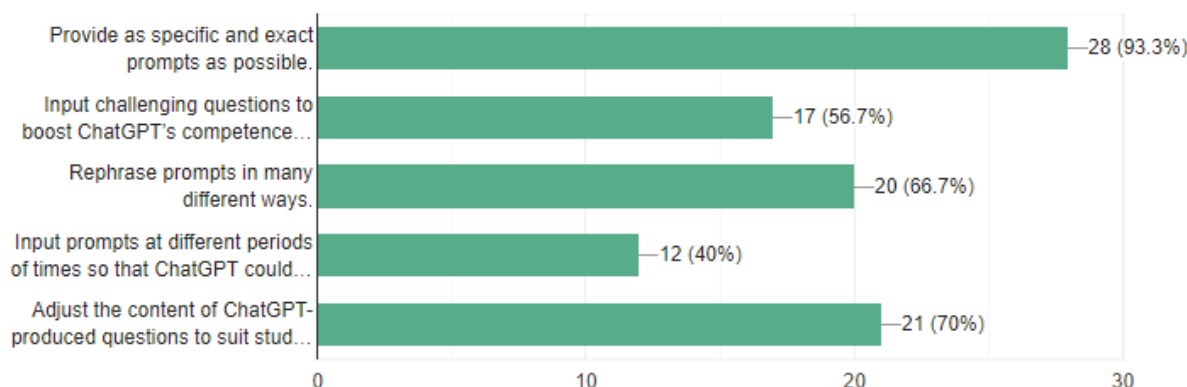
When it comes to other approaches to improve ChatGPT's capabilities in designing English tests and assessments, some lecturers recommended paying attention to the quality of the prompts, placing an emphasis on providing sufficient information concerning students' competence levels, goals of the tests, and kinds of tests for the prompts so that the outcome or test items could adequately and precisely address their requests:

> *"Lecturers need to provide clear instructions of students' levels, proficiency, and objectives of the tests."*

> *"Provide detailed requirements about the test, level, test type, etc."*

Apart from this, one lecturer stated that ChatGPT's capability in designing tests and assessments could be enhanced by inputting various prompts and challenging its comprehension and analysis by posing in-depth questions so that the chatbot can be trained on processing requirements in different ways:

> *"Put in a lot of prompts so that ChatGPT can learn from its history."*

> *"Question depth for ChatGPT"*

Another interesting idea was to integrate an example that illustrates the outcome of the test into the prompt so that ChatGPT was more likely to generate a test that met lecturers' expectations:

> *"Give a sample of a result that you expect to receive from ChatGPT when giving a prompt."*

> *"Give it some examples"*

One lecturer highly appreciated the chatbot for its suitability in creating formative assessments and advised teachers to solely utilize this AI-assisted tool for idea generation, which could help to reduce their reliance on the chatbot. Another lecturer mentioned the need for adjustment of ChatGPT-designed questions prior to distribution to students:

> *"I think it suits the formative assessment. Teachers should use it as a prompt for generating ideas, not too much depend on it."*

> *"Should adjust questions of tests before delivering them to students"*

Two participants emphasized the incorporation of training focusing on instructing them to employ ChatGPT to design tests and assessments in an effective and appropriate manner:

"Provide teachers with specific guidances of how to design tests using ChatGPT."

"My skills in using ChatGPT is still limited, so I think I should take part in many training sessions on how to use ChatGPT for test design."

To foster ChatGPT in tests and assessment design, the other lecturers proposed the critical combination of ChatGPT-produced tests and those tests created by human teachers as well as other tools such as Grammarly:

"Combine human-designed tests with ChatGPT-generated tests"

"Combine Grammarly and other apps with ChatGPT"

Results of the interviews

Theme 1: Lecturers' evaluation of ChatGPT's capabilities in designing English tests and assessment

Potential to generate a variety of test types

All of the interviewees held the belief that ChatGPT was competent in generating a mixed diversity of test formats and quizzes, ranging from multiple choice, True/False, gap-filling, to long or short answers, open or close-ended questions, and so on across a range of skills and disciplines, which significantly led to gains in time efficiency and effort reduction in test designing process. Among these, one lecturer emphasized the tool's usefulness in aiding lecturers with both formative and summative tests:

"I think when ChatGPT was invented, it was programmed with the theories of language testing, so I think it's helpful for teachers to design tests, even the quizzes. We can use them to … uhm… kind of… formative or even the summative tests. Uhm so… I have used it to design the multiple choice questions, even the open-ended, and close-ended questions."

"I used to utilize ChatGPT to design the reading test, and ChatGPT's very useful to help me to design the kind of tasks for example true or false, multiple choice, filling the blanks, short answers, long answers, and writing… It's very excellent in designing the writing topics for students."

"I think there are some benefits when we adopt ChatGPT in designing tests. Firstly, it is fast. By just typing a comment, ChatGPT will create a list of test sentences, so it saves a lot of time… and you can ask ChatGPT to create different types of questions, so it's various…The question types are various, so it really enhances the convenience. It's really handy and very convenient for the teachers… so it saves us a lot of efforts in designing tests."

An effective source of idea generation for testing

Some of the lecturers pointed out ChatGPT's potential to generate outlines or suggestions for various test types with great speed, efficiency, and creativity. The novelty of ChatGPT was

demonstrated in its ability to produce a wide range of activities and tasks that resemble real-life situations, thus stimulating students' interest and engagement in the tasks while simultaneously boosting the efficiency of the assessment process. The two interviewees showed their appreciation of ChatGPT's customization capability, tailoring tests and other forms of evaluation to lecturers' requirements so as to accurately assess students' academic progress:

> *"Moreover, one of the standout features of ChatGPT is its ability to customize questions to match different educational needs, whether teachers require simpler or more challenging tasks, so ChatGPT can address the complexity of these questions. Last but not least, ChatGPT can produce creative and varied task types that might add diversity to assessment. Let me give you an example. Instead of requiring students to make a conversation in pairs for the final test, I was stumped to make a podcast, design a poster, and make a presentation or even make an interview scenario. I think all of them might be more engaging to students, so this variety not only makes assessment more effective but also more engaging to students."*

Theme 2: Lecturers' evaluation of challenges encountered while adopting ChatGPT in English tests and assessment design

Obstacles in generating detailed and higher-order thinking questions

With respect to the challenges associated with the employment of ChatGPT in designing English tests and assessments, two of the interviewed lecturers asserted that ChatGPT could struggle with producing questions that test students' higher-order thinking skills, such as critical thinking and problem-solving since the assessment of these skills was challenging for this AI-assisted tool. They added that ChatGPT was likely to generate critical questions to evaluate students in this respect, but their questions should have been more advanced. Moreover, ChatGPT-produced questions were not as specific as expected, which deterred them from entirely relying on the tool for designing exam questions:

> *"Actually, it depends on the content that the teachers want to ask the students to apply these kinds of skills because to assess the critical thinking skills of students is kind of difficult for ChatGPT to design, I think so…, so that's the reason why we just use ChatGPT to just kind of the prompt or I say the hint to design the kinds of questions, not depend totally on the content of the questions designed by ChatGPT."*

> *"As I said before, I designed the tests, even the quizzes and… I think ChatGPT is just used as a prompt for me to know the kinds of questions I want to ask students, not totally depend the content of ChatGPT because it's very … I can see that ChatGPT's questions are very general, not go detailed, so that's the reason why ChatGPT will be the frame, you know, and I use that frame to redesign the content…"*

Lack of accuracy and up-to-date information

Another obstacle facing the lecturers in utilizing ChatGPT as a tool for tests and assessment design is its infrequent provision of inaccurate responses, leading to the consumption of time to recheck the accuracy of the test items. Producing questions that require an up-to-date source of knowledge is also an obstacle for ChatGPT to deal with:

"… but the most thing I'm worried that if they're tested, answers are not really correct for my students. Yeah so whenever I use ChatGPT to design a test, I have to waste time to read it again and again, again and again… to make sure everything is ok for the task…"

"Talk about the reliability, so as I have checked ChatGPT for the reading test, I think there are still some mistakes inside in terms of the main idea or the details… yeah the accuracy, and the way it creates a set of questions… maybe it… sometimes it overlapped.

"When you ask ChatGPT what is happening now, right now, so it maybe provide you with the wrong information; it could provide you with the information but it's not really correct …"

Unparaphrased or inappropriate questions

Concerning the quality of the questions, a few lecturers revealed ChatGPT's weakness in its incapacity to paraphrase the information in reading passages or draft multiple choice exercises to generate questions that were appropriate to students' linguistic competence, which also entailed a large amount of time for reassessment and refinement to guarantee a perfect match in terms of suitability and pertinence between the output and the learning objectives:

"There's a limitation. For example, I want to create a reading exercise which consists of 10 True/False sentences; however, there are some sentences that were not parphrased, just simply citing the sentences in the reading passage, which means they're too simple. Or sometimes, when I give a multiple choice exercise like checking vocabulary, ChatGPT gave options that were beyond the students' language level… so we have to review and readjust… so it'll take time, and I think that's a problem because it can't completely understand our requirements to design a test which is well suited to our objectives…"

"Well, one of the main challenges I can recognize is… is that I have encountered with using ChatGPT is ensuring the relevance and accuracy of the content it generates. Usually the questions generated by ChatGPT might not perfectly align with specific lesson objectives or course learning outcomes. Let me give you an example. For example, I have just finished designing a final test for non-English majored students. I mean I had to design multiple choice questions for vocab, grammar, and reading skills. So for multiple choice questions, I have to think of… I ask ChatGPT to design 4 options, so it's hard for ChatGPT to choose the correct options that are aligned with the lessons students have already studied in class… so to deal with this, I have to review and refine the AI output. This might involve adjusting the phrasing of questions, checking for appropriateness, and ensuring that each question serves the learning outcome of the course as well as the lesson objectives."

The mismatch between questions and students' language levels

The interview also revealed that the questions suggested by ChatGPT occasionally mismatched students' linguistic competency, which was reflected in its generation of identical questions that made it difficult or impossible for lecturers to categorize students according to their levels:

"In my opinion, so the questions are…uhm… so for example, I can use the same set of ideas, but I asked ChatGPT to create different levels. And when I compared the questions

for different levels, I found that it's ok but not really helpful or has a really distinct difference between … maybe from intermediate and advanced levels, sometimes… so for the beginners, that's ok, but for higher levels, it's kind of similar."

Theme 3: Methods of enhancing ChatGPT's capabilities to design English tests and other forms of assessment activities:

From the lecturers' perspectives, a number of measures could be implemented to foster ChatGPT's testing design competence, among which three key approaches stood out.

Question re-evaluation and redesign

First and foremost, all of the lecturers shared the same viewpoint that ChatGPT should only be utilized as an idea generator for testing or a reference tool that lecturers should refer to when they embark on the design of a test, an exam, or a task for assessment. They thought that subsequent to ChatGPT drafting questions or suggesting ideas for testing activities, lecturers should assume responsibility for double-checking the precision and appropriateness of the questions as well as the provided answers so that readjustments or refinements could be made to the content of the questions that would be used later on. They placed a great emphasis on lecturers' critical sense to select, adapt, and redesign questions so as to satisfy fundamental requirements of a test's learning outcomes and students' levels of language:

"Uhm I think ChatGPT will be an effective tool for us to save time only, and sometimes it will raise some ideas for us to do the questions, and through that maybe we can base on the prompts or the hints from the ChatGPT, so we can know the next step what we will do and what we correct the question so that it goes straight to the point of the lesson we want to check and recheck the understanding of the students…"

"I think we should double-check and revise the questions. If we find a big problem in the questions, we have to read the results again and check …"

"If after I read the products from AI, uh from ChatGPT, so I think all of that's ok and no queries or no questions, so I'll use it, and if I find something not very reliable or something outdated I need to check, so I will use another AI app for double check."

Training ChatGPT through the input of prompts

Another solution mentioned by the lecturers attending the interview was training prompts, highlighting that the prompts to be fed into ChatGPT should be written in the most careful and specific manner, incorporating adequate information such as students' levels of linguistic proficiency, learning objectives, test types, and expected outcomes, etc… Some of them pointed out that the prompts should be made "recursive", which involved the repetition of different rephrased prompts acting as input so as for ChatGPT to process, analyze, and handle the information in the prompts in many ways, and that lecturers had better provide "feedback" for ChatGPT under the form of new prompts so that this AI tool could learn from the information to improve its performance and efficiency, from which it may be geared towards goal achievement of effective testing design. The preparation of prompts for ChatGPT, from one lecturer's point of view, should be a step-by-step process in which ChatGPT was trained to tackle every aspect of the prompts in the most comprehensive manner so as to generate the most

satisfactory output which could objectively and accurately evaluate students' academic performance:

> *"There are two things that need to be taken into consideration. First of all is considering the prompts. I believe that every prompt we need should be included 4 main parts. The first one is the context. I mean that we need to include the background information or the scenario that we set the stage for the questions. The second part is instructions that means the directions that we need to guide the students how to approach and solve the questions. The next part is data. I mean that we have to supply all the facts, the figures, the information, or even sometimes we add a pdf file of the coursebook for ChatGPT in order to ensure that everything needed is readily accessible without extra resources. Last but not least, that is the result, requirements, that mean what we expect to receive from ChatGPT. The second thing is recursive prompt. As usual, when ChatGPT gives me the first result, it's not satisfied with my needs. That's why we need to add more prompts, more detailed prompts … and it seems to be a process; it includes many steps in order that ChatGPT can give me a complete and a relevant result. And the last one is the refine the output. Refine the output it means that we have to edit for the language, the grammar, the structure, the content, also the flow of the test, I mean so."*

> *"I think we have to continuously give ChatGPT guidance so as to train it to the level we want. We have to give it more specific guidelines, and then some feedback. For example, when I give it requirements to design a test, if I'm not satisfied with the output for the first time, I'll give it feedback… and I see it has made some improvements, like if I want the question 2 to be corrected in some respects, I'll give it feedback… and I see ChatGPT has improved it, corrected it, and I continuously re-evaluate the question and provide it with some more feedback … I think there should be training and clear guidelines so that ChatGPT can meet the requirements we expect."*

Training lecturers on how to employ ChatGPT for tests and assessment design

Two lecturers stated that there should be training sessions providing lecturers with meticulous instructions or guidance on how to take the best advantage of ChatGPT for test design, specifying that either the Faculty of Foreign Languages should be held responsible for holding such training courses or lecturers should be engaged in a community whose aim is to share knowledge and experiences as well as offering mutual support so as to enhance their self-assurance in devising tests and assessment with the aid of ChatGPT:

> *"I think that first of all the organizations should have … like … the orientation to instruct teachers how to use them, how to use AI or ChatGPT in particular, in a more effective way because usually teachers nowadays just find it out themselves; they don't have any specific guidelines to utilize ChatGPT in designing tests."*

> *"Teachers also need guidelines. If we have a community for sharing knowledge and sharing the specific or effective prompts for using, I think that it's a good opportunity for teachers to improve themselves in using AI or ChatGPT…"*

Discussion

The study aims to examine the capabilities of ChatGPT in assisting HUFLIT lecturers of English in designing tests and other types of assessment:

Research question 1: What are the potential capabilities of ChatGPT in the design of tests and assessments from the standpoint of lecturers of English at HUFLIT?

The main findings of the research suggest that ChatGPT possesses an excellent capability for quickly generating a vast and impressive variety of questions of varying kinds matching given prompts and pertinent to lecturers' requirements, which is not only time-saving but also energy efficient and assists lecturers in developing their test design competency through serving as a source of suggestions for ideas and structured formats for devising tests and assessment activities. ChatGPT was also highly valued for its provision of very detailed instructions that effectively facilitated the implementation of in-class activities. This finding is in accordance with that of the study conducted by Rasul et al. (2023) who also investigated the potentialities of ChatGPT in designing tests and assessment in his research. The possible explanation for this may lie in the fact that ChatGPT has been specifically trained on massive language models and patterns, which helps it immediately respond to requests demanding linguistic knowledge with ease and speed. Aside from this, it was found that its originality in recommending engaging scenarios or real-life situations aroused students' interest in the tasks, therefore enormously benefiting the assessment process. This result ties well with those reported by Rudolph et al. (2023) and Kolade et al. (2024), who acknowledged the chatbot's capacity to create diverse, authentic scenarios that are potentially advantageous to students in terms of communicative and higher-order thinking skills.

Research question 2: What are the challenges HUFLIT's lecturers of English may encounter regarding ChatGPT's competence in tests and assessment design?

The results from the questionnaire and the interview indicate that the greatest challenge associated with the adoption of ChatGPT in designing English tests and assessments was its inability to draft questions that match the expectations of lecturers if the prompts failed to be sufficiently specific and appropriate. Inaccuracy, mismatch with students' language capacity, failure in generating questions that are specific, updated, less repetitive, and more stimulating and likely to evaluate students' higher-order thinking skills effectively as well as picture-related activities were other obvious shortcomings, which may hinder lecturers from designing reliable and valid tests or assessment tasks. These findings are in alignment with those of Abas et al.'s (2023), Gonsalves (2023), Kostka & Toncelli (2023), Rudolph et al.'s (2023), and Zaiarna et al.'s (2024) research. These limitations exist perhaps due to ChatGPT's inadequate understanding of the prompts given or its shortage of profound specialized knowledge and critical ability to handle lecturers' requirements. After all, ChatGPT was only an AI assistant tool equipped with language training, so it could not supersede the crucial role of human lecturers.

Research question 3: What can lecturers of English at HUFLIT do to enhance ChatGPT's capabilities in tests and assessment design?

Methods to enhance the capabilities of ChatGPT vary, but the most notable ones emerging from the research findings were the training of ChatGPT through the provision of prompts, the reassessment and readjustment of inappropriate questions generated by the chatbot, as well as the coaching of lecturers to adopt ChatGPT as an effective tool for tests design. The former approach involved the meticulous drafting of prompts, which consisted of detailed and exact information on the scenarios, instructions, students' proficiency levels, kinds of tests, examples of expected learning outcomes, and so on, to make sure that the finished product was consistent with lecturers' expectations. Apart from the careful preparation for the input, the repetition of prompts using different wording to convey similar instructions was also of significant importance in improving the efficacy of ChatGPT in generating exam questions or test items. However, the coaching of ChatGPT would not be efficient if there was no continuous feedback about its initial outputs, which means lecturers were supposed to provide comments on the performance of ChatGPT by means of raising in-depth questions to train its critical thinking ability in coping with various requests ranging from simple to complex ones. The latter approach underscored the need for revision, re-evaluation, and refinement of ChatGPT-produced questions if they were found to be unsuitable or imprecise in terms of content or language use. When it comes to devising questions to assess students' higher-order thinking skills, content double-checking should be carried out in order to guarantee the reliability of the questions produced by ChatGPT, which concurs with the findings of Anderson et al.'s (2001) study. The final approach required support from the Faculty of Foreign Languages or the community whose commitment was to offer guidelines or assistance to those wishing to accumulate knowledge and experience in the effective employment of ChatGPT for testing design purposes. This result ties well with the findings of Zaiarna et al.'s (2024) study, when an emphasis was also placed on the advantages of providing guidance or practices to instructors prior to the adoption of ChatGPT in test and assessment design.

Conclusion

In conclusion, the potential capabilities of ChatGPT in aiding lecturers of English at HUFLIT with the design of tests and other forms of assessment have been evaluated in this research paper with the employment of a mixed methods approach. The findings indicate that ChatGPT is a promising AI assistant tool for test design on account of its extraordinary versatility in generating a richer diversity of exam questions, test items, or tasks in accordance with lecturers' demands in a quicker and more detailed manner compared to those drafted by human teachers. Those who take responsibility for designing tests and assessments also substantially benefit from the adoption of the chatbot when referring to ChatGPT as a source of ideas and suggestions for testing. Nevertheless, the chatbot manifests some weaknesses regarding inexactitude, inappropriacy in the content or linguistic usage, the mismatch between the output and learners' language capacity, its failure to produce more specific, more interesting, up-to-date, and less repetitive questions to assess students of distinct levels, questions concerning the use higher-order thinking skills, as well as those dealing with picture activities. It is important, therefore, to train ChatGPT by inputting more precise and detailed prompts which also need to be recursive and paraphrased to express the same ideas at different periods of time. As for lecturers,

they had better critically reassess the precision and appropriateness of the questions generated by ChatGPT in order to be able to make proper readjustments for obtaining the best quality tests or evaluative tasks. Additionally, more training sessions provided by the Faculty of Foreign Languages that instruct lecturers to employ the chatbot for the purpose of designing tests, coupled with lecturers' active engagement in a knowledge-sharing community, should be taken into account so as to maximize the effectiveness of ChatGPT's capabilities in this respect.

Limitations

There are a number of shortcomings regarding several aspects of the research that should not be overlooked. First and foremost, the small sample size involved in the study (only 30 lecturers) could exert an impact on the study, therefore reducing the generalizability and reliability of the findings. Another limitation worth mentioning is although the research employed the quantitative and qualitative approach to reinforce the validity of the data, if tests and assessments of various types and subjects generated utilizing ChatGPT had been collected for analysis and evaluation, more specific information related to the competencies of the chatbot in designing exams and tests would have been able to be obtained, which further revealed fascinating results to benefit those concerned.

Recommendations for further studies

Future studies should include a larger number of participants so as to enhance the precision of the results, thus helping to strengthen the research's representativeness. If an examination of ChatGPT's capabilities to devise tests and assessments was carried out on a particular English subject of different domains (skills, theory, major, etc.), there would be more insightful interpretations that could contribute to the understanding of the chatbot's competency, from which practical measures might be taken to optimize its potential and minimize its downsides.

Acknowledgment

References

Abas, M. A., Arumugam, S. E., Yunus, M. M., & Rafiq, K. R. (2023). ChatGPT and Personalized Learning: Opportunities and Challenges in Higher Education. *International Journal of Academic Research in Business and Social Sciences, 13*(12), 3536-3545.

Ahmed, F., Ali, S., & Shah, R. A. (2019). Exploring Variation in Summative Assessment: Language Teachers' Knowledge of Students' Formative Assessment and Its Effect on their Summative Assessment. *Bulletin of Education and Research, 41*(2), 109-119.

Almasre, M. (2024). Development and Evaluation of a Custom GPT for the Assessment of Students' Designs in a Typography Course. *Education Sciences, 14*, 1-19. https://doi.org/10.3390/educsci14020148

Amin, M. Y. (2023). AI and Chat GPT in Language Teaching: Enhancing EFL Classroom Support and Transforming Assessment Techniques. *International Journal of Higher Education Pedagogies, 4*(4), 1-15. Retrieved from https://doi.org/10.33422/ijhep.v4i4.554

Baskara, F. R., & Mukarto, F. (2023). Exploring the Implications of ChatGPT for Language Learning in Higher Education. *Indonesian Journal of English Language Teaching and Applied Linguistics, 7*(2), 343-358.

Beikmahdavi, N. (2016). Washback in Language Testing: Review of Related Literature First. *International Journal of Modern Language Teaching and Learning, 1*(4), 130-136.

Black, P., & Jones, J. (2006). Formative assessment and the learning and teaching of MFL: sharing the language learning road map with learners. *Language Learning Journal*(34), 4-9.

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 301-318. https://doi.org/10.1080/0969594042000304609

Chandio, M. T., & Jafferi, S. (2015). Teaching English as a Language not Subject by Employing Formative Assessment. *Journal of Education and Educational Development, 2*(2), 151-171.

Fadilah, R., Ayudhia, H. Y., Chairani, V. S., & Afni, F. (2023). Assessment of English Language Teaching for 21st Century: Teachers' Perspectives on Traditional and Alternative Assessment. *JADEs Journal of Academia in English Education, 4*(1), 108-130. Retrieved from https://journal.iainlangsa.ac.id/index.php/jades

Heaven, W. D. (2023, April 7). ChatGPT is going to change education, not destroy it. *MITTechnology Review*. Retrieved from https://www.technologyreview.com/2023/04/06/1071059/-chatgpt-

Jain, N. (2021). Survey Versus Interviews: Comparing Data Collection Tools for Exploratory Research. *The Qualitative Report, 26*(2), 541-554. Retrieved from https://doi.org/10.46743/2160-3715/2021.4492

Kolade, O., Owoseni, A., & Egbetokun, A. (2024). Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon, 10*, 1-21.

Koraishi, O. (2023). Teaching English in the Age of AI: Embracing ChatGPT to Optimize EFL Materials and Assessment. *Language Education & Technology(LET Journal), 3*(1), 55-72.

Kostka, I., & Toncelli, R. (2023). Exploring Applications of ChatGPT to English Language Teaching: Opportunities, Challenges, and Recommendations. *The Electronic Journal for English as a Second Language, 27*(3), 1-19.

Kuphanga, D. (2024). Questionnaires in Research: Their Role, Advantages, and Main Aspects. *Research Gate*, 1-8. https://doi.org/10.13140/RG.2.2.15334.64325

Mansory, M. (2020). The Significance of Non-traditional and Alternative Assessment in English Language Teaching: Evidence From Literature. *International Journal of Linguistics, 12*(5), 210-225.

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

Nguyen, T. P. T. (2023). The Application of ChatGPT in Language Test Design – The What and How. Proceedings of the AsiaCALL International Conference, 4, 104-115. ISSN: 2833-6836; ISBN: 979-8-9870112-4-9. DOI: https://doi.org/10.54855/paic.2348

Pham, V. P. H., & Le, A. Q. (2024). ChatGPT in Language Learning: Perspectives from Vietnamese Students in Vietnam and the USA. International Journal of Language Instruction, 3(2), 59-72. DOI: https://doi.org/10.54855/ijli.24325

Rasul, T., Nair, S., Kalendra, D., Robin, M., Oliveira, F. d., Ladeira, W. J., . . . Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning & Teaching, 6*(1), 1-16.

Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet, 15*, 1-24.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342-363.

Trumbull, E., & Lash, A. (2013). *Understanding Formative Assessment: Insights from Learning Theory and.* San Francisco: WestEd.

Widiastuti, I. A., & Saukah, A. (2017). Formative Assessment in EFL Classroom Practices. *Bahasa dan Seni: Jurnal Bahasa, Sastra, Seni, dan Pengajarannya, 45*(1), 50-63. https://doi.org/10.17977/um015v45i12017p050

Zaiarna, I., Zhyhadlo, O., & Dunaievska, O. (2024). ChatGPT In Foreign Language Teaching And Assessment: Exploring EFL Instructors' Experience. *Information Technologies and Learning Tools, 102*(4), 176-191. https://doi.org/10.33407/itlt.v102i4.5716

Biodata

Huong Thi Thu Le is currently an experienced lecturer working for the Faculty of Foreign Languages, HUFLIT, Vietnam with more than twelve years of teaching experience. She earned a Bachelor's Degree from Ho Chi Minh University of Education and holds a Master's Degree in TESOL from Victoria University. Her interests are primarily novel methodologies to enhance students' learning and achievement, AI tools, and how these can make radical improvements in learners.