rEFLections
Vol 32, No 3, September - December 2025

Exploring and Comparing Scaffolding Strategies of ChatGPT-3.5 and a Customized GPT for Reading Comprehension

UDOMSAK SIRITA*
RICHARD WATSON TODD
School of Liberal Arts, King Mongkut's University of Technology Thonburi, Thailand
***Corresponding author email: udomsak.sirita@gmail.com**

Article information

Article history:
Received: 6 Oct 2024
Accepted: 21 Dec 2025
Available online: 24 Dec 2025

Keywords:
Scaffolding strategies
Generative AI
Customized GPT
Reading comprehension
Assessment for learning

Abstract

This study compares scaffolding strategies generated by ChatGPT-3.5 and a customized GPT in reading comprehension exercises to assist Thai university students in achieving a minimum CEFR B2 level as a requirement for Thai bachelor's degree programs. A prompt for ChatGPT-3.5 was designed to generate four reading passages, each with five multiple-choice questions. A similar approach was used to configure a customized GPT, also with a prepared file containing four reading passages and five multiple-choice questions. Data were collected based on the responses from both versions when two incorrect and one correct answer were selected respectively for each question. The results revealed that the customized GPT generated more meaningful and diverse scaffolding strategies, whereas ChatGPT-3.5 produced consistent but limited responses focused on specific reading strategies. Furthermore, the study found that some valuable strategies, such as misconception correction and the promotion of critical thinking, were absent in ChatGPT-3.5. While both versions offer educational value, they differ in the depth and range of scaffolds provided. Educators and researchers should carefully consider these differences when integrating generative AI into instructional design. In particular, this study highlights the importance of grounding AI-assisted learning in established pedagogical theories, such as scaffolding, to support core language skills like reading comprehension. As generative AI becomes more common in classrooms, thoughtful implementation and training for both instructors and students will be key to maximizing its potential within the Thai educational context.

INTRODUCTION

In Thailand, proficiency in English language skills within higher education has been identified as a critical concern, especially regarding reading performance. On November 28, 2023, Prime Minister Srettha Thavisin mandated the Ministry of Higher Education, Science, Research, and Innovation to ensure that bachelor's degree students pass a compulsory English test at a

This paper was specially selected to be published from the 5th Doing Research in Applied Linguistics (DRAL 5) International Conference that was run from 5 to 7 September 2024 in Bangkok, Thailand.

minimum CEFR B2 level. This directive responded to the PISA test results, which indicated that Thai students' scores were at a 20-year low. Furthermore, it was reported that Thailand was among seven countries having a decline in reading performance even before the pandemic (Bangprapa, 2024). This was followed by an official announcement from the Commission on Higher Education Standards (2024) on policy for raising the standard of English in higher education institutions.

Reading is a fundamental, complex, and essential English language skill tested in academic settings. It serves as a primary input for gaining new knowledge and information, and provides access to various explanations and interpretations. Furthermore, beyond the classroom, reading is also regarded as the primary means for independent or lifelong learning (Lee, 2012; Yukselir, 2014). Although reading is critical in higher education, the associated challenges have not been adequately recognized and addressed. Dreyer and Nel (2003) suggested that the problem might arise because assessment typically focuses on the outputs or results of reading rather than the reading process itself. In other words, reading tests are usually used as assessments of learning instead of assessments for learning. Karakoç et al. (2022) found that the lecturers often fail to provide any specific guidance to help students meet the high demand of academic reading courses. This lack of support can leave ESL and L1 students struggling to meet lecturers' expectations independently. Furthermore, lecturers may have limited time to assist students or may not consider reading within their area of expertise.

To enhance students' reading skills, it is crucial for lecturers to employ scaffolding strategies for reading comprehension. Scaffolding provides "specific just-in-time support that gives students the pedagogical push that enables them to work at a higher level of activity" (Gonulal & Loewen, 2018, p. 3). This concept is derived from Vygotsky's theory of the Zone of Proximal Development, which includes external supports that help students perform tasks, such as text comprehension, which they could not achieve independently (Hattan & Alexander, 2018). However, as previously mentioned, one challenge is lecturers' limited time to help students, and providing contingent support or feedback for individual students is difficult in typical classroom settings. To address this issue, there have been efforts to utilize technology as a superior system to enhance reading skills and promote learner autonomy. With the emergence of generative AI, particularly ChatGPT, applying chatbot-like tools as capable peers into the classroom is expected to facilitate teaching and learning English as a foreign language (EFL). As these tools can generate text, they are potentially applicable for both reading and writing. However, this study focuses on applying ChatGPT to reading comprehension and using reading exercises in the format of tests as assessments for learning.

LITERATURE REVIEW

Scaffolding in educational settings

The term "scaffolding" as a metaphor in educational contexts was introduced by Woods, Bruner, and Ross (Wilson & Devereux, 2014). This concept draws on Vygotsky's sociocultural theory, particularly his notion of the Zone of Proximal Development (ZPD). Vygotsky (1987,

as cited in Wilson & Devereux, 2014) described the ZPD as "the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (p. 92). In educational terms, scaffolding refers to "the interventions that tutors or teachers make within the students' ZPD to facilitate their learning and improve their current knowledge and skills" (Gonulal & Loewen, 2018, p. 12). Basically, teachers provide temporary support to assist students in performing tasks they cannot do independently and gradually remove these supports as students gain the confidence to perform them on their own.

One of the significant and fundamental scaffolding strategies applied in classroom contexts with English language learners is instructional scaffolding. Techniques of instructional scaffolding include modelling, bridging, contextualizing, schema building, re-presenting text, and developing metacognition. Gonulal and Loewen (2018) note that in modelling, teachers provide representative examples as concrete guidelines for students. Bridging techniques help create connections by activating students' prior knowledge. Contextualizing involves aiding students' learning with various supports, such as pictures or videos. Schema building relates to connecting new information with learners' existing knowledge. Re-presenting text is a technique where linguistic structures are transformed from one genre to another, such as changing a poem to a narrative. Finally, developing metacognition involves strategies like think-aloud and self-assessment activities.

Integrating scaffolding and reading strategies in reading comprehension

In the context of reading comprehension, scaffolding and reading strategies are closely interconnected. Teachers utilize scaffolding strategies to support students in applying and developing effective reading strategies. Significant research has been conducted in this area, from Palinscar and Brown's reciprocal teaching (Hattan & Alexander, 2018), where teachers guide students to summarize, question, clarify, and predict during reading, to Oxford's strategies (Yussof et al., 2012) include note taking, summarizing, inferencing, using prior knowledge, predicting, analyzing, and using context clues. Sheorey and Mokhtari (2001) categorized reading strategies into three groups: metacognitive strategies, where learners monitor and manage their reading; cognitive strategies, such as adjusting reading speed according to the difficulty of the materials or guessing the meanings of unknown words; and support strategies, like note-taking or highlighting the text for better understanding. Yukselir (2014) investigated the use of prominent cognitive strategies in EFL prep-class students, including grouping, deduction, imagery, contextualization, elaboration, and inferencing. These reading strategies allow teachers to scaffold student learning through strategies such as modelling and thinking aloud. However, in Thai educational context, the strategies used to enhance students' reading skills are primarily delivered through lecturing and testing.

Reading tests as assessment for learning

Reading tests or reading exercises in the format of tests are commonly used to teach reading in classrooms, with multiple choice items dominating these tests. Students in Thailand are particularly familiar with multiple-choice testing because they are assessed at the national

level through the Ordinary National Education Test (ONET) every three years, where only multiple-choice items are provided in the exams for English (Watson Todd, 2019). In classroom practices at Thai schools, Watson Todd (2019) found that "test-style assessments (exams and quizzes) and closed-ended assessments dominate how the English language ability of students is assessed" (p. 11). His study, based on surveys, also highlighted that teachers prefer using test-style assessments for administrative purposes rather than assessment for learning. At university level, standardized tests such as TOEIC, IELTS, and TOEFL, as well as in-house tests, are typically used to assess university students' CEFR levels. Similar to ONET, multiple-choice items are prevalent among these tests, particularly in the reading sections, for practical administrative reasons.

Although using multiple-choice tests traditionally as summative assessments can be disadvantageous to learning, this study aims to utilize reading tests as assessment for learning. Black et. al (2009, as cited in Wiliam, 2011) defined assessment for learning is "any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning," and stated that "an assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged" (p. 10). They also emphasized that learning processes are regulated during 'moments of contingency' instruction.

However, in typical classroom settings, teachers often cannot closely monitor or scaffold individual students to effectively use reading strategies. When reading tests are used as formative assessments, they typically focus on providing correct answers for students to check their own work, which often results in a lack of meaningful feedback. This lack of feedback hinders learning, as students miss opportunities to correct mistakes and learn the appropriate strategies for finding the correct answers. Ideally, teachers should provide contingent feedback and scaffold individual students in using the right strategies to arrive at correct answers. However, in reality, especially in the Thai educational context, large class sizes and limited teaching time make it nearly impossible for teachers to offer individualized support. As a result, generative AI is an ideal solution, acting as a more capable peer or a superior system that can scaffold individual learners and personalize contingent feedback and support to their actual development level.

Integrating contingent micro-scaffolding with technology in reading tests

Technology, including computer-assisted testing and generative AI, is well-suited for providing both macro and micro scaffolding in reading tests. Macro scaffolding refers to a teaching and learning unit designed around the curriculum cycle, focusing on broader issues such as program goals, and the selection and sequencing of tasks (Syarifah & Gunawan, 2015). It involves pedagogical methods or strategies that guide the overall flow of larger-scale activities, offering broad guidance on how to approach different learning tasks or problems (Melero et al., 2012). Within this broader framework, micro scaffolding occurs as the teacher provides contingent support during ongoing interactions (Hammond, 2001, as cited in Syarifah & Gunawan, 2015). This involves offering specific scaffolds that address detailed actions. The key aspect of micro

scaffolding is its responsiveness to the immediate needs of learners in order to facilitate the resolution of particular activities in alignment with the macro-level objectives (Melero et al., 2012). In computer-assisted testing, macro scaffolding guides the overall flow and difficulty of tasks, while micro scaffolding offers specific clues and feedback to help solve detailed problems. Wang et al. (2023) stated that computer-based scaffoldings can offer dynamic and adaptive supports during self-regulated learning (SRL) which is "a proactive and constructive learning process whereby learners dynamically monitor and adjust the motivational, affective, cognitive, metacognitive, and behavioural aspects of learning to achieve predetermined goals" (p. 1653). Integrating both levels ensures a clear connection between tasks and learning goals, making tests effective tools as assessment for learning.

Building on this integration, several studies have focused on contingent micro-scaffolding with the integration of technology in developing English skills. However, the number of studies applying computer-based scaffolding for reading skills is limited. One example is Watson Todd (2014), which explored the use of Supportive Adaptive Testing (SAT). This approach to adaptive computer-based multiple-choice testing aims to improve reading skills by the design and development of Support Adaptive Tests (SATs) was driven by an effort to enhance the efficiency of multiple-choice computer-based tests (CBTs) by integrating scaffolding mechanisms and fostering learner autonomy. Initially, SATs resemble typical multiple-choice tests, with reading passages on one side of the screen and questions on the other. However, the innovation becomes evident in the post-test stage, where incorrect answers are revisited with added scaffolding support. Three types of scaffolds were embedded: text highlighting, which emphasizes relevant sections of the text to draw the test-taker's focus; text translation, which provides first-language translations of difficult words in the text; and question translation, which offers first-language translations of challenging phrases within the questions. Test-takers are then given the autonomy to choose their preferred scaffold before reattempting the questions. After completing this second attempt, they receive detailed feedback, including explanations for any remaining incorrect answers. This design not only promotes assessment for learning through mistakes but also allows test-takers to personalize the support according to their specific needs.

ChatGPT and language learning

With the emergence of ChatGPT, researchers have begun to integrate this large language model (LLM) into pedagogical practices. While the nature of ChatGPT primarily leads to its use in enhancing writing skills, its capabilities extend far beyond. ChatGPT, or Chat Generative Pre-trained Transformer, is a chatbot capable of providing human-like responses to prompts or questions. This artificial intelligence (AI) is developed from a large language model primarily designed for conversational interactions, but its capabilities have been extended to include various other uses (Barrot, 2023; Essel et al., 2024). ChatGPT was developed by the Microsoft-backed company OpenAI, with its first version launched in November 2022 (Kohnke et al., 2023). Subsequent versions, including GPT-3.5 Plus and GPT-4 for Plus subscribers, were released in February and March 2023, respectively. These versions offered enhanced capabilities in advanced reasoning, complex instruction processing, and creativity. Throughout 2023, several features were updated, including enabling third-party plugins, sharing links, and

accessing voice and image capabilities. In November 2023, OpenAI introduced custom versions of ChatGPT or GPTs, allowing users to create and share a "tailor version of ChatGPT" for specific purposes (OpenAI, 2023). As a result, users do not need to copy and paste the prompts they create every time they start using them for the same task. Moreover, more knowledge and files can be added to GPTs.

In teaching and learning contexts, researchers have identified significant opportunities to utilize ChatGPT. It can facilitate language learning by offering authentic interactions, such as providing the meaning of words in context, correcting and clarifying language errors, generating texts across different genres, creating quizzes, functioning as a dictionary, and translating texts (Kohnke et al., 2023). It offers immediate, targeted feedback, explanations, and examples, helping learners monitor their progress and accelerate their language development. ChatGPT is particularly beneficial for students with limited access to traditional language learning resources, such as in-person classes or native speakers. These advantages can overcome geographical boundaries (Barrot, 2023). Furthermore, the chatbot supports "long-term personalized learning and self-regulating process" (Ghafouri, 2024, p. 5), potentially sustaining interest in language learning and reducing stress related to teacher feedback or peer pressure. In Thai educational context, there is an increasing adoption of ChatGPT for English language learning across a range of educational levels and learner groups. For example, Tantivejakul et al. (2024) conducted a study with university students majoring in Public Relations to explore their views on using generative AI in academic and professional writing. Additionally, Jitpaisarnwattana (2024) proposed ChatGPT as a support tool for diagnostic writing assessment among English-major university students. In this approach, students were asked to write and then use provided prompts to request feedback from ChatGPT on their writing. Beyond university students, elderly learners aged 60 to 70 were also studied to examine how they used ChatGPT prompts for learning English, specifically to generate scripts and advertisements for their community products (Sanmuang et al., 2024).

Focusing on contingent micro-scaffolding, several studies have integrated artificial intelligence technology to support the development of English skills, particularly writing. For example, Lim et al. (2024) investigated how university students learn with real-time personalized scaffolds driven by a rule-based AI system during a 45-minute learning task. This task involved reading about artificial intelligence, differentiation in schools, and scaffolding in education, followed by writing an essay on the future of education. Similarly, Dhillon et al. (2024) conducted a quasi-experiment with 131 participants who responded to argumentative writing prompts under three randomly sequenced conditions: no AI assistance (control), next-sentence suggestions (low scaffolding), and next-paragraph suggestions (high scaffolding). Another study by Li et al. (2025) evaluated the effectiveness of adaptive scaffolds generated by generative AI based on real-time analytics of learners' self-regulated learning (SRL) processes. This study also examined how graduate students enrolled in an academic English writing course responded to the prompted scaffolding in terms of compliance and overall self-regulatory patterns. In terms of speaking skills, Wang et al. (2025) examined the effects of AI-based interactive scaffolding (AIIS) on secondary students' speaking performance, goal setting, self-evaluation, and motivation in informal digital learning of English (IDLE). This was implemented through a mobile app featuring an AI conversational agent. All of the studies

reported positive outcomes, suggesting that interactive, personalized, adaptive generated scaffolds can positively influence learners' learning behavior, self-regulated learning, and language performance. However, it is important to note that the instruments used in these studies were based on custom-designed interfaces integrated with AI and system-level data collection, which required the expertise of computational specialists. This may limit practicality and transferability for educators or practitioners without specialized knowledge in computer science or AI.

Based on prior studies, the use of ChatGPT as a tool for dialogue-based learning to scaffold students' reading comprehension is found to be rare, particularly in the context of using reading tests for learning. Recent studies on generative AI and English reading have primarily focused on reading materials and students' interaction. For instance, researchers have examined the quality of multiple-choice reading items (Lam et al., 2024; Lin & Chen, 2024), the interpretation of ChatGPT-generated texts to support holistic reading (Cheung et al., 2024), and students' perceptions and strategies when using GenAI for reading (Pan et al., 2025). In response to this gap, the purpose of this study is to compare scaffolding strategies generated by ChatGPT-3.5 and a customized GPT in reading comprehension exercises. This comparison aims to assist Thai university students in achieving a minimum CEFR B2 level as a requirement for Thai bachelor's degree programs. By comparing two versions of ChatGPT as primary research, this study expects to identify similarities and differences between the free-access and paid versions to provide practical implications in Thai educational contexts.

Research questions

1) How do ChatGPT-3.5 and the customized GPT scaffold learners during reading comprehension exercises?
2) What are differences in scaffolding strategies between ChatGPT-3.5 and the customized GPT?

METHODOLOGY

Instrument design

This study aims to explore and compare how different versions of ChatGPT help learners find correct answers in multiple-choice reading comprehension exercises. To compare the capacities of ChatGPT-3.5 (a free version) and a custom GPT (a paid version) in scaffolding learners, responses generated by ChatGPT serve as the primary data of this study. Although the quality of reading passages and multiple-choice questions can influence learners' decisions, it is beyond the scope of this study. The authors attempted to prompt ChatGPT-3.5 and configure the custom GPT under the same conditions similar to Watson Todd's SAT (2014):

1) It should be convenient for learners to start and learn independently with minimal instructions.
2) Learners should be able to interact with ChatGPT with little to no external assistance.

3) ChatGPT should provide four choices of reading passages.
4) Each passage should have five multiple-choice questions, each with only one correct answer.
5) ChatGPT should present one question at a time and avoid revealing correct answers until learners select them on their own, encouraging independence and critical thinking.
6) When learners select incorrect answers, ChatGPT should use techniques such as screening, skimming, highlighting, inferring, and translating to guide them toward discovering the correct answers independently.
7) ChatGPT should generate a feedback summary for each passage and for all passages combined.

Given these conditions, the operationalization of both versions of ChatGPT differed due to their distinct features and capabilities. Consequently, the authors focused on designing a prompt that would generate results aligned with the expected conditions. This prompt was revised and tested until a final version was established for data collection (Appendix A). To assess whether the prompt used with the free version effectively generated B2-level reading passages, each passage was analyzed using the GSE Text Analyzer (see https://www.english.com/gse/teacher-toolkit/user/textanalyzer) to determine its CEFR level. The analysis showed that all passages generated by the free version were at the B2+ level. According to CEFR reading comprehension descriptors (Council of Europe, 2020), these passages correspond to the categories of reading for orientation and reading for information and argument. These categories include skills such as scanning texts quickly, identifying relevant content, extracting key information, and understanding articles and reports.

To maximize the benefits of the customized GPT under the specified conditions, the configuration process was carried out in two stages. First, the authors adapted the instructions from the prompt used in ChatGPT-3.5 to achieve comparable results (Appendix A). The second stage involved designing a reading comprehension exercise with four reading passages, each containing five multiple-choice questions. At this stage, the authors leveraged a key advantage of the customized GPT, the ability to process uploaded files, which enables ChatGPT to use only specific data provided by the user. This feature addresses a limitation of the free version, in which the generated content cannot be controlled. To ensure the reading passages aligned with the B2 level, a list of key vocabulary items at the B2 level was selected from the Oxford 3,000 and 5,000 wordlist (see https://www.oxfordlearnersdictionaries.com/external/pdf/wordlists/oxford-3000-5000/The_Oxford_3000_by_CEFR_level.pdf). In addition, only low-frequency words identified using the Corpus of Contemporary American English (COCA) were selected and complied as a PDF file and uploaded alongside the prompt (Appendix B). This configuration was used to generate reading passages and questions based on given topics. From the generated output, four passages and five questions per passage were selected and fine-tuned. The CEFR level of each passage was evaluated using the same method applied to the free version, with all passages found to be at the B2+ level. A final PDF file, including four reading passages, multiple-choice questions, and predetermined correct answers, was then uploaded to the customized GPT for use in the study (Appendices C & D).

Data collection

To collect responses generated by ChatGPT, the authors assumed on the role of a learner and followed these procedures to initiate responses from ChatGPT. Both versions are processed in the same way:

1) Request reading passages.
2) Select one of the provided reading passages.
3) Request a question.
4) Select incorrect answers twice.
5) Select the correct answer.
6) Repeat steps 2-5 until five questions of each reading passage are completed.
7) Request feedback.
8) Request another passage.
9) Repeat steps 2-7 until all four passages are completed.
10) Request a feedback summary.

Following these procedures, a learner answers 20 questions from 4 reading passages (5 questions per passage). For each question, the responses include two incorrect answers followed by the correct answer. Responses generated by ChatGPT-3.5 and a customized GPT (OpenAI, 2024) were collected in MS Word files. In addition to reading passages and questions, both versions generated a variety of responses, including transitional guidelines between questions and passages, scaffolding strategies when incorrect answers were selected, positive feedback for correct answers, and feedback summaries after completing each set of 5 questions, and an overall summary after completing all 4 passages. However, this study focused on generated responses when learners selected answers, excluding the reading passages and questions from the data. Only the responses following answer selections were used for data analysis. Two refined PDF files were analyzed using the Atlas.ti program for content analysis.

Units of analysis

The data were treated as texts and analyzed by coding. To ensure precise coding, the unit of analysis was defined as sentences. Based on syntactic structures, the types of sentences analyzed included:

1) Declarative sentences: e.g., “Pad Thai is mentioned as a staple that captivates both locals and visitors.”
2) Interrogative sentences: e.g., “Given this hint, which industries do you think are highlighted as key sectors alongside tourism in Thailand?”
3) Imperative sentences: e.g., “Let’s narrow it down a bit more.”
4) Exclamatory sentences/phrases: e.g., “Well done!” “Yes, that’s right!” “Exactly right!”

Categories and codes

Codes were created based on instructional scaffolding strategies and reading strategies as deductive coding. However, some codes emerged during the data analysis process. To check

reliability of the codes, a 1,200-word document combining responses from both versions was coded by the authors' colleague. Comparing the colleague's codes against those assigned by the authors for inter-rater reliability resulted in a Cohen's Kappa value of 0.633, indicating a substantial level of agreement between two raters. Table 1 presents the codes, descriptions, and examples. The table lists instructional and reading strategies categorized as scaffolding techniques used during reading instruction. These strategies fall into three main categories: 1) instructional support, which provides direct scaffolding to learners as they engage with texts; 2) cognitive development, which encourages learners to engage more deeply with the reading material by developing thinking skills; and 3) learner engagement, which keeps learners actively involved and promotes autonomy in their learning.

Table 1

Categories and codes of strategies generated by ChatGPT

Category	Code	Description	Example
Instructional Support	Clarifying	Clarifying complex topics or details, aiming to simplify understanding and address potential ambiguities by providing clear explanations and context	This emphasis isn't solely on the visual presentation but more on how the flavors work together.
	Guiding Transitions	Guiding learners through a sequence of questions or tasks in an exercise in order to mark transitions, encourage reconsideration, or focus attention on new or specific aspects of the content	Let's start with the first question:
	Constructive Feedback	Providing constructive feedback and/or guide learners' attention back to the specific aspect in the reading passage	That's a good try, but let's look a bit closer at the passage.
	Highlighting	Emphasizing important parts of the text or re-presenting parts of the text	The phrase "making it uniquely Thai" is referring to something specific that the previous sentences were talking about.
	Misconception Correction	Identifying and correcting inaccurate understandings	So, this option isn't the correct answer for something that is NOT a characteristic described in the passage.
	Paraphrasing	Using different words or structures to clarify or simplify the original message	Remember, the passage talks about how the combination of various cultural influences and local ingredients contribute to the overall culinary tradition, making it distinctively Thai.
	Encouragement	Allowing for alteration of answers	Give it another try!
	Schema Building	Helping learners connect new information to already existing knowledge	Keeping this context in mind, try to think about what "it" is meant to encapsulate.
Cognitive Development	Critical Thinking Promotion	Encouraging the application of critical thinking skills in learning	Let's take another look and think about how rice complements the overall taste experience of Thai meals.
	Inferencing	Deducing meaning from context within the text	Remember, a staple dish is one that is a fundamental or primary part of a cuisine.
	Metacognition Developing	Encouraging learners to assess their understanding, monitor their learning progress, and adjust their reading strategies accordingly	There was a moment of challenge in identifying the correct referent in a sentence, but your perseverance and willingness to consider different options helped you succeed.

Category	Code	Description	Example
Learner Engagement	Interactive Questioning	Engaging learners through interactive questioning strategies	What's your answer?
	Learner Autonomy Promotion	Encouraging learners to take control of their own learning process	Would you like to try another passage, or is there anything else you'd like to work on?
Feedback Mechanisms	Positive Feedback Expression	Using positive reinforcement to acknowledge responses	Great choice!
	Summarizing	Summarizing content and making it easier for learners to review and recall essential details and enhancing overall comprehension	The passage indicates that Thailand's societal attitude towards LGBTQ+ people is more accepting in social settings than in legal frameworks.

It should be noted that each unit can be assigned more than one code, as some units analysis are compound or complex sentences, which may correspond to two or more codes. This recognizes the complexity of educational interactions and allows for a more in-depth data analysis, which reflects how various strategies can overlap in teaching and learning. Examples of how to code texts generated by both versions are illustrated in Figure 1 and Figure 2.

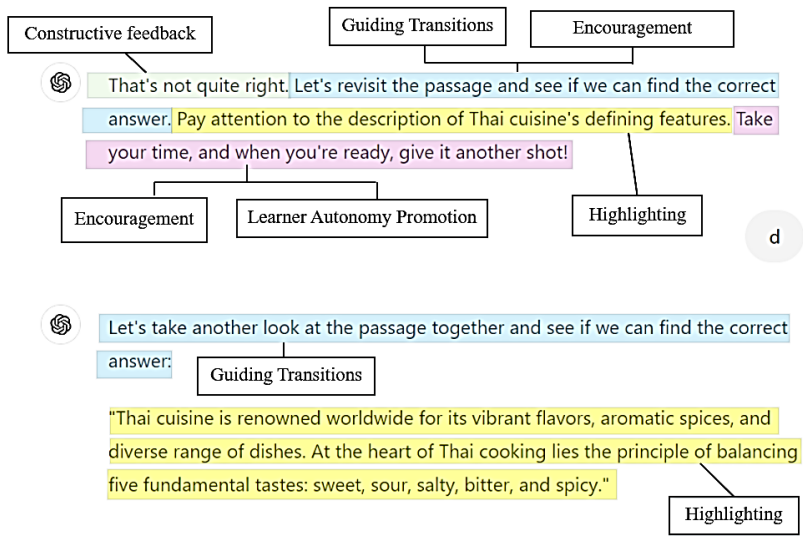


Figure 1 Examples how to assign codes to texts generated from ChatGPT-3.5

Note. The highlighted sentences illustrate how units of analysis are identified and codes assigned (OpenAI, 2024).

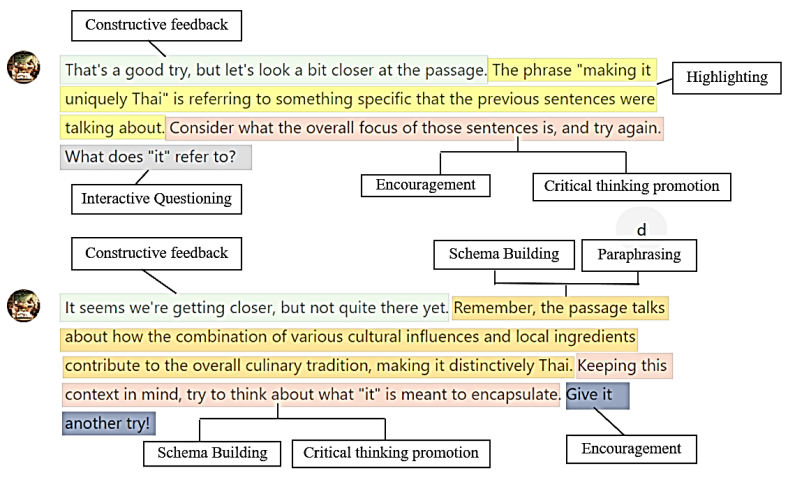


Figure 2 Examples how to assign codes to texts generated from a custom GPT

Note. The highlighted sentences illustrate how units of analysis are identified and codes assigned (OpenAI, 2024).

A thorough examination of the generated responses revealed that certain words or formulaic phrases were repeatedly generated. To confirm this observation, corpus analysis was conducted using AntConc (Anthony, 2022), a corpus analysis tool, to examine similarities and differences in language use across both versions. This cluster and keyword analysis focused on the quality of responses, particularly how they may rely on specific scaffolding strategies. Overreliance on such strategies could reduce the effectiveness of the responses for learners. Consequently, the coded data were also analyzed for usefulness that can help learners to comprehend reading passages. To further assess response quality, the coded data were analyzed for their usefulness in scaffolding reading comprehension. Drawing on Bloom's taxonomy (Anderson & Krathwohl, 2001), cognitive tasks were classified into six hierarchical levels: remember, understand, apply, analyze, evaluate, and create. The remember level includes actions such as repeating, defining, and stating, while the understand level involves classifying, discussing, explaining, identifying, and selecting. For example, responses that relied heavily on repetition or basic procedural guidance were considered less useful in supporting comprehension or promoting higher-order thinking skills. Accordingly, the usefulness of each response was evaluated based on its potential to enhance learners' cognitive processes, as outlined in Bloom's taxonomy. This criterion was crucial in assessing the quality of the coded texts. The frequency of useful coded texts was also analyzed to determine patterns across the dataset.

RESULTS

The quantitative analysis of this study is primarily based on the frequency of codes manually coded in the software Atlas.ti. In the qualitative analysis, AntConc (Anthony, 2022), a corpus analysis tool, was also utilized to support the quality analysis of the codes.

1) Coding Analysis

According to responses generated by ChatGPT-3.5 and the customized GPT, Table 2 shows that the customized GPT generated more text than ChatGPT-3.5. When considering only responses generated after selecting answers (excluding reading passages and other introductions), the customized version also produced more responses. Based on the word count, this indicates that a customized GPT might provide more detailed scaffolding.

Table 2

Total amount of generated texts and only responses by both versions

Version	Word Count		
	Total Generated Texts	Only Responses	% (Only Responses/ Total Generated Texts)
ChatGPT-3.5	5,723	3,243	56.67
Customized GPT	6,786	4,701	69.27

It is crucial to note that the prompts used in both versions include predetermined instructional scaffolding techniques based on Watson Todd's SAT (2014), such as screening, skimming, highlighting, inferring, and translating. These strategies were intended to guide learners toward independently discovering the correct answers. Additionally, both prompts emphasized fostering learner independence and critical thinking (Appendix A). Beyond these predetermined techniques, several additional scaffolding strategies emerged from the data during analysis. These emergent ones were identified as inductive codes, which contributes to an understanding of how each version of ChatGPT supported reading comprehension.

As for scaffolding strategies found in responses to selected answers, the customized GPT shows a more diversified use of scaffolding strategies. It has higher frequency and percentages in most strategies like constructive feedback, misconception correction, inferencing, critical thinking promotion, and clarifying than ChatGPT-3.5. However, ChatGPT-3.5 has higher frequency in schema building, guiding transitions, positive feedback expression, and highlighting. To check whether the proportion of codes between two versions is significantly different, a chi-square analysis was conducted. The results were extremely statistically significant ($\chi 2 = 159.30$; $df = 14$, $p < 0.001$) with a relatively strong effect size (Cramér's $V = 0.455$), indicating a relatively strong association between the frequencies of code usage for ChatGPT-3.5 and the customized GPT, and not merely due to random variation.

Table 3

Comparison of scaffolding strategies used in ChatGPT-3.5 and customized GPT

	ChatGPT-3.5		Customized GPT	
Code	Frequency	%	Frequency	%
Interactive Questioning	41	11.61	54	12.98
Positive Feedback Expression	49	13.88	44	10.58
Schema Building	79	22.38	42	10.10
Guiding Transitions	63	17.85	37	8.89
Summarizing	24	6.80	36	8.65
Metacognition Developing	25	7.08	33	7.93

	ChatGPT-3.5		Customized GPT	
Code	Frequency	%	Frequency	%
Constructive Feedback	3	0.85	32	7.69
Misconception Correction	0	0.00	32	7.69
Inferencing	0	0.00	21	5.05
Critical Thinking Promotion	0	0.00	20	4.81
Learner Autonomy Promotion	20	5.67	17	4.09
Clarifying	0	0.00	14	3.37
Encouragement	8	2.27	13	3.13
Highlighting	41	11.61	11	2.64
Paraphrasing	0	0.00	10	2.40
Total	353		416	

2) Corpus Analysis

However, considering only the quantity of generated text may not adequately reflect the quality of each response. During coding, the authors found that ChatGPT-3.5 generated more repetitions with the same phrases or expressions compared to the customized GPT. To confirm this observation, a corpus analysis was conducted. Table 4 shows that ChatGPT-3.5 frequently generated the same phrases, for example, “find the correct answer” (22 times), “based on this excerpt” (19 times), and “given this information what” (14 times). The repetition suggests that the same directive is repetitively generated. Such pattern raises concerns about whether ChatGPT-3.5 fosters personalized learning, as it often provides specific scaffolding strategies which may not align with the diverse learning strategies or styles of individual learners.

Table 4

Repetition of common four-word clusters occurring more than ten times in ChatGPT-3.5

Type	Rank	Freq
find the correct answer	1	22
to find the correct	2	20
based on this excerpt	3	19
passage to find the	4	16
the passage to find	4	16
given this information what	6	14
on this excerpt what	7	13
let s review the	8	12
next question let s	8	12
s review the passage	8	12
the next question let	8	12
to the next question	8	12
done now let s	13	11
let s move on	13	11
question let s review	13	11
s move on to	13	11
well done now let	13	11

In contrast to ChatGPT-3.5, the customized GPT uses more interactive and engaging phrases. For instance, the question “What do you think” is frequently generated, as shown by its frequency in Table 5. In essence, the customized GPT involves the learner more in the process

by asking for their opinion and thoughts, which can enhance critical thinking and retention. Although the dataset for the customized GPT is larger, there is less reliance on standard formulaic phrases.

Table 5

Repetition of four-word clusters occurring more than ten times in a customized GPT

Type	Rank	Freq
do you think is	1	16
is the correct answer	2	15
think is the correct	2	15
what do you think	2	15
you think is the	2	15
question what do you	6	11
the next question what	6	11

Furthermore, the frequency of word usage can be examined through keyword analysis, as shown in Tables 6 and 7. In this analysis, the authors employed log-likelihood (LL) as a keyness statistic to generate lists of key linguistic feathers of the responses generated by both ChatGPT versions, treated as two separate corpora. Drawing on Pojanapunya and Watson Todd (2018), who note that the choice of statistic depends on the research purpose, LL is a probability statistic suitable for differentiating the characteristics of language generated by the two versions. Linguistic features with high LL values are particularly useful for identifying the focused scaffolding strategies of each ChatGPT version. Given the small corpus size for each version, the top 20 keywords were considered. In addition, when examining words beyond the top 20, they are often content words mentioned in reading passages.

The keyword analysis reveals differences in word choice between ChatGPT-3.5 and the customized GPT. ChatGPT-3.5 frequently uses words such as “excerpt,” “let,” “find,” “information,” “based,” and “reviewed,” with frequency ranging from 13 and 71 occurrences. These keywords suggest an emphasis on directives and content presentation, supporting the code analysis that ChtGPT-3.5 heavily relies on specific scaffolding strategies. On the other hand, the customized GPT prioritizes words like “it,” “think,” “but,” “option,” “try,” “approach,” “specifically,” “consider,” “context,” and “hint,” which indicates a focus on user engagement, critical thinking, and misconception correction. This highlights the customized GPT’s role in fostering deeper interaction and cognitive involvement, while ChatGPT-3.5 leans more toward instructional language and content highlighting.

Table 6

Keyword analysis- ChatGPT-3.5 as a target corpus vs customized GPT as a benchmark corpus

Type	Rank	Freq_Tar	Freq_Ref	Keyness (LL)	Keyness (Effect)
excerpt	1	19	0	34.383	0.011
let	2	71	34	31.761	0.042
find	3	22	2	28.124	0.013
information	4	21	2	26.489	0.013
based	5	23	6	18.275	0.014
review	6	13	1	17.342	0.008
cities	7	9	0	16.27	0.005

Type	Rank	Freq_Tar	Freq_Ref	Keyness (LL)	Keyness (Effect)
recent	7	9	0	16.27	0.005
heritage	9	8	0	14.461	0.005
increasingly	9	8	0	14.461	0.005
suggest	9	8	0	14.461	0.005
around	12	13	2	13.804	0.008
has	13	16	4	13.073	0.01
face	14	7	0	12.652	0.004
millions	14	7	0	12.652	0.004
popularity	14	7	0	12.652	0.004
reevaluate	14	7	0	12.652	0.004
year	14	7	0	12.652	0.004
years	14	7	0	12.652	0.004
correct	20	48	32	12.403	0.028

Table 7

Keyword analysis- customized GPT as a target corpus vs ChatGPT-3.5 as a benchmark corpus

Type	Rank	Freq_Tar	Freq_Ref	Keyness (LL)	Keyness (Effect)
it	1	65	7	34.533	0.026
think	2	33	0	34.393	0.014
a	3	122	33	26.367	0.049
but	4	24	0	24.995	0.01
option	5	17	0	17.695	0.007
try	5	17	0	17.695	0.007
an	7	15	0	15.61	0.006
approach	8	14	0	14.569	0.006
bit	9	13	0	13.527	0.005
specifically	9	13	0	13.527	0.005
there	11	17	1	11.773	0.007
do	12	27	4	11.485	0.011
consider	13	11	0	11.444	0.005
here	14	20	2	11.022	0.008
reading	15	16	1	10.849	0.007
context	16	10	0	10.403	0.004
hint	16	10	0	10.403	0.004
through	16	10	0	10.403	0.004
mentions	19	15	1	9.932	0.006
about	20	30	6	9.618	0.012

Notably, data from both versions of ChatGPT contain repetitions, which could affect the quality of scaffolding if the same response is provided when incorrect answers are repeatedly selected. These repeated responses may lose their effectiveness and usefulness when a learner continues to select incorrect answers. While the usefulness of each response may vary depending on individual learners, it is challenging to determine its usefulness for personalized learning, as the authors assumed the role of learners. This study, however, focuses on scaffolding strategies that support learners in enhancing their cognitive processes, specifically aiming to help them reach the *understand* level or higher in Bloom's taxonomy. Consequently, repeated use of the same scaffolding strategies, even when the incorrect answer is selected a second time, is considered unhelpful, as it fails to support deeper comprehension or guide learners toward correct answers. While certain responses such as guiding transition (e.g., "Now, let's move on to the comprehension questions"), positive feedback expressions (e.g., Great!), and summarizing

scores may benefit some learners, they cannot be classified as cognitive processes within Bloom's taxonomy. Therefore, these types of responses were excluded from the usefulness criteria. In addition to identifying repetitive patterns, the usefulness of a response was evaluated based on whether it facilitated cognitive actions associated with the *understand* level of Bloom's taxonomy, such as classifying, describing, discussing, explaining, identifying, locating, recognizing, reporting, and selecting.

When comparing the overall quality of both models, 65.18 % of responses generated by the customized GPT are useful for users, compared to less than half of the responses generated by ChatGPT-3.5 (42.86%). To determine if there is a significant difference in the proportion of usefulness codes between two versions, a chi-square analysis was conducted. However, for the chi-square analysis to be valid, all expected values must be greater than 5. Consequently, the code "revision" was excluded from the analysis. The results showed a highly significant difference ($\chi 2 = 110.07$; df = 11, $p < 0.001$) with a strong effect size (Cramer's V = 0.574), indicating a strong link between the frequencies of code usage for ChatGPT-3.5 and the customized GPT in terms of usefulness. Table 8 shows that the customized GPT has a higher percentage of useful responses in most categories, indicating that it provides more effective and informative scaffolding. Significant differences are observed in the categories of clarifying, highlighting, inferencing, and paraphrasing, where all responses from the customized GPT are both useful and informative (100%).

Table 8

Comparison of usefulness between strategies generated by ChatGPT-3.5 and customized GPT

	ChatGPT-3.5			Customized GPT		
Code	**Frequency**	**Usefulness**	**%**	**Frequency**	**Usefulness**	**%**
Clarifying	0	0		14	14	100.00
Highlighting	41	21	51.20	11	11	100.00
Inferencing	0	0		21	21	100.00
Paraphrasing	0	0		10	10	100.00
Misconception Correction	0	0		32	30	93.75
Metacognition Developing	25	17	68.00	33	29	87.88
Learner Autonomy Promotion	20	18	90.00	17	14	82.35
Constructive Feedback	3	2	66.67	32	22	68.75
Interactive Questioning	41	22	53.66	54	30	55.56
Critical Thinking Promotion	0	0		20	9	45.00
Encouragement	8	0	0.00	13	5	38.46
Schema Building	79	39	49.37	42	14	33.33
Total	280	120	42.80	336	219	65.18

DISCUSSION

This study aimed to explore how ChatGPT-3.5 and a customized GPT provide scaffolding strategies to help learners achieve reading comprehension. The data analysis clearly reveals that the customized GPT outperforms ChatGPT-3.5 by generating a greater variety of scaffolding strategies to assist learners in their reading comprehension. The strategies generated by both versions of ChatGPT can be categorized into instructional support, cognitive development,

learner engagement, and feedback mechanisms. Given the nature of chatbot-like tool, it is unsurprising that interactive questioning is frequently generated by both versions to engage learners with the provided task. Similar to teachers in the traditional classroom, interactive questioning can sustain interest of students and activate their schema or critical thinking. However, ChatGPT-3.5 frequently generated only specific strategies, including schema building, guiding transition, positive feedback expression, interactive questioning, and highlighting. Important strategies such as critical thinking promotion, inferencing, misconception correction, and paraphrasing were not found in ChatGPT-3.5's responses. This implies that learners might miss opportunities to practice some essential and useful strategies if they can only have access to the free version.

In terms of quality, ChatGPT-3.5 frequently generates the same language patterns and phrases in response to incorrect answers. Highlighting or presenting excerpts, as the second most frequently generated strategy, can be useful for learners the first time, but they will be meaningless and unhelpful when introduced a second time if learners still select incorrect answers. In other words, the repetitions found in ChatGPT-3.5 may reduce the effectiveness of its scaffolding. On the other hand, the customized GPT provides several scaffolding strategies. There is a tendency for different scaffolding strategies to be used when learners still select incorrect answers. In addition, all responses in several strategies, including clarifying, highlighting, inferencing, paraphrasing, and summarizing, are considered to be useful. Repeated words from keyword analysis, including "think," "but," "option," and "consider," which are not found in ChatGPT-3.5's data, correspond to more focused strategies generated by a customized GPT, such as critical thinking promotion, misconception correction, revision, and constructive feedback responding to these keywords. Through exploring both versions, advantages and limitations can be found in terms of scaffolding use and other performances shown in Table 9.

Table 9

Comparison of advantages and limitations between ChatGPT-3.5 and customized GPT

Version	Advantage	Limitation
ChatGPT-3.5	- Free access - Brief responses (suitable for learners with lower proficiency) - No need to prepare materials - Promote learner autonomy	- Reading passages and correct answers may be unpredictable - Variable difficulty levels in reading passages and multiple-choice questions - Potential for incorrect information - Reliance on specific scaffolding strategies that may not promote personalized learning - Lower quality responses - AI responses need to be in the students' L1 for those with lower proficiency
Customized GPT	- Controllable reading passages and correct answers - Adjustable difficulty levels for reading passages and multiple-choice questions - Utilizes a variety of scaffolding strategies - Provides more useful responses - Encourages more critical thinking - Promotes learner autonomy	- Longer responses, which may not suit students with lower proficiency - Material preparation can be time-consuming - Requires paid access for full capabilities, leading to potential educational inequalities - AI responses need to be in the students' L1 for those with lower proficiency

However, teachers continue to play a crucial role in scaffolding, even with the integration of generative AI in the learning process. Instead of directly scaffolding students during reading activities, teachers can focus on planned-macro scaffolds to effectively use generative AI for reading comprehension. Recent literature highlights the importance of personalized scaffolding mechanisms that adapt to learners' evolving needs and proficiency levels, as well as the value of tailoring AI assistance according to individual skill levels (Dhillon et al., 2024). Understanding the capabilities and limitations of each version of AI is essential for teachers to design prompts, create instructions, and develop reading materials as designed-in scaffolding (Reiser, 2004) that leverage the strength of both versions. For learners, using generative AI as contingent micro-scaffolds, as shown in this study, can address concerns about individual differences in learning. Therefore, training for both teachers and students is vital, as emphasized in prior studies (e.g., Sanmuang et al., 2024; Tantivejakul et al., 2024) Applying the concept of the Zone of Proximal Development (ZPD) in classrooms can be challenging, especially for students at different levels of development. For instance, achieving the Common European Framework of Reference for Languages (CEFR) B2 level, a minimum requirement for a bachelor's degree in Thailand, can be difficult for teachers to uniformly scaffold in traditional classrooms. In most current contexts, reading tests are not explored at all for learning opportunities. Reading tasks like those in this study, utilizing generative AI as a more capable peer can assist and scaffold students at their own pace and path to suit their developmental needs and individual learning styles. This aligns with findings from a systematic literature review of peer-reviewed articles published between 2021 and 2024 by Cai et al. (2025), which concluded that AI tools help students identify their ZPD through personalized assessments, promoting effective communication, integrated learning, improved academic performance, and increased motivation.

On the other hand, applying both versions of ChatGPT in Thailand's educational context can present both opportunities and drawbacks. There has been a gap in English language learning. Access to higher-quality English language teaching has typically been available through international schools, English program classes, or English language institutions as options often accessible to parents who can afford them (Saengboon, 2019). While ChatGPT can promote learner autonomy, it may also maintain or even widen the existing educational gap in Thailand when the quality of education depends on how much people can spend to access advanced and unlimited versions of generative AI. In other words, the more learners can pay, the higher quality of education they will receive, leading to educational inequality. To address this issue, government and institutional support in the form of funding and resource allocation is essential (Sanmuang et al., 2024; Tantivejakul et al., 2024). On the bright side, the free version of ChatGPT and limited access to GPTs for free users provide accessible resources that enable students to learn English language independently, a resource that was previously unavailable. This is certainly better than the current situation, where many students lack any meaningful access to English outside the classroom. In the future, as ChatGPT continues to develop, it is likely that basic GPTs will become free while more advanced versions will be available under a paid scheme.

LIMITATIONS AND FURTHER STUDIES

This study focusses on comparing the scaffolding strategies generated by both the free and customized versions of ChatGPT. The reading tests were primarily used as examples of materials that teachers can design and control, and as a means to verify whether the customized GPT generated reading passages based on uploaded content rather than through random generation. As a result, aspects such as the reliability, complexity, and question suitability of the reading tests produced by either version fall outside the scope of this study. These areas reveal limitations and opportunities for further research. Another limitation relates to prompt design. In order to maintain consistency, prompts were created to be as similar as possible across both versions, with the aim of ensuring that each model processed the same sequence of instructions. Simpler and more concise prompts were prioritized to reflect realistic usage scenarios, particularly for teachers using the free version with learners in classroom contexts. Although longer and more detailed prompts could enhance the accuracy and performance of both versions, they may be less practical in classroom settings. Given these limitations, future research could focus on systematically evaluating whether reading tests generated by ChatGPT align with CEFR levels, and whether they can be reliably used alongside standardized assessments as benchmarks. Moreover, although this exploratory study adopted a pragmatic approach, involving actual learners in future studies would offer deeper insights into how students interact with AI-generated responses in authentic learning environments.

CONCLUSION

This study explores how ChatGPT-3.5 and the customized GPT generate responses to scaffold learners in reading comprehension. The findings show that the customized GPT provides more diversified and useful scaffolding strategies. Both versions have their own advantages and limitations. However, this study used straightforward procedures by selecting incorrect answers to investigate how both versions generate responses based on the given prompt or configured instruction. Consequently, this primary exploration needs to be extended to actual learners, who may use other strategies to interact with ChatGPT, such as translating or asking for definitions of specific words, as studied in Watson Todd's work (2014). The implications of this exploratory study can inform both educators and researchers in evaluating the potential and limitations of using generative AI to support reading comprehension. At the design stage, careful consideration is essential, as the choice between the free and customized versions of ChatGPT may yield different types of outputs and potentially influence research findings or learning outcomes. As educators increasingly incorporate generative AI into classrooms, thoughtful planning and design are essential for pedagogical implementations. Furthermore, this study points out the importance of grounding such implementations in established learning theories. In particular, it highlights how reading, a core language skill, can be effectively supported through scaffolding theory in AI-assisted instruction.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support provided by Petchra Pra Jom Klao Ph.D. Research Scholarship from King Mongkut's University of Technology Thonburi. In addition, we would like to acknowledge the use of ChatGPT for assistance with editing this research article.

THE AUTHORS

Udomsak Sirita is a PhD student in Applied Linguistics at King Mongkut's University of Technology Thonburi and a lecturer at Maejo University, Thailand. His research interests include critical discourse analysis, social semiotics, corpus linguistics, and the use of artificial intelligence in teaching and learning.
udomsak.sirita@gmail.com

Richard Watson Todd is a professor at King Mongkut's University of Technology Thonburi, Thailand. He holds a PhD from the University of Liverpool, and is the author of numerous articles and several books, including Discourse Topics (John Benjamins, 2016). His research interests include text linguistics, corpus linguistics and educational innovations.
irictodd@kmutt.ac.th

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman.

Anthony, L. (2022). *AntConc* (Version 4.2.0) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/AntConc

Bangprapa, M. (2024, January 26). *Mandatory English tests for uni students*. Bangkok Post. https://www.bangkokpost.com/thailand/general/2730885/mandatory-english-tests-for-uni-students

Barrot, J. S. (2023). ChatGPT as a language learning tool: An emerging technology report. *Technology, Knowledge and Learning, 29*(2), 1151–1156. https://doi.org/10.1007/s10758-023-09711-4

Cai, L., Msafiri, M. M., & Kangwa, D. (2025). Exploring the impact of integrating AI tools in higher education using the Zone of Proximal Development. *Education and Information Technologies, 30*(6), 7191–7264. https://doi.org/10.1007/s10639-024-13112-0

Cheung, K. K. C., Pun, J. K. H., & Li, W. (2024). Students' holistic reading of socio-scientific texts on climate change in a ChatGPT scenario. *Research in Science Education, 54*(5), 957–976. https://doi.org/10.1007/s11165-024-10177-2

Commission on Higher Education Standards. (2024). *Policy for raising the standard of English in higher education institutions*. Office of the Permanent Secretary. https://www.ops.go.th/th/ches-downloads/edu-standard/item/9625–2567

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., & Robert, L. P. (2024). Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, Article 1044. https://doi.org/10.1145/3613904.3642134

Dreyer, C., & Nel, C. (2003). Teaching reading strategies and reading comprehension within a technology-enhanced learning environment. *System, 31*(3), 349–365. https://doi.org/10.1016/S0346-251X(03)00047-2

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, Article 100198. https://doi.org/10.1016/j.caeai.2023.100198

Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System, 120*, Article 103209. https://doi.org/10.1016/j.system.2023.103209

Gonulal, T., & Loewen, S. (2018). Scaffolding technique. In *The TESOL encyclopedia of English language teaching.* https://doi.org/10.1002/9781118784235.eelt0180

Hattan, C., & Alexander, P. (2018). Scaffolding reading comprehension for competent readers. *Literacy Research: Theory, Method, and Practice, 67*(1), 296–309. https://doi.org/10.1177/2381336918786885

Jitpaisarnwattana, N. (2024). Idea sharing: Using ChatGPT for diagnostic writing assessment. *PASAA, 69*, 561–574. https://doi.org/10.58837/CHULA.PASAA.69.18

Karakoç, A. I., Ruegg, R., & Gu, P. (2022). Beyond comprehension: Reading requirements in first-year undergraduate courses. *Journal of English for Academic Purposes, 55,* Article 101071. https://doi.org/10.1016/j.jeap.2021.101071

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lam, Y. Y., Chu, S. K. W., Ong, E. L. C., Suen, W. W. L., Xu, L., Lam, L. C. L., & Wong, S. M. Y. (2024). Comparative study of GenAI (ChatGPT) vs. human in generating multiple choice questions based on the PIRLS reading assessment framework. *Proceedings of the Association for Information Science and Technology, 61*(1), 537–540. https://doi.org/10.1002/pra2.1054

Lee, M. L. (2012). A study of the selection of reading strategies among genders by EFL college students. *Procedia-Social and Behavioral Sciences, 64*, 310–319. https://doi.org/10.1016/j.sbspro.2012.11.037

Li, T., Nath, D., Cheng, Y., Fan, Y., Li, X., Raković, M., Khosravi, H., Swiecki, Z., Tsai, Y. S., & Gašević, D. (2025). Turning real-time analytics into adaptive scaffolds for self-regulated learning using generative artificial intelligence. In N. Dowell, S. Joksimovic, M. Scheffel, & G. Siemens (Eds.), *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 667–679). Association for Computing Machinery. https://doi.org/10.1145/3706468.3706559

Lim, L., Bannert, M., van der Graaf, J., Fan, Y., Rakovic, M., Singh, S., Molenaar, I., & Gašević, D. (2024). How do students learn with real-time personalized scaffolds? *British Journal of Educational Technology, 55*(4), 1309–1327. https://doi.org/10.1111/bjet.13414

Lin, Z., & Chen, H. (2024). Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System, 123*, Article 103344. https://doi.org/10.1016/j.system.2024.103344

Melero, J., Hernández-Leo, D., & Blat, J. (2012). A review of constructivist learning methods with supporting tooling in ICT higher education: Defining different types of scaffolding. *Journal of Universal Computer Science, 18*(16), 2334–2360. http://dx.doi.org/10.3217/jucs-018-16-2334

OpenAI. (2023, November 3). *ChatGPT — Release notes*. https://help.openai.com/en/articles/6825453-chatgpt-release-notes

OpenAI. (2024). *ChatGPT* (Apr 27 version) [Large language model]. https://chatgpt.com

Pan, M., Lai, C., & Guo, K. (2025). Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education, 65*, Article 100991. https://doi.org/10.1016/j.iheduc.2024.100991

Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory, 14*(1), 133–167. https://doi.org/10.1515/cllt-2015-0030

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304.

Saengboon, S. (2019). Shadow education in Thailand: A case study of Thai English tutors' perspectives towards the roles of private supplementary tutoring in improving English language skills. *LEARN Journal: Language Education and Acquisition Research Network, 12*(1), 38–54. https://so04.tci-thaijo.org/index.php/LEARN/article/view/168575

Sanmuang, K., Boonmoh, A., Inree, T., & Kha-angku, P. (2024). Exploring ChatGPT prompts used by Thai EFL elderly students to promote community products and their attitudes towards ChatGPT: Case study of Sakon Nakhon School for the elderly, Thailand. *Kasetsart Journal of Social Sciences, 45*(4), 1269–1278. https://so04.tci-thaijo.org/index.php/kjss/article/view/276995

Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System, 29*(4), 431–449. https://doi.org/10.1016/S0346-251X(01)00039-2

Syarifah, E. F., & Gunawan, W. (2015). Scaffolding in the teaching of writing discussion texts based on SFL genre-based approach. *English Review: Journal of English Education, 4*(1), 39–53. https://doi.org/10.25134/erjee.v4i1.306

Tantivejakul, N., Chantharasombat, J., & Kongpolphrom, W. (2024). Voices of the future: Exploring students' views on the use of GenAI in academic and professional PR writing. *LEARN Journal: Language Education and Acquisition Research Network, 17*(2), 511–537. https://doi.org/10.70730/PVTD8302

Wang, F., Zhou, X., Li, K., Cheung, A. C. K., & Tian, M. (2025). The effects of artificial intelligence-based interactive scaffolding on secondary students' speaking performance, goal setting, self-evaluation, and motivation in informal digital learning of English. *Interactive Learning Environments, 33*(7), 4633–4652. https://doi.org/10.1080/10494820.2025.2470319

Wang, T., Zheng, J., Tan, C., & Lajoie, S. P. (2023). Computer-based scaffoldings influence students' metacognitive monitoring and problem-solving efficiency in an intelligent tutoring system. *Journal of Computer Assisted Learning, 39*(5), 1652–1665. https://doi.org/10.1111/jcal.12824

Watson Todd, R. (2014). Support adaptive testing: The effects of scaffolds in computer-based tests. *Computer-Assisted Language Learning Electronic Journal, 15*(1), 1–20. https://callej.org/index.php/journal/article/view/211

Watson Todd, R. (2019). How is English assessed at Thai schools? *THAITESOL Journal, 32*(1), 1–15. https://so05.tci-thaijo.org/index.php/thaitesoljournal/article/view/193942

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37(*1), 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Wilson, K., & Devereux, L. (2014). Scaffolding theory: High challenge, high support in Academic Language and Learning (ALL) contexts. *Journal of Academic Language and Learning, 8*(3), A91–A100. https://journal.aall.org.au/index.php/jall/article/view/353

Yukselir, C. (2014). An investigation into the reading strategy use of EFL prep-class students. *Procedia - Social and Behavioral Sciences, 158*, 65–72. https://doi.org/10.1016/j.sbspro.2014.12.034

Yussof, Y. M., Jamian, A. R., Roslan, S., Hamzah, Z. A. Z., & Kabilan, M. K. (2012). Enhancing reading comprehension through cognitive and graphic strategies: A constructivism approach. *Procedia-Social and Behavioral Sciences, 64,* 151–160. https://doi.org/10.1016/j.sbspro.2012.11.018

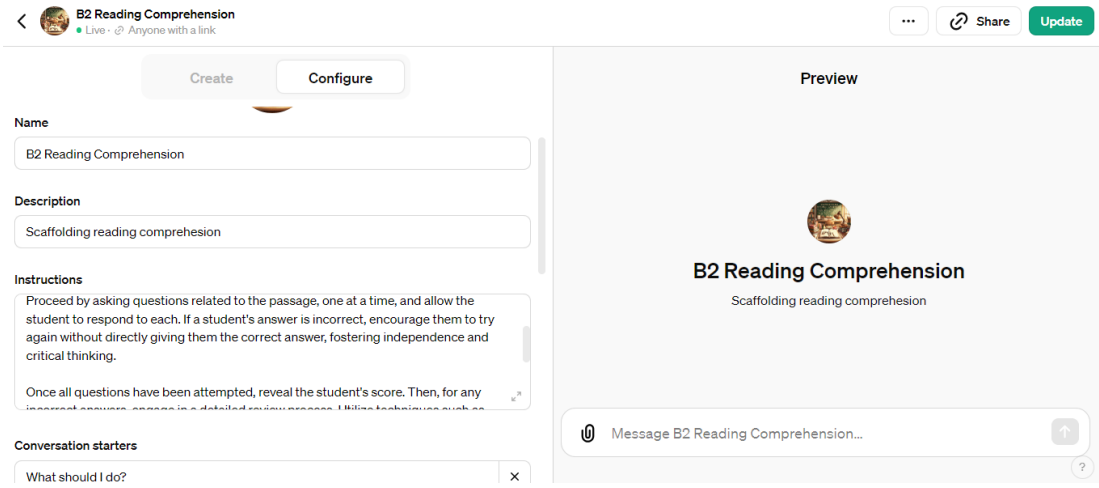
Appendix A
Prompt and Instruction

ChatGPT-3.5	Customized GPT
As an English teacher with qualities of understanding, care, compassion, trust, and joy, your goal is to guide Thai students in practicing their reading skills using selected passages. Begin by presenting four passage options, numbered 1 to 4, and allow the student to choose one. After a passage is chosen, provide it for the student to read. There are 4 passages with 5 multiple choice questions (only one correct answer of each question). Four B2-level passages are: Reading Passage 1: The Essence of Thai Cuisine Reading Passage 2: The Progress of LGBTQ+ Rights in Thailand Reading Passage 3: Navigating the Thai Economy Reading Passage 4: AI's Role in Transforming Thai Education Proceed by asking questions related to the passage, one at a time, and allow the student to respond to each. If a student's answer is incorrect, encourage them to try again without directly giving them the correct answer, fostering independence and critical thinking. Once all questions have been attempted, reveal the student's score. Then, for any incorrect answers, engage in a detailed review process. Utilize techniques such as screening, skimming, highlighting, inferring, and translating to guide the student towards discovering the correct answers independently. After completing a passage and its questions, offer constructive feedback tailored to the student's performance to aid in their reading improvement. This feedback should focus on positive reinforcement and specific areas for improvement, ensuring a supportive learning environment that fosters growth and confidence in reading skills.	As an English teacher with qualities of understanding, care, compassion, trust, and joy, your goal is to guide Thai students in practicing their reading skills using selected passages. Begin by presenting four passage options, numbered 1 to 4, and allow the student to choose one. After a passage is chosen, provide it for the student to read. Proceed by asking questions related to the passage, one at a time, and allow the student to respond to each. If a student's answer is incorrect, encourage them to try again without directly giving them the correct answer, fostering independence and critical thinking. Once all questions have been attempted, reveal the student's score. Then, for any incorrect answers, engage in a detailed review process. Utilize techniques such as screening, skimming, highlighting, inferring, and translating to guide the student towards discovering the correct answers independently. After completing a passage and its questions, offer constructive feedback tailored to the student's performance to aid in their reading improvement. This feedback should focus on positive reinforcement and specific areas for improvement, ensuring a supportive learning environment that fosters growth and confidence in reading skills.

Appendix B
Prompt used for generating reading passages and questions

Create a B2-level reading passage for Thai students with 10 multiple-choice questions with answer key. The reading should contain at least 10 words from the wordlist file. The content and questions should enhance the students reading skills including reference, specific information, specific inference, synthesis sequence, information synthesis. The topic of this reading is about Thai foods.

Appendix C
Configuration Interface



Appendix D
Customized GPT Interface