



ARTICLE



<https://doi.org/10.1057/s41599-025-04919-4>

OPEN

Enhancing student reading performance through a personalized two-tier problem-based learning approach with generative artificial intelligence

Changqin Huang^{1,2}, Yihua Zhong², Yongzhi Li^{3✉}, Xizhe Wang⁴, Zhongmei Han⁴, Di Zhang⁴ & Ming Liu⁵

Reading ability plays a vital role in the academic success of students. Problem-based learning (PBL) helps develop deep engagement with the reading materials and higher-order reading skills. However, conventional PBL (C-PBL) activities ignore differences in students' cognitive levels and fail to provide timely and targeted feedback and guidance to each student. As a result, many students are unable to actively engage in PBL-based reading activities. To address these problems, this study proposes a personalized two-tier PBL (PT-PBL) approach based on generative artificial intelligence (GenAI). It provides a more personalized and refined design for PBL activities to promote personalized reading learning for students. To examine the effectiveness of the proposed approach, 62 college students participated in a quasi-experiment, with the PT-PBL approach in the experimental group and the C-PBL approach in the control group. The results indicate that the PT-PBL approach significantly improves students' reading performance and motivation. In addition, compared to students with lower engagement, this approach is more effective at improving the reading performance of highly engaged students. Interviews with students showed that those who used the PT-PBL approach focused more on reading tasks and reflected more frequently. The main contribution of this study is proposing a novel PT-PBL approach and providing empirical evidence of its effectiveness, while also creating opportunities for future research to further explore the positive impact of GenAI on reading.

¹ College of Education, Zhejiang University, Hangzhou, China. ² Shanghai Institute of Artificial Intelligence for Education, East China Normal University, Shanghai, China. ³ China National Academy of Educational Sciences, Beijing, China. ⁴ Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China. ⁵ Faculty of Education, Southwest University, Chongqing, China. ✉email: liy2@cnaes.edu.cn

Introduction

Reading ability is one of the critical literacies students require for twenty-first-century skills. To develop students' reading skills, teachers employ appropriate teaching approaches to guide students' reading processes and prevent them from reading aimlessly (Spörer et al. 2009). Additionally, the adoption of appropriate teaching approaches to create real-life reading and communication contexts is conducive to fostering deeper engagement with the material (Lin 2018). Consequently, implementing appropriate teaching approaches can help students develop their reading skills.

Problem-based learning (PBL) is a teaching approach that uses real-life problem contexts or examples to drive active learning (Arani et al. 2023; Gao et al. 2022; Liu and Pásztor 2022). By guiding students through real-life problems, PBL helps them establish connections between prior knowledge and practical applications (Orhan 2024). Recently, its potential in language education has attracted the attention of scholars (Arani et al. 2023). Relevant studies have shown that the PBL approach can improve students' reading performance (Lin 2018; Orhan 2024). However, the limitations of applying PBL to large-scale reading instruction are also evident. For instance, the teacher is unable to provide timely and targeted feedback and guidance to each student (Orhan 2024). Moreover, the problems used in PBL are oriented toward all students, ignoring the differences in their cognitive levels, which makes it difficult to meet their personalized needs (Arani et al. 2023; Lin 2018). As a result, many students are unable to actively engage in PBL-based reading activities. Researchers have found that reading motivation is a key factor influencing reading success (Gambrell and Marinak 2009), and using technology (e.g., artificial intelligence) to provide personalized support can enhance students' motivation to read (Hsu et al. 2013; Alrawashdeh et al. 2024). Thus, if the full potential of PBL in reading is to be realized, it is necessary to integrate effective technologies to provide more personalized and targeted support.

Recently, the effects of generative artificial intelligence (GenAI) on reading education have received growing attention (Jeon et al. 2023; Wang et al. 2024). It is an AI technology that uses sophisticated algorithms, models, and rules to learn from large-scale datasets and create new, high-quality content, with great potential to facilitate personalized learning (Corlatescu et al. 2024). Wang et al. (2024) applied GenAI to reading learning, where students received personalized learning content through interaction with GenAI to enhance reading performance. Steiss et al. (2024) employed GenAI as an intelligent assessment tool to provide students personalized feedback and promote reflective behaviors. Consequently, integrating GenAI with PBL can provide students with more opportunities to solve problems in real-life contexts, while providing targeted feedback and guidance, thereby enhancing personalized reading learning. Nevertheless, Kasneci et al. (2023) have also pointed out the shortcomings of GenAI. Students tend to over-rely on GenAI-generated answers, resulting in a lack of deep thinking (Lee et al. 2024). Additionally, GenAI lacks consideration of key student information (e.g., cognitive level), making it difficult to provide the personalization needed for effective learning. In view of this, this study will consider these issues to maximize the potential of GenAI for facilitating personalized reading learning.

Literature review

Technology-assisted reading interventions. Reading skills are universally recognized as fundamental to a student's overall academic success (Alrawashdeh et al. 2024). Proficiency in reading not only serves as a foundation for other academic areas,

but also plays a central role in lifelong learning (Alrawashdeh et al. 2024; Kambara et al. 2023). Previous studies have demonstrated that student engagement mediates the relationship between motivation and reading comprehension (Van Ammel et al. 2021; Ronimus et al. 2022). In other words, when students are sufficiently motivated to read, they tend to focus more on the reading activity, leading to better comprehension of the reading material. Nevertheless, traditional teaching approaches have significant limitations in promoting student engagement in reading (Ronimus et al. 2022). In large classes, teachers struggle to offer personalized reading support based on students' cognitive levels (Rodríguez et al. 2022). Moreover, students often struggle to generate interest in reading due to a lack of real-life contexts. Consequently, addressing the lack of personalized support and real-life contexts in traditional teaching approaches is crucial for improving students' reading outcomes.

In recent years, numerous studies have investigated the benefits of technology-assisted reading interventions (Clinton-Lisell et al. 2023; Egert et al. 2022; Kong et al. 2018; Savva et al. 2022; Yu et al. 2022). Regarding the provision of personalized support, Wang et al. (2024) used AI technology to generate personalized reading questions for students. They found that personalized questions increased students' reading motivation and achievement more than standardized questions. Hsu et al. (2013) developed an adaptive language learning system that significantly improved learning achievement by using a personalized recommendation mechanism to guide students to read materials that match their preferences and knowledge levels. In terms of constructing real-life contexts, Ronimus et al. (2014) used digital game technology to create contextualized reading contexts for students and found that this stimulated students' interest in reading. Clinton-Lisell et al. (2023) compared students' reading performance when using e-texts with interactive features to that of using paper or static e-texts. The results indicated that incorporating more realistic interactive experiences into e-texts was beneficial for improving reading performance and efficiency.

While previous studies have demonstrated the positive impact of technology-assisted reading interventions, few have simultaneously considered personalized support and real-life contexts (Clinton-Lisell et al. 2023; Ronimus et al. 2014; Wang et al. 2024). Moreover, most research has focused on providing either personalized support or real-life contexts from a technological perspective, overlooking the crucial role of teaching approaches in guiding student learning and improving performance. For instance, Wang et al. (2024) used AI technology to generate reading questions tailored to students' personalized needs. However, merely practicing these questions without integrating appropriate teaching approaches may boost short-term reading performance but will not support the long-term development of reading skills. Therefore, this study aims to use effective technology to provide personalized support for students' reading while employing appropriate teaching approaches to guide them to construct knowledge in real-life contexts, thereby fostering deep engagement with the reading material.

PBL in language learning. PBL, as a popular learning and teaching approach, has been investigated across several disciplines (Anggraeni et al. 2023; Arani et al. 2023; Gao et al. 2022; Liu and Pásztor 2022). In this approach, students are placed at the center of the learning process, where they acquire knowledge both individually and collectively, develop their understanding, and enhance their problem-solving skills (Liu and Pásztor 2022). Students can also receive feedback from their peers and teachers regarding their collaboration and learning outcomes (Arani et al.

2023). Several procedural variations of PBL have been proposed in previous research. For example, Edens (2000) divides PBL into three stages: development, exploration, and verification. Barrows (1986) split the PBL into five steps involving issue analysis, data collection, stage integration, summarization, and reflection. Given the diversity of PBL approaches, this study adopts a six-step activity framework, a commonly used and recognized structure in the literature (Aslan 2021; Fidan and Tuncel 2019; Hmelo-Silver 2004). First, the educator presents the problem to the class. Second, students gather relevant details and assess the problem. Third, once they better understand the problem, they generate hypotheses for solving it. Fourth, during the autonomous learning phase, students focus on the problem and their hypotheses. Fifth, after independent study, they use new information to critique their initial assumptions. Finally, students reflect on the learning process and synthesize what they've learned. This widely adopted six-step process represents the conventional form of PBL, referred to in this study as conventional PBL (C-PBL), chosen to provide a clear and consistent framework for comparison.

Recently, studies have shown that PBL enhances students' language learning performance. For example, Lin (2018) incorporated PBL methods into a reading course and significantly enhanced students' academic performance. Chen et al. (2021) found that the PBL approach was effective in improving students' English vocabulary acquisition and motivation while enhancing their problem-solving skills. In addition, Neville and Britt (2007) emphasized the importance of applying the PBL approach to integrating foreign language and engineering. Through this approach, students not only enhance their specialized engineering skills, but they also acquire occupation-specific foreign language skills while solving real-world problems. Furthermore, the PBL approach has been shown to be effective in technology-enhanced educational settings (Arici and Yilmaz 2023; Chen et al. 2021; Donnelly 2010; Fung et al. 2023; Lee 2023). For instance, Chen et al. (2021) incorporated PBL into a virtual learning environment to provide students with an immersive language learning experience. This approach proved to be more effective in motivating students than the C-PBL approach. In another study, Chang et al. (2020) combined PBL with digital games and demonstrated that this approach was effective in enhancing student satisfaction, enjoyment, and motivation. More importantly, they found that this approach greatly facilitated independent learning, thereby improving students' academic performance. In conclusion, the PBL approach helps to improve students' language learning outcomes and is suitable for use in technology-enhanced educational settings.

Most existing studies have designed and investigated the effects of PBL on promoting student learning from a macro perspective, but lack more personalized and refined approaches. In PBL activities, timely and targeted feedback and guidance are critical to deepening understanding as the problem is explored in depth, especially during the students' self-directed learning phase. However, in large classes, it is difficult for teachers to provide such feedback and guidance to each student. In addition, C-PBL activities do not consider differences in students' cognitive levels when designing problems, resulting in some students finding it challenging to cope with (Hmelo-Silver 2004). Therefore, this study aims to integrate effective technologies to guide students' reading in a more personalized and refined way.

GenAI-supported personalized language learning. Researchers are increasingly focusing on the impact of GenAI on language learning (Zhao et al. 2024). Recent studies have demonstrated that GenAI-supported learning systems can enhance personalized language learning for students (Ma and Huo 2023). For example,

Guo and Li (2024) provided personalized assistance to students by involving them in interactions with GenAI-based chatbots, which clarified their writing goals and fostered a more positive attitude towards writing. Meniado et al. (2024) used GenAI as an automated assessment tool for language learning, offering students timely, personalized feedback. This approach addressed the limitations of time-consuming, asynchronous teacher assessments and led to improved student performance. Liu et al. (2024) utilized a GenAI-based learning system to support students in completing multimodal writing tasks, finding that the GenAI-assisted writing process quickly identifies learners' needs and provides meaningful feedback, thus promoting personalized learning. In addition, several studies have reported the benefits of GenAI-supported learning systems in generating learning resources. For instance, Wang et al. (2024) combined GenAI with adaptive learning techniques to develop a personalized reading system that automatically generates questions tailored to individual learners' reading abilities. The results indicated that these personalized questions improved students' reading performance more effectively than standardized questions created by teachers. In another study, Lin and Chen (2024) also investigated the potential of GenAI in generating personalized reading questions. Unlike previous studies, they combined psychometric measurement methods with manual review to validate the quality of GenAI in generating personalized questions. Thus, GenAI plays a crucial role in advancing personalized language learning and has become a key technology in language learning.

However, there are still shortcomings in the use of GenAI in language education. To realize the full potential of GenAI in facilitating language learning, it is necessary to address student over-reliance (Kasneci et al. 2023). GenAI should provide guidance rather than giving direct answers (Lee et al. 2024). Otherwise, it may foster dependency and hinder the development of higher-order thinking skills (Chan and Hu 2023; Ding et al. 2023). Thus, guidance mechanisms should be implemented to prevent GenAI from providing direct answers, and to integrate effective teaching approaches into GenAI-supported learning environments. In addition, while GenAI generates learning resources based on student requests, it lacks consideration of key student information, making it difficult to provide the level of personalization needed for effective learning (Kasneci et al. 2023). Adaptive learning techniques have been shown to be an effective way to facilitate personalized learning (Alrawashdeh et al. 2024). Therefore, using adaptive learning techniques to consider key information about students (e.g., cognitive level) can enhance the personalization of GenAI (Kasneci et al. 2023). With this in mind, this study will address these issues when using GenAI to maximize its potential for facilitating personalized language learning.

In summary, this study aims to combine GenAI with PBL to propose a personalized two-tier PBL (PT-PBL) approach. This approach divides PBL-based reading activities into two stages, guiding students from mastering foundational knowledge to developing higher-order thinking skills through progressively complex and in-depth real-life problem contexts. During the process of solving complex problems, the approach leverages the potential of GenAI to provide more targeted learning support to students, thereby enhancing personalized reading learning. To evaluate the effectiveness of this approach, we aimed to answer the following research questions:

RQ 1. Does the personalized two-tier problem-based learning (PT-PBL) approach improve student learning motivation compared to the conventional problem-based learning (C-PBL) approach?

RQ 2. To what extent does the PT-PBL approach improve the students' reading performance in comparison with the C-PBL approach?

RQ 3. What are the differences in reading engagement among students using different learning approaches, and how does varying engagement affect their reading performance?

RQ 4. How do students feel about their learning under different learning approaches?

Personalized two-tier PBL approach design for reading

Enhancing personalization of ChatGPT-generated reading problems. To realize the full potential of the GenAI-based PT-PBL approach in facilitating personalized reading learning, it is necessary to address the shortcomings of GenAI (as exemplified by ChatGPT).

One of the limitations of ChatGPT is its difficulty in providing the level of personalization needed for effective learning. As a universal conversational LLM, it cannot integrate multiple objective factors directly, such as past performance or learning duration, to accurately assess cognitive states. In addition, model hallucinations and network issues may destabilize prediction results in LLM applications. With reference to Wang et al. (2024), this study developed a customized machine-learning model tailored for cognitive ability prediction tasks. The customized model achieved higher performance (accuracy: 87.4%, AUC: 0.867) compared to ChatGPT-based approaches (accuracy: 78.9%, AUC: 0.806). This confirms that by leveraging the customized model instead of directly relying on ChatGPT's output, significant improvements in prediction performance were achieved.

In detail, the proposed model takes four key features as input: academic knowledge level, effective learning duration, historical performance, and answer precision. Academic knowledge level serves as a measure of students' foundational knowledge and is determined based on past standardized test performance; higher levels suggest greater cognitive processing ability. Effective learning duration measures the time students spend actively engaging with material, focusing on intervals that demonstrate meaningful engagement, which reflects students' ability to concentrate and process information. Historical performance reflects students' accuracy and comprehension in prior tasks, indicating how well they understood and applied the material, with higher scores suggesting stronger mastery and application of knowledge. Answer precision, calculated through semantic similarity between student responses and correct answers, quantifies the relevance and accuracy of responses, indicating the ability to provide precise answers based on understanding.

According to the theory of the zone of proximal development, tasks of appropriate difficulty enhance learning outcomes, while overly difficult or too easy tasks may hinder students' learning experiences and effectiveness. Therefore, the model introduces a cost-sensitive method that imposes prediction loss penalties based on the position of predicted results within a matrix, ensuring through continuous training that more predictions fall within the specified zone of proximal development. Additionally, the model incorporates a forgetting function, drawing on Ebbinghaus's law (Liu et al. 2021), which applies decay to time-related feature values, dynamically reflecting shifts in students' cognitive levels over time. This approach helps improve prediction accuracy and timeliness by accounting for the influence of forgetting on cognitive states. The model undergoes multiple rounds of iterative optimization and parameter tuning to enhance prediction accuracy and stability.

Finally, the model categorizes students' cognitive levels into six tiers, from 1 (the lowest) to 6 (the highest), each aligning with a cognitive objective in Bloom's Taxonomy: recall, understanding, application, analysis, synthesis, and evaluation. Specifically, level 1 tasks require students to recall factual information directly from

the material, while level 2 focuses on comprehension, prompting students to explain or interpret content in their own words. Level 3 involves applying learned knowledge to solve practical problems, and level 4 demands analytical skills to examine relationships or structures within the material. At level 5, students synthesize information from multiple sources to form cohesive solutions, and level 6 requires critical evaluation, asking students to assess arguments or to make judgments based on specific criteria. This structure ensures that each level progressively engages higher cognitive processes, from basic recall to advanced evaluation. To generate questions that correspond to these cognitive levels, prompts are tailored to each predicted level. For example, for level 3 (application), a prompt might be: "Generate a reading problem suitable for students at level 3 (application), focusing on their ability to apply learned knowledge to solve specific problems." By aligning prompts with predicted cognitive levels, ChatGPT is guided to generate questions that match students' cognitive abilities and support individualized learning.

In the PT-PBL approach, ChatGPT is used specifically to automatically generate personalized reading problems for students (see Fig. 1). Teachers collect and prepare essential data—such as the reading material and students' cognitive levels—before uploading it to the system. The reading material defines the content and context of the tasks, while cognitive levels allow for automatic adjustment of task difficulty to match individual students' current learning stages. During task generation, prompt engineering technique are used to structure task descriptions, inputs, and output indicators that help ChatGPT create problems aligned with each student's needs and cognitive objectives.

In addition, we adopted a Turing test-like approach to assess whether the quality of machine-generated problems was comparable to that of expert-written problems, which are generally considered to be of higher quality. Specifically, students were asked to distinguish between machine-generated and expert-written problems, thus providing an indirect measure of the quality of the machine-generated problems. Additionally, three experienced English teachers were invited to conduct a blind, back-to-back assessment of both sets of problems. The teachers independently rated the problems without knowing their source, ensuring objectivity and rigor in the evaluation process. Finally, a questionnaire was administered to students to assess whether the problems met their learning needs and to ensure that the machine-generated problems aligned with their individualized levels. Based on feedback from students and evaluations by teachers, the problem generation process was further refined to enhance the personalization and quality of the reading problems.

Guidance-based ChatGPT-assisted reading aid. Another limitation of ChatGPT is that the direct answers it provides can easily lead to students' over-reliance, thereby hindering their ability to think deeply and solve problemEGs independently. Therefore, referring to Lee et al. (2024), this study employed a guidance-based ChatGPT-assisted reading aid (GCRA) for PBL activities. Unlike the conventional way of using ChatGPT, GCRA does not give answers directly, but rather prompts students to formulate their own answers while providing insightful hints to help them refine their solutions step-by-step.

The GCRA's feedback mechanism is based on a multi-level approach, dynamically generating feedback according to student's performance and learning progress. The following are the three types of feedback provided by GCRA:

- (1) Immediate corrective feedback: This feedback is given when a student's answer contains obvious errors or misunderstandings. For example, if a student misunderstands a key

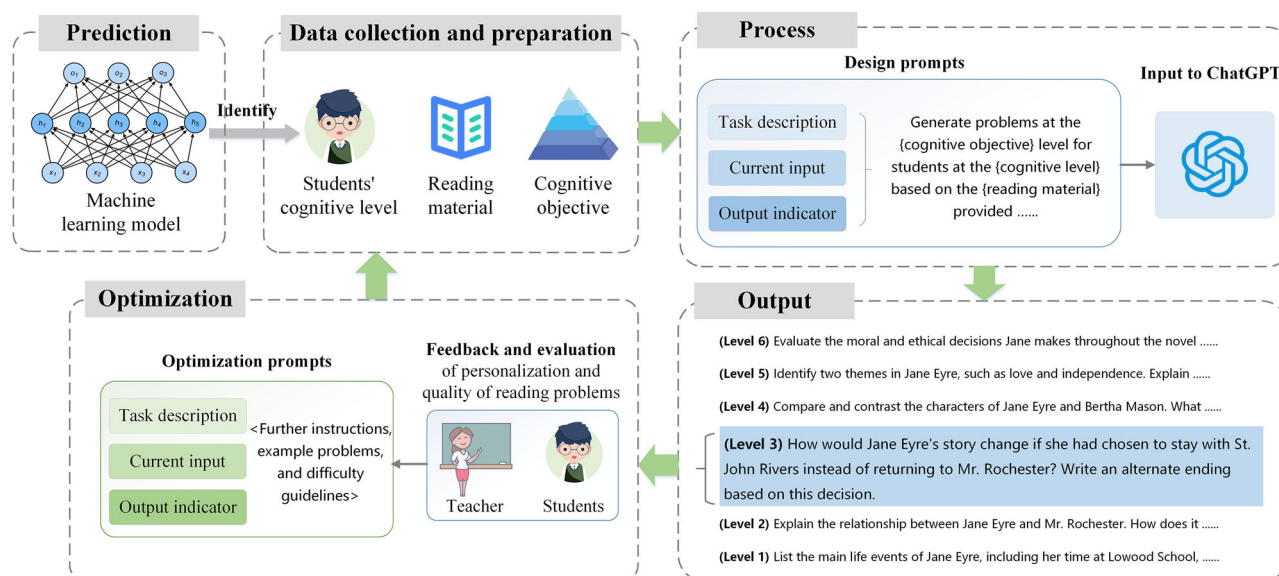


Fig. 1 Example of personalized reading problem generation. It shows the process of using ChatGPT to automatically generate personalized reading problems for learners.

concept, GCRA will offer direct but subtle hints, such as: “Consider re-reading the relevant chapters about Jane Eyre’s childhood. How did her experiences affect her later behavior?”

- (2) Reflective feedback: After a student completes a task, GCRA offers suggestions for improvement based on the student’s learning log. This feedback aims to encourage students to reflect on their entire learning process. For example: “You’ve primarily focused on the characters’ actions, but you could also consider their emotions and psychology. How do Jane Eyre’s emotions influence her decisions?”
- (3) Guided inquiry feedback: When a student’s answer is partially correct, GCRA provides guiding prompts to encourage deeper thinking and analysis. It is important to note that GCRA does not directly correct the student’s mistakes but instead poses exploratory questions to stimulate further thought. For instance, after a student suggests, “If Jane had never met Helen, she might have become colder and more vengeful,” GCRA might respond: “Can you think more deeply about the specific role Helen played in shaping Jane Eyre’s values? How is this reflected in her later actions?”

While the first and second types of feedback provide direct responses based on the student’s current state, the third type, guided inquiry, is more complex and focuses on deeper cognitive engagement and exploration. For a detailed example of this process, see Fig. 2.

Figure 2 shows an example of students using GCRA to assist in solving reading problems. Students employed GCRA to provide guidance when faced with difficult problems in reading. First, they entered questions into GCRA, such as, “If Jane Eyre hadn’t met Helen, how would her character have been affected?” Afterward, GCRA guided students to provide an original solution based on students’ prior knowledge, such as, “She might have become colder and more vengeful.” In Step 4, GCRA entered the student’s question and original solution into ChatGPT, relying on its powerful text comprehension to generate an insightful prompt for the student, such as, “Your answer mentioned that Jane Eyre might become colder and more vindictive, which is great. But can you think more deeply about the specific role Helen played in Jane Eyre’s values? How is this reflected in her later actions?”

After being given the prompt, students reflected and reread the material, thinking about how to add details to refine the solution through deeper comprehension. Upon completion of the reading, GCRA recorded the students’ learning data and created a learning log. The learning log records students’ progress, including the number of attempts at answering questions, the types of errors made, and the cognitive level of the questions. Based on this log, GCRA generates detailed, personalized feedback, providing students with reflective insights and suggestions for improvement. Through GCRA’s feedback and prompts, students not only resolve problems but also engage in reflective learning, refining their understanding through repeated iterations. This guided feedback process helps students engage more deeply with the material and fosters the development of higher-order thinking skills (Lee et al. 2024).

Design and application of the PT-PBL approach in reading.

The PT-PBL approach was designed based on the cognitive goals of Bloom’s Taxonomy (Bloom et al. 1956). The approach divides PBL activities into two tiers (see Fig. 3).

In the first tier, the main purpose of the PBL-based reading activities is to help students grasp the fundamental concepts of the material. Students engage in simpler thinking activities through conventional learning styles under teacher guidance. As students continue to explore reading problems, they engage in more complex thinking activities (e.g., application, analysis, synthesis, evaluation). Thus, in the second tier, personalized reading problems are provided based on students’ cognitive levels. These problems are designed to foster deeper engagement with the material and develop higher-order thinking skills. In addition, when students encounter difficulties, they can use GCRA to obtain timely and targeted feedback to facilitate their problem-solving.

Referring to Fidan and Tuncel (2019) and Aslan (2021), this study divides the reading process based on the PT-PBL approach into six stages: “Problem Presentation”, “Problem Definition”, “Information Gathering and Sharing”, “Personalized Problem Presentation”, “Independent Analysis and Solution”, and “Reflection and Evaluation.” The first three stages are the first tier, and the last three stages are the second tier. Fig. 4 illustrates the reading process based on the PT-PBL approach.

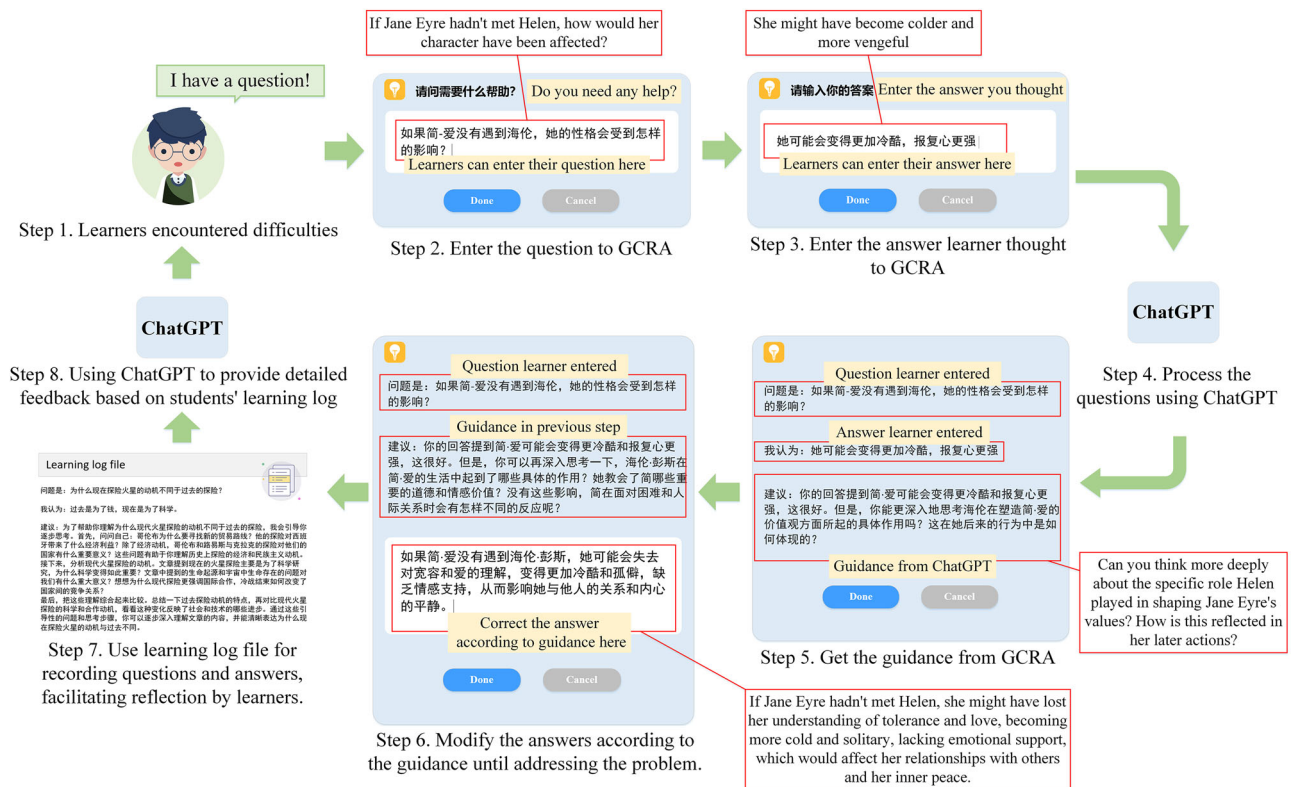


Fig. 2 GCRA-supported problem-solving process in reading. It displays a case in which a student uses GCRA to support problem-solving in reading.

Problem presentation. The teacher prepares the reading material and poses reading problems related to the material, such as “Where was Jane Eyre sent by her aunt? Who did she hate the most there? Why?” These problems are designed to stimulate students’ curiosity and interest in reading.

Problem definition. The teacher guides students in group discussions to clarify the problem and identify the fundamental concepts that need to be understood, such as discussing Jane Eyre’s childhood experiences and listing key events and characters.

Information gathering and sharing. Students read, gather information from the reading material, and share it with their group members. For example, students read relevant chapters of the novel, gather detailed information about Jane Eyre’s childhood, and share their findings with the group.

Personalized problem presentation. The teacher provides ChatGPT with background information about the students’ cognitive level and the reading material, and uses prompt engineering to instruct ChatGPT to generate personalized reading problems. These questions are more in line with students’ cognitive levels than the reading problems posed by the teacher, further examining students’ comprehension of the reading material and aiming to develop deeper understanding and critical thinking. For example, “How would Jane Eyre’s character and fate have been different if she had not met Helen at Lowood School?”

Independent analysis and solution. Students think and solve problems independently. When they encounter difficulties, students can ask GCRA questions and GCRA will provide guiding hints. For example, when a student asks, “How would Jane Eyre’s character have been affected if she had not met Helen?” GCRA might prompt, “Think about how Helen’s behaviour and attitudes

affected Jane Eyre. How did their friendship affect Jane Eyre’s values and beliefs?”

Reflection and evaluation. Students reflect on the effectiveness of the learning process and solutions. During this reflection, they provide some examples similar to the problem situation from real life. Additionally, GCRA provide detailed feedback based on students’ historical learning data. Finally, teachers wrap up and offer insights, encouraging students to engage in a new round of learning.

Methods

Participants. A total of 62 university students (aged 19 or 20) in their second year at a university in China participated in the study. The students were randomly divided into an experimental group (EG) and a control group (CG). The EG consisted of 32 students, 14 males and 18 females, who used the PT-PBL approach. The CG consisted of 30 students, 13 males and 17 females, who used the C-PBL approach. All participants met the inclusion criteria and signed informed consent forms.

Experimental design and procedures. The experimental procedure is illustrated in Fig. 5. To ensure the rigor of the experimental design and minimize the influence of confounding variables, several measures were implemented. All participants used the same online learning system throughout the intervention, with standardized access to identical learning materials, synchronized session schedules, and uniform task requirements. Their learning activities were closely monitored to ensure consistency in the external resources available. The EG and CG were assigned tasks with the same total number of problems, maintaining an equal workload between the groups.

Before the experiment commenced, participants in both groups completed the reading performance and learning motivation pre-tests

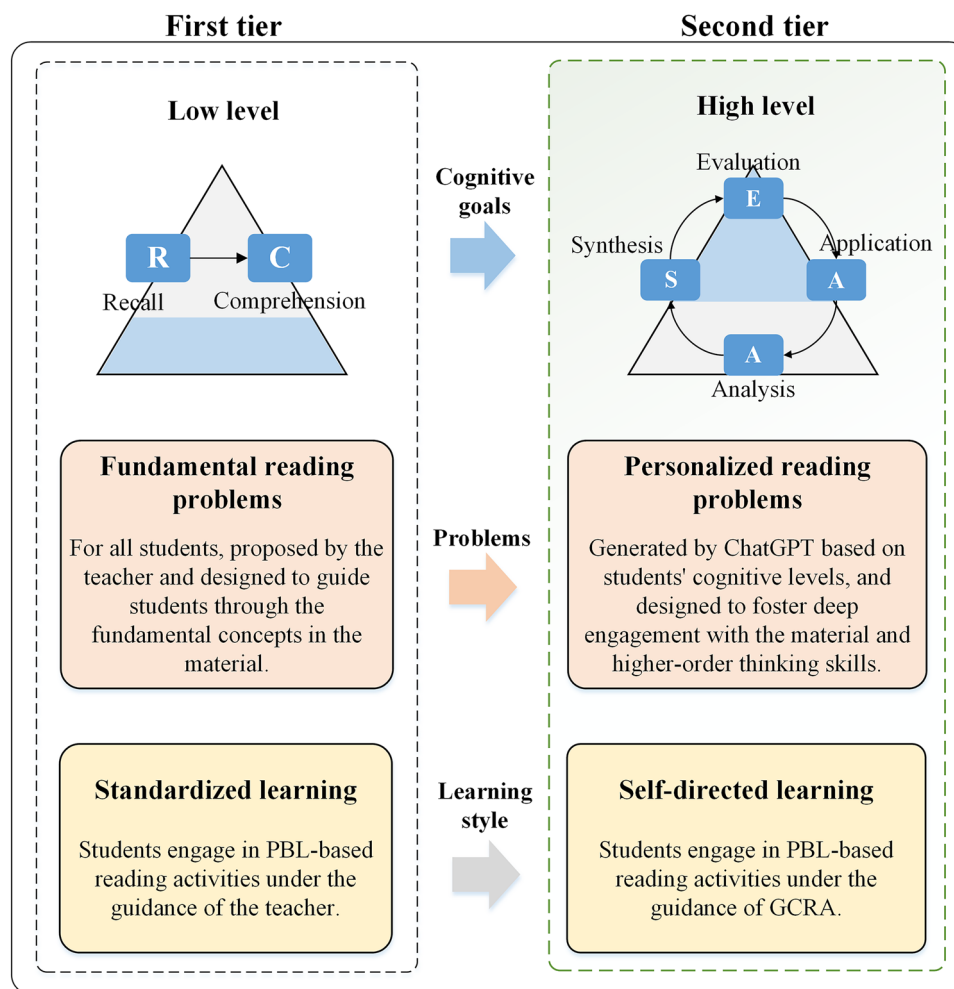


Fig. 3 Design of the PT-PBL approach. It shows the personalized two-tier problem-based learning approach designed based on the cognitive levels of Bloom's taxonomy.

to assess their reading performance and ensure comparability. Following this, students in the EG conducted the reading activity using the PT-PBL approach, and those in the CG did the reading activity using the C-PBL approach. It is worth noting that the EG and CG had the same total number of problems in the reading activity, and both used the online learning system for problem solving. The difference is that the EG received personalized problems when using the system and got detailed and targeted instruction from GCRA, whereas the CG received standardized problems when using the system and got regular instruction. Furthermore, statistical methods, including ANCOVA, were employed to control for baseline differences, such as pre-test scores, and to isolate the effect of the intervention more precisely. ANCOVA was chosen because it allows for the control of potential covariates and reduces error variance, thus providing a more accurate estimate of the intervention's impact. This approach was particularly important for ensuring that the effects measured were not confounded by pre-existing differences between groups.

Subsequently, all participants in both groups completed a post-test assessment of their reading performance, along with a post-test questionnaire on learning motivation and engagement. Additionally, twelve students (six high achievers and six low achievers) were randomly selected from each group for interviews.

Instruments and data collected. The reading performance test comprised six questions: three explicit and three implicit. The test

was designed with reference to Danaei et al. (2020). Explicit questions are those that can be answered directly from the reading material, such as, "According to the reading material, for what reasons do explorers usually venture into uncharted territory?" Implicit questions require students to make inferences based on prior knowledge and clues in the reading material to test whether they have developed a deep comprehension of the reading material, e.g., "Why might science play a more important role in the future exploration of Mars?" Furthermore, explicit questions were categorized as recall of details and cause-and-effect relationships, and implicit questions were categorized as feelings and problem-solving. The test consisted of open-ended questions. Students' reading pre-tests and post-tests were assessed by two experienced language teachers with over a decade of experience in reading instruction. The reading grading teachers were required to use a double-blind, back-to-back assessment method. The Cohen's kappa of the two teachers' grading showed acceptable inter-coder reliability ($k = 0.88$) (Cohen 1968), and the average of the two teachers' scores was used as the final score for the students' reading performance.

The learning motivation scale comprised six items including intrinsic and extrinsic motivation. It was designed with reference to Lin et al. (2021). The phrases were slightly modified based on the context of this study; for instance, "It is the most satisfying thing for me to get a good grade in the reading course." The questionnaire employed a five-point Likert scale (with a minimum score of 1 and a maximum score of 5). The higher

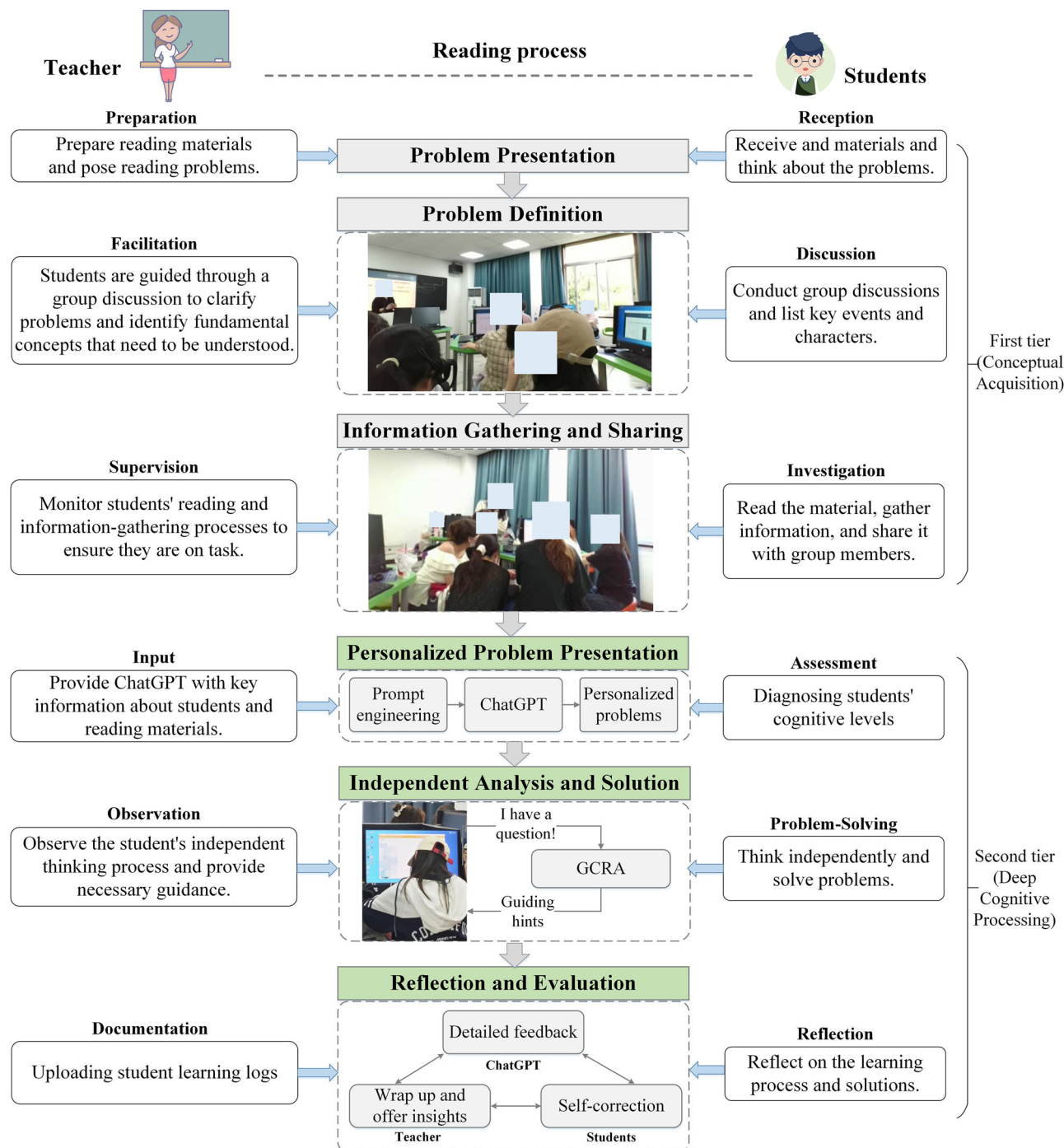


Fig. 4 Reading process based on the PT-PBL approach. It illustrates the reading instruction process in class based on the personalized two-tier problem-based learning approach.

the score, the more motivated the students were. The Cronbach's alpha value of the questionnaire was 0.85.

Students' reading engagement is usually measured through questionnaires. However, since the learning system automatically records useful data, students' engagement can also be measured through their learning records (Huang et al. 2023). Therefore, this study combined questionnaires and learning records to measure students' reading engagement more accurately. Referring to Li et al. (2021), total reading time, total pages read, task completion rate (Xiao and Hew, 2024), login frequency (Lin et al. 2019), and temporarily away (Sun et al. 2023) from the learning record store were utilized, combined with engagement questionnaire scores, to

compare students' reading engagement. Total reading time was recorded as the total amount of time students spent engaged with the reading material during the intervention. Total pages read referred to the number of pages students completed during their reading activities. Task completion rate was calculated as the percentage of tasks completed by each student. Login frequency referred to how often students logged into the learning system. Finally, temporarily away was recorded when students were logged on for more than 20 min without any mouse movement or data recording. The reading engagement questionnaire comprised twenty items including behavioral, emotional, and cognitive engagement. It was designed with reference to Sun et al. (2021).

The phrases were slightly modified based on the context of this study; for instance, “I attempted to engage in reading activities.” The questionnaire employed a five-point Likert scale. The Cronbach’s alpha value of the questionnaire was 0.75. According to Yang et al. (2021), student engagement was categorized into different levels (i.e., high or low engagement) by calculating the mean score of the student engagement questionnaire in each group.

Referring to Lin et al. (2021), an interview outline with five slightly modified questions was developed as follows:

- (1) What do you think you have learned the most when reading using this approach? At what stage did you learn the most? Give concrete examples.
- (2) How is this reading class different from other conventional reading classes you have previously experienced? Do you find it useful? Why?
- (3) What aspects of your reading ability were improved (e.g., critical reading)? Give concrete examples.
- (4) Would you recommend this learning approach to your friends or other students? Why?
- (5) Would you recommend other teachers to use this approach in specific subjects? Why?
- (6) Please summarize the benefits of the adopted approach in learning to read.

Results

Response to RQ 1 Does the PT-PBL approach improve student learning motivation compared to the C-PBL approach? A one-way ANCOVA was used to examine the impact of different learning approaches on students’ learning motivation. The groups using different learning approaches (EG and CG) were considered independent variables, while the learning motivation post-test

and pre-test were the dependent and covariate variables, respectively. ANCOVA was chosen because it allows for the control of potential covariates, such as pre-test scores, ensuring a more precise estimation of the intervention effect. This approach reduces error variance and provides a clearer isolation of the effects of the learning approach, offering a reliable basis for comparing post-test outcomes. The assumption of homogeneity of regression coefficients for extrinsic motivation ($F = 0.475$, $p = 0.493$) and intrinsic motivation ($F = 0.231$, $p = 0.632$) questionnaire scores was confirmed. Table 1 shows the ANCOVA results. Regarding extrinsic motivation, the adjusted means of the EG and CG were 3.57 and 3.28, respectively. Furthermore, there was a significant difference between the EG and CG in extrinsic motivation ($F = 5.967$, $p = 0.018$, $\eta^2 = 0.092$). However, there was no significant difference in intrinsic motivation between the EG and CG.

In addition, students were separated into two subgroups based on the degree of improvement in their motivation (post-test scores minus pre-test scores). Students with an increase in extrinsic motivation equal to or greater than zero were categorized into the Extrinsic Motivation Increase (EM+) subgroup, while the other students were categorized into the Extrinsic Motivation Decrease (EM-) subgroup. Similarly, intrinsic motivation was categorized into the Intrinsic Motivation Increase (IM+) subgroup and Intrinsic Motivation Decrease (IM-) subgroup. In terms of extrinsic motivation, the EM+ subgroup consisted of 25 students (78%) in the EG and 20 students (67%) in the CG. In contrast, the EM- subgroup included 7 students (22%) in the EG and 10 students (33%) in the CG. In terms of intrinsic motivation, the IM+ subgroup consisted of 24 students (75%) in the EG and 16 students (53%) in the CG. In contrast, the IM- subgroup included 8 students (25%) in the EG and 14 students (47%) in the CG.

The results indicated that the PT-PBL approach increased students’ extrinsic motivation more significantly than the C-PBL approach. Although it did not significantly improve students’ intrinsic motivation, it led to an increase in the proportion of students with improved motivation.

Response to RQ 2 To what extent does the PT-PBL approach improve the students’ reading performance in comparison with the C-PBL approach? A one-way ANCOVA was conducted to examine the impact of different learning approaches on students’ reading performance. The groups employing different learning approaches (EG and CG) were considered independent variables, while the post-test and pre-test reading performance were the dependent and covariate variables, respectively. The assumption of homogeneity of regression coefficients for total ($F = 0.205$, $p = 0.652$), explicit ($F = 1.841$, $p = 0.180$), and implicit questions ($F = 0.628$, $p = 0.431$) scores was confirmed. Table 2 shows the ANCOVA results.

For the total score, the adjusted means for the EG and CG were 80.18 and 75.34, respectively. For the implicit questions, the adjusted means for the EG and CG were 45.73 and 41.92, respectively. Additionally, significant differences were found between the EG and CG for both total scores ($F = 11.850$,

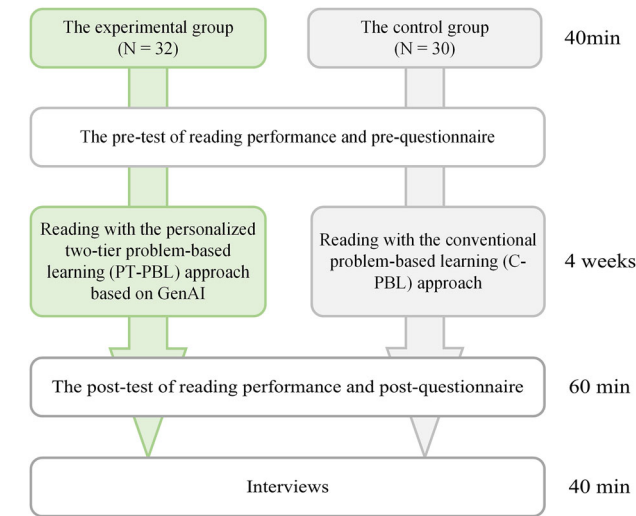


Fig. 5 Experimental procedures. It presents the experimental procedures used in this study.

Table 1 ANCOVA results for learning motivation.									
Variance	Group	N	Mean	SD	Adjusted mean	Std. error	F	p	η^2
Extrinsic motivation	EG	32	3.58	0.46	3.57	0.08	5.967	0.018	0.092
	CG	30	3.27	0.48	3.28	0.09			
Intrinsic motivation	EG	32	2.88	0.45	2.88	0.07	1.410	0.240	0.023
	CG	30	2.76	0.41	2.76	0.08			

Table 2 ANCOVA results for reading performance.

Variance	Group	N	Mean	SD	Adjusted mean	Std. error	F	p	η^2
Total scores	EG	32	80.09	5.03	80.18	0.98	11.850	0.001	0.167
	CG	30	75.43	6.25	75.34	1.01			
Explicit questions	EG	32	34.03	2.52	34.07	0.41	2.731	0.104	0.044
	CG	30	33.13	2.11	33.09	0.42			
Implicit questions	EG	32	45.69	3.35	45.73	0.70	14.130	0.000	0.193
	CG	30	41.97	4.85	41.92	0.73			

Table 3 ANCOVA results for learning records and engagement questionnaire scores.

Variance	Group	N	Mean	SD	Adjusted mean	Std. error	F	p	η^2
Total reading time	EG	32	73.60	16.55	73.73	3.44	13.501	0.000	0.186
	CG	30	55.66	21.97	55.53	3.55			
Total pages read	EG	32	27.59	4.08	27.69	0.48	9.115	0.004	0.134
	CG	30	25.70	4.57	25.60	0.50			
Task completion rate	EG	32	91%	0.08	92%	0.01	22.725	0.000	0.278
	CG	30	89%	0.07	88%	0.01			
Login frequency	EG	32	4.93	2.36	4.71	0.18	8.420	0.005	0.125
	CG	30	3.66	1.72	3.91	0.19			
Temporarily away	EG	32	0.87	0.94	0.93	0.08	16.603	0.000	0.220
	CG	30	1.46	1.13	1.41	0.08			
Engagement questionnaire scores	EG	32	3.46	0.32	3.46	0.64	11.337	0.001	0.161
	CG	30	3.15	0.40	3.15	0.66			

$p = 0.001$, $\eta^2 = 0.167$) and implicit questions ($F = 14.130$, $p = 0.000$, $\eta^2 = 0.193$). However, there was no significant difference between the EG and CG in terms of explicit questions.

The results indicated that the PT-PBL approach significantly improved students' reading performance more than the C-PBL approach, especially in terms of total and implicit question scores.

Response to RQ 3 What are the differences in reading engagement among students using different learning approaches, and how does varying engagement affect their reading performance? A one-way ANCOVA was conducted to test the impact of different learning approaches on students' reading engagement. The groups employing different learning approaches (EG and CG) were considered independent variables, while the learning records in week 4 and learning records in week 1 were the dependent and covariate variables, respectively. Furthermore, the pre-test and post-test scores of the engagement questionnaire were the dependent variables and covariates, respectively. The assumption of homogeneity of regression coefficients for the learning records and engagement questionnaire scores was confirmed. Table 3 shows the ANCOVA results.

In terms of total reading time, the adjusted means for the EG and CG were 73.73 and 55.53, respectively. A significant difference was found between the EG and CG in total reading time ($F = 13.501$, $p = 0.000$, $\eta^2 = 0.186$). In terms of total pages read, the adjusted means for the EG and CG were 27.69 and 25.60, respectively. A significant difference was found between the EG and CG in total pages read ($F = 9.115$, $p = 0.004$, $\eta^2 = 0.134$). In terms of task completion rate, the EG showed a significantly higher completion rate compared to the CG ($F = 22.725$, $p = 0.000$, $\eta^2 = 0.278$). The EG also logged in significantly more frequently than the CG ($F = 8.420$, $p = 0.005$, $\eta^2 = 0.125$). Lastly, the EG showed significantly fewer instances of being temporarily away from the learning platform compared to the CG ($F = 16.603$, $p = 0.000$, $\eta^2 = 0.220$). In terms of engagement

questionnaire scores, the adjusted means for the EG and CG were 3.46 and 3.15, respectively. A significant difference was found between the EG and CG in engagement questionnaire scores ($F = 11.337$, $p = 0.001$, $\eta^2 = 0.161$).

Additionally, Fig. 6 displays the trends in learning records over the 4-week period for both groups. As shown in the upper half of Fig. 6, the EG demonstrated a steady increase in total reading time and total pages read across the 4 weeks. Total reading time progressed from 28.43 min in Week 1 to 73.60 min in Week 4, while total pages read increased from 24.34 to 27.59 pages. In contrast, the CG showed a slower improvement, with total reading time rising from 32.04 min in Week 1 to 55.66 min in Week 4, and total pages read fluctuating from 24.60 pages to 25.70 pages. The lower-left part of Fig. 6 illustrates the trends in task completion rate and login frequency. The EG showed a consistent increase in task completion rate, from 58% in Week 1 to 91% in Week 4, while the CG increased from 64% in Week 1 to 89% in Week 4. In terms of login frequency, the EG logged in more frequently, increasing from 1.65 logins per week in Week 1 to 4.93 logins per week in Week 4. In comparison, the CG logged in from 1.50 times per week in Week 1 to 3.66 times per week in Week 4. Finally, the lower-right part of Fig. 6 demonstrates the instances of students being temporarily away from the learning platform. The EG showed a significant reduction in instances of being temporarily away, from 3.03 instances per week in Week 1 to 0.87 in Week 4. The CG, on the other hand, showed a more gradual decrease, from 3.23 instances in Week 1 to 1.46 in Week 4.

Therefore, by analyzing the learning records and the engagement questionnaire, it can be concluded that students using the PT-PBL approach increased reading engagement more than those using the C-PBL approach.

In addition, a two-factor ANOVA was conducted to investigate the differences in reading performance between students using different learning approaches and different levels of engagement. The different learning approaches and different levels of engagement were considered independent variables, while the pre- and post-test results for reading performance were the



Fig. 6 Four-week trends of learning records. It displays the trends in learning records of the experimental and control groups over four weeks, including total reading time, total pages read, task completion rate, login frequency, and temporarily away.

Table 4 Two-way ANOVA results of different reading engagement levels and pre- and post-reading performance in the EG.													
Group	Variance	RE	N	Pre-test					Post-test				
				Mean	SD	F	p	η^2	Mean	SD	F	p	η^2
EG	Total scores	High	16	65.00	6.94	3.401	0.075	0.102	82.44	3.29	8.660	0.006	0.224
		Low	16	68.88	4.75				77.75	5.46			
	Explicit questions	High	16	28.81	3.41	2.584	0.118	0.079	35.00	1.79	5.397	0.027	0.152
		Low	16	30.50	2.45				33.06	2.82			
	Implicit questions	High	16	36.19	3.87	3.317	0.079	0.100	47.44	2.50	11.719	0.002	0.281
		Low	16	38.37	2.85				43.94	3.23			

RE reading engagement.

dependent variables. There were no significant interaction effects of learning approach and engagement on any of the three dimensions of students' pre- ($F = 0.412$, $p = 0.524$; $F = 0.026$, $p = 0.872$; $F = 1.326$, $p = 0.254$) and post-test ($F = 0.653$, $p = 0.422$; $F = 3.250$, $p = 0.077$; $F = 0.014$, $p = 0.905$) reading performance. Thus, main effects comparisons were conducted.

As shown in Table 4, in the EG, there was no significant difference in the pre-test reading performance of students with different levels of engagement on the dimensions of total, explicit and implicit questions ($F = 3.401$, $p = 0.075$; $F = 2.584$, $p = 0.118$; $F = 3.317$, $p = 0.079$), while there were significant differences in the post-test reading performance of students with different levels of engagement on the dimensions of total, explicit and implicit questions ($F = 8.660$, $p = 0.006$, $\eta^2 = 0.224$; $F = 5.397$, $p = 0.027$, $\eta^2 = 0.152$; $F = 11.719$, $p = 0.002$, $\eta^2 = 0.281$). The results indicate that among students using the PT-PBL approach, those with high levels of engagement tended to perform better than those with low levels of engagement on the dimensions of total score, explicit, and implicit questions.

As shown in Table 5, in the CG, there were no significant differences between students with different levels of engagement on the total score, implicit questions, and explicit questions dimensions of reading performance in both the pre-test ($F = 0.437$, $p = 0.514$; $F = 3.074$, $p = 0.090$; $F = 0.014$, $p = 0.908$) and post-test ($F = 1.175$, $p = 0.288$; $F = 0.029$, $p = 0.866$; $F = 3.729$, $p = 0.064$). The results indicate that there were no significant changes in the performance of students with different levels of engagement on the three dimensions of total scores, implicit, and explicit questions among students using the C-PBL approach.

Figure 7 demonstrates the effects of using different learning approaches and different levels of engagement on students' total scores in reading performance. It can be found that the gap between the pre-test and post-test reading performance of students with different levels of engagement in the EG was larger than that in the CG. The results suggest that the PT-PBL approach may be more effective for students with high engagement and less effective for students with low engagement.

Table 5 Two-way ANOVA results of different reading engagement levels and pre- and post-reading performance in the CG.													
Group	Variance	RE	N	Pre-test					Post-test				
				Mean	SD	F	p	η^2	Mean	SD	F	p	η^2
CG	Total scores	High	15	66.93	7.78	0.437	0.514	0.015	76.67	6.57	1.175	0.288	0.040
		Low	15	68.67	6.52				74.20	5.88			
	Explicit questions	High	15	29.27	3.67	3.074	0.090	0.099	33.07	2.19	0.029	0.866	0.001
		Low	15	31.20	2.18				33.20	2.11			
	Implicit questions	High	15	37.67	4.39	0.014	0.908	0.000	43.60	4.66	3.729	0.064	0.118
		Low	15	37.47	4.99				40.33	4.61			

RE reading engagement.

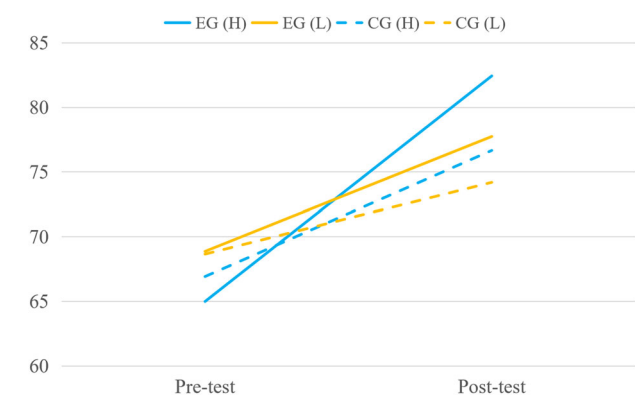


Fig. 7 The gap between the pre-test and post-test reading performance of students with different levels of engagement. It shows the impact of different learning methods and levels of engagement on students' overall reading performance. H refers to high engagement; L refers to low engagement.

Interview results. Semi-structured interviews were analyzed using NVivo 11.4 to understand students' perceptions regarding different approaches. Table 6 shows the analysis results from the interviews, including the themes, coding items, and frequency of mentions. Overall, in the interview results, students in the EG mentioned these aspects more than the CG across the four aspects of supporting personalized learning, promoting reflection, improving learning performance, and increasing learning motivation. Conversely, the CG provided more suggestions for improvement.

In terms of "supporting personalized learning", the EG (n = 22) mentioned it much more frequently than the CG (n = 7). Students in the EG felt that the problems they encountered were more in line with their cognitive level, especially in the self-directed learning phase. In addition, the feedback they received helped them to overcome difficulties.

For example, a student from the EG said, "Compared to traditional learning approaches, this approach presented problems that better matched my cognitive level. Additionally, I received more timely and targeted feedback, which truly helped to improve my problem-solving skills."

Regarding "prompting reflection", the EG (n = 18) mentioned it slightly more often than the CG (n = 12). In the first phase, both groups of students felt that their reflective thinking was enhanced during discussions and interactions with others (e.g., peers, teachers). In the second phase, students in the EG received more detailed and targeted feedback through interactions with ChatGPT, and thus they were more inclined to reflect on and optimize their solutions more frequently.

For instance, an EG student stated, "In the first phase, I thought my solution was pretty good, but after receiving the system's targeted guidance in the second phase, I realized that I could further optimize it."

With regard to "improving learning performance", the EG (n = 24) mentioned it much more frequently than the CG (n = 12). Students in the EG felt that this learning approach promoted in-depth comprehension of the reading material and critical thinking, whereas students in the CG rarely reported on these two aspects.

One EG student said, "Compared to traditional learning approaches, this approach helps me understand characters in the story from different perspectives and allows me to analyse their traits more deeply."

In terms of "increasing learning motivation", it was mentioned more often in the EG (n = 27) than in the CG (n = 7). Students in the EG were inclined to report that this learning approach promoted active reading and eagerness for the next stage of the task, whereas students in the CG seldom reported on any of these aspects.

An EG student stated, "Not only did I not find these problems boring, but I also thought they were quite interesting. I'm looking forward to solving more of them, which gives me a sense of accomplishment."

As for "suggestions for improvement", the CG (n = 18) reported more than the EG (n = 7). Students in the CG felt that the difficulty of the problems was not appropriate (too difficult or too easy). In addition, they reported that the feedback they received was not timely or helpful, whereas the EG students rarely mentioned these aspects.

For instance, a student from the CG said, "Initially, the problems were relatively easy, but later on, they became quite challenging for me. Without timely guidance, I found it difficult to come up with good solutions."

Discussion

Learning motivation. The results showed that students in the EG (PT-PBL approach) were more motivated to learn than students in the CG (C-PBL approach). There may be several reasons for this difference. First, in the PT-PBL approach, the PBL activity offers reading problems based on the students' cognitive level, which ensures that the problems are within a reasonable range of difficulty. This appropriate level of challenge makes students believe that they can independently solve the problem and develop a solution through their efforts. Self-determination theory suggests that a person's motivation to learn is significantly enhanced when they believe they are capable of accomplishing the task in the activity (Reeve 2012). Additionally, after forming a

Table 6 Themes, coding items, and number of occurrences of interview results.

Themes	Coding items	The number of time mentioned	
		EG	CG
Supporting personalized learning	The difficulty of the problem is appropriate.	12	3
	The feedback received is precisely helpful in solving the problem.	10	4
Promoting reflection	Reflect on their understanding of the problem through interaction with others.	10	7
	Promote reflection and optimization of solutions.	8	5
Improving learning performance	This learning approach promotes a deeper comprehension of the material.	11	5
	This learning approach promotes reading critical thinking.	8	2
Increasing learning motivation	This learning approach improves reading scores.	5	5
	This learning approach promotes active reading for me.	12	1
Suggestions for improvement	The interaction in this learning approach makes me look forward to the next tasks.	8	2
	This learning approach is interesting.	7	6
	None	4	6
	The difficulty of the problem is not appropriate.	2	8
	The quality of the content of the feedback is poor.	1	4

solution, students receive personalized feedback through their interactions with ChatGPT. Several studies have shown that using GenAI to offer personalized feedback based on students' performance and task difficulty can help motivate them to learn (Barrot 2023; Su et al. 2023; Wang et al. 2024). Compared with the CG, the feedback received in the EG was more timely and targeted, which helped students better understand their learning, refine their solutions, and enhance their confidence and sense of accomplishment, thereby increasing their motivation to learn. Combined with the results of the interviews, it was found that students in the EG reported more frequently that they were able to read more actively and were eagerly anticipating the next stage of the task. This further confirms the effectiveness of the PT-PBL approach in improving students' motivation.

Reading performance. The results indicated that students in the EG (PT-PBL approach) significantly outperformed students in the CG (C-PBL approach) in reading performance. This can be attributed to the effectiveness of the PT-PBL approach in facilitating learning. Specifically, the PT-PBL approach provides a more fine-grained design of PBL activities and leverages the guiding role of PBL in learning. Starting with the easier problems in the first tier, students ensure they grasp the fundamental concepts before completing the more challenging problems in the second tier. This step-by-step approach helps to enhance students' comprehension of the material (Roll et al. 2018; Roller and Zori 2017). This result is consistent with Aslan (2021), who emphasized the positive impact of incorporating a PBL approach into learning activities. Second, the PT-PBL approach supported students' personalized learning. In contrast to the C-PBL approach, the PT-PBL approach utilizes GenAI to provide students with personalized problems that match their cognitive level. Previous studies have also shown that GenAI-supported personalized learning approaches benefit students' learning outcomes (Abdelhalim 2024; Chen et al. 2022a; El-Sabagh 2021; Tzeng et al. 2024). Third, the PT-PBL approach offers students timely and targeted feedback, which helps students retrieve relevant information or experiences from long-term memory (Li et al. 2023). Interview results showed that students in the EG were inclined to reflect and optimize their solutions after receiving feedback. In addition, they attempted to make connections between new knowledge and existing experiences to achieve deeper comprehension of the material. The experimental results of the implicit questions demonstrated that the PT-PBL approach improved reading performance. The interview results provide additional

evidence that the approach helps to improve reading performance, especially in the areas of deep comprehension and critical thinking.

However, the PT-PBL approach did not significantly improve students' scores on explicit questions in reading performance. A plausible explanation is that explicit questions examine students' direct observation skills. Nevertheless, the PT-PBL approach only offers text-based learning materials during students' reading process. Relevant studies have shown that providing students with more diverse learning materials can improve their observational and perceptual abilities (Chen et al. 2022b; Li et al. 2023). Therefore, educators can explore the integration of multimodal data into the PT-PBL approach in the future to enhance students' observation and perception of learning materials.

Reading engagement and its effect on reading performance.

The results showed that students' reading engagement in the EG was significantly higher than that of students in the CG. Moreover, a significant difference in reading performance emerged between the EG's low- and high-engagement students, whereas no significant difference was observed in the CG. The performance gap between low- and high-engagement students was more pronounced in the EG than in the CG. These results indicate that the PT-PBL approach is more effective in improving the reading performance of highly engaged students than it is for those with low engagement. This variation could stem from the differing capacities of students with varying levels of engagement in sustaining attention during learning tasks (Li et al. 2023). Previous research has indicated that highly engaged students demonstrate increased focus and motivation throughout the learning process (Li et al. 2023). Consequently, they are more likely to be actively involved in learning to comprehend the materials better. Additionally, the PT-PBL approach offers more personalized and challenging tasks during the second tier of PBL activities. According to Van Ammel et al. (2021) and Ronimus et al. (2022), highly engaged students often perceive reading to be enjoyable. They may exert more mental effort when performing challenging reading tasks, which helps them to comprehend the materials. Therefore, students with high levels of engagement are more likely to persist and overcome difficulties when faced with challenging tasks. In contrast, students with low levels of engagement may lack motivation or become distracted and ultimately choose to give up, resulting in poor learning outcomes when faced with challenges. The interview results further supported this conclusion, as highly engaged students expressed

eagerness to tackle challenges in the second tier of PBL activities, which fostered a sense of achievement.

However, low-engagement students in both groups performed nearly identically. This result suggests that the PT-PBL approach failed to improve the reading performance of low-engagement students, and various reasons may account for this outcome. First, owing to the time constraints in the experiment, the enduring effect on low-engagement students needed to be more sufficient to cause significant alterations in their task engagement. Previous research has also suggested that brief interventions often fail to enhance the learning performance of students with low engagement (Yang et al. 2021; Li et al. 2023). Second, despite the personalized problems and feedback offered by the PT-PBL approach, it could have been more effective in benefiting low-engagement students. Interview insights revealed that the negative behaviors of low-engagement students mainly occurred in the second tier of PBL activities. This phenomenon was mainly due to the increased difficulty and complexity of the problems in this tier. Low-engagement students face significant challenges in handling these complex problems because of their limited attention span and patience, which prevents them from reaching the same level of understanding as their highly engaged peers (Jin et al. 2022). Consequently, regardless of whether the PT-PBL or C-PBL approach was used, low-engagement students encountered significant difficulties in achieving excellent reading performance.

Conclusions

In this study, a GenAI-based personalized two-tier PBL approach was proposed. The results of the quasi-experiment revealed that the proposed approach improved students' reading performance, motivation, and engagement more than the C-PBL approach. The interviews provided additional evidence to explain the effectiveness of the proposed approach. Although several previous studies have reported the potential of PBL in reading education, few of them have attempted a more refined design of PBL activities to help students think and comprehend deeply in reading, let alone integrate GenAI into the PBL approach to provide personalized support for students' reading process. The results of the present study provide a new approach that can inspire researchers and teachers to use GenAI for studies and instructional designs in the future.

Although the results of this study have shown the positive effects of integrating GenAI into PBL-based reading activities, there are several limitations and drawbacks. The first was the short duration of the experiment and the relatively small sample size, which could limit the generalizability of the findings. Additionally, learning outcomes between the pre-test and post-test might have been influenced by external factors, such as students' personal study habits or the use of external resources. Although measures such as consistent use of the same learning system, close monitoring of learning activities, and carefully designed pre-test and post-test evaluations were implemented to minimize these influences, these factors could not be entirely eliminated. Moreover, the study primarily utilized engagement indicators such as total reading time and total pages read to evaluate students' behaviors during the intervention period. While these metrics provide useful insights, they may not fully capture the complexity of learning outcomes. For instance, more nuanced patterns such as problem-solving sequences, interactions with specific reading materials, or responses to feedback were not analyzed, potentially leaving out key aspects of how students engage with learning tasks and how these behaviors influence learning outcomes. There was also uncertainty in the GenAI-generated content, especially as the feedback and guidance provided after students engaged in long conversations might not

have been entirely satisfactory. Finally, the PT-PBL approach constructs reading contexts based on real-life problems but lacks multimodal learning materials (e.g., pictures). As a result, students cannot engage with and process information through multiple sensory channels, which may make it difficult for them to understand the materials.

Therefore, future studies should consider increasing the sample size and extending the duration of the experiment to better control for external factors and enhance the robustness of the findings. Additionally, strategies to monitor and mitigate external influences, such as detailed tracking of students' independent study activities, could provide more comprehensive insights into the effectiveness of the PT-PBL approach. Furthermore, expanding the analysis to include more nuanced behavioral patterns, such as problem-solving sequences, engagement in feedback loops, or interactions with specific reading materials, would offer a more holistic understanding of learning processes and their impact on learning outcomes. To address the uncertainty in GenAI-generated content, integrating complementary technologies, such as educational knowledge graphs, could improve the precision and relevance of feedback and guidance. Finally, incorporating multimodal learning content (e.g., images, audio, or video) into PT-PBL activities would cater to diverse learning styles, providing students with a richer and more authentic learning experience that enhances their motivation and comprehension.

Data availability

The datasets generated and/or analyzed during this study are not publicly available due to general data protection regulations, but are available from the corresponding authors on reasonable request.

Received: 10 August 2024; Accepted: 23 April 2025;

Published online: 09 May 2025

References

- Abdelhalim SM (2024) Using ChatGPT to promote research competency: English as a foreign language undergraduates' perceptions and practices across varied metacognitive awareness levels. *J Comput Assist Learn* 40(3):1261–1275. <https://doi.org/10.1111/jcal.12948>
- Alrawashdeh GS, Fyffe S, Azevedo RF et al. (2024) Exploring the impact of personalized and adaptive learning technologies on reading literacy: a global meta-analysis. *Educ Res Rev* 42:100587. <https://doi.org/10.1016/j.edurev.2023.100587>
- Anggraeni DM, Prahani B, Suprpto N et al. (2023) Systematic review of problem based learning research in fostering critical thinking skills. *Think Skills Creat* 49:101334. <https://doi.org/10.1016/j.tsc.2023.101334>
- Arani SMN, Zarei AA, Sarani A (2023) Problem-based language learning: Why aren't teachers using it? *Soc Sci Hum Open* 8(1):100668. <https://doi.org/10.1016/j.ssaho.2023.100668>
- Arici F, Yilmaz M (2023) An examination of the effectiveness of problem-based learning method supported by augmented reality in science education. *J Comput Assist Learn* 39(2):446–476. <https://doi.org/10.1111/jcal.12752>
- Aslan A (2021) Problem-based learning in live online classes: learning achievement, problem-solving skill, communication skill, and interaction. *Comput Educ* 171:104237. <https://doi.org/10.1016/j.compedu.2021.104237>
- Barrot JS (2023) Using ChatGPT for second language writing: pitfalls and potentials. *Assess Writ* 57:100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Barrows HS (1986) A taxonomy of problem-based learning methods. *Med Educ* 20(6):481–486. <https://doi.org/10.1111/j.1365-2923.1986.tb01386.x>
- Bloom BS, Engelhart MD, Furst EJ et al. (1956) Taxonomy of educational objectives: the classification of educational goals. In: *Handbook I: cognitive domain*. David McKay, New York
- Chan CKY, Hu W (2023) Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ* 20:43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chang CS, Chung CH, Chang JA (2020) Influence of problem-based learning games on effective computer programming learning in higher education.

- Educ Technol Res Dev 68:2615–2634. <https://doi.org/10.1007/s11423-020-09784-3>
- Chen CH, Hung HT, Yeh HC (2021) Virtual reality in problem-based learning contexts: effects on the problem-solving performance, vocabulary acquisition and motivation of English language learners. *J Comput Assist Learn* 37(3):851–860. <https://doi.org/10.1111/jcal.12528>
- Chen X, Zou D, Xie H et al. (2022a) Two decades of artificial intelligence in education: contributors, collaborations, research topics, challenges, and future directions. *Educ Technol Soc* 25(1):28–47. <https://www.jstor.org/stable/48647028>
- Chen YT, Li M, Huang CQ et al. (2022b) Promoting deep writing with immersive technologies: an SVVR-supported chinese composition writing approach for primary schools. *Br J Educ Technol* 53(6):2071–2091. <https://doi.org/10.1111/bjet.13247>
- Clinton-Lisell V, Seipel B, Gilpin S et al. (2023) Interactive features of e-texts' effects on learning: a systematic review and meta-analysis. *Interact Learn Envir* 31(6):3728–3743. <https://doi.org/10.1080/10494820.2021.1943453>
- Cohen J (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 70(4):213–220
- Corlatescu DG, Watanabe M, Ruseti S et al. (2024) The automated model of comprehension version 4.0—validation studies and integration of ChatGPT. *Comput Human Behav* 154:108154. <https://doi.org/10.1016/j.chb.2024.108154>
- Danaei D, Jamali HR, Mansourian Y et al. (2020) Comparing reading comprehension between children reading augmented reality and print storybooks. *Comput Educ* 153:103900. <https://doi.org/10.1016/j.compedu.2020.103900>
- Ding L, Li T, Jiang S et al. (2023) Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *Int J Educ Technol High Educ* 20:63. <https://doi.org/10.1186/s41239-023-00434-1>
- Donnelly R (2010) Harmonizing technology with interaction in blended problem-based learning. *Comput Educ* 54(2):350–359
- Edens KM (2000) Preparing problem solvers for the 21st century through problem-based learning. *Coll Teach* 48(2):55–60. <https://doi.org/10.1080/87567550009595813>
- Egert F, Cordes AK, Hartig F (2022) Can e-books foster child language? Meta-analysis on the effectiveness of e-book interventions in early childhood education and care. *Educ Res Rev* 37:100472. <https://doi.org/10.1016/j.edurev.2022.100472>
- El-Sabagh HA (2021) Adaptive e-learning environment based on learning styles and its impact on development students' engagement. *Int J Educ Technol High Educ* 18:53. <https://doi.org/10.1186/s41239-021-00289-4>
- Fidan M, Tuncel M (2019) Integrating augmented reality into problem based learning: the effects on learning achievement and attitude in physics education. *Comput Educ* 142:103635. <https://doi.org/10.1016/j.compedu.2019.103635>
- Fung JTC, Chan SL, Takemura N et al. (2023) Virtual simulation and problem-based learning enhance perceived clinical and cultural competence of nursing students in Asia: a randomized controlled cross-over study. *Nurse Educ Today* 123:105721. <https://doi.org/10.1016/j.nedt.2023.105721>
- Gambrell L, Marinak B (2009) Reading motivation: What the research says. Retrieved from <http://www.readingrockets.org/article/29624>
- Gao X, Wang L, Deng J et al. (2022) The effect of the problem based learning teaching model combined with mind mapping on nursing teaching: a meta-analysis. *Nurse Educ Today* 111:105306. <https://doi.org/10.1016/j.nedt.2022.105306>
- Guo K, Li D (2024) Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: a mixed methods study. *System* 124:103362. <https://doi.org/10.1016/j.system.2024.103362>
- Hmelo-Silver CE (2004) Problem-based learning: What and how do students learn? *Educ Psychol Rev* 16:235–266
- Hsu CK, Hwang GJ, Chang CK (2013) A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Comput Educ* 63:327–336. <https://doi.org/10.1016/j.compedu.2012.12.004>
- Huang AY, Lu OH, Yang SJ (2023) Effects of artificial Intelligence-Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Comput Educ* 194:104684. <https://doi.org/10.1016/j.compedu.2022.104684>
- Jeon J, Lee S, Choe H (2023) Beyond ChatGPT: a conceptual framework and systematic review of speech-recognition chatbots for language learning. *Comput Educ* 206:104898. <https://doi.org/10.1016/j.compedu.2023.104898>
- Jin X, Jiang Q, Xiong W et al. (2022) Effects of student engagement in peer feedback on writing performance in higher education. *Interact Learn Envir* 32(1):128–143. <https://doi.org/10.1080/10494820.2022.2081209>
- Kambara H, Lin YC, Wen H et al. (2023) Investigating American and Chinese college students' reading motivation: a cross-cultural study. *Curr Psychol* 43(12):11330–11342. <https://doi.org/10.1007/s12144-023-04707-w>
- Kasneci E, Seffler K, Küchemann S et al. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ* 103:102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kong Y, Seo YS, Zhai L (2018) Comparison of reading performance on screen and on paper: a meta-analysis. *Comput Educ* 123:138–149. <https://doi.org/10.1016/j.compedu.2018.05.005>
- Lee HY, Chen PH, Wang WS et al. (2024) Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *Int J Educ Technol High Educ* 21:16. <https://doi.org/10.1186/s41239-024-00447-4>
- Lee YH (2023) Achieving success in English medium instruction using computer-mediated terminology pretraining under the problem-based learning pedagogy. *J Comput Assist Learn* 39(3):921–934. <https://doi.org/10.1111/jcal.12777>
- Li H, Majumdar R, Chen MRA et al. (2021) Goal-oriented active learning (GOAL) system to promote reading engagement, self-directed learning behaviour, and motivation in extensive reading. *Comput Educ* 171:104239. <https://doi.org/10.1016/j.compedu.2021.104239>
- Li M, Chen YT, Huang CQ et al. (2023) From motivational experience to creative writing: a motivational AR-based learning approach to promoting Chinese writing performance and positive writing behaviours. *Comput Educ* 202:104844. <https://doi.org/10.1016/j.compedu.2023.104844>
- Lin JW, Tsai CW, Hsu CC et al. (2019) Peer assessment with group awareness tools and effects on project-based learning. *Interact Learn Envir* 29(4):583–599. <https://doi.org/10.1080/10494820.2019.1593198>
- Lin LF (2018) Integrating the problem-based learning approach into a web-based English reading course. *J Educ Comput Res* 56(1):105–133. <https://doi.org/10.1177/0735633117705960>
- Lin YN, Hsia LH, Hwang GJ (2021) Promoting pre-class guidance and in-class reflection: a SQIRC-based mobile flipped learning approach to promoting students' billiards skills, strategies, motivation and self-efficacy. *Comput Educ* 160:104035. <https://doi.org/10.1016/j.compedu.2020.104035>
- Lin Z, Chen H (2024) Investigating the capability of ChatGPT for generating multiple-choice reading comprehension items. *System* 123:103344. <https://doi.org/10.1016/j.system.2024.103344>
- Liu M, Zhang LJ, Biebricher C (2024) Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Comput Educ* 211:104977. <https://doi.org/10.1016/j.compedu.2023.104977>
- Liu S, Xue S, Wu J et al. (2021) Online active learning for drifting data streams. *IEEE Trans Neural Netw Learn Syst* 34(1):186–200. <https://doi.org/10.1109/TNNLS.2021.3091681>
- Liu Y, Pásztor A (2022) Effects of problem-based learning instructional intervention on critical thinking in higher education: a meta-analysis. *Think Skills Creat* 45:101069. <https://doi.org/10.1016/j.tsc.2022.101069>
- Ma X, Huo Y (2023) Are users willing to embrace ChatGPT? Exploring the factors on the acceptance of chatbots from the perspective of AIDUA framework. *Technol Soc* 75:102362. <https://doi.org/10.1016/j.techsoc.2023.102362>
- Meniado JC, Huyen DTT, Panyadilokpong N et al. (2024) Using ChatGPT for second language writing: experiences and perceptions of EFL learners in Thailand and Vietnam. *Comput Educ Artif Intell* 7:100313. <https://doi.org/10.1016/j.caeai.2024.100313>
- Neville DO, Britt DW (2007) A problem-based learning approach to integrating foreign language into engineering. *Foreign Lang Ann* 40(2):226–246. <https://doi.org/10.1111/j.1944-9720.2007.tb03199.x>
- Orhan A (2024) Online or in-class problem based learning: Which one is more effective in enhancing learning outcomes and critical thinking in higher education EFL classroom? *J Comput Assist Learn* 40(5):2351–2368. <https://doi.org/10.1111/jcal.13033>
- Reeve J (2012) A self-determination theory perspective on student engagement. In: *Handbook of research on student engagement*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-2018-7_7
- Rodríguez MF, Nussbaum M, Yunis L et al. (2022) Using scaffolded feedforward and peer feedback to improve problem-based learning in large classes. *Comput Educ* 182:104446. <https://doi.org/10.1016/j.compedu.2022.104446>
- Roll I, Butler D, Yee N et al. (2018) Understanding the impact of guiding inquiry: the relationship between directive support, student attributes, and transfer of knowledge, attitudes, and behaviours in inquiry learning. *Instr Sci* 46(1):77–104. <https://doi.org/10.1007/s11251-017-9437-x>
- Roller MC, Zori S (2017) The impact of instituting process-oriented guided-inquiry learning (POGIL) in a fundamental nursing course. *Nurse Educ Today* 50:72–76. <https://doi.org/10.1016/j.nedt.2016.12.003>
- Ronimus M, Kujala J, Tolvanen A et al. (2014) Children's engagement during digital game-based learning of reading: the effects of time, rewards, and challenge. *Comput Educ* 71:237–246. <https://doi.org/10.1016/j.compedu.2013.10.008>
- Ronimus M, Tolvanen A, Hautala J (2022) The roles of motivation and engagement in computer-based assessment of children's reading comprehension. *Learn Individ Differ* 98:102197. <https://doi.org/10.1016/j.lindif.2022.102197>

- Savva M, Higgins S, Beckmann N (2022) Meta-analysis examining the effects of electronic storybooks on language and literacy outcomes for children in grades Pre-K to grade 2. *J Comput Assist Learn* 38(2):526–564. <https://doi.org/10.1111/jcal.12623>
- Spörer N, Brunstein JC, Kieschke ULF (2009) Improving students' reading comprehension skills: effects of strategy instruction and reciprocal teaching. *Learn Instr* 19(3):272–286. <https://doi.org/10.1016/j.learninstruc.2008.05.003>
- Steiss J, Tate T, Graham S et al. (2024) Comparing the quality of human and ChatGPT feedback of students' writing. *Learn Instr* 91:101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Su Y, Lin Y, Lai C (2023) Collaborating with ChatGPT in argumentative writing classrooms. *Assess Writ* 57:100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Sun FR, Pan LF, Wan RG et al. (2021) Detecting the effect of student engagement in an SVVR school-based course on higher level competence development in elementary schools by SEM. *Interact Learn Envir* 29(1):3–16. <https://doi.org/10.1080/10494820.2018.1558258>
- Sun JCY, Tsai HE, Cheng WKR (2023) Effects of integrating an open learner model with AI-enabled visualization on students' self-regulation strategies usage and behavioral patterns in an online research ethics course. *Comput Educ Artif Intell* 4:100120. <https://doi.org/10.1016/j.caeai.2022.100120>
- Tzeng JW, Huang NF, Chen YH et al. (2024) Personal learning material recommendation system for MOOCs based on the LSTM neural network. *Educ Technol Soc* 27(2):25–42. <https://www.jstor.org/stable/48766161>
- Van Ammel K, Aesaert K, De-Smedt F et al. (2021) Skill or will? The respective contribution of motivational and behavioural characteristics to secondary school students' reading comprehension. *J Res Read* 44(3):574–596. <https://doi.org/10.1111/1467-9817.12356>
- Wang X, Zhong Y, Huang C et al. (2024) ChatPRCS: a personalized support system for English reading comprehension based on ChatGPT. *IEEE Trans Learn Technol* 17:1762–1776. <https://doi.org/10.1109/TLT.2024.3405747>
- Xiao Y, Hew KF (2024) Personalized gamification versus one-size-fits-all gamification in fully online learning: effects on student motivational, behavioral and cognitive outcomes. *Learn Individ Differ* 113:102470. <https://doi.org/10.1016/j.lindif.2024.102470>
- Yang G, Chen YT, Zheng XL et al. (2021) From experiencing to expressing: a virtual reality approach to facilitating pupils' descriptive paper writing performance and learning behaviour engagement. *Br J Educ Technol* 52(2):807–823. <https://doi.org/10.1111/bjet.13056>
- Yu J, Zhou X, Yang X et al. (2022) Mobile-assisted or paper-based? The influence of the reading medium on the reading comprehension of English as a foreign language. *Comput Assist Lang L* 35(1-2):217–245. <https://doi.org/10.1080/09588221.2021.2012200>
- Zhao X, Cox A, Cai L (2024) ChatGPT and the digitisation of writing. *Hum Soc Sci Commun* 11:482. <https://doi.org/10.1057/s41599-024-02904-x>

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62377043 and 62337001) and by the Key Project of Philosophy and Social Science Research of the Ministry of Education of China (Grant No. 24JZD011).

Author contributions

Conceptualization: CH, YZ, YL, XW; methodology: CH, YZ, YL, XW; data curation: YZ, XW; formal analysis: YZ, ZH, DZ; writing—original draft preparation: YZ; writing—review and editing: CH, YZ, YL, XW, ZH, DZ, ML.

Competing interests

The authors declare no competing interests.

Ethical approval

This research was approved by the Human Experimentation Ethics Committee of Zhejiang Normal University (Approval No. ZSRT2024040), with approval granted on March 3rd, 2024. The study was designed and conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and its later amendments.

Informed consent

Informed consent was obtained from all participants on March 6th, 2024, prior to their participation in the survey. Participants were informed about the purpose of the study, the voluntary nature of their participation, and their right to withdraw at any time without consequence. Confidentiality and data protection were emphasized, assuring that all data would be anonymized and stored securely. Measures were taken to ensure no personally identifiable information was collected, and data was encrypted during transmission and storage to protect participants' privacy.

Additional information

Correspondence and requests for materials should be addressed to Yongzhi Li.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025