

# Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction

Jingxi Liu<sup>a</sup>, Daner Sun<sup>a,\*</sup>, Jin Sun<sup>b</sup>, Jingyun Wang<sup>c</sup>, Philip Leung Ho Yu<sup>a</sup>

<sup>a</sup> Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR, China

<sup>b</sup> Faculty of Education, University of Macau, Macao SAR, China

<sup>c</sup> Department of Computer Science, Durham University, Durham, United Kingdom

## ARTICLE INFO

### Keywords:

ChatGPT  
Generative artificial intelligence  
Mathematics word problem solving  
Personalized learning  
Primary mathematics education

## ABSTRACT

Mathematics word problem solving is a critical component of elementary education, yet many students encounter persistent difficulties in this area due to the combined cognitive demands of linguistic comprehension and mathematical reasoning. While previous studies have explored various pedagogical strategies to enhance problem-solving skills, the integration of generative artificial intelligence (GenAI) in this domain remains underexplored. This study introduces the ChatGPT-supported Mathematics Problem-Solving System (ChatGPT-MPS), a GenAI-enabled learning environment designed to support primary students in developing problem-solving strategies and deepening conceptual understanding. To evaluate its effectiveness, a quasi-experimental design was employed involving 104 fifth-grade students, assigned to an experimental group (using ChatGPT-MPS) and a control group (traditional instruction). Both groups completed a pre- and post-tests of MPS and interests scale to evaluate their word problem-solving proficiency and learning interests. Quantitatively data analysis and its results showed that the experimental group exhibited significantly greater improvements in post-test performance compared to the control group. In addition, student feedback revealed increased interest, perceived value, and motivation when engaging with the ChatGPT-MPS learning environment. These findings provide empirical support for the use of GenAI in mathematics education and demonstrate the potential of ChatGPT-MPS to enhance students' mathematical thinking and engagement in problem-solving tasks. The study contributes to the growing body of research on AI-driven personalized learning in primary education, offering insights into the design and implementation of effective GenAI-enabled learning environments.

## 1. Introduction

Artificial Intelligence (AI) is a broad and evolving field that encompasses a range of technologies enabling computers to perform cognitive tasks traditionally associated with human intelligence, including learning, reasoning, and problem-solving (Baker & Smith, 2019; Zawacki-Richter et al., 2019). Rather than denoting a single technology, AI functions as an umbrella term that covers machine learning, natural language processing (NLP), data mining, neural networks, and algorithmic decision-making (Kökver et al., 2024; bib\_kökver\_et\_al\_2024). These technologies allow computational systems to engage in complex processes such as visual and speech recognition, inferencing, and autonomous decision-making (Chen et al., 2020).

Recent breakthroughs in large-scale pre-trained models, most notably GPT and BERT, have significantly extended the capabilities of AI, particularly in NLP domains such as pre-training, fine-tuning, prompting, and text generation (Gardazi et al., 2025). These advancements have transformed AI from a supportive tool into a catalytic force across multiple sectors, including education.

In the field of education, artificial intelligence (AI) has been extensively examined for its potential to enhance learning experiences through adaptive instruction and intelligent tutoring systems (ITS). Early conceptions of AI focused on automating instructional support by enabling systems to independently solve problems, analyze learner behavior, and deliver real-time, human-like feedback (Ipinnaiye & Risquez, 2024). Building on this foundation, Hwang (2003) emphasized

\* Corresponding author. The Education University of Hong Kong, 10 LoPing Road, Tai Po, New Territories, D4-1F-15, Hong Kong SAR, China.

E-mail addresses: [liujingxi53@gmail.com](mailto:liujingxi53@gmail.com) (J. Liu), [dsun@eduhk.hk](mailto:dsun@eduhk.hk) (D. Sun), [jinsun@um.edu.mo](mailto:jinsun@um.edu.mo) (J. Sun), [jingyun.wang@durham.ac.uk](mailto:jingyun.wang@durham.ac.uk) (J. Wang), [plhyu@eduhk.hk](mailto:plhyu@eduhk.hk) (P.L.H. Yu).

<https://doi.org/10.1016/j.caeai.2025.100438>

Received 8 October 2024; Received in revised form 7 June 2025; Accepted 13 June 2025

Available online 13 June 2025

2666-920X/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

AI's capacity to organize instructional content, optimize learning pathways, and dynamically adapt tasks based on students' prior performance—thereby laying the groundwork for personalized and data-informed learning environments. With the advent of advanced large language models (LLMs), such as ChatGPT, the role of AI in education has expanded beyond task-specific automation to encompass broader, generative, and human-like intelligence (Chiarello et al., 2024). This shift has prompted a growing body of empirical research investigating the educational value of generative AI tools like ChatGPT across various instructional contexts. For example, Urban et al. (2024) demonstrated that ChatGPT supports university students in mastering complex mathematical concepts through personalized feedback and structured problem-solving guidance. Similarly, Rutherford et al. (2025) reported that lessons co-developed with ChatGPT reduced students' boredom and elicited fewer negative emotional responses, while Gianos and Delardas (2023) found that ChatGPT improved students' preparedness for standardized mathematics assessments. Collectively, these findings suggest that generative AI can significantly enrich learning by offering interactive, personalized, and context-sensitive support (Barnett et al., 2023), thereby contributing to improved engagement and academic achievement.

Despite the promising progress in AI-enhanced education, current applications often adopt a broad and generalized approach, lacking targeted attention to specific teaching challenges. One such enduring challenge is the instruction of mathematics word problems, which inherently pose learning difficulties due to their requirement for both linguistic comprehension and mathematical reasoning (Vessonen et al., 2024). These challenges are further compounded in traditional instructional settings, particularly in online or self-directed learning environments, where timely scaffolding and individualized support are often lacking. Although various strategies have been introduced to enhance students' problem-solving skills, few studies have systematically explored the potential of ChatGPT to address the unique demands of mathematics word problem solving. Furthermore, there remains a significant gap in the literature concerning the design, implementation, and empirical evaluation of AI-supported learning environments that leverage ChatGPT for this specific purpose.

This research gap underscores the need for targeted investigation into the strategic use of generative AI to support mathematical comprehension and problem-solving, particularly among primary school learners. In response, the present study proposes and evaluates the ChatGPT-supported Mathematics Problem Solving System (ChatGPT-MPS), specifically developed to strengthen students' abilities to navigate the linguistic and cognitive demands of mathematics word problems. By addressing this under-explored area, the study seeks to advance the growing body of research on GenAI in education and provide practical insights into how GenAI can be effectively integrated to tackle domain-specific learning challenges while promoting more inclusive, accessible, and impactful educational practices.

## 2. Literature review

### 2.1. Artificial intelligence in education

Artificial Intelligence (AI) has increasingly emerged as a transformative force in education, driving innovation across a range of domains, including language learning, STEM education, and instructional design (Cheng et al., 2025; Sun et al., 2023). AI-enabled technologies, such as adaptive tutoring systems, intelligent chatbots, and simulation-based learning platforms, have demonstrated substantial potential in facilitating personalized, interactive, and engaging learning experiences (Han et al., 2025). In the context of language education, AI-powered systems have been shown to enhance learners' grammar, vocabulary acquisition, and language fluency by providing real-time, individualized feedback and immersive, context-sensitive environments (Wen et al., 2024). Similarly, in science education, AI contributes

to improved learning outcomes by supporting multimodal representations of scientific content and enabling automated assessment of students' written responses, thereby deepening their conceptual understanding (Cheung et al., 2024; Yang et al., 2025). Collectively, these technologies not only foster enhanced subject mastery but also promote learners' motivation, autonomy, and self-regulated learning across disciplinary boundaries.

In mathematics education, AI has demonstrated particularly significant impact due to the discipline's inherent emphasis on logic, abstraction, and structured reasoning (Li, 2024). These characteristics align well with the precision, adaptability, and data-processing capabilities of AI-driven instructional tools. A growing body of research supports the integration of AI technologies as a means to enhance students' mathematical understanding and achievement (Khazanchi et al., 2024). For instance, the implementation of fuzzy-based convolutional neural networks combined with long short-term memory (LSTM) models in primary school classrooms has yielded notable improvements in student performance by enabling personalized learning experiences (Qiu et al., 2022). In addition, systems such as SAIL and AI-enhanced learning analytics platforms have proven effective in diagnosing students' conceptual difficulties and providing targeted feedback (Draskovic et al., 2018). These developments highlight AI's transformative potential in mathematics education, fostering a shift from traditional instruction toward more responsive, individualized, and data-informed pedagogical practices.

### 2.2. Learning difficulties in mathematics word problem solving

In primary mathematics education, word problem solving plays a critical role in fostering students' mathematical thinking and their ability to apply conceptual knowledge to real-world contexts (Myers et al., 2022). It not only serves as an essential assessment of mathematical understanding but also reflects the overall effectiveness of instructional practices. Word problems are embedded in everyday scenarios ranging from interpreting financial statements and analyzing sports statistics to understanding demographic reports and social events (Roy & Roth, 2018), and thus provide an authentic context for quantitative reasoning. Further study reported that word problem solving is a core component of elementary mathematics curricula, acting as a bridge between classroom learning and real-life application (Wong & Ho, 2017).

Despite its pedagogical significance, mathematics word problem solving remains a persistent challenge for many students, often leading to diminished academic performance and waning motivation. These difficulties commonly referred to as learning difficulties in mathematics word problem solving (Westwood, 2017), arise from a complex interplay of cognitive, linguistic, and foundational mathematical factors. A primary source of difficulty lies in the dual cognitive demands imposed by word problems, which require students to simultaneously decode the linguistic structure of a problem and apply appropriate mathematical operations. Träff et al. (2025) emphasize that effective word problem solving hinges on the construction of accurate mental representations, a process that is heavily reliant on reading comprehension skills. Students with limited proficiency in language processing may misinterpret key elements of the problem or fail to distinguish essential information from extraneous details, thereby undermining their ability to develop effective solution strategies.

In addition to linguistic challenges, deficiencies in fundamental mathematical competencies, such as number sense, arithmetic fluency, and the understanding of relational concepts, pose significant barriers to students' success in solving word problems (Westwood, 2011). Without a solid foundation in basic operations, students often struggle to translate the narrative elements of a problem into appropriate mathematical representations. This issue is especially pronounced among younger learners, who are still in the process of developing a conceptual understanding of mathematical structures and relationships (Vessonen

et al., 2024). Furthermore, limitations in logical reasoning and difficulties in recognizing problem schemas can impede students' ability to select effective solution strategies, particularly when encountering multi-step or non-routine problems. Consequently, many students resort to superficial approaches, such as relying on the identification of numbers and keywords, rather than engaging in meaningful conceptual analysis (Verschaffel et al., 2000).

Together, these cognitive and foundational barriers underscore the need for instructional approaches that simultaneously cultivate mathematical reasoning and language comprehension skills. Effectively addressing these multifaceted challenges calls for the design of learning environments that provide targeted scaffolding, foster metacognitive reflection, and support the development of both linguistic and mathematical literacies. Such integrative approaches are essential for enabling students to construct meaningful representations of problems, apply appropriate solution strategies, and engage in deeper conceptual understanding.

### 2.3. Related studies of AI in mathematics word problem solving

Although AI has been increasingly integrated into educational contexts, the design and development of AI systems specifically tailored to support mathematics word problem solving remains underdeveloped. In response to these limitations, recent innovations in AI-enhanced mathematics learning have begun to explore adaptive and personalized pathways. Hwang and Tu (2021) emphasized the importance of systems capable of delivering continuous access, real-time feedback, and instructional flexibility that evolves with students' progress in mathematics learning. Similarly, Hooshyar et al. (2024) highlighted the growing role of personalized, technology-enhanced learning tools in education. These developments indicate a shift toward more learner-centered AI applications that align more closely with individual learning trajectories. Nonetheless, AI applications specifically targeting mathematics word problem solving still exhibit several shortcomings. Most notably, they often fail to provide dynamic scaffolding and dialogic interaction, both essential for guiding learners through complex, multi-step problem-solving tasks (Kim & Hannafin, 2011). Moreover, these platforms frequently lack the capacity to adapt the linguistic complexity or conceptual rigor of tasks in response to student needs. This limitation is particularly detrimental for primary school students, who require developmentally appropriate support that accounts for both their cognitive and language development. The absence of such adaptive features risks disengaging learners and undermines the potential effectiveness of AI-supported interventions in this domain.

To address these gaps, it is essential to develop AI-powered learning environments that move beyond static, one-size-fits-all instructional models. Such systems must be dynamic, interactive, and attuned to the cognitive and developmental needs of young learners. In response, the present study proposes the ChatGPT-supported Mathematics Problem Solving System (ChatGPT-MPS), a generative AI-enhanced platform specifically designed for primary school students. By integrating ChatGPT with the mathematics curriculum, the system provides personalized, real-time feedback through natural language interaction.

## 3. Research purpose and research questions

The primary aim of this study is to examine the design, development, and implementation of the ChatGPT-MPS, a GenAI-enabled learning environment tailored to support primary school students in deepening their understanding of mathematical concepts, promoting academic inquiry, and strengthening their word problem-solving. The study seeks to generate empirical evidence on the pedagogical value of GenAI in mathematics education, with a particular focus on mathematics word problem solving. To achieve this objective, the following research questions were formulated:

**RQ1:** How was the ChatGPT-MPS system designed and developed, and what methodological strategies were employed to ensure its pedagogical effectiveness and technical robustness?

**RQ2:** What impact did the use of ChatGPT-MPS have on students' performance in solving mathematics word problems compared to traditional instructional methods?

**RQ3:** How did the integration of ChatGPT-MPS affect students' attitudes, motivation, and engagement in mathematics word problem solving?

## 4. System design and technical architecture

### 4.1. Integration with ChatGPT-4 API

The ChatGPT-MPS system was designed to assist primary school students in solving mathematics word problems by leveraging the advanced reasoning and natural language processing capabilities of ChatGPT-4. Among various large language models (LLMs), ChatGPT-4 was selected for its superior performance in mathematical reasoning tasks. Notably, it achieved an average score of 3.50 on the miniGHOSTS dataset, a benchmark indicating a level of mathematical proficiency comparable to that of undergraduate students (Frieder et al., 2023). This level of competence positioned ChatGPT-4 as a robust foundation for developing an intelligent tutoring environment capable of delivering personalized, step-by-step guidance while promoting students' conceptual understanding and problem-solving skills.

### 4.2. System architecture and deployment

The ChatGPT-MPS system adopted a modular architecture that clearly separated the front-end user interface from the back-end processing logic, thereby ensuring efficient interaction and reliable session management. To guarantee stable access to ChatGPT services in mainland China, the system was deployed using a customized cloud server and domain configuration. Secure communication with the official ChatGPT API was achieved via a dedicated foreign server, which facilitated seamless data transmission and supported real-time interaction for up to 100 concurrent users. User queries were processed through this pipeline with predefined constraints, such as maintaining a primary school teacher tone and restricting content to fifth-grade mathematics. The system returned multiple solution pathways for each query, enabling students to compare and reflect on different approaches to problem-solving. The exemplar code of sending and receiving messages via ChatGPT API is below (see Table 1).

### 4.3. Pedagogical features and functional design

#### 4.3.1. Personalized and scaffolding-based learning

The core pedagogical objective of ChatGPT-MPS was to facilitate self-directed learning by offering dynamic, AI-driven scaffoldings tailored to primary students' needs. Through structured guidance and the provision of multiple problem-solving pathways, the system enabled learners to progress incrementally in solving mathematics word problems. By engaging in interactive dialogues with the AI, students could explore alternative strategies and receive real-time, personalized support, thereby reducing dependence on teacher intervention during independent study (see Fig. 1). This design aimed to mitigate cognitive stagnation and foster deeper mathematical reasoning, ultimately supporting the development of a more systematic and conceptual understanding of mathematical principles.

#### 4.3.2. Adaptive feedback and natural language support

One of the system's key strengths was its ability to offer adaptive, real-time feedback aligned with each student's learning progress. When students encountered difficulties, the system dynamically generated explanatory guidance through up to three problem-solving pathways,

**Table 1**  
Exemplar Codes of Sending and Receiving Messages via ChatGPT API.

Action	Code Example
Received:	<pre>{   "model": "x",   "messages": [     {       "role": "system",       "content": "x"     },     {       "role": "user",       "content": "There are 30 people in the art group at Guangming Elementary School. The number of people in the biology group is 1/3 of the art group, and the number of people in the aviation model group is 4/5 of the biology group. How many people are in the aviation model group?"     }   ] }</pre>
Sent:	<pre>{   "id": "x",   "object": "chat.completion",   "created": x,   "model": "x",   "choices": [     {       "index": 0,       "message": {         "role": "assistant",         "content": "There are 30 people in the art group. The number of people in the biology group is 1/3 of the art group, so there are 30 / 3 = 10 people in the biology group. The number of people in the aviation model group is 4/5 of the biology group, so there are 10 * 4/5 = 8 people in the aviation model group."       }     },     "finish_reason": "stop"   ] }, "usage": {   "prompt_tokens": 69,   "completion_tokens": 90,   "total_tokens": 159 } }</pre>

which were retrieved via the API and stored in the Interactive Session Management module. Using natural language processing capabilities, ChatGPT simulated age-appropriate dialogue akin to interactions with a primary school teacher. This interactive element was designed to increase motivation, clarify misconceptions, and support incremental understanding of complex concepts.

**4.3.3. Key functional capabilities**

The ChatGPT-MPS system integrated several key features to enhance its pedagogical effectiveness. It supported multi-solution generation,

allowing up to three distinct solution pathways to be produced for each query. This enabled students to compare and contrast different problem-solving strategies, thereby deepening their conceptual understanding (see Fig. 2., Fig. 3.).

The system also featured interactive session management, which recorded each student’s query and response history to facilitate iterative learning and reflection. In addition, all ChatGPT-generated responses were dynamically adjusted to maintain a tone appropriate for primary school learners, ensuring that the content aligned with age-specific cognitive levels and supported effective comprehension.

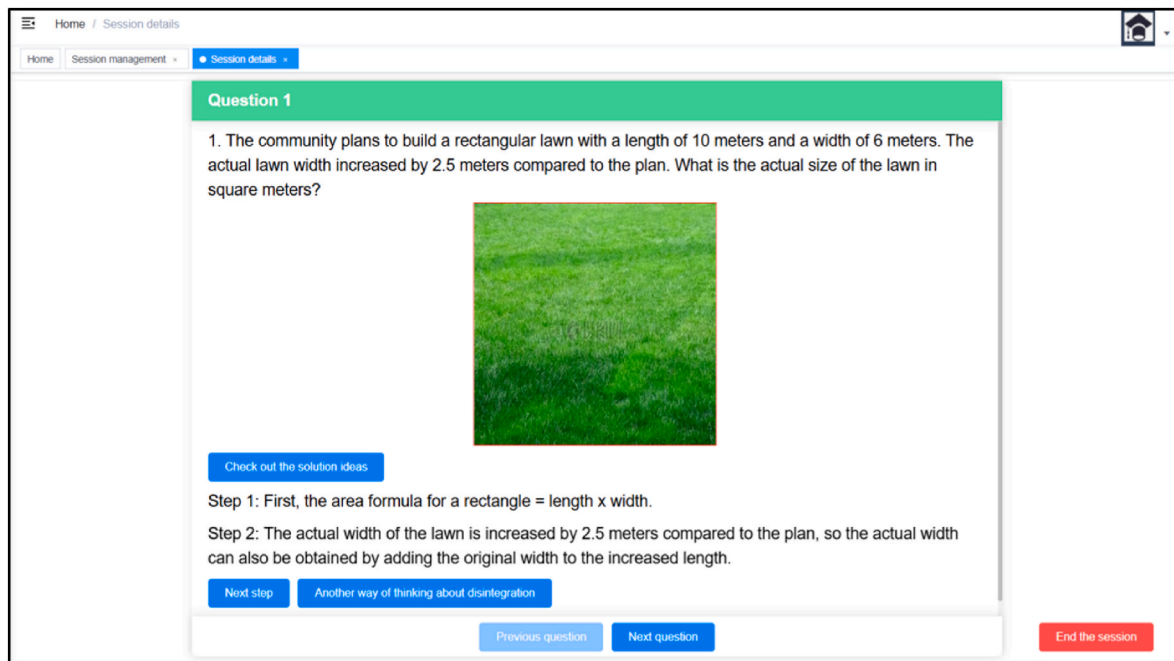


Fig. 1. AI-assisted learning mathematics word problem solving interface.

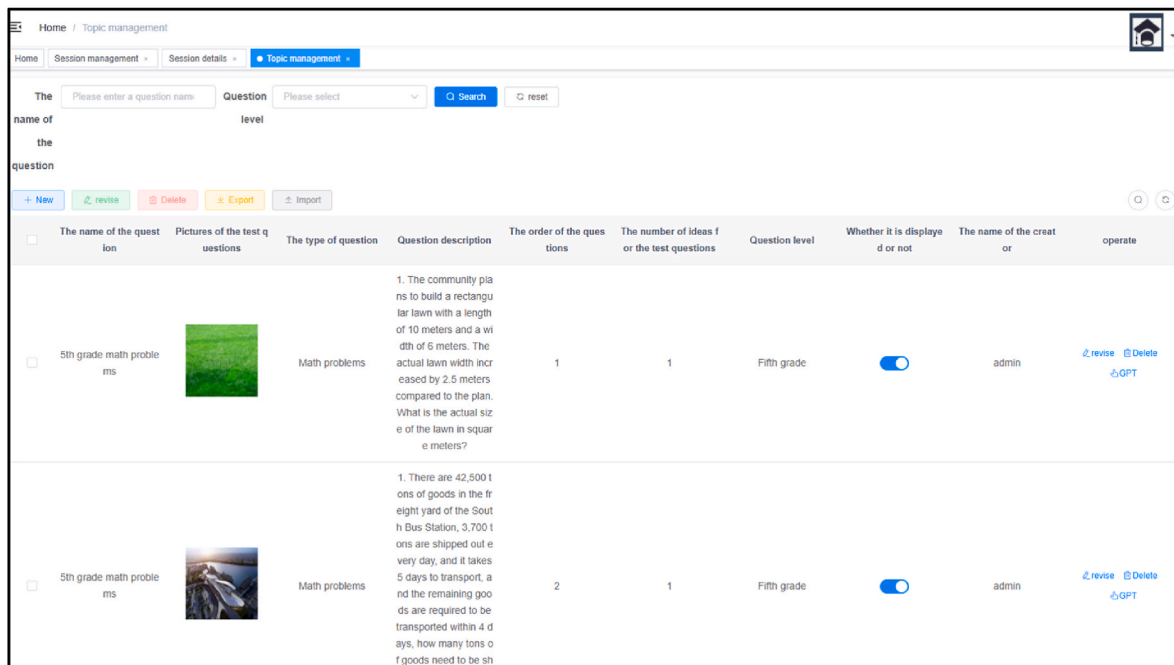


Fig. 2. Interface to enable GPT assist for built-in questions.

#### 4.4. User experience and system performance

Extensive testing demonstrated that the ChatGPT-MPS system consistently delivered fast response times and stable performance, even under high user demand. The system's infrastructure was optimized to ensure smooth operation without noticeable delays or technical disruptions. Users, including both students and educators, reported a high degree of satisfaction with its functionality, particularly praising the system's clear instructional guidance and intuitive interface design.

The visually accessible interface supported easy navigation for young learners, while the real-time feedback maintained student engagement

and enabled immediate clarification of problem-solving steps. Moreover, the system's structured guidance, aligned with real-world mathematical contexts, was noted as particularly helpful in reinforcing conceptual understanding. Participants also suggested that the ability to generate diverse solution pathways was valuable, and recommended further expansion of this feature to enhance learner autonomy and problem-solving flexibility. Overall, the ChatGPT-MPS system offered a reliable, user-centered experience that effectively supported primary students in learning mathematics word problem-solving. The specific underlying logic for key system feature was shown in Table 2.

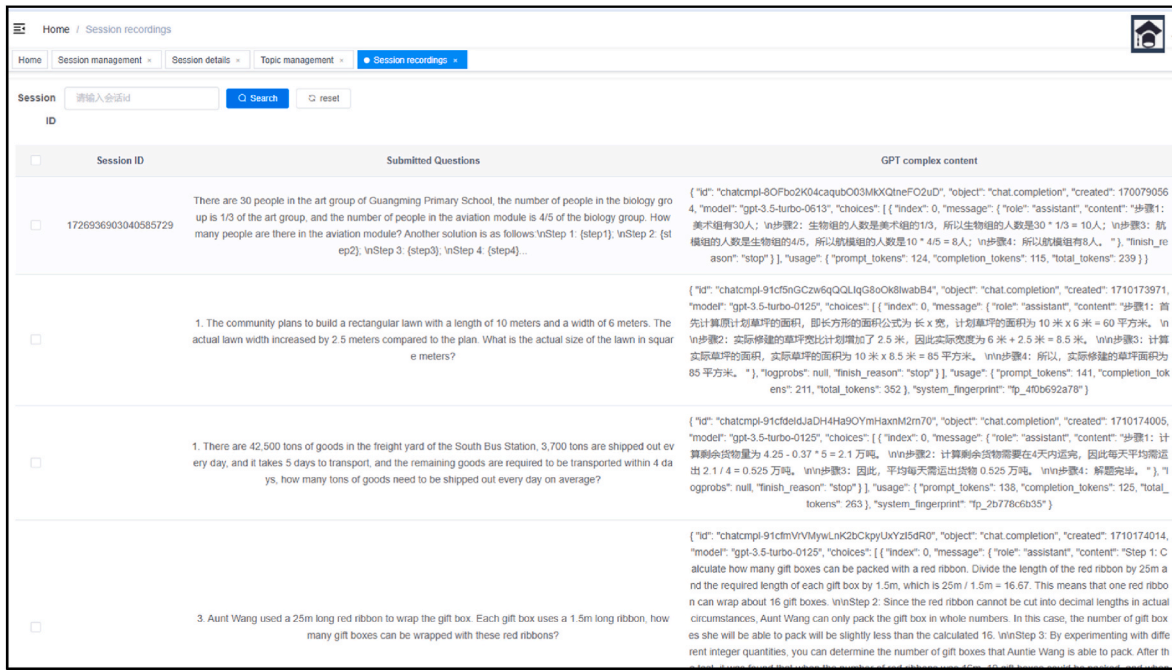


Fig. 3. Structured process for generating AI-Assisted learning approaches.

Table 2  
Underlying logic of the key features of ChatGPT-MPS.

Key Workflow	Work Content
User Input	Input a math word problem from the ChatGPT-MPS user end.
Input Preprocessing	Text cleaning: Remove irrelevant characters or noise, such as extra spaces or special symbols. Intent recognition: Identify the user's question intent, ensuring accuracy for primary school math problems.
ChatGPT API Call	Specify model version: Specify the use of ChatGPT-4/4o for processing. Provide user input after preprocessing. Parameter settings: Include personalized settings such as temperature, depth of knowledge, and maximum generation length. Control response methods for repeated queries.
Response Generation	Once ChatGPT receives the request, it generates an appropriate response based on its training data and the model parameters provided by ChatGPT-MPS. The model generates a step-by-step, structured answer.
Post-Processing	Grammar check: Ensure linguistic correctness. Content optimization: Validate the response is step-by-step and structured; store responses in the database for sequential access. Store up to three solution methods for easy comparison.
Feedback to User	Send the feedback back to the ChatGPT-MPS system and store it on the user end for sequential retrieval.
Keeping Learning Records	Save each problem receiving feedback, along with its various solution methods, in the system database for future retrieval.

5. Methods

5.1. Participants

This study involved 104 fifth-grade students from a public primary school in Guangdong province, China. A quasi-experimental design was employed, in which participants were assigned to experimental and comparison groups based on existing class structures rather than through random allocation. Two intact classes with comparable academic performance were selected. One class (n = 52) was designated as the experimental group and received instruction with the support of the

ChatGPT-MPS system. The other class (n = 52), serving as the comparison group, followed conventional teaching methods without ChatGPT-MPS integration.

Both groups demonstrated similar academic baselines prior to the intervention. Historical data indicated pass rates of approximately 80 %, with achievement rates ranging from 23.1 % to 54 %. Maximum scores recorded in the classes were 100 and 94, while the minimum scores were 18 and 9, respectively, reflecting a wide range of academic abilities but a comparable distribution of performance across groups. In the experimental group, the students had a mean age of 11 years. The gender distribution comprised 22 female students (42.3 %) and 30 male students (57.7 %). The comparison group maintained a similar gender balance, supporting the validity of comparative analysis. In addition, three teachers (two female and one male) from the experimental group's class were involved in the study. These teachers supervised students during the post-task phase and contributed to the system validity evaluation, offering practical insights into the effectiveness and operational viability of the ChatGPT-MPS system within a classroom setting.

5.2. Procedures

The experimental procedure was organized around two parallel tasks designed for the experimental and comparison groups. In the experimental condition, students worked on 12 mathematics word problems using the ChatGPT-MPS system. The session began with a 10-minute instructional briefing to familiarize students with the system's core features and usage protocols. After this orientation, students were allotted 40 minutes to complete the problem set with GenAI-generated support, which provided step-by-step guidance throughout the problem-solving process. Upon task completion, students in the experimental group completed the Student Interest Scale for Using ChatGPT-MPS (SISPS), designed to assess their interest, motivation, and affective responses toward the ChatGPT-MPS learning experience.

In contrast, the comparison group worked on the same set of 12 mathematics word problems under identical time constraints (40 minutes) but without access to the ChatGPT-MPS system. These students completed the task independently, relying solely on their prior knowledge and problem-solving skills. After the session, their written responses were collected for performance evaluation.

### 5.3. Data collection

#### 5.3.1. Pre-and post-MPS tests

To evaluate the impact of the ChatGPT-MPS system on students' ability to solve mathematics word problems, parallel pre- and post-assessments (see [Appendix A](#) and [Appendix B](#)) were meticulously designed. Each assessment consisted of 12 word problems aligned with the fifth-grade mathematics curriculum, covering key topics such as arithmetic operations, fractions, percentages, and basic geometry. The items were developed to comprehensively measure students' mathematical problem-solving proficiency, with careful attention to ensuring equivalence in difficulty, complexity, and structure across both tests. To prevent rote memorization while maintaining conceptual alignment, slight modifications were introduced in the wording of some post-test items.

Multiple strategies were adopted to ensure the validity of the assessment instruments. Content validity was established through strict adherence to the 2022 Curriculum Plan and Curriculum Standards for Compulsory Education. All test items were authored by the Head of Year Mathematics at the participating school to ensure alignment with grade-level learning objectives. Construct validity was ensured by drawing on sample problems from national assessment standards, thereby reflecting authentic, real-world mathematical problem-solving contexts emphasized in the curriculum. Equivalence validity was maintained through careful calibration of item structure, topic distribution, and cognitive demand across both assessments. This parallel test design enabled a reliable comparison of student performance before and after the intervention, ensuring that observed differences could be attributed to the implementation of the ChatGPT-MPS system rather than to test inconsistencies.

#### 5.3.2. Student's interest scale of using ChatGPT-MPS (SISPS)

To assess students' interest in solving mathematics word problems following their use of the ChatGPT-MPS system, this study adapted the Academic Interest Scale for Adolescents (AISA), originally developed by [Luo et al. \(2019\)](#), as the theoretical foundation. Rooted in [Hidi and Renninger's \(2006\)](#) four-phase model of interest development, AISA offers a comprehensive framework for examining how academic interest is initiated, maintained, and deepened in educational contexts. Building on this model, the present study developed the Student Interest Scale for Using ChatGPT-MPS (SISPS), specifically designed to evaluate student interest within the context of AI-assisted mathematics learning.

The SISPS consists of 29 items measured on a five-point Likert scale and has demonstrated strong psychometric properties, including satisfactory reliability and construct validity through preliminary validation procedures (see [Appendix C](#)). The instrument is structured across four dimensions: (a) Interest (7 items), (b) Value (8 items), (c) Knowledge (7 items), and (d) Engagement (7 items). Within this structure, the Interest dimension functions as the dependent variable, capturing students' motivation and curiosity toward using the ChatGPT-MPS system for mathematics word problem solving. The remaining three dimensions: Value, Knowledge, and Engagement, serve as potential predictors, reflecting students' perceptions of the system's relevance, their conceptual understanding gained through its use, and the degree of their active involvement during learning. The reliability coefficient value of each dimension of the research data is higher than 0.8, which indicates that the data reliability quality of SISPS is high and can be used for further analysis ([Appendix D](#)). By integrating the AISA framework into the ChatGPT-MPS learning environment, the SISPS offers a nuanced and multidimensional tool for assessing student motivation and cognitive engagement. It provides valuable insights into how GenAI technologies influence learners' affective responses and foster interest in mathematics within primary education settings.

### 5.4. Data analysis

Besides the description of system design and development to answer the RQ1, appropriate quantitative and qualitative methods were employed to analyze different types of data for addressing RQ2 and RQ3. Both data preprocessing and statistical analyses were aligned with the objectives of each research question, ensuring the validity of the analysis and the rigor of the methodology.

To investigate RQ2, student performance data were collected from pre- and post-task assessments. Prior to analysis, systematic data preprocessing was conducted to ensure data accuracy and suitability for statistical testing. The preprocessing included normality checks for pre- and post-task scores, as illustrated in histograms (see [Fig. 4](#) and [Fig. 5.](#)), as well as outlier detection using boxplot analysis, which revealed three potential outliers in the post-task dataset (no outliers were found in the pre-task dataset), and these outliers were subsequently removed (see [Fig. 6.](#)).

After completing the data preprocessing, a paired-samples *t*-test was conducted to assess the statistical significance of differences in student performance before and after using ChatGPT-MPS. This within-subjects analysis controls for individual baseline differences and reveals the mechanisms of the system's direct impact.

And to address RQ3, data from the Student Interest Scale for Using ChatGPT-MPS (SISPS) were analyzed. Pearson correlation analysis was used to examine the linear relationship between the four dimensions of SISPS and to determine the correlation between value, knowledge, and engagement with the main dependent variable of interest. In addition, the study used regression analysis to demonstrate whether all three variables were significant and to rank the three variables in terms of influence related to Interest. And multicollinearity diagnostics were used to calculate the Variance Inflation Factor (VIF) to ensure that the predictor variables were statistically independent of each other, sufficient to validate the hypothesis. All statistical procedures were performed using SPSS to ensure the reproducibility and reliability of the research questions.

## 6. Results

### 6.1. Impact of ChatGPT-MPS on students' mathematical performance

A paired-samples *t*-test was conducted to examine the impact of the ChatGPT-MPS system on students' ability to solve mathematics word problems. As shown in [Table 3](#), the analysis revealed a statistically significant increase in students' post-task scores ( $M = 72.59$ ,  $SD = 17.45$ ) compared to their pre-task scores ( $M = 56.45$ ,  $SD = 24.37$ ), with a mean improvement of **16.14 points**. This substantial gain suggests that the integration of ChatGPT-MPS had a positive effect on students' mathematics word problem-solving performance. The reduced standard deviation in the post-task scores also indicates a more consistent performance across students following the intervention.

This improvement was statistically significant, as evidenced by the *t*-value of  $-7.156$  with 48 degrees of freedom ( $p < 0.001$ ) (see [Table 4](#)), confirming that the observed gains were unlikely due to chance. Additionally, a strong positive correlation ( $r = 0.763$ ,  $p < 0.001$ ) was found between pre-task and post-task scores (see [Table 5](#)), indicating consistent individual performance patterns across the two assessments. Notably, the reduction in standard deviation in the post-test scores suggests that students' performance became more uniform after using the ChatGPT-MPS system, reflecting not only overall improvement but also enhanced learning stability across the class.

The calculated effect sizes, Cohen's  $d = 1.022$  and Hedges'  $g = 1.014$ , reflect a practical impact of the ChatGPT-MPS system on students' mathematics word problem-solving performance (see [Table 6](#)). These values suggest that the observed improvement was not only statistically significant but also educationally meaningful. The 95 % confidence intervals for both measures (Cohen's  $d$ : 1.365 to  $-0.672$ ;

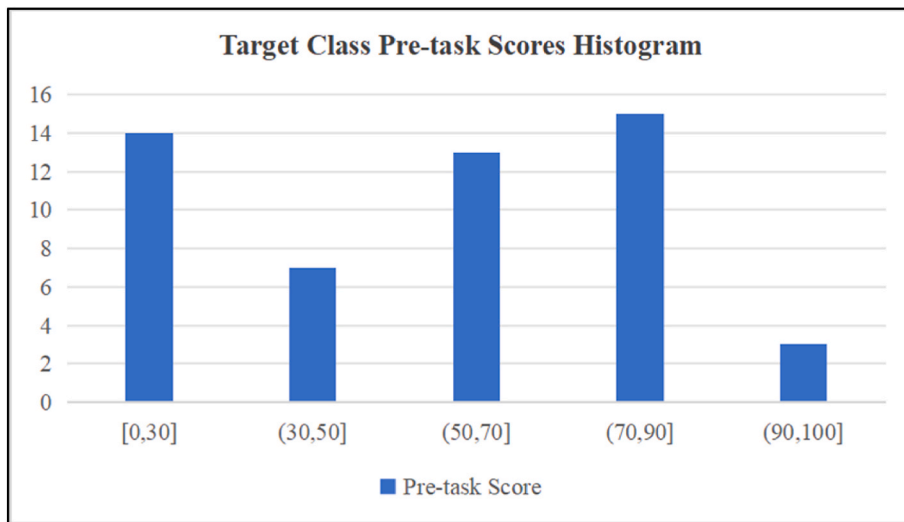


Fig. 4. Histogram of pre-task scores.

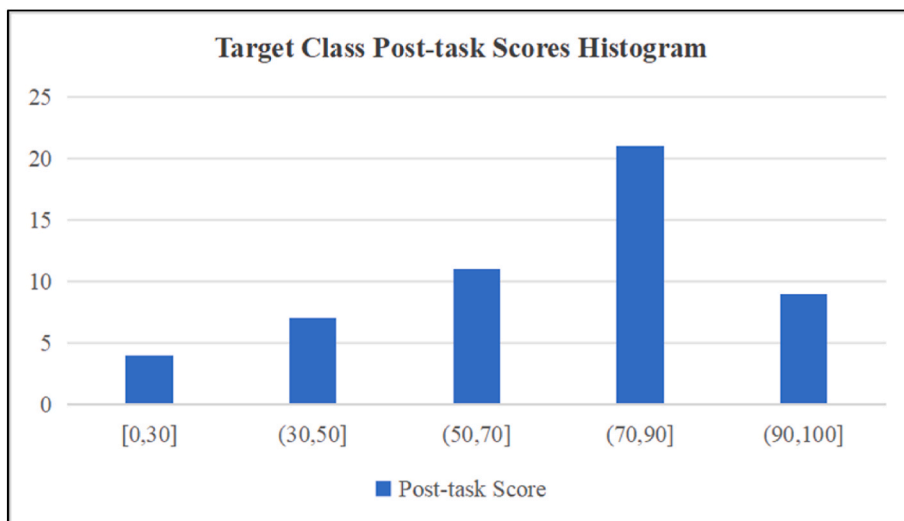


Fig. 5. Histogram of post-task scores.

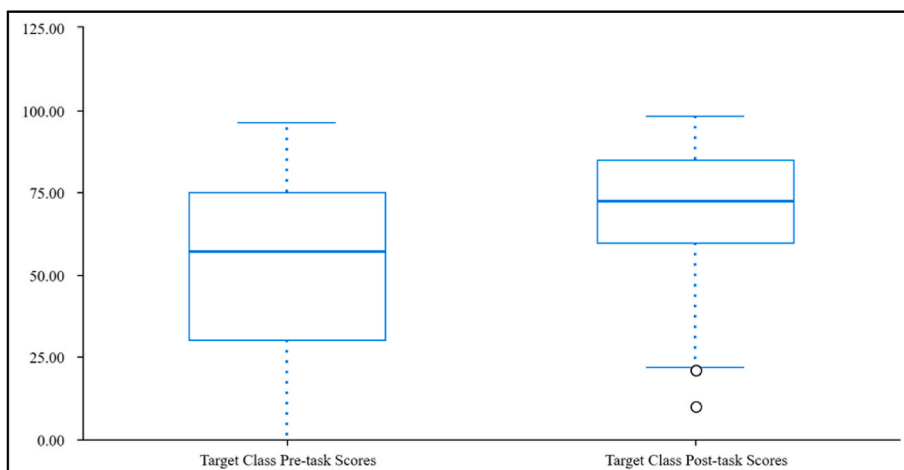


Fig. 6. Box plots of the results of the two tasks.

**Table 3**  
Paired sample statistics.

Pairing	Average	Number	SD	Standard Error Mean
Target Class Pre-task Scores	56.4490	49	24.37302	3.48186
Target Class Post-task Scores	72.5918	49	17.45461	2.49352

**Table 4**  
Paired Samples t-test.

Pairing	Mean Difference	t-value	Degrees of Freedom (df)	Significance (Two-Tailed)
Target Class Pre-task – Post-task Scores	-11.61	-7.156	48	0.000

**Table 5**  
Paired-sample correlation.

Pairing	Number	Correlation	Significance
Target Class Pre-task & Post-task Scores	49	0.763	0.000

**Table 6**  
Paired sample effect size.

Pairing	Effect Size	Point Estimate	95 % Confidence Interval
Target Class Pre-task – Post-task Scores	Cohen’s d	-1.022	-1.365 to -0.672
	Hedges’ g	-1.014	-1.354 to -0.667

Hedges’ g: 1.354 to -0.667) further confirm the robustness and reliability of the effect.

6.2. Correlation between interests and key motivational factors

To explore students’ motivation to engage with the system, Pearson correlation analysis was conducted using the SISPS. The results, presented in Table 7, reveal strong and statistically significant correlations between students’ interest and each of the three contributing dimensions: perceived value of the system ( $r = 0.874, p < 0.01$ ), understanding of system features ( $r = 0.684, p < 0.01$ ), and behavioral engagement ( $r = 0.851, p < 0.01$ ). These findings suggest that students who perceived greater instructional value and gained clearer understanding from the system were more likely to be engaged and express sustained interest in solving mathematics word problems. The strong inter-correlations among value, knowledge, and engagement further indicate a cohesive motivational framework supporting student interaction with the ChatGPT-MPS system (see Table 8).

6.3. Predictors of interest: regression analysis

A multiple linear regression analysis was conducted to examine how

**Table 7**  
Pearson correlation analysis.

Dimension	Interest	Value	Knowledge	Engagement
Interest	1			
Value	0.874**	1		
Knowledge	0.684**	0.702**	1	
Engagement	0.851**	0.865**	0.730**	1

Note:  $p < 0.01$ .

students’ perceptions of value, knowledge, and engagement contributed to their overall interest in using the ChatGPT-MPS system. The results revealed that all three variables were statistically significant predictors of interest, with value emerging as the most influential factor ( $B = 0.736, \beta = 0.684, p < 0.01$ ), followed by knowledge ( $B = 0.613, \beta = 0.548, p < 0.01$ ) and engagement ( $B = 0.428, \beta = 0.377, p < 0.05$ ). These findings highlight the central role of perceived utility in shaping student interest. Students who viewed the system as valuable and informative demonstrated higher motivation and willingness to engage with it.

Furthermore, multicollinearity diagnostics indicated that the predictors were sufficiently independent, as reflected in acceptable Variance Inflation Factor (VIF) and Tolerance values, confirming the robustness of the regression model. These results provide a clear understanding of the motivational dynamics underlying students’ interaction with ChatGPT-MPS and suggest practical implications for future instructional design that prioritizes perceived relevance and cognitive engagement.

7. Conclusions and discussions

This study aimed to explore the role of ChatGPT-MPS in enhancing primary school students’ performance in solving mathematical word problems, focusing on understanding the development and testing process, the impact on students’ problem-solving abilities, and the influence on students’ attitudes and enthusiasm toward mathematics. The development of ChatGPT-MPS involved a systematic and iterative process, beginning with a needs assessment that identified the specific challenges students face in solving mathematical word problems. This initial evaluation informed the design of the system, ensuring it addressed common difficulties. The design and prototyping stages utilized design thinking methodologies, creating prototypes that were refined based on user feedback to enhance functionality and usability.

The results of the paired-samples t-test analysis revealed a significant improvement in students’ performance after using ChatGPT-MPS, with an average pre-task score of 56.45 increasing to 72.59 post-intervention, indicating an average improvement of 16.14 points. The statistical evidence ( $p < 0.001$ ) confirmed that this improvement was not due to random chance, supported by a strong correlation coefficient of 0.763 between pre-task and post-task scores. Several mechanisms contributed to this positive impact, including personalized feedback, which helped students identify and address specific weaknesses. This was consistent with the findings of Pardo et al. (2017) in their study on engineering students, where they found that personalized feedback enabled by learning analytics technologies significantly improved students’ performance. The interactive learning environment fostered deeper engagement and understanding of problem-solving strategies, while the novelty of using AI motivated students to participate actively. This was also evident in the studies by Nazari et al. (2021), where the AI gamified learning frameworks and writing tools they developed and researched were found to enhance students’ motivation and engagement in learning more effectively through innovative approaches. Furthermore, the real-time assistance provided by ChatGPT-MPS clarified misconceptions and guided students through complex problems, leading to improved outcomes.

In addition to enhancing problem-solving abilities, the implementation of ChatGPT-MPS positively influenced students’ attitudes and enthusiasm toward mathematics. The Pearson correlation analysis indicated strong positive relationships between students’ interest in ChatGPT-MPS and the dimensions of value, knowledge, and engagement, particularly emphasizing the significance of perceived value ( $r = 0.874, p < 0.01$ ) in driving student interest. Regression analysis further revealed that perceived value had the most substantial effect on students’ interest, followed by knowledge and engagement. These findings suggest that recognizing the educational worth of ChatGPT-MPS fosters meaningful engagement, because students are more likely to engage

**Table 8**  
Linear regression analysis results (n = 51).

Predictor	B	Std. Error	Beta	t-value	p-value	VIF	Tolerance
Constant	1.175	0.458	–	2.563	0.013*	–	–
Value	0.736	0.112	0.684	6.569	0.000**	1.000	1.000
Knowledge	0.613	0.143	0.548	4.272	0.000**	3.962	0.252
Engagement	0.428	0.146	0.377	2.941	0.005**	3.962	0.252

Note: \*p < 0.05; \*\*p < 0.01.

actively with a tool they perceive as valuable, just as people will have a tendency to choose cognitive tools they find effective (Alexandre et al., 2019).

Reasons for this influence include students' recognition of the tool's relevance, which stems from its ability to make learning relatable and applicable to real-world contexts. Positive learning experiences facilitated by the AI's interactive nature create a feedback loop, where increased interest leads to more active engagement, enhancing overall learning outcomes. Additionally, the deeper understanding gained from using ChatGPT-MPS fosters confidence, making students more willing to tackle challenging problems and acquire enhanced learning skills. Just like Rickles et al. (2019) found in their survey of sixteen high schools, the more deeply ingrained the learning models were in the schools, the stronger the students' ability to delve into certain areas of study.

In conclusion, this study provides compelling evidence that ChatGPT-MPS significantly enhances primary school students' mathematical problem-solving skills and attitudes. The effective development and testing processes ensured the system was tailored to meet students' educational needs. The statistical analyses not only highlight the system's effectiveness in improving performance but also reveal critical factors contributing to student engagement and interest in learning mathematics. Integrating AI tools like ChatGPT-MPS into educational settings presents substantial potential for transforming mathematics education and fostering a more engaging learning experience for students. These insights contribute valuable knowledge to the field of AI in education and underscore the importance of leveraging technology to enhance learning outcomes.

## 8. Implications

The findings from this study yield several important implications for the design and implementation of AI-assisted systems in mathematics education. The success of ChatGPT-MPS underscores the importance of creating intuitive and user-friendly interfaces that cater to the cognitive and developmental needs of primary school learners. Future system development should emphasize adaptive design features that enable personalized learning paths tailored to students' individual abilities, progress levels, and learning styles. Incorporating multi-grade content coverage and interactive elements—such as memory-based scaffolding and contextual hints—could further expand the system's pedagogical utility and responsiveness.

Moreover, the integration of ChatGPT highlights the transformative potential of large language models in supporting mathematics learning by offering dynamic, step-by-step guidance that promotes deeper conceptual understanding and real-time clarification of misconceptions. The interactive, conversational design fosters greater student motivation, persistence, and engagement, which are essential for sustained learning in cognitively demanding tasks such as mathematical word problem solving.

Pedagogically, ChatGPT-MPS contributes to a blended learning model by supplementing traditional instruction with individualized AI feedback, enabling both teacher-facilitated and self-directed learning. It also encourages peer collaboration and reflective thinking, aligning well with contemporary constructivist approaches to mathematics education. As the field moves toward scalable, AI-enhanced instructional tools, future developments should focus on refining system interactivity,

expanding adaptive learning capabilities, and ensuring accessibility across diverse educational contexts.

This study provides empirical support for the meaningful integration of AI into mathematics education and offers a practical framework for the continued design of intelligent systems that promote equitable, engaging, and effective learning experiences.

## CRedit authorship contribution statement

**Jingxi Liu:** Writing – original draft, Software, Methodology, Conceptualization. **Daner Sun:** Supervision, Project administration, Funding acquisition. **Jin Sun:** Supervision, Investigation, Formal analysis. **Jingyun Wang:** Writing – review & editing, Supervision, Resources. **Philip Leung Ho Yu:** Writing – review & editing, Validation, Supervision.

## Statements on open data and ethics

The study was approved by an ethical committee with ID: 2020-2021-0145. Informed consent was obtained from all participants, and their privacy rights were strictly observed. The data can be obtained by sending request e-mails to the corresponding author.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by 1) General Research Fund (GRF) [grant numbers: 18609221] Preparation for Post Pandemic Society: Providing Technology-Supported Adaptive Scaffoldings for Students with Mathematics Learning Difficulties in Hong Kong Primary Schools. 2) CRAC Project (#04A59) of EdUHK, Revitalizing Mathematical Proficiency using AIScaffoldiaMaths: Empowering Students in Low-Resource Families through AI-Powered Supports in Informal Learning Environments.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.caeai.2025.100438>.

## References

- Alexandre, B., Navarro, J., Reynaud, E., & Osiurak, F. (2019). Which cognitive tools do we prefer to use, and is that preference rational? *Cognition*, 186, 108–114. <https://doi.org/10.1016/j.cognition.2019.02.005>
- Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. NESTA. <https://www.nesta.org.uk/report/education-rebooted/>.
- Barnett, M. W., Sipitakiat, A., & Blikstein, P. (2023). Neural syntonicity: A constructionist approach to the development of image recognition tools used to teach students about powerful ideas in artificial intelligence. *International Journal of Information and Education Technology*, 13(12), 1917–1923. <https://doi.org/10.18178/ijiet.2023.13.12.2005>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/access.2020.2988510>

- Cheng, G., Yu, P. L. H., Jia, J., Zheng, Z., & Chen, A. (2025). Personalized stem education empowered by artificial intelligence: a comprehensive review and content analysis. *Interactive Learning Environments*, 1–23. <https://doi.org/10.1080/10494820.2025.2462156>.
- Cheung, K. K., Pun, J., Kenneth-Li, W., & Mai, J. (2024). Exploring students' multimodal representations of ideas about epistemic reading of scientific texts in generative AI Tools. *Journal of Science Education and Technology*, 34(2), 284–297. <https://doi.org/10.1007/s10956-024-10182-0>
- Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on chatgpt. *Technovation*, 133, Article 103002. <https://doi.org/10.1016/j.technovation.2024.103002>
- Draskovic, D., Cvetanovic, M., & Nikolic, B. (2018). Sail—software system for learning AI algorithms. *Computer Applications in Engineering Education*, 26(5), 1195–1216. <https://doi.org/10.1002/cae.21988>
- Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., & Berner, J. (2023). Mathematical capabilities of chatgpt. July 20 *arXiv.org* <https://arxiv.org/abs/2301.13867>.
- Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alshafi, T., & Alshemaimri, B. (2025). Bert applications in natural language processing: A review. *Artificial Intelligence Review*, 58(6). <https://doi.org/10.1007/s10462-025-11162-5>
- Giannos, P., & Delardas, O. (2023). Performance of CHATGPT on UK standardized admission tests: Insights from the BMAT, TMUA, LNAT, and tsa examinations. *JMIR Medical Education*, 9. <https://doi.org/10.2196/47737>
- Han, I., Ji, H., Jin, S., & Choi, K. (2025). Mobile-based artificial intelligence chatbot for self-regulated learning in a hybrid flipped classroom. *Journal of Computing in Higher Education*. <https://doi.org/10.1007/s12528-025-09434-8>
- Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 111–127. <https://doi.org/10.1207/s15326985Sep4102.4>.
- Hooshyar, D., Weng, X., Sillat, P. J., Tammets, K., Wang, M., & Hämäläinen, R. (2024). The effectiveness of personalized technology-enhanced learning in higher education: A meta-analysis with association rule mining. *Computers & Education*, 223, Article 105169. <https://doi.org/10.1016/j.compedu.2024.105169>
- Hwang, G.-J. (2003). A conceptual map model for developing intelligent tutoring systems. *Computers & Education*, 40(3), 217–235. [https://doi.org/10.1016/s0360-1315\(02\)00121-5](https://doi.org/10.1016/s0360-1315(02)00121-5)
- Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, 9(6), 584. <https://doi.org/10.3390/math9060584>
- Ippinaiye, O., & Riquez, A. (2024). Exploring adaptive learning, learner-content interaction and student performance in Undergraduate Economics classes. *Computers & Education*, 215, Article 105047. <https://doi.org/10.1016/j.compedu.2024.105047>
- Khazanchi, R., Di Mitri, D., & Drachsler, H. (2024). The effect of ai-based systems on mathematics achievement in rural context: A quantitative study. *Journal of Computer Assisted Learning*, 41(1). <https://doi.org/10.1111/jcal.13098>
- Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (teles): Bridging research and theory with practice. *Computers & Education*, 56(2), 403–417. <https://doi.org/10.1016/j.compedu.2010.08.024>
- Kökver, Y., Pektaş, H. M., & Çelik, H. (2024). Artificial intelligence applications in education: Natural language processing in detecting misconceptions. *Education and Information Technologies*, 30(3), 3035–3066. <https://doi.org/10.1007/s10639-024-12919-1>
- Li, M. (2024). Integrating Artificial Intelligence in primary mathematics education: Investigating internal and external influences on teacher adoption. *International Journal of Science and Mathematics Education*, 23(5), 1283–1308. <https://doi.org/10.1007/s10763-024-10515-w>
- Luo, Z., Dang, Y., & Xu, W. (2019). Academic Interest Scale for adolescents: Development, validation, and measurement invariance with Chinese students. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02301>
- Myers, J. A., Witzel, B. S., Powell, S. R., Li, H., Pigott, T. D., Xin, Y. P., & Hughes, E. M. (2022). A meta-analysis of mathematics word-problem solving interventions for elementary students who evidence mathematics difficulties. *Review of Educational Research*, 92(5), 695–742. <https://doi.org/10.3102/00346543211070049>
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), Article e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2017). Using learning analytics to change the provision of personalised feedback. *British Journal of Educational Technology*, 50(1), 128–138. <https://doi.org/10.1111/bjet.12592>
- Qiu, Y., Pan, J., & Ishak, N. A. (2022). Effectiveness of artificial intelligence (AI) in improving pupils' deep learning in primary school mathematics teaching in Fujian Province. *Computational Intelligence and Neuroscience*, 1–10. <https://doi.org/10.1155/2022/1362996>, 2022.
- Rickles, J., Zeiser, K. L., Yang, R., O'Day, J., & Garet, M. S. (2019). Promoting deeper learning in high school: Evidence of opportunities and outcomes. *Educational Evaluation and Policy Analysis*, 41(2), 214–234. <https://doi.org/10.3102/0162373719837949>
- Roy, S., & Roth, D. (2018). Mapping to declarative knowledge for word problem solving. *Transactions of the Association for Computational Linguistics*, 6, 159–172. [https://doi.org/10.1162/tacl\\_a\\_00012](https://doi.org/10.1162/tacl_a_00012)
- Rutherford, T., Rodrigues, A., Duque-Baird, S., Veng, S., Mykita-Chomsky, R., Cao, Y., Chisholm, K., & Bergwall, E. (2025). "I just think it is the way of the future": Teachers' use of chatgpt to develop motivationally-supportive math lessons. *Computers and Education: Artificial Intelligence*, 8, Article 100367. <https://doi.org/10.1016/j.caeai.2025.100367>
- Sun, D., Zhan, Y., Wan, Z. H., Yang, Y., & Looi, C.-K. (2023). Identifying the roles of technology: A systematic review of STEM education in primary and secondary schools from 2015 to 2023. *Research in Science & Technological Education*. <https://doi.org/10.1080/02635143.2023.2251902>, 1–25.
- Träff, U., Skagerlund, K., & Skagenholt, M. (2025). Development of early acquired arithmetic skills and development of reading comprehension facilitate the development of later developing arithmetic skills. *Educational Psychology*, 45(2), 127–147. <https://doi.org/10.1080/01443410.2025.2463930>
- Urban, M., Dèchtèrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). CHATGPT improves creative problem-solving performance in University Students: An Experimental Study. *Computers & Education*, 215, Article 105031. <https://doi.org/10.1016/j.compedu.2024.105031>
- Verschaffel, L., Greer, B., & Corte, E. de (2000). *Making sense of word problems*. Swets & Zeitlinger Publishers.
- Vessonen, T., Dahlberg, M., Hellstrand, H., Widlund, A., Korhonen, J., Aunio, P., & Laine, A. (2024). Task characteristics associated with mathematical word problem-solving performance among elementary school-aged children: A systematic review and meta-analysis. *Educational Psychology Review*, 36(4). <https://doi.org/10.1007/s10648-024-09954-2>
- Wen, Y., Chiu, M., Guo, X., & Wang, Z. (2024). ai-powered vocabulary learning for lower primary school students. *British Journal of Educational Technology*, 56(2), 734–754. <https://doi.org/10.1111/bjet.13537>
- Westwood, P. (2011). The problem with problems: Potential difficulties in implementing problem-based learning as the core method in Primary School Mathematics. *Australian Journal of Learning Difficulties*, 16(1), 5–18. <https://doi.org/10.1080/19404158.2011.563475>
- Westwood, P. (2017). *Numeracy and learning difficulties: Approaches to teaching and assessment*. David Fulton Publishers.
- Wong, T. T.-Y., & Ho, C. S.-H. (2017). Component processes in arithmetic word-problem solving and their correlates. *Journal of Educational Psychology*, 109(4), 520–531. <https://doi.org/10.1037/edu0000149>
- Yang, J., Latif, E., He, Y., & Zhai, X. (2025). Fine-tuning CHATGPT for automatic scoring of written scientific explanations in Chinese. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-025-10199-z>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on Artificial Intelligence Applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1). <https://doi.org/10.1186/s41239-019-0171-0>