Tech. Edu. J. 20(1): 157-178, Winter 2026

Technology of Education Journal
(TEJ)
Homepage: jte.sru.ac.ir

ORIGINAL RESEARCH PAPER

Evaluating pre-service teachers' mathematical modelling task performance: a comparison with ai responses

***N. Azadi*[1]*, E. Reyhani*[1]*, M. Ghorbani*[*,1]*, R. Ebrahimpour*[2]*, S. Haghjoo*[3]**

[1] *Department of Mathematics, Faculty of Science, Shahid Rajaee Teacher Training University, Tehran, Iran*
[2] *Center for Cognitive Science, Institute for Convergence Science and Technology (ICST), Sharif University of Technology, Tehran, Iran; School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*
[3] *Farhangian University, Bushehr, Iran*

Received: 16 August 2025
Reviewed: 25 September 2025
Revised: 28 October 2025
Accepted: 19 December 2025

KEYWORDS:
Mathematical Modelling
AI (AI)
ChatGPT
PSTs

* Corresponding author
mghorbani@sru.ac.ir
(+98912) 4074398

ABSTRACT

Background and Objectives: Mathematical modelling has increasingly become a central component of mathematics education, as it helps students connect abstract mathematical concepts to real-world situations. Despite its importance, many elementary pre-service teachers (PSTs) face challenges when attempting to design or solve modelling problems, partly due to their limited exposure to such tasks during their teacher education programs. With the rapid development of AI (AI), new opportunities have emerged to compare human problem-solving approaches with machine-generated solutions. The purpose of this study is to examine how elementary PSTs and an AI system perform on mathematical modelling tasks. Specifically, the study seeks to understand the strengths and weaknesses of PSTs' modelling competencies, while also exploring the potential contributions of AI as a supportive tool in mathematics education. The objective is not only to evaluate performance but also to address the broader significance of integrating AI into teacher education, highlighting both opportunities and risks.

Methods: A convenience sample of 50 PSTs participated in the study. Participants were asked to complete a researcher-developed test that included mathematical modelling tasks designed to evaluate different aspects of modelling competence. The face and content validity of the instrument were reviewed and confirmed by two experts in mathematics education, ensuring the appropriateness and relevance of the items. Data were collected in written form and analyzed using both quantitative and qualitative approaches. The PSTs' responses were examined based on the competency framework, which provides structured criteria for assessing modelling skills such as understanding the problem, using mathematics, interpreting results, and validating solutions. AI-generated responses were also collected for the same tasks, and these were compared with the human responses in terms of strategy, accuracy, creativity, and solution quality. The study followed a qualitative design with deductive (directed) content analysis, which allowed the researchers to systematically code and interpret the data while drawing comparisons across human and AI performance.

Findings: The findings revealed several important patterns. Many PSTs demonstrated weaknesses in mathematical modelling, particularly in the stages of validation and interpretation. Their answers were often incomplete or lacked sufficient reasoning. However, the participants performed relatively better in the criterion of 'use of mathematics', indicating that they could carry out mathematical procedures but struggled to connect them meaningfully to real-world contexts. No statistically significant differences were found between participants based on gender or academic background, suggesting that the difficulties with modelling are widespread among elementary PSTs regardless of their demographic characteristics. In contrast, the AI system produced solutions that were generally more coherent and well-structured, particularly in terms of validation and creativity. The AI was able to provide step-by-step explanations and generate multiple approaches to solving the same problem, which could serve as a model for PSTs. Nevertheless, the AI responses were not flawless, and in some cases, over-simplifications or assumptions were made without sufficient justification.

Conclusion: The study concludes that while PSTs have some foundational skills in mathematical modelling, their overall competencies are insufficient for solving complex modelling tasks independently. This finding points to a need for teacher education programs to place greater emphasis on developing modelling skills through practice, reflection, and guided instruction. At the same time, the study highlights the potential of AI as a valuable educational resource. AI can accelerate the learning process by providing worked examples, alternative solutions, and creative perspectives on modelling tasks. However, the findings also caution against over-reliance on AI. If PSTs depend too heavily on AI-generated solutions without critical evaluation, their own problem-solving and reasoning abilities may be undermined. Therefore, AI should be integrated into teacher education thoughtfully, with clear guidelines for its use as a supportive rather than substitutive tool. By balancing human learning with technological support, teacher educators can better prepare future teachers to engage their students in meaningful mathematical modelling activities.

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
47	3	13

مقاله پژوهشی

ارزیابی عملکرد دانشجو معلمان در تکالیف مدل‌سازی ریاضی: مقایسه با پاسخ‌های هوش مصنوعی

نفیسه آزادی[۱]، ابراهیم ریحانی[۱]، مجتبی قربانی[،۱]، رضا ابراهیم‌پور[۲]، سعید حق جو[۳]*

[۱] *گروه ریاضی، دانشکده علوم، دانشگاه تربیت دبیر شهید رجایی، تهران، ایران*

[۲] *مرکز علوم‌شناختی، مؤسسه علوم و فناوری‌های همگرا (ICST)، دانشگاه صنعتی شریف، تهران، ایران؛ مدرسه علوم‌شناختی، پژوهشگاه دانش‌های بنیادی (IPM)، تهران، ایران*

[۳] *دانشگاه فرهنگیان، بوشهر، ایران*

چکیده

پیشـینه و اهداف: مدل‌سـازی ریاضـی به طور فزاینده‌ای به بخش مرکزی آموزش ریاضـی تبدیل شـده اسـت، زیرا به دانش‌آموزان کمک می‌کند مفاهیم انتزاعی ریاضـی را به موقعیت‌های واقعی پیوند دهند. باوجود اهمیت آن، بسـیاری از دانشـجو معلمان ابتدایی در طراحی یا حل مسـائل مدل‌سـازی با چالش‌هایی مواجه می‌شـوند که بخشـی از آن ناشـی از مواجهه محدود آنان با این‌گونه فعالیت‌ها در طول دوره‌های تربیت‌معلم است. با توسعه سریع هوش مصنوعی فرصت‌های تازه‌ای برای مقایسه رویکردهای حل مسئله انسانی با راه‌حل‌های تولیدشده توسط ماشین به وجود آمده است. هدف این مطالعه بررسی نحوه عملکرد دانشـجو معلمان ابتدایی و یک سـامانه هوش مصـنوعی در انجام تکالیف مدل‌سـازی ریاضـی اسـت. به طور خاص، این پژوهش به دنبال درک نقاط قوت و ضـعف شـایسـتگی‌های مدل‌سـازی دانشـجو معلمان اسـت، درحالی‌که هم‌زمان نقش بالقوه هوش مصنوعی به‌عنوان یک ابزار حمایتی در آموزش ریاضی را نیز بررسی می‌کند. هدف پژوهش صرفاً ارزیابی عملکرد نیست، بلکه تأکید بر اهمیت گسترده‌تر ادغام هوش مصنوعی در آموزش معلمان است که هم فرصت‌ها و هم مخاطرات آن را برجسته می‌سازد.

روش‌ها: نمونه‌ای در دسـترسـی شـامل ۵۰ دانشجو معلم در این پژوهش شـرکت کردند. از شـرکت‌کنندگان خواسـته شـد آزمون محقق-سـاختی را تکمیل کنند که شـامل تکالیف مدل‌سـازی ریاضـی برای ارزیابی جنبه‌های مختلف شـایسـتگی مدل‌سازی بود. روایی صوری و محتوایی ابزار توسط دو متخصص آموزش ریاضی بررسی و تأیید شد تا تناسب و ارتباط سـؤالات تضـمین شـود. داده‌ها به‌صـورت نوشـتاری جمع‌آوری و با رویکردهای کمّی و کیفی تحلیل شـدند. پاسـخ‌های دانشجو معلمان بر اساس چارچوب شایسـتگی گریفراس و ماس بررسی گردید؛ چارچوبی که معیارهای ساختاریافته‌ای برای ارزیابی مهارت‌های مدل‌سازی مانند درک مسئله، به‌کارگیری ریاضیات، تفسیر نتایج و اعتبارسـنجی راه‌حل‌ها ارائه

تاریخ دریافت: ۲۵ مرداد ۱۴۰۴
تاریخ داوری: ۰۳ مهر ۱۴۰۴
تاریخ اصلاح: ۰۶ آبان ۱۴۰۴
تاریخ پذیرش: ۲۸ آذر ۱۴۰۴

واژگان کلیدی:
مدلسازی ریاضی
هوش مصنوعی
چت جی‌پی‌تی
دانشجو معلمان

* نویسنده مسئول
mghorbani@sru.ac.ir
۰۹۱۲-۴۰۷۴۳۹۸

می‌دهد. همچنین پاسخ‌های تولیدشده توسط سامانه هوش مصنوعی برای همان تکالیف جمع‌آوری و از نظر راهبرد، دقت، خلاقیت و کیفیت راه‌حل با پاسخ‌های انسانی مقایسه شدند. پژوهش حاضر از طراحی کیفی با تحلیل محتوای استقرایی هدایت‌شده بهره گرفت که به پژوهشگران امکان داد داده‌ها را به طور نظام‌مند کدگذاری و تفسیر کنند و هم‌زمان مقایسه‌هایی میان عملکرد انسانی و هوش مصنوعی انجام دهند.

یافته‌ها: یافته‌ها الگوهای مهمی را نشان دادند. بسیاری از دانشجو معلمان ضعف‌هایی در مدل‌سازی ریاضی از خود بروز دادند، به‌ویژه در مراحل اعتبارسنجی و تفسیر. پاسخ‌های آنان اغلب ناقص یا فاقد استدلال کافی بود. بااین‌حال، مشارکت‌کنندگان در معیار «به‌کارگیری ریاضیات» عملکرد نسبتاً بهتری داشتند، به این معنا که می‌توانستند فرایندهای ریاضی را اجرا کنند؛ اما در برقراری ارتباط معنادار آن ها با زمینه‌های واقعی مشکل داشتند. تفاوت معناداری از نظر آماری بین شرکت‌کنندگان بر اساس جنسیت یا پیشینه تحصیلی مشاهده نشد که نشان می‌دهد فارغ از ویژگی‌های جمعیت‌شناختی، دشواری‌های مربوط به مدل‌سازی در میان دانشجو معلمان ابتدایی رایج است. در مقابل، سامانه هوش مصنوعی راه‌حل‌هایی منسجم‌تر و ساختارمندتر ارائه داد، به‌ویژه در زمینه اعتبارسنجی و خلاقیت. هوش مصنوعی قادر بود توضیحات گام‌به‌گام و رویکردهای متعددی برای حل یک مسئله تولید کند که می‌تواند به‌عنوان الگو برای دانشجو معلمان مورداستفاده قرار گیرد. بااین‌حال، پاسخ‌های هوش مصنوعی نیز بی‌نقص نبودند و در برخی موارد ساده‌سازی بیش از حد یا فرضیات بدون توجیه کافی در بین پاسخ ها مشاهده شد.

نتیجه‌گیری: نتایج نشان داد اگرچه دانشجو معلمان ابتدایی دارای برخی مهارت‌های پایه‌ای در مدل‌سازی ریاضی هستند، اما شایستگی کلی آنان برای حل مستقل تکالیف پیچیده مدل‌سازی ناکافی است. این یافته بیانگر ضرورت توجه بیشتر برنامه‌های تربیت‌معلم به توسعه مهارت‌های مدل‌سازی از طریق تمرین، بازاندیشی و آموزش هدایت‌شده است. درعین‌حال، پژوهش بر پتانسیل هوش مصنوعی به‌عنوان منبعی ارزشمند در آموزش تأکید دارد. هوش مصنوعی می‌تواند با ارائه مثال‌های حل‌شده، راه‌حل‌های جایگزین و دیدگاه‌های خلاقانه، فرایند یادگیری را تسریع کند. بااین‌وجود، یافته‌ها هشدار می‌دهند که اتکای بیش از حد به هوش مصنوعی می‌تواند توانایی‌های استدلالی و حل مسئله دانشجو معلمان را تضعیف کند؛ بنابراین، هوش مصنوعی باید به‌صورت سنجیده و با دستورالعمل‌های روشن به‌عنوان ابزار حمایتی و نه جایگزین در آموزش معلمان به کار گرفته شود. با ایجاد توازن میان یادگیری انسانی و پشتیبانی فناورانه، مربیان معلم می‌توانند معلمان آینده را بهتر برای درگیرکردن دانش‌آموزان در فعالیت‌های معنادار مدل‌سازی ریاضی آماده سازند.

Introduction

Teaching and learning mathematical modeling are a key area in mathematics education, enhancing students' understanding, critical thinking, problem-solving skills, and adaptability to real-world contexts [1-6]. Developing these skills early supports both academic success and future professional competencies. Mathematical modelling is challenging to teach and learn, but teachers with appropriate training and support can successfully design and implement modelling activities [7-9]. Modelling enhances understanding, retention, and positive attitudes toward mathematics [10], highlighting the need to assess PSTs' modelling competencies.

The 21st-century educational landscape has been transformed by technology, especially artificial intelligence (AI) [11]. Future math education will likely feature blended learning environments, enhanced computational thinking, data literacy, problem-solving, critical thinking, and interdisciplinary connections [12]. Considering the importance of mathematical modelling in the education of elementary education pre-service teachers (PSTs), as well as the advancements in technology and the need to pay attention to AI as a field that impacts education, particularly in mathematics education, this study focuses on the learning process and modelling competency of elementary education PSTs in solving mathematical modelling tasks, rather than on the instruction or teaching methods of modelling. Accordingly, the study aims first to assess the modelling competency of elementary education PSTs in solving mathematical modelling tasks; then to examine the responses of ChatGPT to mathematical modelling questions; and finally, to compare the performance of PSTs with that of AI, ChatGPT. Therefore, the research questions are stated as follows:

- How do elementary education PSTs perform in each of the seven criteria for mathematical modelling competencies?

- How does the performance of elementary education PSTs compare to that of ChatGPT in responding to different types of modelling problems?

- Is there a difference in the modelling competencies of elementary education PSTs based on gender and educational background?

Review of the Related Literature

Mathematical Modeling

Blum and Borromeo Ferri conceptualize mathematical modelling as a process of transition between the real world and the world of mathematics [3]. According to Blum and Leiß, the seven-step process of mathematical modelling can be depicted using a modelling cycle [13], as shown in Figure 1.

The modelling cycle begins with understanding the situation and constructing a situational model [13-15]. The problem is then simplified and structured to form a real model, identifying key variables and relationships [14-15]. This real model is converted into a mathematical model for calculations and analysis, producing mathematical results that require interpretation in context. Validation follows, ensuring logical consistency and accuracy, often involving comparison of models and critical reflection [13][15]. Finally, the solution is communicated for external evaluation and discussion, completing the ideal modelling cycle, though iterative movement between stages is often necessary [16-17].

AI

Advancements in information and communication technologies have led to scientific transformations, such that in recent years, AI technology has been present in almost all scientific fields. According to Wang and Sun, AI is a new field in technology designed for researching and developing theories, methods, technologies, and application systems to simulate, enhance, and advance human intelligence [18]. Recently, initiatives and projects have been launched in various countries to introduce AI and machine learning into the education system [19].

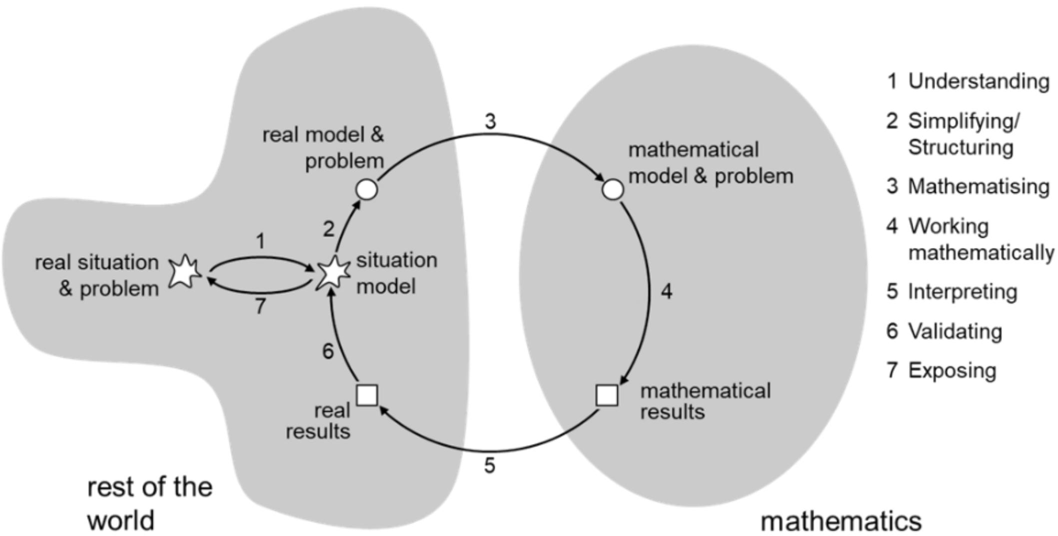


Fig. 1: Diagram of the mathematical modelling cycle proposed by [13]

One of the innovations with the potential to bring about profound changes in the education system is ChatGPT [20], which is an advanced language-processing AI developed by OpenAI [12]. ChatGPT is available in versions GPT-3, GPT-3.5, and GPT-4. Mathematics is continually employed to design innovative AI algorithms that enhance both power and efficiency. Additionally, it can aid in identifying and understanding the limitations of AI [21]. AI-based chatbots, like ChatGPT, can solve basic mathematical and logical problems, but they may not provide reliable results for more complex issues [22].

AI and Mathematics Education

AI, capable of independent thinking and creativity, enhances learning efficiency and reduces study time [18], [23]. Generative AI models like ChatGPT have transformed education, changing how students approach mathematical problem-solving [24]. Generative AI, as a pioneering technology, has challenged traditional teaching and learning methods, and the performance of AI-based chatbots like Google Bard (now Gemini), ChatGPT, and Bing Copilot (now Copilot) in solving mathematical problems, especially in educational settings, has attracted significant attention [25]. These systems have demonstrated their ability to solve basic and moderately complex math problems. However, they may not always provide the best or most efficient solution. Teachers must ensure that students benefit from this valuable technology without becoming overly reliant on it for solving complex problems [26]. It is anticipated that AI will bring significant transformations to mathematics education. Therefore, teachers and instructors must stay informed about the latest advancements and adapt to them, while remaining committed to delivering high-quality education [12].

Chatbots are increasingly recognized as effective learning tools in education. They provide quick, clear, and personalized support, helping students understand concepts, solve problems step-by-step, improve math skills, and explore unique solutions [22] [27]. Chatbots can enhance math education, enable learning at an individual pace, and make mathematics more accessible, while students take a more active role in learning. However, their use should complement, not replace, traditional learning [21].

Parra et al. showed that AI can help students explore geometry concepts under teacher supervision [25], though chatbots may struggle with real-world tasks and over-reliance may reduce human interaction [26]. Lee and Yeo developed an AI chatbot acting as a virtual student to support PSTs' questioning and teaching practice [28].

Spreitzer et al. found that different ChatGPT versions possess basic problem-solving skills, but effectiveness declines with task complexity [26]. Yang et al. reported stronger modelling competencies among German PSTs compared to China and Hong Kong [29]. Nguyen and Tran observed improvements in modelling skills among Vietnamese PSTs [30]. Geiger et al. proposed a framework to enhance teachers' modelling competencies holistically [17].

Dasari et al. showed that students relying solely on AI performed worse than those taught by teachers or assisted by ChatGPT, emphasizing both AI's potential and teachers' essential role [31]. Gerber et al. highlighted the importance of integrating technology training into teacher education [32]. Guerrero–Ortiz demonstrated that technological environments support PSTs in designing modelling tasks, understanding real-world problems, performing complex calculations, and applying mathematical concepts [33].

Research on elementary mathematical modelling, particularly teacher support and AI-related skills, remains limited [9][19]. Therefore, more research must be conducted in the area of teaching mathematical modelling at the elementary level. According to Wei et al., teaching mathematics at the elementary level is closely related to authentic real-world issues [34], which is one of the key characteristics of mathematical modelling activities.

Given the background of the present study, the key to this lies in the hands of teachers. To enhance the teaching and learning of mathematical modelling for elementary students, it is essential to develop teachers in this area. Therefore, the focus of this research is on examining the modelling competencies of pre-service elementary teachers. Additionally, the need for research in the field of AI prompted the authors to compare the responses of pre-service elementary teachers with those of advanced AI technology.

Methods

Participants

A random sample of 50 pre-service elementary teachers, from the cohorts of 2021 to 2023, was selected to answer the research questions (Table 1). Of these PSTs, 22 were female, and 28 were male. The sample consisted of 15 students from the 2021 cohort, 20 from the 2022 cohort, and 15 from the 2023 cohort of the university.

Instruments

To ensure usability, the mathematical modeling questions were reviewed by elementary education experts and pilot-tested with a small group of pre-service teachers. Necessary adjustments were made to enhance clarity, reliability, and validity, ensuring the results accurately reflect PSTs' competencies in mathematical modeling. The pilot test showed acceptable reliability (Cronbach's alpha = 0.8). Three researcher-developed problems, aligned with the mathematical modelling competency criteria, were provided to PSTs. The problems allowed multiple solution strategies. The participants had 24 hours to complete the assignments.

Procedure

The responses of PSTs to the questionnaire items were scored based on the mathematical modelling competency framework by Greefrath and Maaß [15]. Subsequently, the questions were answered by ChatGPT AI, and its responses were compared with those of the PSTs. After various descriptions were made, the significance of gender and educational background of the pre-service elementary teachers was examined, and questions one to three were answered completely. The significance of the differences was analyzed using SPSS software. Before conducting the main test, a pilot test consisting of six mathematical modelling problems (including three levels of modelling assignments) was administered.

Table 1: Number of Participants

Year of Entry	Girls	Percentage of Girls	Boys	Percentage of Boys	Total
Entry 2021	8	53%	7	47%	15
Entry 2022	6	30%	14	70%	20
Entry 2023	8	53%	7	47%	15
Total	22	44%	28	56%	50

Table 2 shows the discrimination and difficulty indices of the pilot test questions. According to Table 2, questions 1, 3, and 5 were considered for the main test with minor modifications. Table 3 shows the discrimination and difficulty indices of the pilot test questions. These indices were calculated based on the percentage of correct responses from different groups of participants, the top and bottom 25% of respondents, commonly used in item analysis. According to Table 1, questions 1, 3, and 5 were selected for the main test with minor modifications.

Table 2: Discrimination and Difficulty Index

Question	Discrimination Index	Difficulty Index
1	0.733	0.8
2	0.5	0.3
3	0.6	0.4
4	0.16	0.29
5	0.61	0.12
6	0.28	0.11

Design

This study employed a mixed-methods approach (a combination of quantitative and qualitative methods) of the sequential explanatory type. In this way, the main focus of the work is on qualitative results, and the use of quantitative results aims to shape the qualitative data [35]. This study used a qualitative design based on deductive (directed) content analysis. Qualitative content analysis is "An approach to documents that emphasizes the role of the investigator in the construction of the meaning of and in texts" [36]. "DQA is an attempt to encourage researchers to test, refine, reformulate, refute, and replace theoretical models qualitatively" [37]. In deductive qualitative content analysis, the analysis and interpretation of data are based on categories and themes that the researcher has defined prior to data collection. In other words, the data analysis is guided by existing theoretical frameworks [38].

To evaluate the modelling competencies of pre-service teachers in this study, an analytical framework was employed that includes specific criteria for assessing performance in solving modelling tasks. In this regard, the criteria proposed by Greefrath and Maaß for evaluating modelling competencies were selected as the basis for the analysis [15]. Table 3 shows the criteria for evaluating modelling competencies presented by Greefrath and Maaß [15].

Greefrath and Maaß employed the modelling competency framework in their research [15]. As shown in Table 3, the criteria for evaluating modelling competencies used in their study include understanding, gathering information, mathematization, use of mathematics, interpretation, validation, and discussion of results.

Kang and Noh classified modelling problems into three levels based on ambiguity and information completeness [39]:

- Level 1 is clearly defined with minimal ambiguity,
- Level 2 contains some ambiguity and incomplete information, and
- Level 3 is open-ended with incomplete or excessive information, requiring students to analyze, strategize, and implement solutions.

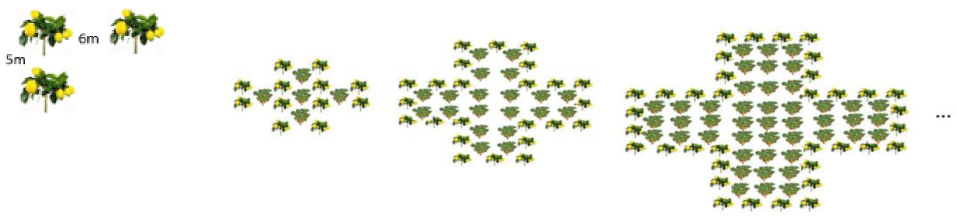
The researcher-developed questionnaire was designed according to these three levels and aligned with the mathematical modelling competency criteria [39], as detailed in Table 4.

Based on the mathematical modelling framework [15] (understanding the problem, gathering information, mathematization, using mathematics, interpretation, validation, and discussion of results), the responses were scored. The scoring followed the classification by Spreitzer et al. [26] as shown in Table 5.

Table 3: Criteria for Evaluating Modelling Competencies [15]

Competence	Description	Competence
Understanding	PSTs construct a mental model of a specific situation and understand the question.	Understanding
Gathering Information, Analyzing Resources (Transforming)	Separating important information from unimportant information in a real situation.	Gathering Information, Analyzing Resources (Transforming)
Mathematization	Translating real situations into appropriate mathematical models (sentence, equation, shape, etc.).	Mathematization
Use of Mathematics	Working with the mathematical model.	Use of Mathematics
Interpretation	Relating the results of the model to the real situation and achieving real outcomes.	Interpretation
Validation	Checking the appropriateness of real results and comparing and evaluating different models concerning a real situation.	Validation
Discussion of (Potentially) Contradictory Results	Linking the model's responses to the real situation and answering the question.	Discussion of (Potentially) Contradictory Results

Table 4: Test Questions

Row	Test Questions
1	A school with 282 students has decided to plant the maximum number of orange and lemon saplings on agricultural land with the following pattern, with the help of the students (the planting pattern for the saplings is illustrated below from left to right. The first figure of the pattern begins with four orange saplings and 12 lemon saplings surrounding them, and this continues in subsequent figures. The horizontal distance between the saplings is 6 meters, and the vertical distance between two saplings is 5 meters. Lemon saplings are placed all around the field, while orange saplings are positioned in the center of the field. It is known that the circumference of the agricultural land is 264 meters. a) How many orange and lemon saplings should the school prepare for planting in this agricultural land? b) How many students will you divide into groups for planting each sapling? 
2	The school is looking to purchase gifts for outstanding students. There are two options to choose from at the store: each pair of ping pong paddles costs 100,000 z, and a chess set costs 150,000 z. For every 5 pairs of ping pong paddles purchased, you get one pair free, and if you buy 2 chess sets, you get one chess set free. If the school needs a certain number of gifts, which option is better [40]?
3	Recently, charitable organizations have started campaigns to collect bottle caps. You might wonder why bottle caps. One reason is their small size, as they can be easily gathered without requiring a lot of space. These caps are made of high-quality materials, and after recycling, they are used in various industries to produce office and even commercial supplies. On the other hand, the body of the bottles is made of "polyethylene terephthalate" (265 degrees Celsius) [41], and the labels around them are mostly made of "PVC." However, the bottle caps are made of a different material, "PE polyethylene" (with a melting point

Row	Test Questions
	of 143.3 degrees Celsius) [42]. Each of the materials mentioned has a different melting point, and since it is not possible to predict two melting points in a recycling machine, putting them together can cause issues. Both plastic caps and plastic bottles are recyclable, but due to their different melting points, they cannot be recycled together. Considering the costs and time involved, in some cases, the relevant centers may decide against recycling the caps, and both the caps and labels end up being discarded. These factors make bottle caps unique materials that are unwanted by both the public and recyclers, while also being distinct and special materials from which good raw materials can be derived. Collecting these caps helps protect the environment, encourages citizen responsibility, and aids those in need. Many elderly individuals and patients require wheelchairs but cannot afford them. The charitable efforts that use the proceeds from bottle cap sales to purchase wheelchairs for such individuals significantly improve their lives. Engaging in charitable work and spreading this practice can enhance individuals' sense of responsibility. How many days should pass for your university to be able to procure a high-quality wheelchair for a nursing home?

Table 5: Scoring Method for PSTs' and ChatGPT's Responses

Score	0	1	2	3
Description	Not Attempted	Attempted but incorrect	Attempted but incomplete	Completed and sufficient

Based on the selected scoring method, the maximum score for the questionnaire is 84 points.

Results and Findings

Responses of the PSTs

Table 6 shows the minimum and maximum scores obtained by the PSTs, along with the standard deviation and mean. The overall mean score is 26.16, with a standard deviation of 21.53.

In Table 7, the average total scores of PSTs were compared across the two variables of gender and academic background. It was observed that the average total score of male PSTs (27.35) was higher than that of female PSTs (24.63). The average scores of PSTs admitted in 2021 (30.46) were higher than those admitted in 2022.

Additionally, Figure 2 displays the scores of the PSTs. The highest score achieved was 68, obtained by three PSTs. Five individuals received a score of one, which is the lowest score. None of the PSTs received a score of zero. Fourteen individuals scored above half of the maximum score (i.e., a score of 42), while 36 individuals scored below half of the maximum score.

Table 6: Statistical Information of PST Teachers' Scores

Number	Minimum Score	Maximum Score	Mean	Standard Deviation
50	1	63	26.16	21.53

Table 7: The Mean and Standard Deviation of PSTs' Scores Across Different Variables

	Gender		Educational Background		
	Girl	Boy	2021	2022	2023
Mean Scores	24.63	27.35	30.46	27.1	21.78
Standard Deviation	23.08	19.72	21.47	22.64	17.85

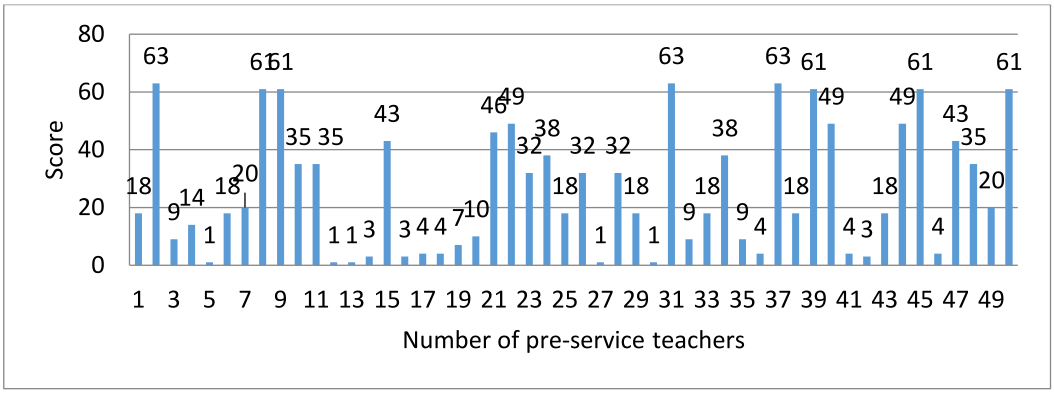


Fig. 2: Status of PSTs ' Scores in the Exam

The percentage of responses from PSTs to each of the exam questions is shown in Table 8. In this analysis, since a score of zero does not reflect the ability to evaluate the thinking of the PSTs, the scores of 1, 2, and 3 were examined for each question.

Table 8 shows that PSTs performed best in the 'use of mathematics' criterion for question 2 (score 1, 48%), indicating basic procedural engagement, but overall, participation in 'validation' was extremely low (0% for questions 1b, 2, and question 3), revealing a major gap in evaluating results.

In question 1a, most PSTs scored well in 'understanding the problem', 'gathering information', and 'mathematization' (score 3, ~42%), identifying context and relevant quantities, though few articulated assumptions or reflected on alternatives. Interpretation and discussion were minimal (score 1, 2%), showing limited connection between mathematical work and real-world meaning.

Question 1b mirrored this pattern, with good scores in problem understanding (36%) but no engagement in validation. Question 2 showed frequent procedural use of mathematics (score 1, 48%), with some algorithmic reasoning but limited contextual consideration. Question 3 exhibited generally low participation, with partial engagement in criteria such as gathering information, mathematization, use of mathematics, and interpretation (score 2, 10%), and many responses lacked coherent modeling strategies or depth.

Overall, students demonstrated stronger performance in early modeling phases like problem understanding and mathematization, but consistently struggled with validation and interpretation. This pattern suggests a fragmented understanding of the modeling cycle, where modeling is seen more as a problem-solving routine than as an iterative reasoning process.

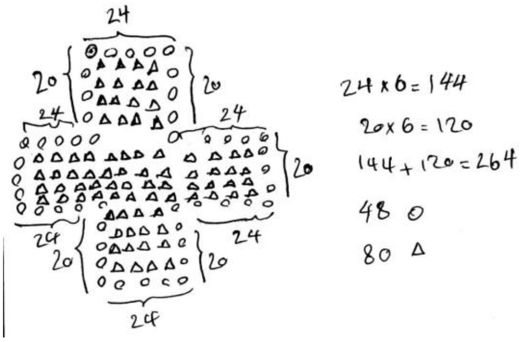
A sample of PSTs’ written responses is shown in Table 9, illustrating common reasoning patterns and recurring difficulties in transitioning from mathematical formulation to meaningful interpretation.

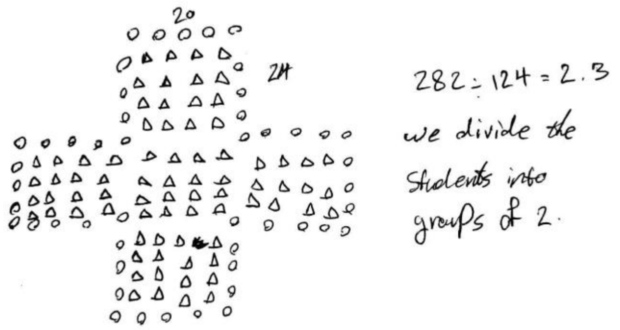
A review of the average scores in each criterion of modeling competency, along with the qualitative analysis of PSTs’ responses (Table 9), reveals that student teachers demonstrated relatively better performance in the 'use of mathematics' criterion, while their weakest area was clearly 'validation'. This finding is consistent with what emerged in Table 8, where performance across criteria appeared interrelated. In many cases, PSTs who struggled to understand the problem also faced difficulties in subsequent phases such as information gathering, mathematization, and interpretation. This pattern suggests that modeling competencies are not isolated skills but function as a connected cognitive process.

Table 8: Percentage of PSTs' Responses to the Mathematical Modelling Competency Criteria

	Mathematical Modelling Competency Criteria																				
	Understanding			Gathering Information			Mathematization			Use of Mathematics			Interpretation			Validation			Discussion		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1a	6	14	42	6	14	42	4	14	42	12	14	36	2	6	32	20	16	20	2	6	32
1b	24	2	36	10	22	30	24	8	24	24	12	20	24	18	20	0	32	20	20	18	20
2	16	22	6	32	22	6	40	14	6	48	20	6	4	24	6	0	20	6	26	20	6
3	0	8	8	0	10	6	0	10	6	0	10	6	6	10	6	0	2	6	4	6	6

Table 9: A Sample of PSTs' Responses to the Questionnaire Questions

The criteria for mathematical modelling competency.	Score	Question	Sample response from PSTs
Understanding the problem: The PST drew a model of how to plant trees according to the situation of the problem, and therefore understood the question well.	3	1a	
Gathering information: The student's response indicates a lack of complete understanding in searching for information and the ambiguous assumptions of the problem.	2	3	"We have about 1,900 students, and including the administrative staff, we are around 2,000 people in total. We are provided with drinks along with lunch twice a week, so approximately 4,000 bottle caps are collected".
Mathematization: The student-teacher did not appropriately transform the simplified real-world situation, along with the information and assumptions of the problem, into correct mathematical expressions to form a mathematical model. Since the number of gifts required for the students was not specified in the question, the student should have considered different scenarios for the number of gifts but only examined one scenario.	1	2	"The ping-pong racket is more cost-effective because the price of each one is less than 100 z, whereas the price of each chess set is 100 z. Chess: 1 + 2 = 3 Ping-pong racket: 1 + 5= 6"
Use of mathematics: The student presented an incomplete solution, which ultimately did not lead to a conclusion with proper reasoning.	2	3	"The majority of bottle usage at the university is for malt drinks and yogurt drinks served with lunch, which happens twice a week. There are approximately 1,500 students, and around 3,000 bottle caps are collected weekly. Additionally, around 1,000 mineral water bottle caps can be collected occasionally. In total, 4,000 bottle caps are gathered, and it seems that within one month, a wheelchair could be purchased.

<table>
<tr><td>Interpretation: The student correctly solved the problem mathematically and then provided an appropriate interpretation of the mathematical results to relate them to real-world outcomes.</td><td>3</td><td>2</td><td>Ping-pong is cheaper.
Let’s assume we need 10 gifts.
For ping-pong: 10 = 4 we buy + 1 free + 5 we buy.
For chess: 10 = 1 we buy + 2 we buy + 1 we take + 2 we buy + 1 we take + 2 we buy.
Without using discounts, we spend 900 z on ping-pong. For chess, we have:
1,050 = 1,500 × 7.</td></tr>
<tr><td>In the validation phase, a student grouped participants but did not check the results for real-world consistency, specify group sizes, or assess the number of groups. Proper validation requires carefully reviewing the problem and ensuring the feasibility of the solution.</td><td>2</td><td>1b</td><td>
</td></tr>
<tr><td>Discussion of results: The student related the obtained results to the real-world situation, thus answering the question.</td><td>3</td><td>2</td><td>5 pairs of ping-pong = 1 pair free
2 chess sets = 1 free
If we need a total of 50 gifts:
<table><tr><td>Pairs of ping-pong</td><td>free</td></tr><tr><td>5</td><td>1</td></tr><tr><td>50</td><td>10</td></tr></table><table><tr><td>Chess sets</td><td>free</td></tr><tr><td>2</td><td>1</td></tr><tr><td>50</td><td>25</td></tr></table>For 7.5 million z, we get 3.7 million gifts (better).
For 5 million z spent, we get 1 million gifts.</td></tr>
</table>

For example, effective validation of a mathematical model depends on one’s ability to interpret results meaningfully within the context. Many PSTs failed to critically evaluate the plausibility of their solutions or consider whether their model realistically represented the original situation. Instead, they often accepted the numerical output as final, without checking it against real-world expectations or proposing alternative explanations.

This lack of reflective thinking indicates a procedural view of modeling, where reaching an answer is prioritized over reasoning about the model's validity or broader implications. The responses suggest that PSTs may perceive modeling as a linear process — once the math is done, the task is complete — rather than as a cyclical inquiry where interpretation and validation guide refinements to the model. Moreover, the absence of alternative responses or justifications in many answers points to limited metacognitive engagement. Students rarely questioned their assumptions or evaluated whether different modeling approaches could yield more realistic or insightful results. This cognitive gap leads not only to less sophisticated models but also to answers that may be mathematically correct yet practically unfeasible. The findings underscore the need for more explicit emphasis on interpretation and validation in preservice teacher training, as these components are essential for developing modeling competencies that extend beyond symbolic manipulation.

In Tables 10 and 11, the significance of the total score differences among PSTs has been examined. Since the calculated P-value is greater than 0.05, the observed differences in gender and academic background of the PSTs in responding to the questionnaire questions are not significant.

Table 10: The significance of the gender variable in the total score

	Variable	Mann-Whitney U	P-Value
Total score	Gender	293.500	0.776

Table 11: The significance of the academic background variable in the total score

	Variable	Kruskal-Wallis H	P-Value
Total score	Academic background	1.760	0.415

ChatGPT's responses to the exam questions

To obtain more suitable responses from ChatGPT using the mathematical modelling cycle and emphasizing the criteria for mathematical modelling competency, this research utilized neutral prompts based on the study by Spreitzer et al. [26], such as 'verify your calculations', 'provide validation for the task', or are there alternative solutions?' to minimize the impact of external factors. At this stage, the questionnaire questions were presented to ChatGPT based on the prompts provided by Spreitzer et al. [26]. Table 12 shows the scores achieved by ChatGPT in response to the questionnaire questions.

As observed, ChatGPT achieved a total score of 54 in response to the mathematical modelling questions. Table 12 shows that ChatGPT performed poorly in responding to Question 1(a), misunderstanding the question (score of 1), which resulted in an inability to identify the correct pattern for planting saplings. This error also affected the subsequent criteria (all criteria in Question 1(a), except for the interpretation criterion, scored 1). ChatGPT's mistake in determining the appropriate pattern and number of saplings in part (a) impacted the mathematical modelling competency criteria in Question 1(b) as well, leading to incorrect responses. Thus, while the understanding of the problem, information gathering, mathematical formulation, and interpretation were appropriately conducted, they were based on an incorrect assumption, resulting in a score of 2. The criteria for the use of mathematics (incorrect mathematical solution), validation (inadequate validation), and discussion (providing an incorrect answer) received a score of 1. In Question 2, there was only one mistake: failing to consider all scenarios for the number of gifts needed for the school, as ChatGPT only examined one scenario. However, it correctly executed this scenario and achieved a score of 2 across all mathematical modelling criteria. GPT was successful in answering question 3. It also scored 3 in all the competency criteria for modelling. The responses of GPT to the mathematical modelling questions in Table 13, along with the analysis of the answers, are provided.

Analysis of GPT's responses across the three modeling problems reveals both potential and limitations in its modeling competency, particularly when compared to PSTs. One of the most prominent weaknesses in GPT's performance was observed in Problem 1, which required identifying a visual pattern. GPT struggled to accurately process and interpret the shape-based data and misidentified the fourth pattern. This suggests a limitation in visual-spatial reasoning, a key component in pattern recognition. Furthermore, GPT incorrectly calculated the number of trees and proposed grouping students into units of 4 or 5, overlooking the requirement that 282 students be organized into groups of 2 and 3 — a sign of insufficient contextual reasoning and interpretation of constraints embedded in the problem.

Table 12: Scores achieved by ChatGPT in each criterion of mathematical modelling competency

	Mathematical Modelling Competency Criteria						
	Understanding	Gathering Information	Mathematization	Use of Mathematics	Interpretation	Validation	Discussion
1a	1	1	1	1	2	1	1
1b	2	2	2	1	2	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3

Table 13: Responses of ChatGPT to the mathematical modelling exam questions.

AI Response	Mathematical Modelling Competency Framework for GPT Responses

1

Let's analyze the image provided and determine the number of orange and lemon seedlings, as well as how to divide the students into groups.

Part a: Number of Orange and Lemon Seedlings

From the image, it seems the arrangement of the seedlings forms a specific pattern where the lemon seedlings are placed around the perimeter and the orange seedlings are in the middle.

Counting Seedlings

1. **Counting Lemon Seedlings:**
   - Lemon seedlings are around the edge. To count these, we need to identify the outermost layer of the grid.
2. **Counting Orange Seedlings:**
   - Orange seedlings are in the inner part of the field, arranged in a cross pattern.

Image Analysis

By examining the image, we can count the number of seedlings manually.

- Each side of the cross pattern of orange seedlings appears to be composed of 7 seedlings.
- There are 4 sections of the cross, plus the central block.

Let's count the orange seedlings first:

- Each arm of the cross has 7 seedlings.
- There are 4 arms, so $4 \times 7 = 28$ seedlings.
- Additionally, there is a central block with 1 seedling.
- Total number of orange seedlings: $28 + 1 = 29$.

Now, counting the lemon seedlings around the perimeter:

- Each side of the square around the cross has 8 seedlings on each side of the cross, making it 9 on each straight side.

Total number of lemon seedlings:

- 4 corners, each counted twice (once for each adjoining side): $4 \times 8 = 32$.
- Subtract the 4 corners counted twice: $32 - 4 = 28$.

Part b: Dividing Students into Groups

Given:

- Total number of students: 282.
- Total number of seedlings (orange + lemon): $29 + 32 = 61$.

To plant each sapling, students are divided into groups:

- Each group should handle one seedling.
- Number of students per seedling: $\frac{282}{61} \approx 4.62$.

This means each group should consist of approximately 4 or 5 students, rounded for practical purposes.

Summary

- **a) Number of Seedlings:**
  - Lemon seedlings: 32.
  - Orange seedlings: 29.
- **b) Groups of Students:**
  - Each group should ideally consist of 4 or 5 students, considering practical organization.

If there are alternative solutions or different patterns in the image that suggest a different number of seedlings, the above method provides a validated calculation based on the given distances and arrangement.

Understanding: GPT failed to fully understand the first problem, recognizing the general placement of saplings but misjudging their number and ignoring key patterns, reflecting an incomplete grasp of the task.

Information Gathering: GPT poorly executed the information-gathering step, failing to identify the correct pattern, misusing data, and ignoring the distances between saplings.

Mathematical Formulation: As a result of the previous steps and inconsistent assumptions, the mathematical model was not well formed, and the quantities and dimensions were not used correctly.

Use of Mathematics: GPT produced inaccurate mathematical results due to incorrect data use and misunderstanding the problem, leading to errors in calculating saplings and student groupings.

Interpretation: Overall, GPT demonstrates good performance in interpretation and explanations, effectively relating mathematical results to real-world contexts.

Validation: In this problem, validation was not performed well, and alternative solutions were not provided.

Discussion: Given the incorrect results obtained from the use of assumptions and inconsistent solutions, the final answer is also incorrect.

2

Let's work through this problem step-by-step, verifying the calculations and comparing the two options to determine which is the better deal based on the given promotions.

Step 1: Clarify the Promotions and Prices

- **Ping pong paddles**:
  - Each pair costs **100,000 z**.
  - For every **5 pairs of paddles**, you get **1 pair free**.
- **Chess sets**:
  - Each set costs **150,000 z**.
  - For every **2 chess sets**, you get **1 chess set free**.

Step 2: Determine Effective Cost for Each Option

Option 1: Ping Pong Paddles

For every **5 pairs of paddles** purchased, the school gets **1 free pair**. This means that for every 6 pairs received, the school only pays for 5 pairs.

1. **Cost of 6 pairs of paddles**:

$$\text{Cost of 6 pairs} = 5 \times 100,000\,z = 500,000\,z$$

- So, the **effective cost per pair** of ping pong paddles (with the discount) is:

$$\text{Effective cost per pair} = \frac{500,000\,z}{6} \approx 83,333.33\,z$$

Option 2: Chess Sets

For every **2 chess sets** purchased, the school gets **1 free chess set**. This means for every 3 chess sets received, the school only pays for 2 sets.

1. **Cost of 3 chess sets**:

$$\text{Cost of 3 chess sets} = 2 \times 150,000\,z = 300,000\,z$$

- So, the **effective cost per chess set** (with the discount) is:

$$\text{Effective cost per chess set} = \frac{300,000\,z}{3} = 100,000\,z$$

Step 3: Comparison of Effective Costs

- The effective cost per **pair of ping pong paddles** is **83,333.33 z**.
- The effective cost per **chess set** is **100,000 z**.

Conclusion for Best Option:

- Based on the **effective cost**, **ping pong paddles** offer a **cheaper per-unit price** (83,333.33 z) compared to chess sets (100,000 z).
- Therefore, **ping pong paddles** would be the **better option** if the school wants to maximize the number of gifts for a given budget.

Validation and Alternative Considerations:

- If the school is concerned with **variety** or **student preference**, chess sets might still be a worthwhile option, even though they are slightly more expensive per unit.
- **Budget Constraints:** If the school has a fixed budget, the analysis would hold, as the effective cost favors ping pong paddles unless the school has a particular reason to prefer chess sets (e.g., if fewer, higher-value gifts are desirable).

Thus, the **ping pong paddles** option provides more gifts for the same amount of money due to the more favorable promotion.

Understanding: GPT understood the problem and its assumptions well but overlooked the unspecified number of students for gift preparation.

Information Gathering: GPT correctly used the relevant information except for the unspecified number of gifts needed for the students.

Mathematical Formulation: GPT failed to complete the mathematical model because it overlooked the number of students, though considering different scenarios could have led to a correct model.

Use of Mathematics: The solutions used are correct. However, not considering all possible scenarios for the number of students led to incomplete results.

Interpretation: The presentation of incomplete results led to an inadequate interpretation. However, within the scope of the obtained mathematical result, a suitable interpretation was provided.

Validation: GPT showed creativity and suggested alternative solutions in the validation section but did not consider all scenarios, making its validation incomplete.

Discussion: Given the solutions and validation of the responses, GPT could have corrected its answer and paid more attention to various scenarios regarding the number of students.

3

To determine how many days it would take for your university to procure a high-quality wheelchair for a nursing home through a bottle cap collection campaign, we can break down the process and perform some calculations. Here's how it might work:

1. Understanding the Collection Process

Assumptions:

- **Weight of a single bottle cap:** Approximately 2.5 grams.
- **Recycling value of polyethylene (PE):** Typically between $0.50 and $1.00 per kilogram.
- **Cost of a high-quality wheelchair:** Typically ranges from $500 to $1,500, depending on quality and features.

2. Calculating Funds from Bottle Caps

Let's assume your university aims to collect a significant number of bottle caps in a certain timeframe. For this example, we will consider a collection of 1 million bottle caps.

Weight Calculation:

- Total weight of caps:

$$1,000,000\,\text{caps} \times 2.5\,\text{g} = 2,500,000\,\text{g} = 2,500\,\text{kg}$$

Revenue Calculation:

- If we assume a mid-range price of $0.75 per kilogram for recycling:

$$2,500\,\text{kg} \times 0.75\,\text{USD} = 1,875\,\text{USD}$$

3. Cost of the Wheelchair

Assuming the cost of a high-quality wheelchair is around $1,200.

4. Timeframe for Collection

To estimate how long it would take to collect 1 million bottle caps, we need to consider:

- **Participation Rate:** How many students are involved in the collection effort.
- **Collection Rate:** How many caps can be realistically collected per student per day.

Example Estimation:

- Let's assume **100 students** participate and each collects **50 caps per day**.

Total caps collected per day:

$$100\,\text{students} \times 50\,\text{caps/student} = 5,000\,\text{caps/day}$$

To collect 1 million caps:

$$\frac{1,000,000\,\text{caps}}{5,000\,\text{caps/day}} = 200\,\text{days}$$

5. Validation of Task

If your university can maintain this collection rate, it would take approximately **200 days** to collect enough caps to fund one wheelchair at $1,200, assuming a stable participation rate and commitment from students.

6. Alternative Solutions

Here are some alternative methods to potentially expedite the process or raise funds more efficiently:

- **Direct Monetary Donations:** Encouraging students and faculty to donate directly could lead to faster funding for the wheelchair. For example, if 120 people donate $10 each, you would reach $1,200 quickly.
- **Collecting Larger Quantities of Plastics:** Expanding the campaign to include entire bottles (not just caps) can yield more revenue, as the plastic in bottles can have a higher recycling value.
- **Sponsorships or Grants:** Seeking sponsorship from local businesses or applying for grants can provide immediate funding without relying solely on bottle cap collection.
- **Awareness Campaigns:** Running awareness campaigns to motivate more students to participate or organize events that facilitate the collection can increase the number of caps collected per day.

Conclusion

To procure a high-quality wheelchair for a nursing home, it could take approximately **200 days** if your university collects 1 million bottle caps at the specified rate. However, alternative methods like direct donations or expanding the recycling effort could significantly reduce the time needed to achieve this goal.

Understanding: In the first stage, GPT demonstrated its understanding by summarizing the information needed to solve the problem. Therefore, it has a suitable grasp of this issue.

Information Gathering: When writing the steps of the work, the assumptions that would be used were stated, such as the price of the wheelchair or the number of collected doors, etc.

Mathematical Formulation: In this stage, the collected information was organized and categorized correctly.

Use of Mathematics: Appropriate mathematical solutions were applied to the data.

Interpretation: Suitable interpretations were made for each stage of the mathematical results, and the results were related to the real world.

Validation: The price of the wheelchair and daily bottle use should be examined, showing GPT can guide humans in evaluating responses against reality.

Discussion: By expressing suitable statements, the result obtained in the previous section was effectively related to the real world, providing an appropriate answer to the question.

Although GPT did correctly interpret the intent that each group should plant one sapling, it failed to fully integrate this understanding into a consistent model. It also neglected spatial relationships in the diagram (such as distances between lemon and orange saplings), which further weakened its information-gathering stage. The failure to account for relevant spatial and quantitative information led to an inadequate mathematical model and ultimately to incorrect solutions. Despite these flaws, GPT rated its own solution as valid, reflecting a lack of metacognitive oversight — it did not question or revise its output in light of inconsistencies.

In Problem 2, GPT's performance improved. While it ignored the open-ended nature of the student population and considered only a single scenario, its use of mathematics and interpretation within that scenario was appropriate. Interestingly, GPT showed signs of creative reasoning in the validation phase — including comparing gift options, referencing economic considerations, and suggesting a bulk purchase strategy. This indicates GPT's potential to introduce alternative perspectives, although it still lacked full awareness of the problem's openness and did not explore multiple interpretations unless explicitly prompted.

Problem 3 demonstrated GPT's best performance. It correctly identified key variables such as the price of the wheelchair, the number of students, and the number of bottles used. GPT's discussion highlighted practical considerations and economic reasoning. However, it still exhibited a tendency to rely on straightforward data extraction and lacked deeper questioning or back-checking of assumptions — essential components of robust modeling.

Across the three tasks, GPT's strengths were most apparent in structured quantitative reasoning and creative validation, while its weaknesses lay in visual interpretation, flexible problem framing, and self-correction. GPT did not demonstrate the ability to identify and revise errors unless guided explicitly — a limitation for its role as an autonomous modeling agent.

When comparing GPT's responses (Table 13) with those of the PSTs (Table 8), key differences emerge. In Question 1a, PSTs outperformed GPT in problem comprehension, information gathering, and mathematization. Their written responses showed a better grasp of visual data and a more grounded sense of realism. However, in Question 1b, GPT surpassed PSTs in mathematical formulation and interpretation, suggesting stronger algorithmic reasoning. The two groups performed similarly in validation, though GPT's validation was more imaginative while PSTs were more grounded in real-world feasibility.

In Question 2, PSTs and GPT showed comparable performance in problem understanding and validation. GPT outperformed in information use and mathematical modeling, possibly due to its ability to generate well-structured solutions, but PSTs demonstrated more critical reflection on data limitations. In Question 3, since most PSTs scored zero, GPT clearly performed better overall.

Broadly speaking, PSTs demonstrated stronger performance in using mathematics but consistently underperformed in validation — confirming the findings from earlier sections. GPT, by contrast, showed limitations in recognizing structure or patterns within problems and lacked the capacity to correct its own flawed solutions. However, its creativity in the validation stage and structured reasoning suggest that GPT can serve as a helpful guide in modeling education.

It is important to emphasize, however, that GPT's responses should not be used in isolation. While GPT can offer diverse perspectives and facilitate initial reasoning, it cannot yet simulate the full spectrum of human metacognition, flexibility, or contextual judgment. Therefore, GPT is best employed as a pedagogical tool to support PSTs, not as a replacement for critical human evaluation or collaborative reasoning. Over-reliance on its outputs could mislead students into accepting incomplete or inappropriate models without proper scrutiny.

Discussion

The study found that PSTs performed best in the 'use of mathematics' criterion but weakest in 'validation', indicating struggles with applying, evaluating, and validating results. These weaknesses may stem from limited prior knowledge and insufficient support [43]. PSTs' competencies are interdependent, suggesting that comprehensive training across all modelling skills is essential to improve their overall performance.

This study highlights AI tools like ChatGPT as practical supports in teacher education. ChatGPT can guide PSTs through mathematical modelling stages, provide step-by-step explanations, alternative approaches, timely feedback, and simulate real-world scenarios, especially supporting areas like validation, while fostering critical thinking and greater cognitive skills. Durandt et al. found that students receiving mathematical modelling instruction showed the greatest growth in competency [44]. Similarly, this study indicates that elementary PSTs have the potential to solve modelling problems and use the modelling cycle, but a lack of formal courses and resources led to incomplete answers, highlighting their weaknesses. No significant differences were observed based on gender or academic background.

Comparison of PSTs' responses with ChatGPT showed that the AI provided step-by-step answers with good explanations but struggled in Problem 1 (pattern recognition) and Question 2 (considering all scenarios), while performing well in Question 3. This highlights the interdependence of modelling competency criteria, where weaknesses in one area affect others. Although ChatGPT offered useful ideas for validation [15], it did not revise its responses accordingly. AI shows limitations in holistic understanding of geometry, as seen in its errors with the first problem [21][25][45]. Consistent with Spreitzer et al. [26], GPT performs poorly on low- to medium-complexity modelling tasks. ChatGPT offered alternative solutions not seen in PSTs' responses, suggesting its role as a supportive tool, especially for validation. AI should complement, not replace, teachers, enhancing student engagement, learning efficiency, and critical thinking while maintaining the essential benefits of in-person classroom interaction [46-47].

Therefore, integrating AI tools like ChatGPT into teacher education curricula can provide practical and meaningful support in modelling education. For example, such tools can be used in workshops or courses to simulate modelling tasks, provide instant feedback, and encourage reflective discussions. This practical integration not only leverages AI's strengths but also maintains the essential role of human educators in guiding and contextualizing learning.

AI serves as an effective educational tool for teaching topics like mathematical modelling. According to Yang et al., AI education can enhance creative thinking and problem-solving skills in learners. In the current study, considering the creativity demonstrated by ChatGPT in solving mathematical modelling problems—especially in the validation phase—it can be argued that ChatGPT can be utilized as a complementary educational tool in teaching mathematical modelling [29]. Similarly, Spreitzer et al. identified two key outcomes for ChatGPT: assisting with simple modelling problem-solving and serving as a supplementary educational resource [26].

It can be argued that pre-service teachers can utilize ChatGPT as a supportive tool for solving modelling problems, particularly in enhancing and improving their modelling competencies. Furthermore, according to

Spreitzer et al., excessive reliance on technology, especially AI, may lead to a decrease in human interaction during the learning process [26]. Therefore, it is essential to pay special attention to the guiding role of the teacher and the cognitive abilities of students, while also reviewing the validity of AI-generated solutions.

Conclusions

The findings of this study reveal that while elementary pre-service teachers (PSTs) possess some basic competencies in mathematical modelling, they face significant challenges in areas such as validation, interpretation, and the meaningful connection of mathematical processes to real-world contexts. These weaknesses suggest that current teacher education programs may not provide sufficient opportunities for PSTs to engage with modelling tasks in a structured and reflective manner. At the same time, the analysis of AI responses shows that tools like ChatGPT can offer structured solutions, creative perspectives, and validation strategies, though they also suffer from limitations such as oversimplifications and gaps in holistic understanding. Together, these findings emphasize the need for a balanced approach in teacher education—one that develops human competencies while making strategic use of AI to scaffold learning.

In light of these insights, the integration of the mathematical modelling cycle into teacher training programs is strongly recommended, using both innovative instructional methods and AI as a supportive tool. Teacher educators should design tasks that require PSTs to reflect, validate, and critique both their own solutions and AI-generated ones, thereby fostering deeper reasoning and critical thinking. The findings also suggest that AI tools like ChatGPT can serve as effective supportive resources by providing step-by-step guidance, creative problem-solving strategies, and immediate feedback. Educators can incorporate these tools into workshops or courses to simulate real-world modelling scenarios, encourage reflective discussions, and enhance student engagement. However, careful pedagogical guidance is necessary to prevent over-reliance on AI and ensure that human reasoning and creativity remain central in the learning process.

One limitation of the present study is the relatively small sample size, which may limit the generalizability of the findings. The study was conducted with the participation of [number] pre-service teachers, which was sufficient for qualitative analysis; however, the results should be generalized to other groups with caution. There are several suggestions for further research: use of larger, more diverse samples of pre-service teachers to improve generalizability and study demographic effects on modelling skills; conducting longitudinal research to assess long-term impacts of AI-assisted instruction on problem-solving, reasoning, and teaching; exploring different educational and cultural contexts (e.g., middle/high schools, various curricula) to test broader applicability; and running comparative studies on different AI supports (chatbots, VR, videos) to identify the most effective approaches.

Authors' Contribution

All authors have the same contribution.

Acknowledgements

The authors would like to express their sincere gratitude to Shiraz Farhangian University for their support and collaboration during this research.

Conflict of Interest

The authors have no conflicts of interest.

References

[1] Greefrath G, Carreira S. Mathematical applications and modelling in mathematics education. In: Wang J, editor. Proceedings of the 14th International Congress on Mathematical Education. Vol. 1. Singapore: World Scientific; 2024. p. 417–21. doi:10.1142/9789811287152_0046.

[2] Schukajlow S, Kaiser G, Stillman G. Empirical research on teaching and learning of mathematical modelling: a survey on the current state-of-the-art. ZDM Math Educ. 2018;50:5–18. doi:10.1007/s11858-018-0933-5.

[3] Blum W, Borromeo Ferri R. Mathematical modelling: can it be taught and learnt? J Math Model Appl. 2009;1(1):45–58.

[4] Fulford GR. Mathematical modelling using scenarios, case studies and projects in early undergraduate classes. Int J Math Educ Sci Technol. 2024;55(2):468–79. doi:10.1080/0020739X.2023.2244490.

[5] Rosanti A, Aisyah N, Darmawijoyo, Lestari S. High school students' modelling abilities in arithmetic sequences using body weight contexts. Al-Jabar J Pendidik Mat. 2024;15(2):309–21. doi:10.24042/ajpm.v15i2.19372.

[6] Oficiar JRG, Ibañez ED, Pentang JT. Metacognitive awareness as a predictor of mathematical modeling competency among preservice elementary teachers. Int J Educ Methodol. 2024;10(2):1079–92. doi:10.12973/ijem.10.2.1079.

[7] Zbiek RM, Peters SA, Galluzzo B, White SJ. Secondary mathematics teachers learning to do and teach mathematical modelling: a trajectory. J Math Teach Educ. 2024;27(1):55–83. doi:10.1007/s10857-022-09550-7.

[8] Vorhölter K, Greefrath G, Borromeo Ferri R, Leiss D, Schukajlow S. Mathematical modelling. In: Jahnke HN, Hefendehl-Hebeker L, editors. Traditions in German-speaking mathematics education research. Cham: Springer; 2019. p. 91–114. doi:10.1007/978-3-030-11069-7_4.

[9] Turner EE, Bennett AB, Granillo M, Ponnuru N, Roth McDuffie A, Foote MQ, et al. Authenticity of elementary teacher-designed and implemented mathematical modelling tasks. Math Think Learn. 2024;26(1):47–70. doi:10.1080/10986065.2022.2028225.

[10] Blum W, Galbraith PL, Henn HW, Niss M, editors. Modelling and applications in mathematics education: the 14th ICMI Study. New York: Springer; 2007. doi:10.1007/978-0-387-29822-1.

[11] Vose JM, Peterson DL, Fettig CJ, Halofsky JE, Hiers JK, Keane RE, et al. Fire and forests in the 21st century: managing resilience under changing climates and fire regimes in USA forests. In: Greenberg CH, Collins BS, editors. Fire ecology and management: past, present, and future of US forested ecosystems. Cham: Springer; 2021. p. 465–502. doi:10.1007/978-3-030-73267-7_12.

[12] Supriyadi E, Kuncoro KS. Exploring the future of mathematics teaching: insights with ChatGPT. Union J Ilm Pendidik Mat. 2023;11(2):305–16. doi:10.30738/union.v11i2.14898.

[13] Blum W, Leiß D. Investigating quality mathematics teaching: the DISUM project. In: Bergsten C, Grevholm B, editors. Developing and researching quality in mathematics teaching and learning. Linköping: Swedish Society for Research in Mathematics Education; 2007. p. 3–16.

[14] Bali M. An analysis of Grade 12 national examinations marking memoranda on the topic of measurement in Mathematical Literacy using a mathematical modelling framework [master's thesis]. Cape Town: University of the Western Cape; 2017.

[15] Greefrath G, Maaß K. Diagnosis and evaluation in mathematical modelling. In: Greefrath G, Maaß K, editors. Modelling competences: diagnosis and evaluation. Cham: Springer; 2020. p. 1–19.

[16] Borromeo Ferri R. On the influence of mathematical thinking styles on learners' modelling behaviour. J Math Didakt. 2010;31(1):99–118. doi:10.1007/s13138-010-0009-8.

[17] Geiger V, Galbraith P, Niss M, Delzoppo C. Developing a task design and implementation framework for fostering mathematical modelling competencies. Educ Stud Math. 2022;109(2):313–36. doi:10.1007/s10649-021-10039-y.

[18] Wang Y, Sun J. Design and implementation of virtual reality interactive product software based on AI deep learning algorithm. Adv Multimed. 2022; 2022:9104743. doi:10.1155/2022/9104743.

[19] Kim S, Jang Y, Choi S, Kim W, Jung H, Kim S, et al. Analyzing teacher competency with TPACK for K–12 AI education. KI Künstl Intell. 2021;35(2):139–51. doi:10.1007/s13218-021-00731-9.

[20] Rawas S. ChatGPT: empowering lifelong learning in the digital age of higher education. Educ Inf Technol. 2024;29(6):6895–908. doi:10.1007/s10639-023-12114-8.

[21] Dahal N, Luitel BC, Pant BP. Exploring capabilities and limitations of generative AI chatbots in solving math algorithm problems. In: Proceedings of the 15th International Congress on Mathematical Education; 2024 Jul 7–14; Sydney, Australia. 2024.

[22] Plevris V, Papazafeiropoulos G, Jiménez Rios A. Chatbots put to the test in math and logic problems: a comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google Bard. AI. 2023;4(4):949–69. doi:10.3390/ai4040048.

[23] de Oliveira Fornasier M. The impact of the introduction of AI in advocacy: skills and professional ethics necessary for the future lawyer. Rev Fac Direito UFPR. 2021;66(2):69–94. doi:10.5380/rfdufpr.v66i2.73458.

[24] Rane N. Enhancing mathematical capabilities through ChatGPT and similar generative artificial intelligence: roles and challenges in solving mathematical problems. SSRN Electron J. 2023. doi:10.2139/ssrn.4603237.

[25] Parra V, Sureda P, Corica A, Schiaffino S, Godoy D. Can generative AI solve geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish. Int J Interact Multimed Artif Intell. 2024;8(5):65–74. doi:10.9781/ijimai.2024.02.009.

[26] Spreitzer C, Straser O, Zehetmeier S, Maaß K. Mathematical modelling abilities of artificial intelligence tools: the case of ChatGPT. Educ Sci. 2024;14(7):698. doi:10.3390/educsci14070698.

[27] Cevher AY, Yıldırım S. Design and use of chatbots for educational purposes: a study on student opinions. İleri Teknol Çalış Derg. 2023;1(2):76–83. doi:10.5281/zenodo.10445240.

[28] Lee D, Yeo S. Developing an AI-based chatbot for practicing responsive teaching in mathematics. Comput Educ. 2022;191:104646. doi:10.1016/j.compedu.2022.104646.

[29] Yang X, Schwarz B, Leung IK. Pre-service mathematics teachers' professional modelling competencies: a comparative study between Germany, Mainland China, and Hong Kong. Educ Stud Math. 2022;109(2):409–29. doi:10.1007/s10649-021-10064-x.

[30] Nguyen ATT, Tran D. Quantitative reasoning as a lens to examine changes in modelling competencies of secondary preservice teachers. Math Educ Res J. 2025;37(1):27–54. doi:10.1007/s13394-023-00481-x.

[31] Dasari D, Hendriyanto A, Sahara S, Suryadi D, Muhaimin LH, Chao T, et al. ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. Front Educ. 2024;8:1295413. doi:10.3389/feduc.2023.1295413.

[32] Gerber S, Quarder J, Greefrath G, Siller HS. Pre-service teachers' pedagogical content knowledge for teaching simulations and mathematical modelling with digital tools. In: Hodgen J, Geraniou E, Bolondi G, Ferretti F, editors. Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education. Bolzano: Free University of Bozen-Bolzano and ERME; 2022. p. 1051–8.

[33] Guerrero-Ortiz C. Implicaciones de los ambientes digitales en los procesos de modelización. In: Actes du septième symposium d'Étude sur le Travail Mathématique; 2023; Strasbourg, France. Strasbourg: IREM de Strasbourg; 2023. p. 683–92. [In Spanish].

[34] Wei Y, Zhang Q, Guo J. Can mathematical modelling be taught and learned in primary mathematics classrooms? A systematic review of empirical studies. Educ Sci. 2022;12(12):923. doi:10.3390/educsci12120923.

[35] Plano Clark VL, Creswell JW. Understanding research: a consumer's guide. 2nd ed. Boston: Pearson; 2015.

[36] Bryman A. Social research methods. 5th ed. Oxford: Oxford University Press; 2016.

[37] Gilgun JF. Qualitative research and family psychology. J Fam Psychol. 2005;19(1):40–50. doi:10.1037/0893-3200.19.1.40.

[38] Kibiswa NK. Directed qualitative content analysis (DQlCA): a tool for conflict analysis. Qual Rep. 2019;24(8):2059–79. doi:10.46743/2160-3715/2019.3778.

[39] Kang O, Noh J. Teaching mathematical modelling in school mathematics. In: Proceedings of the 12th International Congress on Mathematical Education; 2012 Jul 8–15; Seoul, South Korea. Seoul: COEX; 2012. p. 145–57.

[40] Anhalt CO, Cortez R. Developing an understanding of mathematical modelling in secondary teacher preparation. J Math Teach Educ. 2016;19(6):523–45. doi:10.1007/s10857-015-9309-8.

[41] Daubeny RDP, Bunn CW, Brown CJ. The crystal structure of polyethylene terephthalate. Proc R Soc Lond A Math Phys Sci. 1954;226(1167):531–42. doi:10.1098/rspa.1954.0273.

[42] Xiao Y, Dai X, Wang K, Zhou G. High melting point of linear, spiral polyethylene nanofibers and polyethylene microspheres obtained through confined polymerization by a PPM-supported Ziegler-Natta catalyst. ChemistryOpen. 2020;9(11):1173–80. doi:10.1002/open.202000290.

[43] Hwang GJ, Tu YF. Roles and research trends of artificial intelligence in mathematics education: a bibliometric mapping analysis and systematic review. Mathematics. 2021;9(6):584. doi:10.3390/math9060584.

[44] Durandt R, Blum W, Lindl A. Fostering mathematical modelling competency of South African engineering students: which influence does the teaching design have? Educ Stud Math. 2022;109(2):361–81. doi:10.1007/s10649-021-10068-7.

[45] Wardat Y, Tashtoush MA, AlAli R, Jarrah AM. ChatGPT: a revolutionary tool for teaching and learning mathematics. Eurasia J Math Sci Technol Educ. 2023;19(7):em2286. doi:10.29333/ejmste/13272.

[46] Su J, Ng DT, Chu SK. Artificial intelligence literacy in early childhood education: the challenges and opportunities. Comput Educ Artif Intell. 2023;4:100124. doi:10.1016/j.caeai.2023.100124.

[47] Elbanna S, Armstrong L. Exploring the integration of ChatGPT in education: adapting for the future. Manag Sustain Arab Rev. 2024;3(1):16–29. doi:10.1108/MSAR-03-2023-0016.

AUTHOR(S) BIOSKETCHES

Nafise Azadi is a Ph.D. student of mathematics education under the supervision of Professor Modjtaba Ghorbani.

Azadi, N., Ph.D. student, Department of Mathematics, Faculty of Science, Shahid Rajaee Teacher Training University, Tehran, I. R. Iran
nafise.azadi@sru.ac.ir

Ebrahim Reyhani received his bachelor's degree in mathematics education from Kharazmi University, his master's degree in pure mathematics from Shahid Beheshti University, and his PhD in mathematics education from Moscow State University. He was serving as an associate professor at Shahid Rajaee Teacher Training University when he passed away in December 2024.

Reyhani, E., Associate Professor, Department of Mathematics, Faculty of Science, Shahid Rajaee Teacher Training University, Tehran, Iran
e_reyhani@sru.ac.ir

Modjtaba Ghorbani is a professor of pure mathematics at Shahid Rajaee Teacher Training University. His interests include algebra and mathematics education.

Ghorbani, M., Professor, Department of Mathematics, Faculty of Science, Shahid Rajaee Teacher Training University, Tehran, I. R. Iran
mghorbani@sru.ac.ir

Reza Ebrahimpour is a professor of cognitive neuroscience at the Sharif University of Technology. His interests include neuroscience, AI, and neuroeducation.

Ebrahimpour, R. Professor, Center for Cognitive Science, Institute for Convergence Science and Technology, Sharif University of Technology, Tehran, I. R. Iran
ebrahimpour@sharif.edu

Saeid Haghjoo is currently a Lecturer in the Department of Mathematics at Farhangian University, Bushehr. His interests include teaching and learning of mathematics. He holds a Ph.D. in mathematics education from Shahid Rajaee Teacher Training University.

Haghjoo, S., PhD., Mathematics Education, lecturer at Farhangian University in Bushehr, Iran
s.Haghjoo@sru.ac.ir

Citation (Vancouver): Azadi N, Reyhani E, Ghorbani M, Ebrahimpour R, Haghjoo S. [Evaluating pre-service teachers' mathematical modelling task performance: a comparison with ai responses] *Tech. Edu. J. 2026; 20(1): 157-178*

https://doi.org/10.22061/tej.2026.12519.3299