



Should I move outdoors or ask AI? Influence of setting on performance, interest and self-efficacy in solving Fermi problems

Carlos Segura¹ · César Gallart¹ · Irene Ferrando¹ · Lluís Albarracín²

Received: 2 May 2025 / Revised: 9 June 2026 / Accepted: 3 August 2026
© The Author(s) 2026

Abstract

Fermi problems (FPs) are suitable for introducing real-world mathematical tasks at the primary level. However, primary school teachers often have difficulties solving FPs and rarely use them in their classrooms. Hence, promoting the successful implementation of FPs in pre-service teachers' training is necessary. To this end, we aimed to examine whether specific FP-solving settings could facilitate pre-service teachers' performance, interest and self-efficacy. The study is divided into two phases. First, pre-service teachers solve near-context FPs in one of three different settings: working in the classroom (indoors), working at the actual problem site (outdoors), or working with access to ChatGPT (AI-support). In the second phase, all pre-service teachers solve more complex FPs with the same AI support. Results indicate that the problem-solving setting significantly influenced outcomes. The outdoor condition led to the highest performance in both phases, with effects persisting into the second phase. Both the outdoor and the AI-supported settings promoted interest, with effects also carrying over to the second phase. Although self-efficacy was initially lower in the outdoor setting, it showed the greatest improvement later on. These findings suggest that outdoor and AI-supported approaches can benefit pre-service teachers' own FP-solving performance, interest, and self-efficacy, and may also contribute to preparing them to implement FPs in schools.

Keywords Fermi problems · ChatGPT · problem-solving setting · interest · performance · self-efficacy · real-world problems · outdoors · Artificial Intelligence

✉ Irene Ferrando
Irene.ferrando@uv.es

¹ Dpto. Didáctica de las Matemática, Universitat de València, València, Spain

² Dpto. de Didáctica de la Matemática y les Ciències Experimentals, Universitat Autònoma de Barcelona, Barcelona, Spain

Introduction

For decades, mathematics education has emphasised the importance of connecting mathematical content with real-world situations, not only to increase the relevance of the discipline but also to foster transferable competencies such as mathematical reasoning and decision-making in complex context (Geiger et al., 2018; Maaß et al., 2019). Approaches that engage students in solving real-world problems have been recognised as essential for achieving these goals and have been incorporated into many mathematics curricula worldwide (Cevikbas et al., 2022). Despite this curricular emphasis, activities that involve mathematising real-world situations remain relatively uncommon in classroom practice, particularly in primary education (Borromeo Ferri, 2021; English, 2023).

Among the activities that engage students with real-world problems, Ferrando and Albarracín (2021) have shown that the so-called Fermi problems (FPs) are particularly suitable for primary education, as they can be accessible to learners at this stage. FPs present a real and open-ended situation, without data, which requires estimating a quantity only achievable by involving assumptions, estimations and simple calculations (Ärleback, 2009). An example could be to make a reasoned estimate of how many kilograms of waste the school generates in a year. To solve these problems, the real situation must be structured, identifying and quantifying the relevant variables (Peter-Koop, 2009; Robinson, 2008).

Although FPs can provide a useful entry point for engaging students with real-world mathematical tasks, their use remains challenging for primary school teachers. Thus, Segura and Ferrando (2023) found that some pre-service teachers have difficulties in solving FPs based on estimating a large number of elements in a delimited area. In addition, some studies (Böswald & Schukajlow, 2023; García-Cerdá et al., 2024; Krawitz & Schukajlow, 2018; Schukajlow et al., 2022) have found that solving real-world problems is related to lower values in some affective aspects (interest and self-efficacy) than solving intramathematical or word problems. It seems clear that teacher education is key to bringing real-world problems into the classroom, equipping pre-service teachers with the tools and skills to link real-world phenomena and mathematics (Geiger et al., 2022).

In this context, problem-solving settings can be used in initial teacher education to support pre-service teachers when solving FPs. This includes both cognitive and affective dimensions. Studies by Jablonski (2023a, 2023b) and Segura et al. (2023) have shown that an outdoor setting—in which the solver can observe and interact with the actual location in which the situation posed by the FP takes place—encourages the adaptation of strategies to the context and has a positive impact on the solution process. However, Hartmann and Schukajlow (2021) concluded that it is the problem and not the problem-solving setting that is important for the development of interest and emotions.

In addition to outdoor settings, digitally supported settings may also shape how pre-service teachers solve FPs and experience the solving process. In recent years, AI systems have attracted increasing educational attention (Lo, 2023; Zhang & Aslam, 2021). Studies suggest that Large Language Models (LLMs) are capable of solving mathematical problems (Schorcht et al., 2024), including real-world problems (López

Simó & Rezende, 2024; Spreitzer et al., 2024). Consequently, providing access to an LLM-based conversational system, such as ChatGPT, during the solution process could be considered a problem-solving setting with the potential to enhance pre-service teachers' performance or their emotions (Noster et al., 2024). An AI-supported setting could be particularly useful for solving more complex FPs involving situations outside solvers' immediate environment, which cannot be addressed outdoors.

Although both the outdoor setting and the AI-supported setting may be beneficial for solving FPs located in solvers' immediate environment—hereafter referred to as *near-context FPs*—it remains unclear which of the two would produce a greater positive effect on performance and key affective variables, such as interest and self-efficacy. Furthermore, it is unknown whether previous engagement in either setting affects pre-service teachers' subsequent performance, interest, or self-efficacy when they later solve more complex FPs involving situations outside their immediate environment—hereafter referred to as *distant-context FPs*—with AI support. To address these questions, we prepared a two-phase study. First, pre-service teachers solved near-context and accessible FPs across three settings: working in the classroom (indoors), working at the actual problem site (outdoors), or working with the option to consult ChatGPT (AI-supported). Then, they solved distant-context and complex FPs with the same AI support.

Theoretical framework

Fermi problems: nature and educational value

Numerous studies have highlighted the importance of real-world problem solving for helping students make meaning of mathematical content (Geiger et al., 2018; Gravemeijer & Doorman, 1999) and connect it to other disciplines, especially STEM (Maaß et al., 2019). Unlike a standard word problem, in which the real-world context is simply a backdrop for applying a procedure (Krawitz & Schukajlow, 2018; Relensmann & Schukajlow, 2017), solving a real-world problem requires an explicit process of mathematisation. That is, to solve a real-world problem, one must identify the relevant elements and regularities of the situation, quantify them to transform them into mathematical objects and relationships, and use them to develop a strategy that leads to the solution of the problem (García-Cerdá et al., 2024).

Real-world problems occupy a central role in mathematics curricula in many countries (Cevikbas et al., 2022; Schukajlow et al., 2018). However, their actual presence in classrooms remains limited, particularly in primary education (Borromeo Ferri, 2021; English, 2023). Previous studies involving school students (e.g., Ferrando & Albarracín, 2021) and pre-service teachers (Peter-Koop, 2009) have shown that certain FPs can be suitable for introducing this type of activity at this educational stage. FPs are real-world problems that present a missing-data situation and are therefore open-ended; solving them requires estimating a quantity that can only be obtained through reasoned assumptions, partial estimates, and simple calculations (Ärleback, 2009; Sriraman & Knott, 2009). In this sense, developing strategies for solving FPs involves mathematising the real situation (Robinson, 2008; Peter-Koop, 2009).

Variation in Fermi Problems: contextual distance and task complexity

Ferrando and Albarracín (2021) showed that FPs involving the estimation of a large number of elements distributed over a surface can be suitable for introducing real-world problem solving in primary education. Solving such problems requires only basic knowledge of measurement and estimation of lengths and areas (Segura et al., 2026), proportional reasoning, and arithmetic calculations (Ärlebäck, 2009).

However, these FPs can vary substantially in the demands they place on solvers. Albarracín et al. (2022) point out two particularly relevant dimensions: contextual distance and complexity. Regarding the former, near-context FPs present situations that are familiar to the solver and located in their immediate environment (Fig. 1), which in some cases even allows them to be solved at the actual problem site (Jablonski, 2023a, 2023b; Segura et al., 2023). Distant-context FPs, in contrast, involve less familiar situations (Fig. 2), located outside the solver's immediate environment and therefore potentially requiring greater extra-mathematical knowledge (Albarracín et al., 2022).

Concerning problem complexity, we follow an approach that focuses on the complexity of the situation posed by the problem (Knabbe et al., 2025), drawing on Maaß's (2010) classification in terms of the number of assumptions, variables, and

Explain step-by-step how you would solve the following problems, considering (and estimating) what data you would need to measure to solve them and describing the procedures to arrive at the solution.

Problem 1 How many people can take shelter under the porch at the entrance to the faculty if it rains?



Fig. 1 Problem 1 of the four near-context and accessible FPs of Phase 1

Now consider the following problem. Explain, step by step, how you would solve it, the measurements, data and procedures you would use to reach a solution.

Problem 6 How many people can fit inside this furniture store in Valencia at the time of maximum influx of customers?



Fig. 2 Problem 6 of the two distant-context and complex FPs of Phase 2

quantitative relationships that solvers must consider to develop a solution strategy. Within this dimension, we use the term *accessible* to refer to low-complexity FPs, that is, problems that can be solved by considering a limited set of assumptions and relationships.

In this sense, we refer to FPs such as the one shown in Fig. 1 as *accessible*, because their solution can be reached directly through one of three straightforward strategies identified by Ferrando and Albarracín (2021): a base-unit approach, based on comparing the total area with the area occupied by one element; a linearisation strategy, based on estimates by rows and columns; or a density strategy, based on estimating a smaller portion and scaling proportionally.

By contrast, complex FPs, such as the one illustrated in Fig. 2, require numerous realistic assumptions (Krawitz et al., 2018) and the consideration of multiple variables before developing a solution strategy. Unlike accessible FPs, which can be approached directly through base-unit, linearisation, or density strategies, complex FPs must be decomposed into several subproblems. For example, estimating the capacity of a large furniture store—a dynamic system with constant inflows and outflows and an irregular space—requires considering variables such as parking space, average number of people per vehicle, customers arriving on foot, usable shop area, or the presence of trolleys, among others. As explained in Albarracín et al. (2022), accessible and near-context FPs can be used as preliminary engagement activities to later address more complex distant-context FPs, a process that can also be combined with different settings that may enhance performance or affective aspects.

Pre-service teachers' performance in solving FPs

Implementing real-world problems is a challenge for primary school teachers (Greefrath et al., 2022; Wess et al., 2021), who need to master the competencies required to teach them effectively (Borromeo Ferri, 2018). One of these competences concerns teachers' knowledge for teaching problem solving (Chapman, 2013), including the capacity to analyse tasks, develop solution strategies, and mobilise concepts and procedures (Chapman, 2015; Schoenfeld, 1985). Consequently, pre-service teachers need experience solving problems from the solver's perspective before introducing them successfully in classrooms (Thompson, 1985). In this study, however, we focus on their performance when solving FPs, rather than on problem-solving knowledge, competence, or ability as broader constructs.

In this sense, Segura and Ferrando (2023) defined pre-service teachers' performance in solving FPs in terms of errors. They distinguished between *strategic error*, which prevents a valid strategy to obtain the estimate, and *mathematical error*, where a valid strategy is devised but an error occurs during its implementation. Strategic errors relate to difficulties in mathematising the real-world situation, such as identifying relevant variables or their relationships (Klock & Siller, 2020; Segura & Ferrando, 2023). In a study with pre-service teachers, Segura et al. (2026) found that some mathematical errors were associated with difficulties in measurement and estimation of lengths and areas, which are important for successfully solving these problems.

In Segura and Ferrando (2023) framework, performance is low when strategic errors are made, medium when mathematical errors occur, and high when no errors are observed. Segura et al. (2023) further refined the analysis of error-free solutions by distinguishing between schematic and complete solutions. A schematic solution consists of a valid step-by-step description of the procedures needed to obtain the estimate, without carrying them out with numerical values; a complete solution involves the careful execution of the strategy with numerical estimates and measurements to compute a final value that can be validated. Using this distinction, Segura et al. (2023) found that many pre-service teachers did not reach the highest level of performance, even when solving near-context and accessible FPs. In this study, we combine these two characterisations into four performance levels.

Interest and self-efficacy in problem solving

When solving problems, solvers' cognitive activity is accompanied by beliefs, emotions and attitudes (McLeod, 1989; Schoenfeld, 2013). Empirical studies confirm that this affective system is related to problem-solving performance, with interest and self-efficacy playing a central role (Gasco & Villarreal, 2014; Nuutila et al., 2020).

Interest is a psychological state involving a positive emotional connection between a person and an object, fostering engagement (Hidi and Renninger, 2006). Here, we focus on situational or task-specific interest, that is, temporary engagement triggered by activity features (Harackiewicz et al., 2016). Such interest may relate to the perceived importance or usefulness of the problem context (Schulze Elfringhoff & Schukajlow, 2021), and enhance cognitive engagement and perseverance (Ainley et al., 2002; Rellensmann & Schukajlow, 2017; Schukajlow et al., 2012).

Bandura (1997) defined self-efficacy as individuals' beliefs in their capabilities to organise and execute the actions required to attain desired outcomes. In this study, task-specific self-efficacy refers to solvers' confidence in their ability to solve the proposed FPs. Self-efficacy predicts performance (Honicke & Broadbent, 2016; Talsma et al., 2018) and, like situational interest, can vary according to task features (Street et al., 2024; Usher & Pajares, 2009). Students have been found to report lower interest and self-efficacy in real-world problems than in other problem types (Krawitz & Schukajlow, 2018; Rellensmann & Schukajlow, 2017; Schukajlow et al., 2012). However, pre-service teachers report higher interest in real-world problems than students (García-Cerdá et al., 2024), suggesting that motivational effects may depend on participants' experience. In this regard, previous interest and mastery experiences may stabilise or increase solvers' interest and self-efficacy in similar future problems (Bandura, 1997; Fryer & Shum, 2025; Usher & Pajares, 2009).

Outdoor problem-solving setting

Effective training that supports pre-service teachers cognitively and affectively when solving real-world problems is key to bringing these problems to primary education (Geiger et al., 2022). While the classroom (indoors) is the traditional problem-solving setting, it requires solvers to mentally reconstruct the situation before mathematising it and developing a strategy (Sriraman & Knott, 2009). For near-context real-world

problems, outdoor settings at the actual problem location may provide more affordances for successful problem solving (e.g., Buchholtz, 2021; Hartmann & Schukajlow, 2021; Jablonski, 2023a; Segura et al., 2023).

An outdoor problem-solving setting involves addressing the problem in the physical space where the real-world situation unfolds (Cross, 1997; Vale et al., 2019). Working outdoors can foster connections and the application of mathematical skills in meaningful contexts (Richardson, 2004). Indeed, studies show that solving real-world problems outdoors supports the identification and estimation of data in situ, thereby facilitating the mathematisation of the situation (e.g., Buchholtz, 2021). In studies with secondary school students, Jablonski (2023b, 2023a) found that outdoor solvers, compared with those reconstructing objects indoors from pictures, had more opportunities to perceive dimensions and shapes, measure or estimate directly, and obtain more realistic and accurate solutions.

Similarly, Segura et al. (2023) found that pre-service teachers solving near-context, accessible FPs in the outdoor setting developed more complete solutions than those working indoors (see Sect. 2.3), suggesting limitations in mentally recreating even familiar real-world situations. Gallart et al. (2025) later reported that the outdoor setting promoted more realistic assumptions, thereby increasing accuracy. Producing more complete and accurate solutions may have implications for performance.

Few studies have examined whether outdoor and indoor settings differ in affective factors such as interest or self-efficacy. In studies with students, Hartmann and Schukajlow (2021) and Jablonski (2023a) found no setting effect on interest, concluding that the problem itself, rather than the setting, is crucial for interest and emotions. Evidence on self-efficacy is scarcer: in science education, Lindemann-Matthies et al. (2011) found that pre-service teachers reported lower self-efficacy outdoors, whereas Ayotte-Beaudet et al.'s (2017) review found no conclusive differences between indoor and outdoor problem solving.

AI-supported problem-solving setting

Recent research suggests that digitally supported problem solving also affects cognitive outcomes in real-world problems, although effects on affective factors remain unclear (Cevikbas et al., 2023; Quarder et al., 2025). In a study with high-school students, Barlovits and Ludwig (2023) found that using digital tools improved performance in near-context real-world problems by supporting data collection and processing, but did not affect interest or self-efficacy. Similarly, studies on tools such as 3D printing and apps show support for phases such as mathematisation and validation, but no clear effects on interest or self-efficacy (Jablonski, 2023a, 2024; Jablonski et al., 2023). Finally, Quarder et al. (2025) reported that simulation-based digital tools enhance pre-service teachers' mastery of competencies for teaching real-world problem solving.

The adoption of digital tools has positioned Artificial Intelligence (AI) as a critical driver of educational transformation (Zhang & Aslam, 2021). Its impact has accelerated significantly with the emergence of large language models (LLMs), which have opened new opportunities (Lo, 2023) but have also generated substantial ethical and operational challenges (Yu, 2023). These challenges also affect

higher education and pre-service teacher training (Korseberg & Elken, 2025; Lund & Wang, 2023). ChatGPT, introduced in 2022 as a widely adopted LLM-based conversational AI model (Korseberg & Elken, 2025), is now pervasive in mathematics education, where students and teachers use it, among other purposes, to support problem solving (Pepin et al., 2024).

Fock and Siller (2025) note that most studies on the use of AI in STEM education focus on technical evaluation and tool development, identifying strengths in basic tasks but weaknesses in complex reasoning. Schorcht et al. (2024) similarly found that LLMs can solve some mathematical problems depending on domain and complexity. Beyond these technical issues, AI-based tools may pose risks to students' autonomy and critical thinking (Fock & Siller, 2025; Yu, 2023), especially when uncritical use creates a black-box risk: students may rely on automatically generated results without engaging in the problem-solving process (Cevikbas et al., 2023).

Consequently, concerns remain about whether AI-based tools, including LLM-based conversational systems, support solvers' cognitive and affective processes, given the scarcity of empirical studies with students and teachers and the limited attention to non-cognitive variables (Fock & Siller, 2025). Canonigo (2024) documents improvements in mathematical conceptual understanding and self-efficacy when AI tools are integrated with instructional design and teacher support. However, Dasari et al. (2024) found that exclusive reliance on ChatGPT led to poorer performance in mathematical problem-solving, despite positive perceptions of AI and greater confidence. On the other hand, Noster et al. (2024) found that pre-service teachers lacked prompting knowledge when using ChatGPT, leading to incorrect answers and responses that did not meet their expectations.

Regarding real-world mathematical problems, Spreitzer et al. (2024) have found that contextual complexity can compromise solution reliability. This pattern aligns with Fock and Siller's (2025) scoping review, which documents limitations of AI-based tools in tasks requiring spatial reasoning and mathematization from images or real-world contexts. López-Simó and Rezende (2024) observed similar limitations when LLM-based conversational systems such as ChatGPT address FPs, as responses are variable and may include errors related to imprecise estimations, incorrect concept use, or miscalculations. Despite these challenges, Gallart et al. (2025) found that pre-service teachers with access to ChatGPT developed more complete solutions (Segura et al., 2023) than those solving such problems in classroom settings without digital tools.

Research questions and hypotheses

RQ1. What is the effect of the setting on pre-service teachers' performance in solving near-context and accessible FPs?

H1: Pre-service teachers' performance in solving near-context and accessible FPs will be higher when solving problems outdoors or with AI support than when solving the problem indoors.

Justification: Studies have found that students and pre-service teachers solve real-world problems more successfully outdoors than indoors (Buchholtz, 2021;

Jablonski, 2023b; Segura et al., 2023). Similarly, digital tools and AI-based tools have been shown to support problem-solving success (Barlovits & Ludwig, 2023; Canonigo, 2024; Gallart et al., 2025; Quarder et al., 2025). We have no empirical basis to hypothesise whether differences exist between solving FPs outdoors or with AI support, although some studies point to the limitations of AI-based tools for solving real-world problems (Fock & Siller, 2025; López-Simó & Rezende, 2024; Spreitzer et al., 2024).

RQ2. Does previous experience influence pre-service teachers' performance in solving distant-context and complex FPs with AI support?

H2: The previous problem-solving setting will not significantly influence pre-service teachers' performance when solving distant-context and complex FPs with AI support. However, when comparing performance between Phase 1 and Phase 2, we expect the greatest improvement among those who previously solved near-context and accessible FPs indoors.

Justification: No previous studies establish whether the setting in which pre-service teachers solve near-context and accessible FPs affects their subsequent performance when solving distant-context and complex FPs with AI support. However, because H1 predicts lower initial performance in the indoor setting, we anticipate an interaction effect whereby participants who previously solved near-context and accessible FPs indoors will show the greatest performance gains when solving distant-context and complex FPs with AI support.

RQ3. What is the effect of the setting on pre-service teachers' interest and self-efficacy in solving near-context and accessible FPs?

H3: There will be no significant differences in pre-service teachers' interest and self-efficacy when solving near-context and accessible FPs across settings.

Justification: Research on outdoor versus indoor problem solving has found no differences in interest (Hartmann & Schukajlow, 2021; Jablonski, 2023a). Regarding self-efficacy, evidence is inconclusive when comparing outdoor and indoor settings (Ayotte-Beaudet et al., 2017), although some research suggests it may be somewhat lower outdoors (Lindemann-Matthies et al., 2011). Results are also inconclusive when using digital tools; while Barlovits and Ludwig (2023) found no differences in self-efficacy when solving real-world problems with apps, some studies suggest that AI-supported problem solving can increase confidence (Canonigo, 2024; Dasari et al., 2024).

RQ4. Does previous experience influence pre-service teachers' interest and self-efficacy in solving distant-context and complex FPs with AI support?

H4: The previous problem-solving setting will not significantly influence pre-service teachers' interest and self-efficacy when solving distant-context and complex FPs with AI support, although both will be overall higher in the second phase.

Justification: Given that our hypothesis establishes that there will be no significant differences in pre-service teachers' interest and self-efficacy when solving near-context and accessible FPs across settings (H3), we expect that the type of previous experience will not influence these affective factors when solving distant-context and complex FPs. However, as the problems in the first phase are accessible for pre-service teachers (Segura et al., 2023, 2026), previous research suggests that successful

problem-solving experiences can increase both interest and self-efficacy across all groups (Fryer & Shum, 2025; Usher & Pajares, 2009).

Method

Context and participants

The study was conducted with a sample of 137 pre-service teachers in their third year of the Degree in Primary School Education at the Faculty of Teacher Training (Universitat de València, Spain). The average age of the participants is 22.2 years. The participants were finishing their degree, having all received homogeneous training in arithmetic, geometry, measurement, statistics and probability. All participants were informed about the purpose of the research and voluntarily participated in the study. Before starting, all participants registered on ChatGPT (version 3.5, free access) and were informed that they needed to bring a personal computer with internet access.

Design and procedure

To answer the research questions, we conducted a quasi-experimental study with a convenience sample of pre-service teachers. The study is divided into two phases, conducted in two one-hour sessions. In Phase 1, participants were randomly assigned to three problem-solving settings: indoor setting, outdoor setting, or AI-support setting. 55 participants were assigned to the indoor setting, 41 were assigned to the outdoor setting, and 41 were assigned to the AI-supported setting. All participants answered an online questionnaire about their background solving real-world problems to ensure comparability between problem-solving settings.

In the first phase, for one hour, participants in each setting solved the same four near-context, accessible FPs (see 4.3.2). In all three settings, participants received the problems on paper and had to develop an individual handwritten solution, but the conditions differed across settings. In the indoor setting, all 55 participants solved the FPs in the classroom without any support. The 41 participants in the outdoor setting solved the FPs outdoors, in the actual location of the problems, being able to perceive the environment's dimensions in situ and move freely without restrictions. The 41 participants in the AI-supported setting solved the FPs indoors, but were able to access ChatGPT from their laptops, although they were not allowed to use any web browser or any other digital tool.

Then, in Phase 2, for one hour, participants solved two more complex, distant-context FPs (see 4.3.2). All participants solved these two problems under the same AI-support conditions: indoors, developing the individual handwritten solutions, and being able to consult the system used in the AI-supported setting.

In Fig. 3, we present a diagram that summarises the two-phase study design.

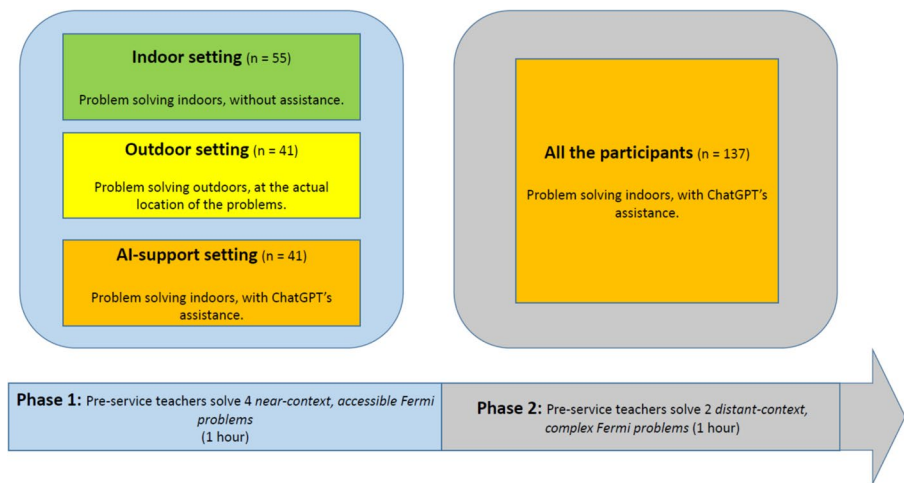


Fig. 3 Research design

Tasks and instruments

Background questionnaire on real-world problem solving

The background questionnaire on solving real-world problems comprises six items rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). It assesses the frequency with which students solved real-world problems in their mathematics classes, including characteristics such as under-determination, over-determination, and open-endedness, allowing for multiple solutions or different results. The questionnaire has been validated in previous studies (Rellensmann & Schukajlow, 2017).

Fermi problems

We administered a performance assessment consisting of six FPs. In Phase 1, participants had to solve four near-context, accessible FPs. Problems were presented in paper format with a statement, a photograph of their actual location, and a blank box for written solutions (see Fig. 1). All statements required participants to estimate a large number of elements on a rectangular area located in the Faculty of Teacher Training—a daily, familiar context for the students. These four FPs have been validated in previous studies (Segura & Ferrando, 2023; Segura et al., 2023, 2026), confirming their accessibility for pre-service teachers (see Sect. 2.2). All statements included explicit instructions for participants to document their assumptions, measurements, procedures, and calculations leading to the solution, rather than providing only the numerical estimate.

In Phase 2, participants had to solve two distant-context, complex FPs (see Fig. 2). Problems were presented in the same paper format as Phase 1. These FPs featured contexts that are not part of participants' daily experience and are distant from their immediate environment. They are more complex than those in Phase 1, as they

require a strategy with a greater number of assumptions, variables, and quantitative relationships (see Sect. 2.2).

Interest and self-efficacy questionnaire

Following each of the six problem statements, two items were included for participants to self-assess—before solving the problems—their interest and self-efficacy regarding each problem. We used items developed in the study by Schukajlow et al. (2012), translated into Spanish and validated in previous works (García-Cerdá et al., 2024). The item assessing interest was “I think it is interesting to solve this problem”; and for self-efficacy, “I am confident that I can solve this problem”. Participants marked their degree of agreement on a 5-point Likert scale (1=strongly disagree, 5=strongly agree).

Measures

Interest and self-efficacy scales

Two affective dependent variables were derived from participants’ responses to the items regarding the problems: interest and self-efficacy. For each of these variables, we computed separate scale scores for both Phase 1 and Phase 2 by averaging participants’ responses across the FPs in each phase: the four near-context and accessible FPs in Phase 1 and the two distant-context and complex FPs in Phase 2. Table 1 presents the reliability of the interest and self-efficacy scales, measured by Cronbach’s alpha coefficient, for the FPs in Phase 1 and Phase 2. Although we must note the limitation of using only one item per scale per problem, reliability ranged from acceptable to good across all measures.

Performance measurement

Categories for coding performance were based, on the one hand, on Segura and Ferrando’s (2023) distinction between strategic error and mathematical error, explained in subsection 2.3. On the other hand, we have also considered Segura et al.’s (2023) distinction between a schematic and a complete solution (justified in Sect. 2.3). Accordingly, the two researchers coded the pre-service teachers’ performance in solving FPs on a scale of 0–3, as shown in Table 2.

Two researchers coded the solutions of 164 problems (20% of the total) independently to calculate the agreement with Cohen’s kappa and obtained $\kappa=0.92$.

Table 1 Reliability (Cronbach’s α) for interest and self-efficacy scales

Scale		Number of items	Cronbach’s α
Interest	Near-context, accessible FPs	4	0.80
	Distant-context, complex FPs	2	0.68
Self-efficacy	Near-context, accessible FPs	4	0.83
	Distant-context, complex FPs	2	0.72

Table 2 Coding, description and examples of categories of performance in solving FPs

Performance (0–3)	Description	Example
0	Strategic error.	Blank solutions or incomplete mathematical models because variables or relationships between variables are not defined; e.g. in the Porch problem (Fig. 1), a solver proposes to estimate the porch area and the area occupied by a person, but does not relate both variables by a division of measurements.
1	Valid strategy, but with mathematical errors.	Strategy with measurement or measurement estimation errors (unit conversion, confusion of width and area, implausible estimates, etc.) or calculation errors (inversion error, etc.). For example, in the Porch problem, after obtaining that the area of the porch is 200 m^2 , it is written that this area must be divided by the width of a person, 0.5 m (instead of the area they occupy).
2	Valid strategy without mathematical errors, but schematic.	A valid and correct solution scheme is, for example: “First, I would estimate the width and length of the porch (in metres) and multiply them to get the area. Then I would consider that one square metre fits 3 people. Finally, to get the total number of people, I would multiply the 3 people/ m^2 by the m^2 of the porch”.
3	Valid strategy without mathematical errors and complete.	A correct and complete solution would be: “Between each column of the porch there are about 7 metres, as there are 4 spaces between columns, I estimate the length of the porch to be 28 metres. The width will be about 7 metres as well. Therefore, the area of the porch is about $72 \times 8 = 196 \text{ m}^2$. I am going to assume that an average person occupies a square area of 0.5 metres on a side, i.e. 0.25 m^2 . Then the total number of people is.”

Therefore, the level of inter-rater agreement was very good. The coding of the remaining solutions was shared between both coders.

In addition, Cronbach’s alpha coefficient was calculated to determine the reliability of the performance scale (0–3). When calculated separately by phase, internal consistency was acceptable for the four FPs in Phase 1 ($\alpha=0.71$), but lower for the two FPs in Phase 2 ($\alpha=0.56$). This lower value should be interpreted with caution, given the small number of FPs included in Phase 2. Across the six FPs, the performance scale showed moderate internal consistency ($\alpha=0.66$). Moreover, the coding of performance according to strategic errors, mathematical errors, and the completeness

of valid solutions is grounded in previous studies that have used and validated this analytical framework with pre-service teachers solving FPs (Segura et al., 2026).

Data analysis

In this study, we have two nominal independent variables: the problem-solving setting (indoors, outdoors, AI-support) and the type of FPs (near and accessible; distant and complex). From the instruments and measures described above, we obtain three dependent scale variables: performance (0–3), interest (1–5), and self-efficacy (1–5). To answer RQ1 and RQ2, we fixed the problem variable (near-context and accessible FPs) and studied the effect of setting on the dependent variables using one-way ANOVA and post-hoc comparisons. If we fix the independent variable on the distant context and complex FPs, then the setting corresponds to previous experience. Again, a mixed ANOVA analysis (between-subject and within-subject main effects, interaction effect) and post-hoc comparisons allow us to answer RQ3 and RQ4, analysing the influence of previous experience on the dependent variables when participants solve distant-context and complex FPs. These analyses assume normality and homoscedasticity in the distribution of the dependent variables but remain robust in cases where this is not satisfied (Rasch & Guiard, 2004). Table 3 presents the results of tests for normality and homoscedasticity for the variables involved in this study.

When homogeneity of variance was confirmed, standard ANOVA was conducted despite any violations of normality. When homoscedasticity was violated, Brown–Forsythe ANOVA with Games-Howell post-hoc analyses were employed.

Results

Preliminary analyses

Table 4 presents the results of the background questionnaire on solving real-world problems for participants assigned to each of the three settings.

A one-way ANOVA analysis revealed no significant differences in background solving real-world problems between participants in the three settings ($F(2, 136) = 1.86, p = .16$), confirming that the three groups were comparable.

Performance in solving near-context and accessible FPs

Table 5 shows pre-service teachers' average performance (0–3) in solving near-context and accessible FPs in the three problem-solving settings.

We observe in Table 5 that the average performance is higher in the outdoor and AI-supported settings than in the indoor setting. Using the Brown–Forsythe ANOVA, we confirm that the setting has a large effect on pre-service teachers' performance in solving near-context and accessible FPs, with $F(2, 119.50) = 24.32, p < .001$ and $\eta^2 = 0.24$. Indeed, post-hoc Games-Howell analyses revealed that pre-service teachers' performance indoors is significantly lower than outdoors ($p < .001$) and in the AI-supported setting ($p < .001$).

Table 3 Normality and homoscedasticity analyses

Variable	Phase	Setting	Normality (Shapiro-Wilk test)	Homoscedasticity (Levene test)
Performance	1	Indoors	No ($p < .001$)	No ($p < .001$)
		Outdoors	No ($p < .001$)	
		AI-support	No ($p < .001$)	
	2	Previous indoors	No ($p < .001$)	No ($p < .05$)
		Previous outdoors	No ($p < .001$)	
		Previous AI-support	No ($p < .001$)	
Interest	1	Indoors	Yes ($p = .34$)	Yes ($p = .75$)
		Outdoors	Yes ($p = .12$)	
		AI-support	Yes ($p = .06$)	
	2	Previous indoors	No ($p < .05$)	Yes ($p = .60$)
		Previous outdoors	No ($p < .05$)	
		Previous AI-support	No ($p < .05$)	
Self-efficacy	1	Indoors	No ($p < .05$)	No ($p < .05$)
		Outdoors	No ($p < .05$)	
		AI-support	Yes ($p = .06$)	
	2	Previous indoors	No ($p < .05$)	No ($p < .05$)
		Previous outdoors	No ($p < .05$)	
		Previous AI-support	Yes ($p = .07$)	

Table 4 Mean scores and standard deviations of background in solving real-world problems

Problem-solving setting	Background solving real-world problems
Indoors	3.30 (0.52)
Outdoors	3.46 (0.39)
AI-supported	3.47 (0.56)

Table 5 Average performance (and standard deviation) in near-context and accessible FPs

Problem-solving setting	Performance in near-context and accessible FPs
Indoors	1.59 (0.98)
Outdoors	2.42 (0.55)
AI-supported	2.47 (0.55)

Performance in solving distant-context and complex FPs

Table 6 presents the average performance scores (0-3) of pre-service teachers in solving distant-context and complex FPs with AI support according to previous experience in Phase 1 of the study.

We found a small effect of previous experience on subsequent performance solving distant-context and complex FPs with AI support (Brown–Forsythe’s $F(2, 133.99) = 3.03$, $p < .05$ and $\eta^2 = 0.04$). Table 6 shows that pre-service teachers with previous outdoors experience achieved the highest performance in Phase 2. Post-hoc comparisons revealed that this difference was significant compared to pre-service teachers who had solved them indoors ($p < .05$).

In addition, we conducted a mixed ANOVA to analyse whether there are differences in performance between previous experience (Phase 1) and experience solving complex distant-context FPs with AI support (Phase 2). The main effect (within-subjects) shows that there are indeed differences in overall performance between the two phases ($F(1, 134) = 6.84$, $p < .01$, $\eta^2 = 0.05$), with higher overall performance in Phase 2. There is also an interaction effect with the type of previous experience ($F(2, 134) = 6.00$, $p < .01$, $\eta^2 = 0.08$). Hence, post-hoc comparisons indicate that only pre-service teachers who solved near-context FPs indoors improve significantly in Phase 2 ($p < .001$), as can be clearly seen in Fig. 4. Conversely, those with previous experience in the AI-supported setting were the only group to show a slight decrease in performance.

Interest and self-efficacy in solving near-context and accessible FPs

Table 7 presents the average interest and self-efficacy ratings (1-5) of pre-service teachers in solving near-context and accessible FPs in the three problem-solving settings.

As shown in Table 7, there are clear differences in interest across the three problem-solving settings. A one-way ANOVA revealed a significant effect of setting on pre-service teachers’ interest in solving near-context and accessible FPs ($F(2,$

Table 6 Average performance (and standard deviation) in distant-context and complex FPs

Previous experience	Performance in distant-context and complex FPs
Indoors	2.21 (0.95)
Outdoors	2.62 (0.76)
AI-supported	2.35 (0.75)

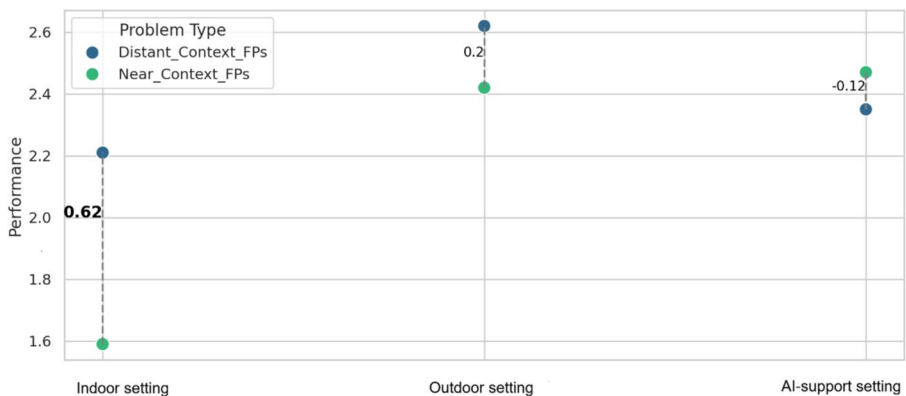


Fig. 4 Performance differences between Phase 1 and Phase 2

Table 7 Average interest and self-efficacy (and standard deviation) in near-context and accessible FPs

Problem-solving setting	Interest in near-context FPs	Self-efficacy in near-context FPs
Indoors	2.95 (0.95)	2.55 (1.09)
Outdoors	3.69 (0.79)	2.08 (0.90)
AI-supported	3.54 (0.91)	2.72 (0.72)

136)=9.41, $p < .001$, medium to large effect size, $\eta^2=0.12$). Post-hoc comparisons indicated that pre-service teachers' interest indoors was significantly lower than outdoors ($p < .001$) and in the AI-supported setting ($p < .01$).

There are also significant differences in self-efficacy depending on the setting (Brown–Forsythe's $F(2, 130.19)=5.54$, $p < .01$, and medium effect size $\eta^2=0.07$). Post-hoc comparisons indicated that pre-service teachers' self-efficacy outdoors was significantly lower than indoors ($p < .05$) and in the AI-supported setting ($p < .01$).

Interest and self-efficacy in solving distant-context and complex FPs

Table 8 presents the average interest and self-efficacy ratings (1-5) of pre-service teachers in solving distant-context and complex FPs according to previous experience in Phase 1 of the study.

We found a strong effect of previous experience on pre-service teachers' interest when solving distant-context and complex FPs with AI support ($F(2, 136)=10.74$, $p < .001$, $\eta^2=0.14$). Post-hoc comparisons revealed that pre-service teachers who previously had solved near-context FPs indoors showed significantly lower interest

Table 8 Average interest and self-efficacy (and standard deviation) in distant-context and complex FPs

Previous experience	Interest in distant-context FPs	Self-efficacy in distant-context FPs
Indoors	3.25 (0.91)	2.79 (1.13)
Outdoors	3.98 (0.74)	2.71 (1.11)
AI-supported	3.83 (0.76)	2.85 (0.81)

in Phase 2 than those who had solved them outdoors ($p < .001$) and with AI support ($p < .01$).

Furthermore, the within-subjects main effect of the mixed ANOVA shows that overall interest was higher in Phase 2 than in Phase 1 ($F(1,134) = 15.56, p < .001, \eta^2 = 0.10$). However, there is no interaction effect with the type of previous experience ($F(2, 134) = 0.006, p = .99, \eta^2 = 0.00$) because interest increased significantly in a similar way in all of them (around 0.3, as can be seen in Fig. 5).

Regarding pre-service teachers' self-efficacy solving distant-context and complex FPs with AI support, there is no influence of the previous experience in Phase 1 (Brown–Forsythe's $F(2, 125.52) = 0.21, p = .81$ and $\eta^2 = 0.003$).

However, the within-subjects main effect of the mixed ANOVA indicates that overall self-efficacy was higher in Phase 2 than in Phase 1 ($F(1,134) = 15.35, p < .001, \eta^2 = 0.10$). There is a trend of interaction effect with the type of previous experience ($F(2, 134) = 2.84, p = .06, \eta^2 = 0.04$) because self-efficacy only increased significantly (0.63, see Fig. 6) among those who had the previous experience outdoors ($p < .001$).

Discussion

In the following, we address the research questions, comparing our results with the hypotheses.

What is the effect of the setting on pre-service teachers' performance in solving near-context and accessible FPs?

Our results confirm H1: pre-service teachers' performance in solving near-context and accessible FPs is higher when the problems are solved outdoors or with AI support than when they are solved indoors. These findings are consistent with previous studies suggesting a facilitating role of both outdoor settings and digital tools (Barlovits & Ludwig, 2023; Quarder et al., 2025).

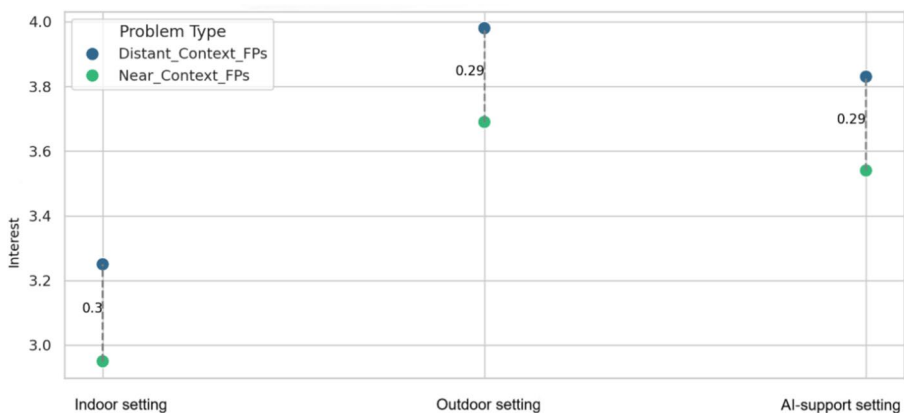


Fig. 5 Differences in interest between Phase 1 and Phase 2

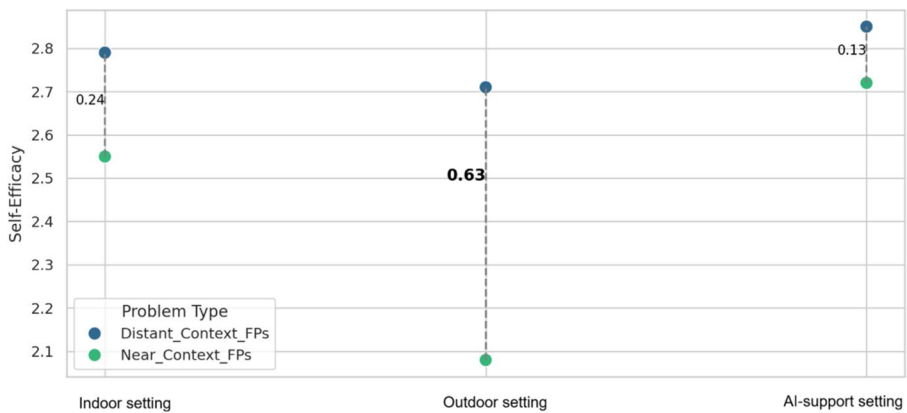


Fig. 6 Differences in self-efficacy between Phase 1 and Phase 2

One possible explanation for the superior performance of pre-service teachers solving near-context FPs at the actual location is that the outdoor setting allowed direct data collection using non-conventional units (e.g., steps) and perception of real dimensions, which can facilitate mental benchmarks for estimation (Jablonski, 2023b; Segura et al., 2026). Physically traversing the space—rather than mentally reconstructing it from photographs—may make magnitude perception more intuitive (e.g., distinguishing area from perimeter, width from surface) and support mathematization (Buchholtz, 2021; Jablonski, 2023a, 2024; Segura et al., 2023). Thus, the outdoor setting may support the use of measurement and estimation knowledge that has been identified as challenging for many pre-service teachers and is important for successfully solving FPs (Segura et al., 2026).

Higher performance in the AI-supported setting suggests several factors, some of which are potentially problematic. First, outperforming the indoor condition suggests that pre-service teachers may have employed appropriate prompting techniques for near-context and accessible FPs. For instance, they could prompt the AI system about how many people fit in a square meter or a column's width to support their solution to the problem in Fig. 1. Notably, Noster et al. (2024) found pre-service teachers often struggle with effective prompting in mathematical problem-solving, yielding incorrect answers. Thus, in the case of near-context FPs, the task characteristics may have facilitated more successful prompting.

Second, these results also suggest that access to an LLM-based conversational system may provide useful assistance for solving accessible FPs. While Schorcht et al. (2024) and Spreitzer et al. (2024) showed ChatGPT can solve real-world mathematical problems depending on complexity, Fock and Siller (2025), and López-Simó and Rezende (2024) warned of limitations in spatial skills and mathematising reality, including risks of imprecise estimates, calculation errors, or unreliable solutions. Here, performance in the AI-supported setting matched outdoor solving and exceeded indoors, suggesting that the AI system may have supported magnitude perception and data quantification (Gallart et al., 2025), or even allowed some participants to solve FPs by delegating part of the process.

If the AI system can generate complete solutions to these problems, this introduces a third problematic aspect: improved performance will only translate into improved learning if pre-service teachers use AI support responsibly and autonomously (Cano-nigo, 2024; Fock & Siller, 2025; Yu, 2023), without fully delegating problem-solving to it (Cevikbas et al., 2023). To examine this, qualitative or mixed-methods studies are needed to analyse participants' interactions with the AI system through their prompts (Noster et al., 2024).

Does previous experience influence pre-service teachers' performance in solving distant-context and complex FPs with AI support?

In the first part of our H2, we expected pre-service teachers' performance on distant-context and complex FPs with AI support to be equivalent across settings from Phase 1. However, results show that those who had previously solved near-context and accessible FPs outdoors performed significantly better in Phase 2.

One possible interpretation of this transfer is that the outdoor setting, as discussed in Sect. 6.1, may promote strategies based on more realistic assumptions (Gallart et al., 2025; Segura et al., 2023). Solving Fermi problems outdoors may require solvers to engage more directly with the physical situation than solving them indoors or with AI support, as perceiving real space involves more than mentally reconstructing the situation from photographs or relying on AI-generated guidance. This may lead to more complete solutions rather than schematic ones (Gallart et al., 2025; Jablonski, 2023b; Segura et al., 2023). From this perspective, participants with previous outdoor experience may have been better prepared for complex FPs like the furniture store capacity problem (Fig. 2), which requires considering a larger number of variables and assumptions (Albarracín et al., 2022).

This finding also contrasts with the expectation that participants in the AI-supported setting in Phase 1, through prior prompting practice, would outperform the other groups in Phase 2, where all participants solved distant-context and complex FPs with AI support. It is possible that these participants developed some prompting-related familiarity during Phase 1. However, this possible advantage did not translate into higher performance in Phase 2 compared with the outdoor group. One possible explanation is that the outdoor setting may have fostered greater autonomy, whereas some participants in the AI-supported setting may have relied too heavily on the AI system, potentially limiting learning (Dasari et al., 2024). In this sense, meaningful gains may depend on integrating AI support into their own problem-solving process rather than delegating the process to the system (Canonigo, 2024). Future qualitative or mixed-methods studies could examine this issue further by analysing participants' prompts and their interactions with the AI system.

Regarding the second part of our H2, pre-service teachers in the indoor setting showed the largest performance gain in Phase 2. This was statistically expected, as they started from significantly lower Phase 1 scores but reached levels similar to those of participants in the AI-supported setting in Phase 2. Thus, previous experience in the AI-supported setting did not appear to provide an additional advantage in this comparison. This may suggest that pre-service teachers can develop effective prompting while working directly on complex FPs with AI support, although this

should be confirmed through analysis of prompt–response interactions. Therefore, while AI support may remain valuable as real-world problem complexity increases (Spreitzer et al., 2024), Phase 2 gains alone do not confirm equivalent development in pre-service teachers' FP-solving performance, given the black-box risk from uncritical digital tool use (Cevikbas et al., 2023; Fock & Siller, 2025).

What is the effect of the setting on pre-service teachers' interest and self-efficacy in solving near-context and accessible FPs?

Our H3 posited that there would be no differences in pre-service teachers' interest and self-efficacy when solving near-context and accessible FPs across settings, but our results contradict these expectations.

Regarding interest, it is significantly lower among indoor setting participants, differing from prior studies that found no differences in students' interest in solving real-world problems indoors, outdoors, or with digital tools (Hartmann & Schukajlow, 2021; Jablonski, 2023). One explanation for this difference is that our study focuses on pre-service teachers, whereas those studies involved secondary students, who have been found to show lower interest in real-world problems (García-Cerdá et al., 2024). In this sense, solving problems outdoors emphasizes their authenticity, thereby activating pre-service teachers' prior interest in real-world mathematical problems. Greater interest in solving FPs with AI support may stem from the novelty of this digital tool compared to those in other studies (e.g., 3D printing in Jablonski, 2023b, 2024). Overall, our findings align with Hidi and Renninger's (2006) emphasis on the learning setting's role in arousing interest.

Regarding self-efficacy, we found it significantly lower among pre-service teachers who solved problems outdoors, despite their superior performance. We expected similar self-efficacy across settings, as in prior studies (Ayotte-Beaudet et al., 2017; Barlovits & Ludwig, 2023), though works like Canonigo (2024) and Dasari et al. (2024) report greater confidence with AI support. Our findings do not fully align with these either, as self-efficacy in the AI-supported setting is not significantly higher than indoors.

Lower self-efficacy when solving FPs outdoors may arise from greater awareness of the challenge in the actual location, where the scenario may appear more complex than when it is mentally reconstructed from photographs. Indeed, Segura et al. (2023) demonstrated that solving FPs outdoors leads participants to incorporate unanticipated contextual features that were overlooked during prior indoor solving: for example, in Fig. 1's problem, outdoor solvers noticed revolving doors and bins occupying space—details not obvious from the photo, despite its familiarity. Since outdoor settings may require more realistic assumptions and thorough mathematization (Gallart et al., 2025), it is understandable that solvers may feel less confident (Lindemann-Matthies et al., 2011). Yet this lower self-efficacy does not impair interest or performance, suggesting that such challenge awareness is beneficial.

Does previous experience influence pre-service teachers' interest and self-efficacy in solving distant-context and complex FPs with AI support?

Although the first part of our H4 predicted that self-efficacy and interest would remain similar across all settings in Phase 2, we found that previous experience was associated with subsequent interest. Pre-service teachers from the indoor setting experienced lower interest than the other settings, and this lower interest persists when solving distant-context and complex FPs with AI support, despite similar gains across the three settings. Since those who solved near-context and accessible FPs in outdoor and AI-supported settings showed higher interest and better performance initially, prior situational interest (Harackiewicz et al., 2016) and success may predict sustained situational interest, aligning with recent findings by Fryer and Shum (2025).

Regarding self-efficacy, our results align with H4: there is no evidence that self-efficacy from the previous experience transferred to later self-efficacy when solving distant-context and complex FPs. This occurs because the self-efficacy gain is much greater for outdoor setting participants, who started from lower levels but equalised with the other two settings in Phase 2.

Our findings confirm the second part of H4, as both interest and self-efficacy were globally higher in Phase 2. The overall increase in interest may be explained by distant-context and complex FPs—such as a concert in a football stadium or the capacity of a furniture store—aligning better with participants' sources of interest (Schulze Elfringhoff & Schukajlow, 2021) than the near-context and accessible FPs—like estimating people on a porch or counting blades of grass in a garden. These distant and complex problems might be perceived as more relevant to their lives or to society (Maaß et al., 2019). Future studies could examine this further by analysing task value across both types of experiences (Böswald & Schukajlow, 2023; Krawitz & Schukajlow, 2018).

Additionally, the overall increase in both interest and self-efficacy in Phase 2 may also be explained by participants' previous experience with accessible FPs, which enabled them to build more stable values for these affective variables (Bandura, 1997; Fryer & Shum, 2025; Usher & Pajares, 2009).

Conclusions, limitations, and future directions

Our results show that the setting shapes pre-service teachers' performance, interest, and self-efficacy when solving real-world problems. For near-context, accessible Fermi problems, performance and interest are higher outdoors or with AI support than indoors, whereas self-efficacy is lower outdoors. In the subsequent phase, when solving more complex Fermi problems with AI support, only those who previously worked outdoors exhibit superior performance; interest differences persist, with lower levels among those who started indoors, while the self-efficacy of those who worked outdoors levels out with the rest. Taken together, these findings suggest that combining outdoor real-world problem solving with AI-supported approaches can be beneficial for pre-service teachers' own task-solving performance, interest, and self-

efficacy. More cautiously, this combination may also be relevant for preparing them to implement Fermi problems in schools, provided that teacher education explicitly promotes responsible use and preserves autonomous decision-making.

These conclusions should be interpreted with caution. Our sample comprised pre-service teachers from a single Spanish university; extending the study to other institutions would improve representativeness. The results should also be supplemented with qualitative work that examines which features of the setting influence solvers' cognitive and affective processes, and how. Accordingly, the single-item, per-problem self-reports of interest and self-efficacy should be complemented with richer observations of affective dimensions. Regarding performance, our quantitative design does not reveal the specific role played by AI in problem solving. Future research should adopt qualitative or mixed-methods approaches (e.g., analyses of prompt–response logs) to study interactions with the AI system during the solution process. Finally, in line with recent work by Bastani et al. (2025)—which finds that requiring solvers to use AI support can depress subsequent performance once they stop using it, compared with those who never used it in the first place—it would be valuable to track pre-service teachers' Fermi problem-solving performance and autonomy longitudinally.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by C. Segura and C. Gallart. The first draft of the manuscript was written by C. Segura and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The authors gratefully acknowledge the support received from Grant No. PID2021-126707NB-I00, funded by MCIU/AEI/10.13039/501100011033 and by “ERDF A way of Making Europe”, as well as from the Special Research Actions Grant UV-INV-AE-3668116 and from the Generalitat Valenciana through project CIGE/2024/64.

Data availability Data from this study are available in Zenodo (DOI: <https://doi.org/10.5281/zenodo.15316429>) and can be shared.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethical approval Participants in the study, being of legal age, gave their consent after being informed of the purpose of the study, how the data would be collected and analysed (quantitative analysis of aggregated, anonymous data), and the publication of the results of this research. All ethical considerations were respected, and approval was obtained from the University Committee of Ethics and Human Research (UV-INV_ETICA-3909285).

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

References

- Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 94(3), 545–561. <https://doi.org/10.1037/0022-0663.94.3.545>
- Albarracín, L., Segura, C., Ferrando, I., & Gorgorió, N. (2022). Supporting mathematical modelling by upscaling real context in a sequence of tasks. *Teaching Mathematics and its Applications*, 41(3), 183–197. <https://doi.org/10.1093/teamat/hrab027>
- Ärleback, J. B. (2009). On the use of realistic Fermi problems for introducing mathematical modelling in school. *The Mathematics Enthusiast*, 6(3), 331–364. <https://doi.org/10.54870/1551-3440.1157>
- Ayotte-Beaudet, J. P., Potvin, P., Lapierre, H. G., & Glackin, M. (2017). Teaching and learning science outdoors in schools' immediate surroundings at K-12 levels: A meta-synthesis. *Eurasia Journal of Mathematics Science and Technology Education*, 13(8), 5343–5363. <https://doi.org/10.12973/eurasia.2017.00833a>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Co.
- Barlovits, S., & Ludwig, M. (2023). Effective or not? The impact of mobile learning on students' interest, self-efficacy, and performance in outdoor mathematics education. In Drijvers, P., Csapodi, C., Palmér, H., Gosztönyi, K., & Kónya, E. (Eds.). *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*. Alfréd Rényi Institute of Mathematics and ERME.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Borromeo Ferri, R. (2018). *Learning how to teach mathematical modeling in school and teacher education*. Springer. <https://doi.org/10.1007/978-3-319-68072-9>
- Borromeo Ferri, R. (2021). Mandatory Mathematical Modelling in School: What Do We Want the Teachers to Know? In F. K. S. Leung, G. Stillman, G. Kaiser & K. L. Wong (Eds.), *Mathematical Modelling Education in East and West*, 103–117. Springer. https://doi.org/10.1007/978-3-030-66996-6_9
- Böswald, V., & Schukajlow, S. (2023). I value the problem, but I don't think my students will: preservice teachers' judgments of task value and self-efficacy for modelling, word, and intramathematical problems. *ZDM – Mathematics Education*, 55, 331–344. <https://doi.org/10.1007/s11858-022-01412-z>
- Buchholtz, N. (2021). Processos de modelação dos alunos envolvidos em trilhos matemáticos. *Quadrante*, 30(1), 140–157. <https://doi.org/10.48489/quadrante.23699>
- Canonigo, A. M. (2024). Levering AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of computer assisted learning*, 40(6), 3215–3229. <https://doi.org/10.1111/jcal.13065>
- Cevikbas, M., Greefrath, G., & Siller, H. S. (2023). Advantages and challenges of using digital technologies in mathematical modelling education—A descriptive systematic literature review. *Frontiers in Education*, 8, 1142556. <https://doi.org/10.3389/educ.2023.1142556>
- Cevikbas, M., Kaiser, G., & Schukajlow, S. (2022). A systematic literature review of the current discussion on mathematical modelling competencies: State-of-the-art developments in conceptualizing, measuring, and fostering. *Educational Studies in Mathematics*, 109, 205–236. <https://doi.org/10.1007/s10649-021-10104-6>
- Chapman, O. (2013). Mathematical-task knowledge for teaching. *Journal of Mathematics Teacher Education*, 16, 1–6. <https://doi.org/10.1007/s10857-013-9234-7>
- Chapman, O. (2015). Mathematics teachers' knowledge for teaching problem solving. *LUMAT: International Journal on Math. Science and Technology Education*, 3(1), 19–36. <https://doi.org/10.31129/umat.v3i1.1049>
- Cross, R. (1997). Developing Maths Trails. *Mathematics Teaching*, 158, 38–39.
- Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education*, 8, 1295413. <https://doi.org/10.3389/educ.2023.1295413>
- English, L. D. (2023). Multidisciplinary modelling in a sixth-grade tsunami investigation. *International Journal of Science and Mathematics Education*, 21, 41–65. <https://doi.org/10.1007/s10763-022-10303-4>

- Ferrando, I., & Albarracín, L. (2021). Students from grade 2 to grade 10 solving a Fermi problem: Analysis of emerging models. *Mathematics Education Research Journal*, 33, 61–78. <https://doi.org/10.1007/s13394-019-00292-z>
- Fock, A., & Siller, H. S. (2025). Generative artificial intelligence in secondary STEM education in the light of Human Flourishing: a scoping literature review. *International Journal of STEM Education*, 12. <https://doi.org/10.1186/s40594-025-00589-5>
- Fryer, L. K., & Shum, A. (2025). Math task experiences and motivation to learn more: How prior knowledge and interest interact with Task-Interest & Task-Difficulty perceptions and feed a desire to reengage. *British Educational Research Journal*, 51(1), 49–72. <https://doi.org/10.1002/berj.4057>
- Gallart, C., Segura, C., & Albarracín, L. (2025). Impacto del entorno de resolución de problemas de Fermi en la complejidad y flexibilidad de las estrategias de maestros en formación. *Contextos Educativos*, 35, 139–164. <https://doi.org/10.18172/con.6480>
- García-Cerdá, C., Segura, C., & Ferrando, I. (2024). Influence of the type of mathematical problems on students' and pre-service teachers' interest and performance. A replication and elaboration study. *Implementation and Replication Studies in Mathematics Education*, 4(1), 125–159. <https://doi.org/10.1163/26670127-bja10017>
- Gasco, J., & Villarroel, J. D. (2014). The motivation of secondary school students in mathematical word problem solving. *Electronic Journal of Research in Educational Psychology*, 12(1), 83–106. <https://doi.org/10.14204/ejrep.32.1>
- Geiger, V., Galbraith, P., Niss, M., & Delzoppo, C. (2022). Developing a task design and implementation framework for fostering mathematical modeling competencies. *Educational Studies in Mathematics*, 109(2), 313–336. <https://doi.org/10.1007/s10649-021-10039-y>
- Geiger, V., Stillman, G., Brown, J., Galbraith, P., & Niss, M. (2018). Using mathematics to solve real world problems: The role of enablers. *Mathematics Education Research Journal*, 30(1), 7–19. <https://doi.org/10.1007/s13394-017-0217-3>
- Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics*, 39(1–3), 111–129. <https://doi.org/10.1023/A:1003749919816>
- Greefrath, G., Siller, H. S., Klock, H., & Wess, R. (2022). Pre-service secondary teachers' pedagogical content knowledge for the teaching of mathematical modelling. *Educational Studies in Mathematics*, 109(2), 383–407. <https://doi.org/10.1007/s10649-021-10038-z>
- Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy insights from the behavioral and brain sciences*, 3(2), 220–227. <https://doi.org/10.1177/237273221665555>
- Hartmann, L. M., & Schukajlow, S. (2021). Interest and Emotions While Solving Real-World Problems Inside and Outside the Classroom. In: Leung, F. K. S., Stillman, G. A., Kaiser, G., Wong, K. L. (Eds.), *Mathematical Modelling Education in East and West*, 153–163. Springer. https://doi.org/10.1007/978-3-030-66996-6_13
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4
- Honick, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational research review*, 17, 63–84. <https://doi.org/10.1016/j.edurev.2015.11.002>
- Jablonski, S. (2023a). Indoors vs. Outdoors: Student Perception of Different Modelling Settings. *Research in Integrated STEM Education*, 1(3), 403–421. <https://doi.org/10.1163/27726673-bja00016>
- Jablonski, S. (2023b). Is it all about the setting? A comparison of mathematical modelling with real objects and their representation. *Educational Studies in Mathematics*, 113, 307–330. <https://doi.org/10.1007/s10649-023-10215-2>
- Jablonski, S. (2024). Challenges in geometric modelling—A comparison of students' mathematization with real objects, photos, and 3D models. *EURASIA Journal of Mathematics Science and Technology Education*, 20(3), em2414. <https://doi.org/10.29333/ejmste/14321>
- Jablonski, S., Barlovits, S., & Ludwig, M. (2023). How digital tools support the validation of outdoor modelling results. *Frontiers in education*, 8, 1145588. <https://doi.org/10.3389/feduc.2023.1145588>
- Klock, H., & Siller, H.S. (2020). A time-based measurement of the intensity of difficulties in the modelling process. In H. Wessels, G.A. Stillman, G. Kaiser, & E. Lampen (Eds.), *International Perspectives on the Teaching and Learning of Mathematical Modelling*, 163–173. Springer. https://doi.org/10.1007/978-3-030-37673-4_15

- Knabbe, A., Leiss, D., & Ehmke, T. (2025). Reality-Based Tasks with Complex-Situations: Identifying Sociodemographic and Cognitive Factors for Solution. *International Journal of Science and Mathematics Education*, 23, 97–120. <https://doi.org/10.1007/s10763-024-10463-5>
- Korseberg, L., & Elken, M. (2025). Waiting for the revolution: How higher education institutions initially responded to ChatGPT. *Higher Education*, 89(4), 953–968. <https://doi.org/10.1007/s10734-024-01256-4>
- Krawitz, J., & Schukajlow, S. (2018). Do students value modelling problems, and are they confident they can solve such problems? Value and self-efficacy for modelling, word, and intra-mathematical problems. *ZDM – Mathematics Education*, 50, 143–157. <https://doi.org/10.1007/s11858-017-0893-1>
- Krawitz, J., Schukajlow, S., & Van Dooren, W. (2018). Unrealistic responses to realistic problems with missing information: what are important barriers? *Educational Psychology*, 38(10), 1221–1238. <https://doi.org/10.1080/01443410.2018.1502413>
- Lindemann-Matthies, P., Constantinou, C., Lehnert, H. J., Nagel, U., Raper, G., & Kadji-Beltran, C. (2011). Confidence and perceived competence of preservice teachers to implement biodiversity education in primary schools: Four comparative case studies from Europe. *International Journal of Science Education*, 33(16), 2247–2273. <https://doi.org/10.1080/09500693.2010.547534>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13, 410. <https://doi.org/10.3390/educsci13040410>
- López-Simó, V., & Rezende, M. F. (2024). Challenging ChatGPT with Different Types of Physics Education Questions. *The Physics Teacher*, 62, 290–294. <https://doi.org/10.1119/5.0160160>
- Lund, B. D. & Wang, T. (2023) Chatting about ChatGPT: How May AI and GPT Impact Academia and Libraries? *Library Hi Tech News*. <https://doi.org/10.2139/ssrn.4333415>
- Maaß, K. (2010). Classification scheme for modelling tasks. *Journal für Mathematik-Didaktik*, 31(2), 285–311. <https://doi.org/10.1007/s13138-010-0010-2>
- Maaß, K., Geiger, V., Ariza, M. R., & Goos, M. (2019). The Role of Mathematics in interdisciplinary STEM education. *Zdm*, 51, 869–884. <https://doi.org/10.1007/s11858-019-01100-5>
- McLeod, D.B. (1989). Beliefs, Attitudes, and Emotions: New Views of Affect in Mathematics Education. In: McLeod, D.B., Adams, V.M. (Eds.) *Affect and Mathematical Problem Solving*, 245–258 Springer. https://doi.org/10.1007/978-1-4612-3614-6_17
- Noster, N., Gerber, S., & Siller, H. S. (2024). Pre-Service Teachers' Approaches in Solving Mathematics Tasks with ChatGPT. *Digital Experiences in Mathematics Education*, 10, 543–567. <https://doi.org/10.1007/s40751-024-00155-8>
- Nuutila, K., Tapola, A., Tuominen, H., Kupiainen, S., Pásztor, A., & Niemivirta, M. (2020). Reciprocal predictions between interest, self-efficacy, and performance during a task. *Frontiers in Education*, 5(1), 1–13. <https://doi.org/10.3389/feduc.2020.00036>
- Peter-Koop, A. (2009). Teaching and Understanding Mathematical Modelling through Fermi-Problems. In B. Clarke, B. Grevholm & R. Millman (Eds.), *Tasks in primary mathematics teacher education*, 131–146. Springer. https://doi.org/10.1007/978-0-387-09669-8_10
- Quarder, J., Greefrath, G., Gerber, S., & Siller, H. S. (2025). Pedagogical content knowledge for simulations and mathematical modelling with digital tools: a quasi-experimental study with pre-service mathematics teachers. *ZDM – Mathematics Education*, 1–15. <https://doi.org/10.1007/s11858-025-01673-4>
- Rasch, D. & Guizard, V. (2004). The robustness of parametric statistical methods. *Psychology Science*, 46, 175–208.
- Rellessmann, J., & Schukajlow, S. (2017). Does students' interest in a mathematical problem depend on the problem's connection to reality? An analysis of students' interest and pre-service teachers' judgments of students' interest in problems with and without a connection to reality. *ZDM – Mathematics Education*, 49(3), 367–378. <https://doi.org/10.1007/s11858-016-0819-3>
- Richardson, K. (2004). Designing math trails for the elementary school. *Teaching Children Mathematics*, 11, 8–14. <https://doi.org/10.5951/TCM.11.1.0008>
- Robinson, A. W. (2008). Don't just stand there—teach. *Fermi problems! Physics Education*, 43(1), 83–87. <https://doi.org/10.1088/0031-9120/43/01/009>
- Segura, C., & Ferrando, I. (2023). Pre-service teachers' flexibility and performance in solving Fermi problems. *Educational Studies in Mathematics*, 113, 207–227. <https://doi.org/10.1007/s10649-023-10220-5>
- Segura, C., Ferrando, I., & Albarracín, L. (2023). Does collaborative and experiential work influence the solution of real-context estimation problems? A study with prospective teachers. *The Journal of Mathematical Behavior*, 70, 101040. <https://doi.org/10.1016/j.jmathb.2023.101040>

- Segura, C., Gallart, C., & Ferrando, I. (2026). Influence of pre-service primary school teachers' prior knowledge of measurement and measurement estimation in solving modelling problems. *Journal of Mathematics Teacher Education*, 29, 433–458. <https://doi.org/10.1007/s10857-025-09685-3>
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.
- Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, 10(1), 9–34. <https://doi.org/10.54870/1551-3440.1258>
- Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the problem – investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education*, 9, 1386075. <https://doi.org/10.3389/educ.2024.1386075>
- Schukajlow, S., Kaiser, G., & Stillman, G. (2018). Empirical research on teaching and learning of mathematical modelling: A survey on the current state-of-the-art. *ZDM – Mathematics Education*, 50, 5–18. <https://doi.org/10.1007/s11858-018-0933-5>
- Schukajlow, S., Krawitz, J., Kanefke, J., & Rakoczy, K. (2022). Interest and performance in solving open modelling problems and closed real-world problems. In C. Fernández, S. Linares, A. Gutiérrez, & N. Planas (Eds.). *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 3, pp. 403–410. PME.
- Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics*, 79(2), 215–237. <https://doi.org/10.1007/s10649-011-9341-2>
- Schulze Elfiringhoff, M., & Schukajlow, S. (2021). O que torna um problema de modelação interessante? Fontes de interesse situacional em problemas de modelação. *Quadrante*, 30(1), 8–30. <https://doi.org/10.48489/quadrante.23861>
- Spreitner, C., Straser, O., Zehetmeier, S., & Maaß, K. (2024). Mathematical Modelling Abilities of Artificial Intelligence Tools: The Case of ChatGPT. *Education Sciences*, 14, 698. <https://doi.org/10.3390/educsci14070698>
- Sriraman, B., & Knott, L. (2009). The mathematics of estimation: Possibilities for interdisciplinary pedagogy and social consciousness. *Interchange*, 40(2), 205–223. <https://doi.org/10.1007/s10780-009-9090-7>
- Street, K. E. S., Malmberg, L. E., & Schukajlow, S. (2024). Students' mathematics self-efficacy: a scoping review. *ZDM – Mathematics Education*, 56, 265–280. <https://doi.org/10.1007/s11858-024-01548-0>
- Talsma, K., Schütz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, 61, 136–150. <https://doi.org/10.1016/j.lindif.2017.11.015>
- Thompson, P. W. (1985). Experience, problem solving, and learning mathematics: Considerations in developing mathematics curricula. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives*, 189–243. Erlbaum.
- Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34(1), 89–101. <https://doi.org/10.1016/j.cedpsych.2008.09.002>
- Vale, I., Barbosa, A., & Cabrita, I. (2019). Mathematics outside the classroom: Examples with preservice teachers. *Quaderni di Ricerca in Didattica*, 2(3), 138–142.
- Wess, R., Klock, H., Siller, H.S., & Greefrath, G. (2021). Professional Competence for Teaching Mathematical Modelling. In *Measuring Professional Competence for the Teaching of Mathematical Modelling. International Perspectives on the Teaching and Learning of Mathematical Modelling*, 21–36. Springer. https://doi.org/10.1007/978-3-030-78071-5_2
- Yu, H. (2023). Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontier in Psychology*, 14, 1181712. <https://doi.org/10.3389/fpsyg.2023.1181712>
- Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computer and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>