International Electronic Journal of Mathematics Education
2026, 21(2), em0873
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Research Article

OPEN ACCESS

Training mathematics teachers' error behavior through ChatGPT-based error situations

Katja Lenz [1], Julia Sirock [2*]

[1] University of Education Schwäbisch Gmünd, Schwäbisch Gmünd, GERMANY
[2] University of Education Heidelberg, Heidelberg, GERMANY
*Corresponding Author: sirock@ph-heidelberg.de

Citation: Lenz, K., & Sirock, J. (2026). Training mathematics teachers' error behavior through ChatGPT-based error situations. *International Electronic Journal of Mathematics Education, 21*(2), em0873. https://doi.org/10.29333/iejme/18101

ARTICLE INFO

Received: 28 Jul. 2025

Accepted: 28 Dec. 2025

ABSTRACT

This study examines the potential of ChatGPT as a training tool for prospective mathematics teachers, focusing on error behavior and instructional strategies for solving word problems. ChatGPT consistently generated incorrect solutions, mirroring common learner error patterns. Based on the responses of prospective teachers (N = 26) to these errors, three types of guidance emerged: co-constructive, directive, and non-responsive. These types represent varying instructional strategies guiding learners, as identified in previous research. In addition, participants evaluated the realism and usefulness of ChatGPT-based interactions. While many found the tool valuable for practicing guidance techniques and anticipating learners' misconceptions, they noted limitations at the same time, including a lack of emotional nuance. Overall, the findings emphasize both the opportunities and limitations of using artificial intelligence-based dialogue systems in teacher education.

Keywords: artificial intelligence in education, ChatGPT, error behavior, teacher-student interaction, teacher education, mathematics education, word problem-solving

INTRODUCTION

Errors are an integral part of the learning process, as making and overcoming them constitutes a fundamental mechanism of conceptual change and cognitive development (Steuer, 2014; Steuer et al., 2013). The way how teachers approach errors in the classroom critically influences how learners perceive them and fosters learners' willingness to explore and transform them into learning opportunities (Heinze, 2004; Santagata, 2005). A supportive error climate, characterized by open dialogue, reflection, and constructive feedback, fosters deeper understanding, whereas punitive or dismissive responses can discourage engagement and hinder learning (Tulis, 2013). Therefore, teachers play a decisive role in determining whether errors are perceived as failures to be avoided or as opportunities for growth.

Research on classroom discourse has revealed a wide spectrum of teacher responses to learners' errors, ranging from direct correction or redirection to peers to adaptive practices, such as prompting self-correction or initiating reflective discussion (Hiebert et al., 2003; Oser & Spychiger, 2005; Santagata, 2005). Adaptive responses, such as guiding learners toward recognizing and repairing their own errors, are associated with more effective and lasting learning gains (Anderson et al., 2004; Heimbeck et al., 2003; Tulis, 2013). In contrast, maladaptive responses, including ignoring errors, showing frustration, or immediately providing the correct answer, may limit opportunities to learn. Nevertheless, empirical findings indicate that actual error events occur relatively infrequently in classroom settings, with only a handful of errors typically observed per lesson (Heinze, 2004; Meyer et al, 2006; Oser & Spychiger, 2005; Santagata, 2005), limiting the opportunities for prospective teachers to practice such responses in authentic contexts. This scarcity of authentic error events poses a challenge for teacher education, as it limits opportunities for prospective teachers to practice adaptive responses in authentic classroom contexts. Consequently, teacher education must create alternative learning environments.

Recent developments in artificial intelligence (AI) present new opportunities for such training scenarios. AI-based tools can simulate classroom discourse and generate plausible learners' errors, thereby providing prospective teachers with practice opportunities (Kortenkamp & Dohrmann, 2023). Such simulated discourse environments can bridge the gap between theoretical knowledge about error management and its practical application, as well as allow prospective teachers to experience, reflect on, and refine their responses to learners' errors in a safe and controlled setting.

Building on this potential, the present study examines ChatGPT as a training tool for prospective teachers in mathematics education. In addition to the already known learner-likeness, the behavior of prospective teachers in conversation with ChatGPT is also considered focusing on instructional strategies for solving word problems. Furthermore, the study also looks at the perception of prospective teachers regarding authenticity, motivation and support.

Professional Development of Teachers

It is generally recognized that professional skills form an important basis for the work of teachers in the classroom and are therefore an essential prerequisite for university education quality and learners' performance (Scheunpflug et al., 2006). These skills distinguish novices from experts, who are described as highly experienced and qualified teachers (Stahnke & Blömeke, 2021). It has been confirmed that experts, in comparison to novices, have access to more extensive, better organized, and at the same time more effective knowledge in teaching situations and can transfer this knowledge to their everyday activities (Berliner, 1991; Bromme, 1992; Stahnke & Blömeke, 2021).

The foundation of this professional knowledge is established during education at university and further enhanced through practical experience (cf. Riese & Reinhold, 2012). It can be assumed that prospective teachers already have a basis of expertise regarding the subject content they wish to teach at the beginning of their education at university (Shulman, 1986). Furthermore, prospective teachers develop professional competencies during their university education and the subsequent practical phase through a wide range of (subject-specific) learning opportunities and systematic feedback. Therefore, university education forms the basis for acquiring professional skills (Besser & Krauss, 2009; Bromme, 2008; Kennedy et al., 2008).

Teachers' professional skills are in a constant state of development that is not necessarily linear (Bastian et al., 2022; Baumert & Kunter, 2011, 2006). Thus, university education, with a subsequent practical phase, is the most important element of the formal learning program, in which prospective teachers develop professional skills (cf. Cochran-Smith & Zeichner, 2005; Terhart, 2001, 2008). Although professional knowledge is acquired during university education, it is further deepened in the subsequent practical phase, where it unfolds its operational potential (Bastian et al., 2022; Baumert et al., 2011; Kunter et al., 2011).

Simulations offer an opportunity to teach practical skills during university education. Using simulative methods in teacher education can be well justified in terms of professional theory, and their effectiveness has been proven many times before. A review published in 2022 synthesized 13 effectiveness studies of simulative methods and found positive effects on the professional skills of prospective teachers in the areas of beliefs, self-efficacy, knowledge, communicative skills, and reflective competence (Ade-Ojo et al., 2022). However, there is little corresponding debate regarding teacher education in German-speaking countries, even though such discussions are occurring more on an international scale.

Practice simulations, such as social experiments, role plays, and simulation games, provide valuable opportunities to connect theoretical concepts with practical applications in professional contexts (Kadel et al., 2023). Prospective teachers can experiment in a safe environment and develop strategies for future professional challenges (Kadel et al., 2023). One objective of practice-simulating arrangements is to enhance the theoretical knowledge gained about various types of tasks, such as complex word problems, while also experiencing common misconceptions in a realistic context.

Research on ChatGPT in Mathematics Education

Recent research in mathematics education increasingly explores the integration of AI in mathematics teaching and learning, particularly large language models (LLMs) such as ChatGPT. Three major directions can be distinguished:

(1) ChatGPT's performance in solving mathematical problems,

(2) the support ChatGPT can offer to learners, and

(3) its use as a tool in the professional development of mathematics teachers.

Research on ChatGPT's mathematical performance reveals that the model can successfully solve routine and procedural tasks, but it often struggles with multi-step reasoning, abstraction, and context interpretation (Caylan-Ergene et al., 2025; Dahal, 2024; Frieder et al., 2023b; Xuan-Quy & Ngoc-Bich, 2023). Studies demonstrate that while ChatGPT's accuracy improves with well-structured prompts (Arora et al., 2022; Schorcht et al., 2023), newer model iterations (Dahal, 2024; Ergene & Ergene, 2025) and the use of external computational tools such as Wolfram Alpha, its explanations frequently remain superficial or contain subtle logical errors (Helfrich-Schkarbanenko, 2023; Spannagel, 2023). In particular, a study by Caylan-Ergene et al. (2025) found that ChatGPT produced correct solutions to only about two-thirds of statistics problems posed by prospective teachers. Further analysis revealed that ChatGPT performed significantly better on purely symbolic problems than contextualized problems, whereas four major error types were identified (calculation errors, data omission, generating non-existing data, and misinterpreting correct calculations).

These findings indicate progress in computational reliability but underscore the continuing need for critical human evaluation of AI-generated mathematical reasoning. In the field of learner support, this highlights the complementary role that AI can play when used as an interactive learning aid rather than an autonomous problem solver. ChatGPT can support learners in mathematics by providing immediate explanations, step-by-step guidance, and adaptive feedback, which can enhance engagement and self-directed learning (Asare et al., 2023; Ergene & Caylan-Ergene, 2024). Studies show that its use can reduce cognitive load and encourage metacognitive reflection, helping learners articulate their reasoning processes (Contel & Cusi, 2025). However, its effectiveness depends on learners' ability to critically evaluate AI responses, as unreflective use may reinforce misconceptions rather than deepen understanding (Heung & Chiu, 2025).

In order to effectively guide such AI-supported learning, teachers need to develop AI literacy. This is essential not only for teaching learners to use AI responsibly but may also foster teachers' own professional development. Through AI-based tools such as ChatGPT, teachers can enhance their professional practice by using the technology for lesson planning (Huget & Buchholtz, 2024; Şimşek, 2025; Yanar & Ergene, 2025) and engaging in simulated teaching dialogues that allow them to practice scaffolding, providing feedback, and developing diagnostic reasoning in realistic learning contexts (Gíslason, 2023; Kortenkamp & Dohrmann, 2023).

This study is situated at the intersection of teacher education and instructional interaction**.** While previous studies have examined ChatGPT's ability to solve mathematical problems, fewer have explored its potential as a simulation tool for developing pedagogical guidance skills**.** By analyzing how prospective teachers interact with ChatGPT's erroneous solutions to word problems, this research extends the current discourse from assessing AI's correctness and usefulness to investigating how AI can serve as a pedagogical instrument in teacher education.

ChatGPT as a Simulated Learner: Developing Instructional Guidance Through AI-Generated Errors

LLMs, such as ChatGPT, offer new opportunities for practice-based simulations in teacher education. Research has shown that ChatGPT often produces errors similar to those made by learners (Kortenkamp & Dohrmann, 2023; Schorcht et al., 2023). This learner-like error behavior allows prospective teachers to engage in authentic instructional dialogues, providing an opportunity to practice identifying misconceptions and guiding problem-solving processes. The techniques used to help ChatGPT reach correct solutions during interaction, such as providing hints or reformulating prompts, mirror common scaffolding strategies in classroom discourse and thus offer valuable training experiences (Arora et al., 2022; Schorcht et al., 2023).

In the context of word problems, ChatGPT introduces a potentially transformative development for mathematics education. Generative language models, such as ChatGPT, provide the capability to solve word problems directly, eliminating the need for users to first convert the scenario into a formal mathematical model (Spannagel, 2023). While this feature increases accessibility, it also heightens the risk of error: ChatGPT often misinterprets problem contexts or applies inappropriate algorithms, resulting in plausible yet incorrect reasoning (Ammel, 2023; Bryden, 2023; Frieder et al., 2023a, 2023b; Helfrich-Schkarbanenko, 2023; Kortenkamp & Dohrmann, 2023).

Several studies have shown that ChatGPT can be directed toward correct solutions using specific prompts (Arora et al., 2022; Schorcht et al., 2023). This interactive process closely resembles teacher-student dialogue, where targeted questions and feedback promote conceptual understanding (Gíslason, 2023). Developing these dialogic competencies is essential for effective teaching. However, traditional opportunities to practice these skills, such as field placements, are often limited in availability and consistency. ChatGPT could provide a complementary learning environment by serving as a conversational partner that simulates learner-like reasoning within a controlled, repeatable setting (Gíslason, 2023). The error patterns produced by ChatGPT enhance its pedagogical value in teacher education. Its tendency to correct errors locally without revising the entire reasoning process mirrors common learner behaviors, allowing prospective teachers to practice error analysis and responsive feedback (Kortenkamp & Dohrmann, 2023). Nevertheless, important differences remain: ChatGPT's responses lack emotional depth and spontaneity, it may abruptly change strategies or mechanically persist through problem-solving tasks without displaying frustration or confusion (Gíslason, 2023).

Overall, ChatGPT provides a valuable simulated practice environment for prospective mathematics teachers. By engaging in interactive dialogues, they can rehearse strategies such as prompting, scaffolding, and error-sensitive feedback. Although these interactions are artificial, they offer realistic, practice-oriented scenarios that strengthen diagnostic reasoning and communication skills essential for effective mathematics instruction.

Solving Word Problems

Solving word problems is a critical aspect of mathematics education because it requires applying mathematical knowledge to real-world contexts. This skill enables learners to use mathematics authentically and improves their general problem-solving abilities (Verschaffel et al., 2020). Despite its importance, solving word problems remains a considerable challenge for many learners (Daroczy et al., 2015).

This can be attributed to the additional cognitive demands imposed by word problems. Learners must translate real-life scenarios into mathematical models, solve them, and then interpret the results within the original context (Blum & Leiss, 2005). A crucial intermediate step in this process involves constructing a "situation model"–which is a mental representation of the real-world scenario, built from prior knowledge and logical inference (Greefrath et al., 2013; Leiss et al., 2010; Verschaffel et al., 2020). However, this step often poses a major hurdle for learners. Research has shown that learners often bypass the construction of a situation model by depending on surface-level cues, such as numbers or keywords, to derive solutions, often without a comprehensive understanding of the problem context (Hegarty et al., 1995; Verschaffel et al., 2020; Vicente et al., 2022). Moreover, learners tend to neglect key stages in their problem-solving processes, such as interpreting and validating their results (Verschaffel et al., 2000). While such strategies may occasionally yield correct answers for simple problems, they are inadequate and often misleading when applied to more complex, non-routine word problems (Vicente et al., 2022). Non-routine word problems are mathematical tasks in real-life contexts for which the solution path is not immediately apparent and cannot be addressed through the routine application of standard procedures. Instead, they require creative thinking and heuristic strategies to interpret the problem context and devise an appropriate solution approach (Elia et al., 2009; Pantziara et al., 2009). These problems are typically characterized by a complex mathematical structure, the absence of a clear algorithmic solution path, and the potential for multiple valid strategies (Vincente et al., 2022). Consequently, solving such problems requires advanced problem-solving competencies beyond the mere application of known algorithms (Elia et al., 2009; Pólya, 2010).

Given these findings, it is essential to support learners develop the necessary skills to effectively engage with complex word problems. Success in solving complex word problems relies heavily on metacognitive skills and the use of heuristic strategies (Pantziara et al., 2009; Vicente et al., 2022) such as decomposition, analogical reasoning, reducing unfamiliar problems to familiar ones, and reasoning forward or backward through the task (Pólya, 2010). However, these heuristics do not typically arise spontaneously. Instead, they must be explicitly taught and practiced under the guidance of competent educators using thoughtfully designed tasks. Therefore, assisting learners in solving complex word problems is a central core responsibility of mathematics teachers that should be a focus of teacher education.

Research Questions

ChatGPT's responses are often inaccurate. When applied to word problems, its output reflects common errors made by learners, as identified in mathematics education research. This similarity suggests that ChatGPT could serve as a valuable training tool for teacher education, as it requires users to verify responses and guide solution processes through interaction. In this study, we explore the potential of using ChatGPT in the education of prospective teachers, with a focus on word problem-solving. In concrete terms, the following research questions (RQs) are addressed:

RQ1. To what extent does ChatGPT provide learner-like responses that can be used for teacher professionalization?

RQ2. What behavior do prospective teachers show when guiding ChatGPT through solving word problems?

RQ3. How do prospective teachers perceive ChatGPT as a useful simulation tool for teacher education?

MATERIALS AND METHODS

Sample

In Germany, prospective teachers undergo a structured university education program that typically begins with a Bachelor of arts in education. This degree equips them with essential knowledge in mathematics as well as effective teaching methodologies. The regular study period is 6 semesters for primary and secondary school teacher education and education for special needs. Afterwards, the prospective teachers pursue a master's degree, which takes an additional 2 semesters for primary school education and 4 semesters for secondary school education and education for special needs, and focuses on advanced mathematics education, didactics, and research. The mathematics educational studies also include teaching practice through internships and longer-term school-based experiences. In our study, N = 26 prospective teachers from the Bachelor's degree program in primary education and education for special needs participated. The study was conducted as part of a seminar at Heidelberg University of Education; participation was voluntary. The average age of the participants was *mean (M) (standard deviation [SD])* = 22.07 (1.90) years (range 20-27 years). At the time of the survey, the participants were, on average, in their 4th semester (range 2nd-7th semester). The sample was predominantly female (85% female, 15% male, and 0% non-binary).

Data Gathering

In the context of a subject-specific didactics course, a learning activity focused on solving word problems with ChatGPT (browser-based version 3.5) was implemented on January 21, 2024. This activity followed a session that addressed the goals and functions of problem-solving in mathematics, as well as the different task formats and their distinctions. The prospective teachers also had prior knowledge regarding common support strategies and typical errors made by learners when solving word problems. However, this prior knowledge was not explicitly tied to the type of problem-based tasks used in the study. During the seminar session, the functionality of ChatGPT was first explained, and an example of how to use AI to solve word problems was demonstrated. Following this, the prospective teachers were tasked with solving a complex word problem using ChatGPT and guiding the solution process (see **Figure 1**). The participants worked alone or in pairs to foster a productive learning environment and encourage peer discussion. For this reason, only 21 chats are available for evaluation, even though 26 participants were involved. The problem used in the study has been previously tested with ChatGPT and resulted in incorrect solutions, which made it suitable for the study. Immediately after completing the task, the prospective teachers were surveyed digitally using the browser-based questionnaire software SoSciSurvey (Leiner, 2019). For the reliability analysis, Cronbach's alpha was calculated to assess the internal consistency of the scale. The internal consistency of the questionnaire is good (cf. Taber, 2018), with Cronbach's alpha = .81 (5 items). The survey addressed two main aspects:

(1) the authenticity of the conversations with ChatGPT in relation to interactions with learners and

(2) the motivation and support provided by the conversations with ChatGPT in relation to their future profession.

Task: A snail in a 20 m deep well wants to go up to the meadow. It continuously crawls up 5 m during the day and slides back down 2 m at night while sleeping. On how many days does it reach the edge of the well?
Work assignment: 1. Imagine that ChatGPT is your student. 2. Have ChatGPT solved the word problem. To do this, copy the text of the task into the input field. 3. Give ChatGPT suggestions on how it can improve the solution. Guide ChatGPT to fully understand the word problem and to find a correct and reasoned solution.

Figure 1. Word problem and work assignment for prospective teachers (Source: Authors' own elaboration)

Regarding the authenticity, motivation, and support of conversations with ChatGPT in relation to potential interactions with learners was assessed using five items in total.

Authenticity

- "I can easily imagine myself engaging in a teacher-student conversation while interacting with ChatGPT." (realistic)

- “The responses from ChatGPT in the conversation appear learner-like.” (learner-like)

Motivation and support

- “I find the simulated teacher-student interaction with ChatGPT helpful for my future career.” (helpful)
- “I consider using ChatGPT as a learning opportunity for teacher-student interaction to be meaningful.” (useful)
- “I find interacting with ChatGPT motivating within the context of my teacher education.” (motivating)

Participants were asked to indicate their level of agreement with the provided statements using a 6-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = somewhat agree, 5 = agree, 6 = strongly agree).

Although both data strands were collected concurrently, they were analyzed separately and therefore do not constitute a typical mixed-methods design (Creswell & Plano Clark, 2018) but rather provide complementary insights. The qualitative data were used to analyze ChatGPT’s errors and to describe and categorize prospective teachers’ guidance during word problem-solving. The quantitative data captured participants’ self-reported perceptions and attitudes toward ChatGPT and were enriched with open-ended questions. These open responses were used to further explain the quantitative results and to gain deeper insight into participants’ evaluative judgments.

Data Analysis

To answer **RQ1**, the response behavior of ChatGPT was analyzed qualitatively. The responses were then compared to learner solutions documented in previous research. This comparison aimed to identify common patterns or errors in ChatGPT’s problem-solving approach that align with typical learners’ behavior when solving mathematical tasks, particularly word problems.

Concerning **RQ2**, the chat transcripts were analyzed qualitatively using MAXQDA (2025), with categories being inductively derived from the data (**Table 1**). By examining the content of the interactions in detail, the researchers developed a coding system that reflects the nuances of the responses provided by the prospective teachers. This inductive approach helps ensure that the categories are grounded in the actual data, allowing for a more comprehensive and context-specific understanding of the participants’ experiences and behaviors during the task. This enables the identification of recurring patterns in instructional behavior and the derivation of a typology of pedagogical support strategies. MAXQDA (2025) facilitated a structured and systematic approach to organizing and interpreting the data, supporting the overall analysis of how participants engaged with the AI tool in solving mathematical word problems. The coding process followed a consensual qualitative approach (Kuckartz & Rädiker, 2022). Two researchers independently analyzed the interview data, assigned text segments to codes, and developed inductive categories. After this independent phase, the coders compared their results to identify agreements and discrepancies. Disagreements were discussed in light of the theoretical framework to refine category definitions and achieve shared understanding. In each case, a consensus could be reached. Intercoder reliability was assessed based on the percentage of agreement, yielding an overall agreement of 90.20%. Agreement levels across individual codes ranged from 84.21% to 97.67%, indicating high and acceptable reliability of the coding. The in-depth discussions and negotiated consensus among coders are considered appropriate to ensure validity in the qualitative content analysis conducted, as the aim was to develop theoretically grounded, context-sensitive interpretations, rather than to test a predefined coding scheme.

Table 1. Code system–Categories and brief descriptions

Code	Definition	Examples from transcripts
Processing of the task by ChatGPT		
Correct solution	ChatGPT gives a mathematically and conceptually accurate answer.	“The snail reaches the top on day 6 before sliding back.”
Incomplete solution	ChatGPT provides a response that lacks a final answer or stops mid-way.	“The snail climbs 3 meters per day ...” [but no conclusion is given].
Error	An incorrect solution or flawed reasoning by ChatGPT.	“It takes 6.67 days, so the snail reaches the top on day 7.”
Reaction and guidance of prospective teachers		
Error reference (general)	The prospective teacher points out that the solution is not correct.	“Unfortunately, this is not correct. Try again.”
Error reference (specific)	The prospective teacher explicitly refers to a specific error in the solution.	“That’s not correct, the snail doesn’t fall back on day 7.”
Solution hint (General)	The prospective teacher makes a general suggestion without revealing the answer. For example, they might advise decomposing the problem or provide reasoning prompts (overarching level).	“What happens at the end of each day? Let’s look at each step.” or “Are you sure that’s correct?”
Solution hint (Specific)	The prospective teacher provides a specific prompt toward the correct answer.	“Does the snail slide back after reaching 20 meters?”
Providing solution	The prospective teacher provides the correct solution.	“I tell you the solution … How can you explain it?”
Other		
Emotions	Emotional expressions (positive or negative) from the prospective teacher.	“With pleasure, we are a good team.” “Are you even listening?”

For **RQ3**, the individually collected quantitative data were analyzed using IBM SPSS statistics (version 29) (IBM Corp, 2022). The responses of the prospective teachers to the five items (see section *data gathering*) were analyzed descriptively based on the absolute frequencies. The absolute frequencies provide a clear understanding of the distribution of responses, showing how many participants agreed, disagreed, or selected a neutral option for each item. This descriptive analysis helps to illustrate the overall trends and patterns in the responses, offering insights into how the prospective teachers perceived the usefulness, motivation, and authenticity of their interactions with ChatGPT in the context of teacher education.

Given that the study relied on a single word problem task, a short one-session implementation, and a convenience sample, the design necessarily offers an exploratory rather than a comprehensive perspective on prospective teachers' interactions with ChatGPT.

RESULTS

RQ1. Likeness of ChatGPT's Responses to Those of the Learners

In our study, ChatGPT provided incorrect answers in all cases except one. Upon analyzing its solutions, the following primary error strategies can be identified:

Aggregated daily progress approach

In this strategy, a simplified average daily progress of three meters is assumed (5 meters climbed during the day minus 2 meters slid back at night). This value is then used in a division or multiplication operation to determine the total time required to reach 20 meters. Although this algorithm is mathematically valid regarding average calculation, it is conceptually flawed. The error lies in the assumption that only 3 meters of the distance are covered each day, without considering that, on Day Six, the snail has already climbed 5 meters per day to reach 20 meters before sliding down at night. For example, the following approach was taken: progress = 3 meters per day, 20 meters ÷ 3 meters/day = 6.67 days → implies that the snail reaches the top on the seventh day.

Step-by-step daily progress approach

This strategy involves a detailed, incremental analysis of each individual day and night. The progress is cumulatively calculated, including the nightly slippage, by considering every day and night cycle separately. While this method allows for a more systematic examination, it remains susceptible to conceptual errors, specifically in estimating when the nightly regression occurs. A common misconception is the belief that the snail slides back at the end of day six, even though it has already exited the well by that point. For example:

- Day 1: climbs 5 m, night: -2 m → 3 m total
- Day 2: climbs 5 m, night: -2 m → 6 m total

...

- Day 6: climbs 5 m, night: -2 m → 18 m total (the well is exited before any nighttime slippage)
- Day 7: climbs 5 m → reaches 23 m, finish

These two strategies were also identified in learners' solutions in previous research. For example, Schreiber (2010) describes how several learners approached the task by conceptualizing the total distance as a series of equal-length segments, thereby employing a strategy aligned with the aggregated daily progress approach. In this method, learners relied on the assumption of a fixed daily progress rate and used this value as a divisor to estimate the total duration required to reach the goal. In addition, another group of learners demonstrated a different strategy, working with pairs of numerical values that represented the snail's position at the end of each day and night cycle. The number of days needed was then inferred from the number of recorded position pairs. This approach aligns with what we refer to as the step-by-step daily progress approach, as it entails a sequential, day-by-day reconstruction of the scenario, based on cumulative distance tracking.

RQ2. Behavior of Prospective Teachers When Guiding ChatGPT Through Solving Word Problems

As stated above, ChatGPT provided incorrect answers to the given tasks, prompting participants to provide appropriate prompts for support. **Table 1** shows the inductively derived categories of ChatGPT´s processing of the task and the reactions and guidance of the prospective teachers, created through MAXQDA (2025) coding. The code system captured the different aspects of the interactions between prospective teachers and ChatGPT. The coding categories reflect content-related themes such as error identification, problem-solving strategies, instructional prompts, and the emotional or linguistic characteristics of the dialogue. **Table 1** presents the code categories and brief descriptions with examples.

Based on the inductively determined coding, three different types of guidance could be identified from the conversation management and the response patterns of the prospective teachers.

Co-constructive guidance

This type of guidance demonstrates a stimulating approach that encourages further work and reflection. During the conversation, the teacher facilitates the problem-solving process by promoting reflection and independent thinking instead of providing direct answers. Thoughtful questions support learning and help learners find solutions through their own thinking. Typical for the co-constructive guidance are suggestions such as "Go through each step individually" or "Think again: Does the snail only move 3 meters further each day or are there days when it also moves higher?"

Five of the 21 documented conversations could be assigned to this type of conversational guidance.

Directive guidance

This approach determines the learning process by pointing out errors and suggesting specific paths to the correct answer. For example, the answer is formulated as a question that only needs to be confirmed by the learner (here, by ChatGPT), or the solution is given down to the very last step so that it is now relatively obvious to the learner.

In this study, the directive guidance included questions such as “Doesn’t the snail reach the end on day 6? When it no longer sleeps but crawls out directly at the top of the well when it arrives” or statements such as “But the snail arrives on the 6th day before night falls. So, it doesn’t have to slide down anymore” are particularly striking.

Nine of the 21 documented conversations could be assigned to this type of conversation management.

Non-responsive guidance

This type is characterized by providing unresponsive and task-unspecific hints. Often, the conversation ends before a correct solution to the task is found, leaving the solution provided by ChatGPT incorrectly and largely unedited.

A teacher who provides non-responsive guidance is one who lacks persistence, providing initial stimulation but withdrawing before a complete solution is reached, especially if progress is slow.

Four out of the 21 documented conversations could be assigned to this type of conversation.

In addition to the three established guidance types, three exceptional cases were identified that could not be categorized into any of the three types for various reasons. In two of these cases, ChatGPT came to a solution too quickly to identify the type of guidance: Once immediately, with the first response, and once after only a single prompt. The third case showed/exhibited characteristics of all three types and could therefore not be assigned to one.

RQ3. Prospective Teachers’ Perceptions on the Use of ChatGPT

RQ3 was divided into two parts for the analysis: On the one hand, the authenticity of the dialogue with ChatGPT was focused on, and on the other hand, motivational aspects and benefits for the later profession. **Figure 2** shows the exact distribution of responses on the 6-point Likert scale for the two relevant items (see section data gathering for the specific formulation of the items).

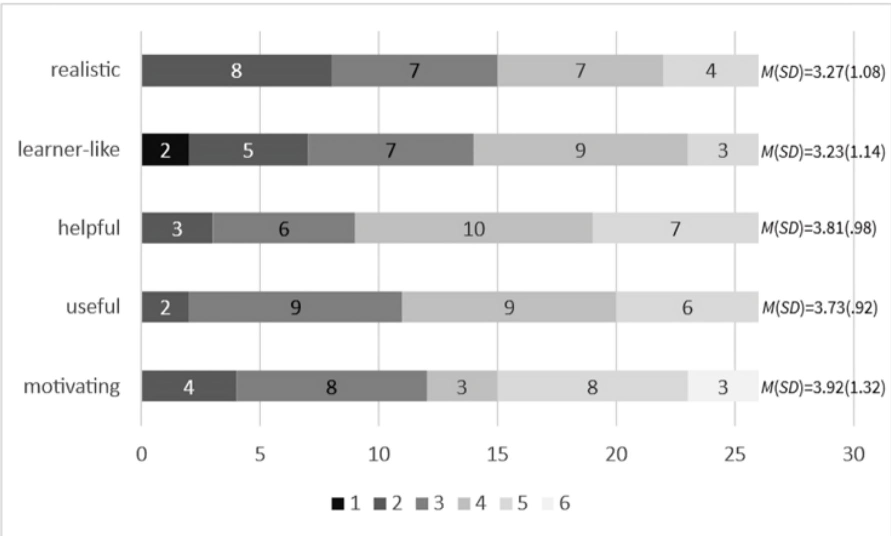

Figure 2. Distribution of responses (Source: Authors’ own elaboration)

The authenticity of the dialogue with ChatGPT was evaluated using the two items “realistic” and “learner-like”. Overall, the responses paint a mixed picture, which is also reflected in the qualitative statements made by the prospective teachers (see statements below).

In response to an open-ended question about the authenticity of interacting with ChatGPT, the participants made statements such as:

- “The solution steps and the descriptions of the calculation methods.”
- “The results provided by ChatGPT.”
- “ChatGPT gives answers based on errors that I could imagine occurring in a real-world context.”

At the same time, some participants expressed a critical view of the realism of ChatGPT’s responses, as illustrated by the following statements:

- “I don’t find ChatGPT very learner-like because it offers many complex solutions.”
- “You’re talking to a computer, it’s hard to imagine that a child is behind it.”
- “I think learners would phrase things differently and have more emotions than ChatGPT, which also needs to be particularly considered.”

Motivation and support were assessed using the items “helpful”, “meaningful”, and “motivating” (**Figure 2**).

Positive comments about motivation and support mainly refer to the opportunity to gain insight into the learners’ perspective, to see where problems may arise, and to get new ideas and try out ways of addressing them.

In response to the associated open-ended question, participants primarily noted that they gained new insights into potential forms of assistance and ideas for continuing their work:

- “New ideas on how to guide learners toward the correct solution.”
- “Adjusting questions and feedback to suit children and their errors.”
- “You learn to give helpful support […]. By experimenting with ChatGPT, you can find the best way to offer help, which can be applied to children who struggle with solving word problems.”

Additionally, it was mentioned that interacting with ChatGPT can help anticipate potential comprehension barriers:

- “By interacting with ChatGPT, you can recognize which tasks are difficult to understand. You can also observe different solution paths from ChatGPT and see where errors are made in the tasks. This can help guide lesson preparation and offer support.”

At the same time, these participants acknowledged the limitations of conversations with ChatGPT, which primarily address the lack of emotion and irrational reactions from learners. This is evident, for example, in the following comments:

- “One has to be careful not to forget that it’s an AI, which always acts logically, while learners don’t necessarily behave that way.”
- “However, I think it’s important to note that every child is also individual and emotions play a role.”
- “On the other hand, emotions are not an issue with ChatGPT, even after repeated corrections.”

The comments from the prospective teachers highlight both the potential of ChatGPT dialogues and the challenges and differences when compared to conversations with learners. These insights align with findings from previous research on the use of AI in teacher professionalization (Kortenkamp & Dohrmann, 2023; Schorcht et al., 2023).

DISCUSSION AND CONCLUSIONS

Implementing ChatGPT as a didactic tool for the professionalization of mathematics teachers provided valuable insights into the opportunities and limitations of using generative language models in teacher education.

ChatGPT-Based Error Situations

The analysis revealed that ChatGPT’s problem-solving approaches closely mirrored common learners’ strategies, notably the *aggregated daily progress* and *step-by-step daily progress* strategies. These findings align with prior research by Schreiber (2010), who documented similar solution patterns among learners tackling analogous word problems. This resemblance suggests that ChatGPT can effectively approximate learners’ reasoning processes and errors, offering a valuable opportunity for teacher education focused on diagnosing and supporting learners in problem-solving contexts. Analyzing the support strategies of the participating prospective teachers in more detail, three distinct types of guidance were identified: co-constructive guidance, directive guidance, and non-responsive guidance. These categories align with established patterns of instructional responses described in previous research on teacher-student interactions and feedback (e.g., Hattie & Timperley, 2007; Mercer & Littleton, 2007).

The co-constructive type reflects adaptive error management behavior as described by Tulis (2013), where teachers engage learners in reflective dialogue, promote metacognitive questioning, and emphasize the learning potential of errors. Similarly, studies by Anderson et al. (2004) and Heimbeck et al. (2003) have linked such reflective practices to enhanced conceptual understanding and learner autonomy.

By contrast, the directive guidance identified in this study mirrors less adaptive, but still prevalent forms of error management found in classroom research. Comparable to the direct correction described by Oser and Spychiger (2005) and Santagata (2005), this type of response emphasizes teacher control and immediate remediation of learners’ errors. While such behavior may efficiently clarify misconceptions, it risks constraining learners’ cognitive engagement and limiting opportunities for self-correction (Hattie & Donoghue, 2016). Finally, the non-responsive type, characterized by limited or absent follow-up to learner errors, echoes the ignoring category found in classroom observations by Hiebert et al. (2003) and Oser and Spychiger (2005).

Notably, the identification of these three types of guidance in our study suggests that established patterns of teacher reactions to learners’ errors can be replicated even within an artificial, simulated learning environment. Although the interaction with ChatGPT did not involve real learners, the error scenarios elicited similar instructional behaviors to those observed in real classrooms. This indicates that AI-based simulations can capture essential features of pedagogical decision-making and thereby serve as a valuable proxy for real-life teaching practice in early stages of teacher education.

The use of AI as a simulated practice environment thus holds particular promise for teacher education. As Steuer (2014) and Tulis (2013) argue, developing a constructive error climate requires explicit reflection and practice in responding to errors. However, this is often limited by the rarity of observable errors in real classrooms (Oser & Spychiger, 2005; Santagata, 2005) and the infrequency of practical training phases. Integrating AI-generated classroom discourse into teacher education allows prospective teachers to encounter, analyze, and respond to typical learners’ misconceptions in a low-risk environment.

Importantly, this approach enables engagement with non-adaptive strategies, such as overly directive or non-responsive reactions, at an early stage of professional development, without risking negative consequences for actual learners.

However, despite this promising alignment, ChatGPT's frequent provision of incorrect answers underscores the model's conceptual limitations. This emphasizes the necessity for critical engagement by prospective teachers when utilizing AI tools, as uncritical acceptance may reinforce misconceptions about AI (for an overview of existing misconceptions, Dilling & Hermann, 2024). Therefore, while ChatGPT can serve as a valuable didactic tool, its use must be accompanied by reflection on the function of AI tools.

Prospective Teachers' Perceptions of ChatGPT

Participants expressed ambivalent views regarding the authenticity and utility of ChatGPT. While many regarded the tool as a meaningful and motivating way to practice dialogic guidance, others questioned its resemblance to genuine classroom interactions. Echoing findings by Gíslason (2023), the absence of non-verbal cues and emotional responsiveness limited the naturalness of the exchanges, which in turn reduced the perceived realism of the simulated discourse. Furthermore, the participants' limited teaching experience likely influenced their perceptions of ChatGPT's learner-likeness and instructional relevance. Consistent with research on technology acceptance in education (Venkatesh et al., 2003), individual differences such as prior knowledge, technological self-efficacy, and attitudes toward AI (Koehler & Mishra, 2009; Lenz, 2024; Thurm et al., 2017) likely shaped these divergent evaluations, ranging from enthusiasm to scepticism.

However, the perceived learner-likeness of ChatGPT must be interpreted with caution. Because the dialogue unfolded within a single continuous session, the model's later responses were unavoidably shaped by its earlier errors and by the interventions of the prospective teachers. What appeared to participants as persistence or misconception-driven reasoning was, in reality, a form of statistical recall rather than genuine cognitive continuity. ChatGPT does not learn across turns in a human sense; it generates text conditioned on prior context. As a result, its apparent learning trajectory reflects contextual patterning, not developmental change. This distinction is pedagogically crucial: prospective teachers must recognize that AI systems simulate cognition through linguistic probability, not through understanding or conceptual reasoning. Without this awareness, there is a risk of overattributing cognitive or emotional depth to AI behavior, thereby reinforcing misconceptions about AI and learning processes (Dilling & Hermann, 2024).

Limitations and Future Research

The present study also faces limitations, each of which suggests directions for further research.

First, the sample constituted a relatively small convenience sample of 26 prospective teachers, the majority of whom were female. In addition, no further covariates were collected that could have provided insight into participants' prior school experience, levels of digital literacy, familiarity with or affinity toward AI technologies, or underlying teacher beliefs and pedagogical knowledge. As a result, differences in instructional guidance types cannot be meaningfully linked to individual characteristics or levels of teaching expertise. Future research should therefore aim to recruit larger and more diverse samples, including greater gender diversity, and incorporate systematic measures of digital competence, teaching experience, and pedagogical orientations to examine potential moderating effects.

A further methodological constraint concerns the use of the browser-based version of ChatGPT, which limits reproducibility because key technical details, such as the underlying model and parameter settings (e.g., temperature, top-p, and context length), are not disclosed and may change over time. Consequently, identical prompts may yield different responses, which should be considered when interpreting the findings. Beyond these technical constraints, particular caution is also required when interpreting ChatGPT's perceived learner-likeness. Because each interaction unfolded within a single continuous dialogue session, the model's later responses were necessarily conditioned by its earlier outputs as well as by the interventions provided by the prospective teachers. What participants may have interpreted as persistent misconceptions or stable reasoning patterns was, in fact, a function of statistical context retention rather than genuine cognitive continuity. ChatGPT does not learn across turns in a human sense; instead, it generates responses probabilistically based on prior textual context (cf. section "Prospective Teachers' Perceptions of ChatGPT"). This limitation underscores the need for caution when attributing learner-like cognitive processes to current generative AI systems.

At the same time, the findings must be situated within the rapidly evolving landscape of AI development. Generative language models are continuously refined, with ongoing improvements in training data, reasoning accuracy, and adaptive feedback mechanisms. Consequently, the types and frequencies of conceptual errors observed in this study may change in future model iterations. The present results therefore represent a snapshot of a dynamic technological moment rather than a stable characterization of AI behavior. One notable limitation of the present study is the absence of multimodal interaction. At the time of data collection, it was not feasible to integrate visual or diagrammatic representations into the AI-mediated dialogues. However, research in mathematics education has long demonstrated the value of heuristic tools and visual representations, such as sketches, for supporting problem-solving processes (Hembree, 1992; Sturm, 2018). Recent advances in AI applications now make the incorporation of multimodal inputs–such as images, visual models, or diagrams–increasingly feasible. Future studies should therefore explore how multimodal AI-supported simulations and advances in AI capabilities affect interactional patterns, perceived learner realism, and instructional decision-making, potentially enabling closer alignment with authentic classroom practices.

Finally, while AI-mediated simulations offer promising opportunities for teacher education, their pedagogical relevance ultimately depends on their relationship to real teaching contexts. Future research should therefore combine AI-assisted training environments with longitudinal designs and live classroom observations to examine whether patterns observed in simulated interactions translate into actual instructional practices. Such research would contribute to a more nuanced understanding of how generative AI can be meaningfully integrated into teacher education while respecting both its current limitations and its evolving potential.

Author contributions: KL: conceptualization, data curation, formal analyis, investigation, methodology, project administration, resources, validation, visualization, writing – original draft, writing – review & editing; **JS:** conceptualization, data curation, formal analyis, methodology, resources, validation, visualization, writing – original draft, writing – review & editing. Both authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the research study did not require approval by an ethics committee in accordance with national standards of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). According to these standards, ethical approval is required when participants are exposed to risk or significant physical or emotional stress, are not fully informed about the aims and procedures of the study, or when the study involves patients, (f)MRI, electrical or magnetic stimulation, or psychopharmacological investigations. The authors further stated that none of these conditions applied to the present study, which involved minimal risk, fully informed participants, and the anonymous collection of data; therefore, formal ethics approval was not required. All students provided verbal informed consent prior to participation. Data were collected and analyzed anonymously, ensuring that participants' rights, privacy, and confidentiality were fully respected throughout the research process.

AI statement: The authors stated that ChatGPT was used as part of the learning environment, and chat transcripts generated through students' interactions with ChatGPT were analyzed as study data. AI-based tools were additionally used for language editing. The authors retain full responsibility for all analyses and interpretations.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Ade-Ojo, G. O., Markowski, M., Essex, R., Stiell, M., & Jameson, J. (2022). A systematic scoping review and textual narrative synthesis of physical and mixed-reality simulation in pre-service teacher training. *Journal of Computer Assisted Learning, 38*(3), 861-874. https://doi.org/10.1111/jcal.12653

Ammel, R. (2023). Wie gut ist ChatGPT als Mathe-Coach? Teil 1: Textaufgaben [How good is ChatGPT as a math coach? Part 1: Word problems]. *Mathegym*. https://mathegym.de/blog/wie-gut-ist-chatgpt-als-mathe-coach

Anderson, A., Hamilton, R. J., & Hattie, J. (2004). Classroom climate and motivated behaviour in secondary schools. *Learning Environments Research, 7*, Article 211e225. https://doi.org/10.1007/s10984-004-3292-9

Arora, S., Narayan, A., Chen, M. F., Orr, L., Guha, N., Bhatia, K., Chami, I., Sala, F., & Ré, C. (2022). *Ask me anything: A simple strategy for prompting language models*. arXiv. https://doi.org/10.48550/arXiv.2210.02441

Asare, B., Arthur, Y. D., & Boateng, F. O. (2023). Exploring the impact of ChatGPT on mathematics performance: The influential role of student interest. *Educational Science and Management, 1*(3), 158-168. https://doi.org/10.56578/esm010304

Bastian, A., Kaiser, G., Meyer, D., Schwarz, B., & König, J. (2022). Teacher noticing and its growth toward expertise: An expert-novice comparison with pre-service and in-service secondary mathematics teachers. *Educational Studies in Mathematics, 110*(2), 205-232. https://doi.org/10.1007/s10649-021-10128-y

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competence of teachers]. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469-520. https://doi.org/10.1007/s11618-006-0165-2

Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV [COACTIV's competency model]. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 29-53). Waxmann. https://doi.org/10.1007/s35834-011-0017-x

Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV): Ein Forschungsprogramm [Professional competence of teachers, cognitively activating teaching and the mathematical competence of students (COACTIV): A research program]. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 7-25). Waxmann. https://doi.org/10.31244/9783830974338

Berliner, D. C. (1991). Educational psychology and pedagogical expertise: New findings and new opportunities for thinking about training. *Educational Psychologist, 26*(2), 145-155. https://doi.org/10.1207/s15326985ep2602_6

Besser, M., & Krauss, S. (2009). Zur Professionalität als Expertise [On professionalism as expertise]. In O. Zlatkin-Troitschanskaia (Ed.), *Beltz-Bibliothek. Lehrprofessionalität: Bedingungen, Genese, Wirkungen und ihre Messung* (pp. 71-82). Beltz.

Blum, W., & Leiss, D. (2005). Modellieren im Unterricht mit der "Tanken"–Aufgabe [Modeling in the classroom using the "refueling" task]. *Mathematik Lehren, 128*, 18-21.

Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens* [The teacher as expert: On the psychology of professional knowledge]. H. Huber.

Bromme, R. (2008). Kompetenzen, Funktionen und unterrichtliches Handeln von Lehrer/innen [Competencies, functions and teaching practices of teachers]. In B. Rendtorff (Ed.), *Schule, Jugend und Gesellschaft* (pp. 244-256). Kohlhammer.

Bryden, K. (2023). Math is the worst enemy of ChatGPT. *Crast*. https://crast.net/265780/math-is-the-worst-enemy-of-chatgpt/

Caylan-Ergene, B., Ergene, O., & Kontorovich, I. (2025). Analysing the validity of ChatGPT's solutions to statistical problems. In T. Fujita (Ed.), *Proceedings of the British Society for Research into Learning Mathematics* (pp. 1-6). BSRLM.

Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). Studying teacher education: The report of the AERA panel on research and teacher education. *American Educational Research Association*. https://permalink.obvsg.at/AC08413669

Contel, F., & Cusi, A. (2025). Investigating the role of ChatGPT in supporting metacognitive processes during problem-solving activities. *Digital Experiences in Mathematics Education, 11*, 167-191. https://doi.org/10.1007/s40751-024-00164-7

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research*. SAGE.

Dahal, N. (2024). Exploring capabilities and limitations of generative AI chatbots in solving math algorithm problems. In *Proceedings of the 15th International Congress on Mathematical Education* (pp. 1-4). ICME.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00348

Dilling, F., & Herrmann, M. (2024). Using large language models to support pre-service teachers' mathematical reasoning, an exploratory study on ChatGPT as an instrument for creating mathematical proofs in geometry. *Frontiers in Artificial Intelligence, 7*. https://doi.org/10.3389/frai.2024.1460337

Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Mathematics Education, 41*, 605-618. https://doi.org/10.1007/s11858-009-0184-6

Ergene, O., & Ergene, B. C. (2025). AI chatbots' solutions to mathematical problems in interactive e-textbooks: Affordances and constraints from the eyes of students and teachers. *Education and Information Technologies, 30*, 509-545. https://doi.org/10.1007/s10639-024-13121-z

Frieder, S., Berner, J., Petersen, P., & Lukasiewicz, T. (2023a). *Large language models for mathematicians*. arXiv. https://doi.org/10.48550/arXiv.2312.04556

Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., & Berner, J. (2023b). *Mathematical capabilities of ChatGPT*. arXiv. https://doi.org/10.48550/arXiv.2301.13867

Gíslason, I. (2023). What can mathematics teacher-students learn from dialogue with AI? In *Proceedings of the 13th Congress of the European Society for Research in Mathematics Education*. CERME. https://doi.org/10.13140/RG.2.2.10592.05121

Greefrath, G., Kaiser, G., Blum, W., & Borromeo Ferri, R. (2013). Mathematisches Modellieren–Eine Einführung in theoretische und didaktische Hintergründe [Mathematical modeling–An introduction to theoretical and didactic backgrounds]. In R. Borromeo Ferri, G. Greefrath, & G. Kaiser (Eds.), *Realitätsbezüge im Mathematikunterricht. Mathematisches Modellieren für Schule und Hochschule: Theoretische und didaktische Hintergründe* (pp. 11-37). Springer. https://doi.org/10.1007/978-3-658-01580-0_1

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *npj Science of Learning, 1*, Article 16013. https://doi.org/10.1038/npjscilearn.2016.13

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*(1), 18-32. https://doi.org/10.1037/0022-0663.87.1.18

Heimbeck, D., Frese, M., Sonnentag, S., & Keith, N. (2003). Integrating errors into the training process: The function of error management instructions and the role of goal orientation. *Personnel Psychology, 56*, 362-366. https://doi.org/10.1111/j.1744-6570.2003.tb00153.x

Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I. Theoretische Grundlegung, Methode und Ergebnisse einer Videostudie [Dealing with errors in classroom discussions in lower secondary school: Theoretical foundations, methodology, and results of a video study]. *Journal für Mathematik-Didaktik, 25*(3/4), 221-244. https://doi.org/10.1007/BF03339324

Helfrich-Schkarbanenko, A. (2023). *Mathematik und ChatGPT: Ein Rendezvous am Fuße der technologischen Singularität* [Mathematics and ChatGPT: A rendezvous at the foot of the technological singularity]. Springer. https://doi.org/10.1007/978-3-662-68209-8

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education, 23*(3), 242-273. https://doi.org/10.2307/749120

Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement: A systematic review and meta-analysis study. *Computers & Education: Artificial Intelligence, 8*, Article 100361. https://doi.org/10.1016/j.caeai.2025.100361

Hiebert, J., Gallimore, R., Garnier, H., Givvin Bogard, K., Hollingsworth, H., Jacobs, J., et al. (2003). Teaching mathematics in seven countries: Results from the TIMSS 1999 video study. U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Huget, J., & Buchholtz, N. (2024). ChatGPT als Reflexionsinstrument zur Förderung von Unterrichtsplanungskompetenzen von Lehramtsstudierenden [ChatGPT as a reflection tool to promote lesson planning skills in student teachers]. In P. Ebers, F. Rösken, B. Barzel, A. Büchter, F. Schacht, & P. Scherer (Eds.), *Beiträge zum Mathematikunterricht 2024. 57. Jahrestagung der Gesellschaft für Didaktik der Mathematik* (pp. 925-932). WTM. https://doi.org/10.37626/GA9783959872782.0

IBM Corp. (2022). IBM SPSS statistics for windows (version 29). *IBM Corp*. https://www.ibm.com/products/spss-statistics

Kadel, J., Buschmann, C., Haas, S., Meßner, M. T., & Adl-Amini, K. (2023). Planspiele und simulative Methoden in der Lehrkräftebildung–Ein Literaturüberblick [Simulation games and simulation methods in teacher education–A literature review]. *Zeitschrift Für Hochschulentwicklung, 18*, 19-39. https://doi.org/10.21240/zfhe/SH-PS/02

Kennedy, M. M., Soyeon, A., & Jinyoung, C. (2008). The value added by teacher education. In M. Cochran-Smith (Ed.), *Handbook of research on teacher education: Enduring questions in changing contexts* (pp. 1247-1271). Routledge. https://doi.org/10.4324/9780203938690-134

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60-70.

Kortenkamp, U., & Dohrmann, C. (2023). Pre-service teacher training with AI: Using ChatGPT discussions to practice teacher-student discourse. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education*.

Kuckartz, U., & Rädiker, S. (2022). *Methoden, Praxis, Computerunterstützung* [Methods, practice, computer support]. Beltz Juventa.

Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2011). Die Entwicklung professioneller Kompetenz von Lehrkräften [The development of professional competence of teachers]. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (pp. 55-68). Waxmann. https://doi.org/10.31244/9783830974338

Leiner, D. J. (2019). SoSci survey (version 3.1.06). *soSci*. www.soscisurvey.de

Leiss, D., Schukajlow, S., Blum, W., Messner, R., & Pekrun, R. (2010). The role of the situation model in mathematical modelling, task analyses, student competencies, and teacher interventions. *Journal für Mathematik-Didaktik, 31*(1), 119-141. https://doi.org/10.1007/s13138-010-0006-y

Lenz, K. (2024). Einstellungen von angehenden Grundschullehrkräften zum Einsatz digitaler Werkzeuge im Mathematikunterricht [Attitudes of prospective primary school teachers towards the use of digital tools in mathematics lessons]. In P. Ebers, F. Rösken, B. Barzel, A. Büchter, F. Schacht, & P. Scherer (Eds.), *Beiträge zum Mathematikunterricht 2024. 57. Jahrestagung der Gesellschaft für Didaktik der Mathematik* (pp. 393-396). WTM. https://doi.org/10.37626/GA9783959872782.0

MAXQDA. (2025). MAXQDA software für qualitative Datenanalyse, 1989-2025 [MAXQDA software for qualitative data analysis, 1989-2025]. *MAXQDA*. https://www.maxqda.com/

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Routledge. https://doi.org/10.4324/9780203946657

Meyer, L., Seidel, T., & Prenzel, M. (2006). Wenn Lernsituationen zu Leistungssituationen werden: Untersuchung zur Fehlerkultur in einer Videostudie [When learning situations become performance situations: An investigation into the culture of error in a video study]. *Schweizerische Zeitschrift für Bildungswissenschaften, 28*, 21-41. https://doi.org/10.24452/sjer.28.1.4717

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft: Zur Theorie des negativen Wissens und zur Praxis der Fehlerkultur* [Learning is painful: On the theory of negative knowledge and the practice of a culture of error]. Beltz.

Pantziara, M., Gagatsis, A., & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics, 72*(1), 39-60. https://doi.org/10.1007/s10649-009-9181-5

Pólya, G. (2010). *Schule des Denkens: Vom Lösen mathematischer Probleme. Sammlung Dalp* [School of thinking: On solving mathematical problems. Dalp collection]. Francke.

Riese, J., & Reinhold, P. (2012). Die professionelle Kompetenz angehender Physiklehrkräfte in verschiedenen Ausbildungsformen [The professional competence of prospective physics teachers in various forms of education]. *Zeitschrift für Erziehungswissenschaft, 15*(1), 111-143.https://doi.org/10.1007/s11618-012-0259-y

Santagata, R. (2005). Practices and beliefs in mistake-handling activities. A video study of Italian and U.S. mathematics lessons. *Teaching and Teacher Education, 21*(5), 491-508. https://doi.org/10.1016/j.tate.2005.03.004

Scheunpflug, A., Baumert, J., & Kunter, M. (2006). Schwerpunkt: Professionelle Kompetenz von Lehrkräften [Focus: Professional competence of teachers]. *Zeitschrift für Erziehungswissenschaft, 9*, 465-468. https://doi.org/10.1007/s11618-006-0164-3

Schorcht, S., Baumanns, L., Buchholtz, N., Huget, J., Peters, F., & Pohl, M. (2023). Ask smart to get smart: Mathematische Ausgaben generativer KI-Sprachmodelle verbessern durch gezieltes prompt engineering [Ask smart to get smart: Improving the mathematical output of generative AI language models through targeted prompt engineering]. *Mitteilungen Der Gesellschaft Für Didaktik Der Mathematik, 115*, 12-23.

Schreiber, C. (2010). *Semiotische Prozess-Karten–Chatbasierte Inskriptionen in mathematischen Problemlöseprozessen* [Semiotic process maps–Chat-based inscriptions in mathematical problem-solving processes]. Waxmann.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

Şimşek, N. (2025). Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT. *Education and Information Technologies, 30*(8), 11391-11462. https://doi.org/10.1007/s10639-024-13258-x

Spannagel, C. (2023). Hat ChatGPT eine Zukunft in der Mathematik [Does ChatGPT have a future in mathematics]? *Mitteilungen Der Deutschen Mathematiker-Vereinigung, 31*(3), 168-172. https://doi.org/10.1515/dmvm-2023-0055

Stahnke, R., & Blömeke, S. (2021). Novice and expert teachers' situation-specific skills regarding classroom management: What do they perceive, interpret and suggest? *Teaching and Teacher Education, 98*, Article 103243. https://doi.org/10.1016/j.tate.2020.103243

Steuer, G. (2014). *Fehlerklima in der Klasse: Zum Umgang mit Fehlern im Mathematikunterricht* [The climate surrounding mistakes in the classroom: How to deal with errors in mathematics lessons]. Springer. https://doi.org/10.1007/978-3-658-05293-5

Steuer, G., Rosentritt-Brunn, G., & Dresel, M. (2013). Dealing with errors in mathematics classrooms. Structure and relevance of perceived error climate. *Contemporary Educational Psychology, 38*(3), 196-210. https://doi.org/10.1016/j.cedpsych.2013.03.002

Sturm, N. (2018). *Problemhaltige Textaufgaben lösen: Einfluss eines Repräsentationstrainings auf den Lösungsprozess von Drittklässlern* [Solving problematic word problems: The influence of representation training on the problem-solving process of third graders]. Spektrum Akademischer Verlag. https://doi.org/10.1007/978-3-658-21398-5

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Terhart, E. (2001). *Lehrerberuf und Lehrerbildung: Forschungsbefunde, Problemanalysen, Reformkonzepte. Beltz Pädagogik* [The teaching profession and teacher education: Research findings, problem analyses, reform concepts. Beltz pedagogy]. Beltz.

Terhart, E. (2008). Die Lehrerbildung [Teacher training]. In K. S. Cortina (Ed.), *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick* (pp. 745-772). Rowohlt-Taschenbuch-Verlag.

Thurm, D., Klinger, M., Barzel, B., & Rögler, P. (2017). Überzeugungen zum Technologieeinsatz im Mathematikunterricht: Entwicklung eines Messinstruments für Lehramtsstudierende und Lehrkräfte [Beliefs about the use of technology in mathematics education: Development of a measurement instrument for student teachers and teachers]. *Mathematica Didactica, 40*(1), 19-36. https://doi.org/10.18716/ojs/md/2017.1230

Tulis, M. (2013). Error management behavior in classrooms: Teachers' responses to student mistakes. *Teaching and Teacher Education, 33*, 56-68. https://doi.org/10.1016/j.tate.2013.02.003

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478. https://doi.org/10.2307/30036540

Verschaffel, L., Schukajlow, S., Star, J., & van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*, 1-16. https://doi.org/10.1007/s11858-020-01130-4

Vicente, S., Verschaffel, L., Sánchez, R., et al. (2022). Arithmetic word problem solving. Analysis of Singaporean and Spanish textbooks. *Educational Studies in Mathematics, 111*, 375-397. https://doi.org/10.1007/s10649-022-10169-x

Xuan-Quy, D., & Ngoc-Bich, L. (2023). Investigating the effectiveness of ChatGPT in mathematical reasoning and problem solving: Evidence from the Vietnamese national high school graduation examination. *ArXiv.* Vorab-Onlinepublikation. https://doi.org/10.48550/arXiv.2306.06331

Yanar, A. N., & Ergene, Ö. (2025). Integrating artificial intelligence in education: How pre-service mathematics teachers use ChatGPT for 5E lesson plan design. *Journal of Pedagogical Research, 9*(2), 158-176. https://doi.org/10.33902/JPR.202533163