Beliefs and belief-change of in-service mathematics teachers on the use of generative artificial intelligence

MARC HERRMANN, FREDERIK DILLING & INGO WITZKE, SIEGEN

***Zusammenfassung**: Der Beitrag befasst sich mit den Beliefs von Mathematiklehrer:innen zum Einsatz von generativer Künstlicher Intelligenz (KI) für die Planung von Unterricht sowie als Lehr- und Lernwerkzeug im Unterricht. Auf der Basis von halbstandardisierten Interviews konnte eine Vielzahl unterschiedlicher Beliefs identifiziert werden. Die interviewten Lehrkräfte waren mehrheitlich Novizen im Bereich von generativer KI. Nach einer explorativen Phase, in welcher die Lehrkräfte erste Erfahrungen mit dem Einsatz generativer KI zur Planung von Mathematikunterricht sammeln konnten, wurden weitere Interviews geführt. Anhand von drei Fallbeispielen zeigt der Beitrag auf, dass dies zu teilweise grundlegenden Änderungen der Beliefs führen kann.*

***Abstract:** The article examines mathematics teachers' beliefs about the use of generative artificial intelligence (AI) for lesson planning as well as its use as a teaching and learning tool in the classroom. Based on semi-structured interviews, a wide variety of beliefs were identified. The teachers interviewed were, for the most part, novices in the field of generative AI. After an exploratory phase in which the teachers were able to gain initial experience with using generative AI to plan mathematics lessons, further interviews were conducted. Using three case studies, the article shows that this can, in some cases, lead to fundamental changes in beliefs.*

1. Introduction

The release of ChatGPT 3.5 in November 2022 has sparked a boom in the field of artificial intelligence (AI) and caused ripples in the educational landscape. It quickly became clear that this new technology has the potential to fundamentally change teaching and learning in schools and universities, creating both opportunities and challenges (Kasneci et al., 2023). Since then, educational research has taken a close look at generative AI such as ChatGPT – also in the context of mathematics education. In this research, at least three perspectives on the subject have emerged:

- analyses of the mathematical and pedagogical capacities of generative AI language models (Schorcht et al., 2024; Sjuts, 2024; Wardat et al., 2023)
- generative AI as a tool for teachers, e.g. to plan and reflect on lessons (Buchholtz & Huget, 2024; Noster et al., 2024; Schorcht et al., 2025),
- and generative AI as a tool for students to support their learning processes (Dilling et al., 2024b; Dilling & Herrmann, 2024; Yoon et al., 2024).

This article focuses on the second perspective: the role this technology plays for teachers. Because the tool's mathematical and pedagogical capabilities shape teachers' decisions about employing it in their work—and their views of the opportunities and challenges it poses for students' learning—the discussion also touches on the two remaining perspectives.

Mathematics teachers are considered an essential factor for the successful implementation of digital technologies in the classroom (Clark-Wilson et al., 2020). This is due to the responsibility they bear for the design of the lesson and the selection and preparation of classroom resources such as digital technologies. Teachers' beliefs about the use of digital technologies guide such design processes. Empirical studies have shown that beliefs and self-efficacy beliefs about teaching with technology as well as basic epistemological beliefs about mathematics influence the modes of technology use in the classroom (Thurm & Barzel, 2022).

Furthermore, it could be shown that corresponding beliefs can be changed by professional development interventions (Thurm & Barzel, 2020). In addition to more general beliefs about the use of digital technologies in mathematics education, specific beliefs about certain technologies or approaches are also considered important (Dilling et al., 2024a), especially when it comes to such a fundamentally new technology as generative AI. This paper therefore examines the beliefs of teachers from different school types regarding the use of generative AI in lesson preparation and classroom teaching, as well as the potential changes in these beliefs when teachers gain initial experience with the technology.

In the following, we first provide a theoretical background to describe teachers' beliefs and belief changes in the context of generative AI as well as a short overview of recent research on that topic. The research questions for this article are derived from these considerations, along with the methodology and framework of the study. The presentation of the results and the discussion are structured along with the research questions. Finally, a conclusion is drawn and initial implications for teacher training are formulated.

2. Theoretical Framework

2.1 Generative AI in mathematics education

Generative AI is a specific form of AI that has been developed especially for creating texts, images or other representations. While AI in general has been used for data analysis and similar contexts (including educational settings) for several decades, generative AI is a new field of AI development. In the area of generative AI, currently some of the most notable tools for educational settings are large language models (LLMs) and image generating AI (Buchholtz et al., 2024). When talking about generative AI, the most notable applications like LLMs or rather their implementations in chatbots, such as ChatGPT by OpenAI are usually implicitly meant. While the field of generative AI encompasses more than just LLMs, a big part of the discussion is centered around the latter, especially among non-experts (Scantamburlo et al., 2025), such as the teachers who participated in this study. For this reason, the theoretical framework of this article will also focus primarily on large language models.

Such large language models are an application of artificial intelligence capable of imitating human language or creating images from a text or image input by the user, a so-called prompt. To do so, these models were trained using machine learning techniques and employ a special kind of deep neural network, called a transformer (Vaswani et al., 2017) to generate the output. These models are trained on huge sets of training data (often comprising of data from multiple sources, like archived pages from the internet or text corpora) to predict the next word for a given sequence of words. Repeating this, whole sentences or even longer texts can be produced by a single prompt. From this operating principle it needs to be stated, that the models do not follow strict predefined rules for logic, nor do they rely on a database of verified information. The generation of a response solely relies on the frequency and patterns with which certain words and combinations of words occur in the training data, or rather the parameters of the model, that mirror these.

In the educational field, the research on large language models as a tool for teachers is still in its early stages, due to the short timeframe since the broad public release of such language models. First papers contrast the opportunities and challenges provided by the implementation of the models in educational settings (Kasneci et al., 2023). Specifically for mathematics education, a scoping survey by Pepin et al. (2025) lists benefits and challenges associated with the use of LLMs. A summary of their results will be presented below for framing the article's subsequent discussion.

Supporting the understanding of basic mathematical concepts, acting as a mathematical search engine and support in arithmetic or algebra are seen as benefits, while a lack of deep understanding of mathematical concepts, inaccurate solutions and problems with more complex reasoning tasks are seen as challenges of LLMs. Considering the interaction of mathematics teachers with LLMs the possibilities to support their design of lessons and giving reflective feedback are beneficial, while the varying quality of generated lessons and a potential bias towards teacher-centered approaches are challenging. In the area of assessment and task design, adapting tasks for different students, providing personalized feedback and semi-automated grading are mentioned as benefits, while ethical or legal concerns and potential biases in assessment are risks associated with this use of LLMs.

According to Pepin et al. (2025), when considering student interaction and personalized learning, LLMs can support self-regulated learning, provide real-time feedback, and assist learners with different first languages, but they may also foster an over-reliance on the technology. In the area *of* critical thinking challenges, LLMs can foster critical thinking, reflection and creativity, but misconceptions in AI-generated answers may pose a risk. Furthermore, LLMs can enable collaborative learning and cooperation by supporting group projects and project-based learning, but may pose challenges for group dynamics and participation. Lastly, teachers' and students' attitudes towards LLMs and student outcomes are generally positive, embracing its potential for personalized learning. However, concerns about privacy and ethical issues are proving challenging.

In mathematics education, initial studies have emerged that build on the topics from Pepin et al.'s (2025) scoping review, examining LLMs both as a tool for teachers (e.g., Noster et al., 2024) and as a tool for students (e.g., Dilling et al., 2024b). The studies also look at mathematical processes such as problem solving and proving and showcase specific potentials and challenges the technology might offer (Dilling & Herrmann, 2024; Schorcht et al., 2024; Yoon et al., 2024). In addition to the systematic examination of the technology by researchers as it can be seen above, it is also important to investigate teachers' perspectives and approaches, since they are responsible for its actual implementation in mathematics instruction.

2.2 Teachers' beliefs about technology

This article adopts the theoretical construct of beliefs—following Philipp (2007) to capture how teachers perceive generative AI and its applications in teaching and learning, defining beliefs as follows:

> Psychologically held understandings, premises, or propositions about the world that are thought to be true. Beliefs are more cognitive, are felt less intensely, and are harder to change than attitudes. Beliefs might be thought of as lenses that affect one's view of some aspect of the world or as dispositions toward action. Beliefs, unlike knowledge, may be held with varying degrees of conviction and are not consensual. Beliefs are more cognitive than emotions and attitudes. (p. 259)

In this study, beliefs capture the understandings, premises and propositions the teachers hold about generative AI in the teaching and learning of mathematics. For mathematics education, beliefs are not a new concept, as they have been used to describe learners' and teachers' behavior for decades (e.g., Schoenfeld, 1985, Philipp, 2007). Nevertheless, while beliefs or similar cognitive and affective constructs have been a subject of research for so long, still many different terms and constructs such as *perceptions, attitudes* or *values* are used and it often remains unclear how they relate to one another (Weygandt, 2021). Following the definition of Philipp (2007) we use beliefs as a construct for our study in contrast to *attitudes* as we are focusing on cognitive aspects.

While the definition of Philipp (2007) already gives a distinction between beliefs and other affective constructs, it does not touch upon the difference between beliefs and knowledge too much. To make a distinction here, we refer to the notions of subjective and objective knowledge (Pehkonen & Pietilä, 2003). Objective knowledge is formal knowledge accepted and shared by a community (e.g. mathematics teachers), while subjective knowledge is informal knowledge of an individual, which it does not necessarily share with anyone else. Using these notions, beliefs in a cognitive sense can be understood as part of the subjective knowledge.

While many studies of teachers' beliefs in mathematics education focused on the beliefs about mathematics (e.g., Grigutsch et al., 1998), there are also some about the teaching and learning of mathematics with technology. Thurm and Barzel (2022) have summarized the most often identified belief groups in a literature review regarding technology usage:

- Beliefs about the role of technology to support discovery based learning
- Beliefs about the role of technology to support multiple representations
- Beliefs about time needed to teach with technology
- Beliefs about the loss of by-hand skills
- Beliefs about mindless working when teaching with technology
- Beliefs about the time point of technology use

These belief groups reflect important aspects of the use of digital technologies in mathematics teaching, but they relate in particular to traditional digital mathematics tools such as graphing calculators, spreadsheets or dynamic geometry software. Other empirical studies have shown that different beliefs arise in relation to other technologies (e.g. Dilling & Vogler, 2022, related to online learning platforms; Dilling et al., 2024a related to visual programming; Kamodi & Garegae, 2019 related to spreadsheets, Saralar-Aras, 2022 related to GeoGebra). For this reason, it is important to also examine teachers' beliefs in the context of generative AI, which is a technology that differs fundamentally from previous mathematics tools and is the subject of intense public debate. Furthermore, it is important to investigate how such beliefs evolve as experience with the technology grows.

2.3 Describing belief changes through instrumental genesis

When a new technology – like a chatbot powered by a large language model – is first introduced to users, they are not immediately aware of all functions this *artifact* may provide and how they might use it for their needs (Trouche et al., 2019). An artifact refers to any kind of physical object used by people for a kind of task or in any way. The exploration of an artifact's functions and the formation of utilization schemes to use the artifact is a process that unfolds gradually. This process of forming utilization schemes and exploring the artifact, making it a useful tool for the user is referred to as instrumental genesis by Verillon & Rabardel (1995). The resulting instrument consists of two sub-systems: the artifact itself, being material or symbolic and the utilization schemes created by the subject or appropriated from social utilization schemes (Verillon & Rabardel, 1995). For this reason, no artifact can be an instrument by itself, as it can only become an instrument through the process of using it in specific situations.

The duality of the instrument is represented in the process of instrumental genesis, consisting of two sub-processes: instrumentalization (directed towards the artifact) and instrumentation (directed towards the subject). As described by Trouche (2004) the process of instrumentalization is directed towards the artifact and may consist of different stages as the discovery and selection of its functions, the personalization towards the user and the transformation, wherein applications of the tool unintended by its developers may arise. The process of instrumentation is directed towards the user and describes how an artifact may shape its user by the opportunities and constraints it provides. This includes the formation of utilization schemes, which define how the user can utilize the artifact to achieve a set goal. This is also associated with a development or change of person's beliefs regarding the artifact or instrument, which are increasingly expanded and differentiated by the instrumentation process.

The research on belief development of teachers shows contradictory results regarding the ease of changing beliefs. In general, a distinction is made between central and peripheral beliefs. Central beliefs are held with a stronger conviction in contrast to peripheral beliefs. Research suggests that the central beliefs seem to be more stable against changes in contrast to peripheral beliefs (Grootenboer, 2008; Liljedahl et al., 2012). The assumed link of centrality to stability against change creates a conceptual problem: Labeling a belief as central or peripheral in the literature is often done by evaluating a beliefs stability against change, resulting in circular reasoning, where stability and centrality can be used to define the other (first noted by Liljedahl et al., 2012; see also Eichler et al., 2023). Furthermore, Liljedahl et al. (2012) criticize the general use of stability of beliefs as a concept in the existing literature, as it is not defined clearly in most studies and used with a variety of different meanings. Although the concept of stability of beliefs and its use in literature remain ambiguous, authors agree that a ‘strong impact’ is necessary to change teachers’ beliefs (Eichler et al., 2023, p. 1493).

3. Research Questions

The study presented in this article seeks to identify mathematics teachers’ beliefs about teaching and learning with generative AI. Accordingly, our first research question is:

1) What beliefs do mathematics teachers hold about using generative AI in mathematics teaching and learning?

Considering the previously mentioned process of instrumental genesis and the literature on beliefs, the formation and change of beliefs can be described as a gradual process. As generative AI is such a new emerging technology, it seemed likely to us that many of the teachers willing to participate in our study would be novices in using generative AI either in general or at least for their professional activities. Further evidence for this hypothesis is the near absence of generative AI in both teacher-education programs and professional-development courses. Nonetheless, insights into this early stage of belief formation—and the belief shifts that may accompany it—could greatly inform and enhance teacher training.

For this reason, all participants in this study were asked to describe their experiences with generative AI and those who had little to no experience were asked to participate in an exploration phase in the weeks afterwards. The participating teachers were asked to choose at least one specific way to incorporate generative AI into their out-of-class teaching practice (e.g., lesson planning, worksheet creation) and to carry it out, thereby gaining meaningful first-hand experience with the technology. Our aim is to determine whether—and how—their beliefs shift through this exploration. This leads to the second research question:

2) After a brief exploratory period in which novices first use generative AI for teaching and learning mathematics, do their beliefs change—and if so, in what ways?

After outlining our more exploratory research questions that were derived from the literature on generative AI, literature on beliefs and studies from mathematics education considering other technologies available at that time, we also want to take the chance and integrate studies published while this project was underway and during peer review, so we can interpret our results in light of this emerging body of work. This tries to account for the rapid changes and frequent publications in the field of generative AI in education.[1] Although this structure is unusual in a research article, we present these studies after our research questions to emphasize that the more recent literature did not inform our research questions or methodology. We review it here to situate and discuss our findings against the newer evidence.

Beege et al. (2024) surveyed 102 STEM teachers in Germany, finding that competence in using ChatGPT correlated positively with both current and future ChatGPT use, as well as perceived benefits for teaching quality. Despite acknowledged risks, teachers continued to use ChatGPT. The study only identifies correlations between factors, but does not identify causal relations. ElSayary (2024) studied 40 private-school teachers in the UAE using a mixed-methods approach. Teachers saw ChatGPT's main benefits in lesson planning (idea generation, differentiation), with fewer mentions of assessment and feedback. While some valued its tutorial potential, concerns included misuse and negative learning effects. Results are limited by the small sample of private-school teachers with previous training on the use of AI.

In a recent study Busuttil & Calleja (2025) investigate the beliefs of 86 Maltese secondary school mathematics teachers in a professional development session. Teachers – most with little prior ChatGPT experience described benefits for personalized learning (71%), engagement (56%), exploration (51%) and language support (21%). Considering mathematical topics, most teachers considered ChatGPT useful for statistics, probabilities, arithmetics and algebra, while only few considered it suitable for geometry. Field notes indicated perceived benefits for lesson planning and a growing need to adapt to students' AI use. Reported outcomes included increased teacher confidence and openness to integrating ChatGPT.

Overall, existing studies rely largely on predefined items and provide only partial insights into teachers' beliefs about generative AI. ElSayary's (2024) findings are based on a narrow, special sample, while Busuttil & Calleja (2025) offer only brief evidence of belief change during short PD. Thus, current literature gives some indications of perceived benefits and evolving openness but leaves the structure of teacher beliefs underexplored. Therefore, it can be concluded that the published literature during the progression of our study can be helpful in discussing our findings, while it only touches upon our research questions marginally.

4. Methods & Materials

4.1 General methodology

The present study aims to gain a first understanding of teachers' beliefs about generative AI for teaching and learning mathematics, trying to find core constructs and categories as well as changes to novices' beliefs after a first exploration phase with the tool. For this kind of exploratory study with the goal of forming hypotheses, a qualitative approach is chosen. Beliefs are individual for every person, but there may be shared beliefs or categories of beliefs among groups of people – teachers in this case – that can be identified by aggregating and comparing the beliefs of individuals. For this reason, a case study with several cases (according to Yin, 2018) is conducted, where each teacher with their own beliefs is a single case.

The case study is suitable for the given research questions, as it allows us to describe the beliefs of each case (individual teacher) and how they may change through the exploration phase, by only looking at each case individually. Additionally, it facilitates identifying commonalities and differences between the cases in a cross-case analysis, generating categories of shared beliefs. As beliefs are mental constructs that cannot be observed directly, they have to be reconstructed in some way. For this study, teachers are asked directly about their beliefs. These beliefs, which teachers report directly are also called *professed beliefs* and can be distinguished from *attributed beliefs*, which are the beliefs reflected in teachers' practices (Speer, 2005). The chosen method to facilitate teachers expressing their beliefs is the implementation of semi-structured interviews.

In semi-structured interviews, a predetermined set of open questions gives the interview a general structure and focus on relevant topics, while also retaining the flexibility to further explore interesting aspects mentioned. This way, the teachers are asked to state their opinion on multiple different areas of teaching and learning and the role of generative AI for them. The stated opinions are either beliefs or showcase an underlying belief, which enables coding them with much less room for interpretation than in other survey formats, such as reconstructing beliefs from behavior in teaching situations.

These semi-structured interviews are conducted with every participant and with the same set of questions. Considering the second research question, those teachers who consider themselves novices in the use of generative AI, either in general or for their professional activities as teachers, are asked to participate in an exploration phase, after which another semi-structured interview with an adjusted set of questions is performed. In the next sections, the exploration phase, the questions for the semi-structured interviews, the timeline and participants and the data analysis are described in more detail.

4.2 The exploration phase

The exploration phase was a period of one to four weeks, in which the participants had time to interact with generative AI for their professional activities as teachers. If a participant at the first interview was suitable for this phase (i.e. describes having no experience with generative AI whatsoever or for their professional activities), the interviewer in the first interview asked them if they would volunteer to participate in the exploration phase and describes the purpose of it. A participant declaring to also participate in the exploration was then asked to specify at least one professional activity in that time frame, where they would use generative AI as a support. This activity could be the planning of a lesson, the design of resources for a lesson or any other kind of teaching-related activity. Trying to keep this exploration close to what teachers would experience without this study, no recommendations on which software, for what professional activity or how the generative AI should be used, were given to the participants.

Although the beliefs of teachers that are of interest for this study also include uses of generative AI beyond teacher activities, such as students using it, it was still decided to limit this exploration to teacher activities outside of the classroom. This intentional exclusion of AI in the students' hands was justified by ethical considerations. If a teacher is a novice in the use of generative AI, it is not only debatable if they are able to create a meaningful learning setting for students to use generative AI in, but it furthermore is also questionable if such a teacher would be able to predict and mitigate potential harmful interactions with AI. Consequentially, we deemed it neither logical nor ethical to ask teachers to create a learning environment, in which students would interact with generative AI during the exploration phase.

After the exploration phase a second interview was conducted, in which the participants were asked some of the questions from the first interview but also some additional questions considering the experiences of the exploration phase and their consequences.

4.3 Questions for the guided interviews

The semi-structured interviews are structured by questions around multiple topics, which will be listed below. While these questions give the general structure for the interview, follow-up questions reacting to participant answers are also possible. The main topics are given below, with sub-questions that are asked if the participants do not give this information already in their answers in brackets:

1) Previous experience with AI in private and school context (Naming a specific example and used AI systems)
2) Opportunities of AI in education (Differences between mathematics and other subjects; differences between teacher and student use)
3) Challenges of AI in education (same sub-questions as for opportunities)
4) Changes in learning in school through AI in the future
5) Usage scenarios of AI for the planning and development of mathematics lessons of this participant (Specific usage scenarios, attributed suitability of the technology for different mathematical topics)
6) Support needs to use AI in school

The decision for this set of questions was made to touch upon relevant topics for generative AI in educational settings while also keeping the length of the interviews short enough for teachers to be willing to participate. The topics were selected based on theoretical considerations and questions from belief interviews about other technologies in mathematics education, which yielded meaningful results before (Dilling & Vogler, 2022). For those teachers participating in the exploration phase, a second interview is conducted afterwards, to examine changes to their beliefs. The main topics for those questions are given below in a similar manner to those for the first interview:

1) Experience with AI during the exploration phase (General impression and specific use)

2) Changes of beliefs towards AI in education (Changes of viewed opportunities and challenges)
3) Specific opportunities perceived during the exploration phase
4) Specific challenges and problems during the exploration phase
5) Influence of using AI on their teaching of mathematics
6) AI for different mathematical topics
7) Support needs to use AI in school
8) Future role of AI in mathematics education (Changes through the exploration phase)

The full interview guide for both interviews is given in the appendix.

4.4 Timeline and participants

To describe and understand a new theoretical construct, as the structure of teacher beliefs about the use of generative AI for the teaching and learning of mathematics, the process of data collection requires to collect sufficient information to establish the categories and the properties of these categories. For the Grounded Theory approach Glaser and Strauss (1967) defined the concept of theoretical saturation, which is partially applicable to this study, as the inductive qualitative content analysis is methodologically related to Grounded Theory. The concept of theoretical saturation aims to fully develop all categories, only ending data sampling, once new data provides no new properties. This kind of saturation is not the goal of this study, due to the nature of beliefs. For a given sub-level category, it would be possible to identify numerous slightly different beliefs, given enough time and participants. As the present study seeks to contribute to our understanding of the overarching structure of teacher beliefs, not all manifestations of all categories are of central relevance. Instead, the focus lies on the categories themselves and selected manifestations that serve to illustrate them. Therefore, a different kind of saturation is necessary.

In reviewing literature on theoretical saturation in qualitative research, Saunders et al. (2018) identify four different categories of saturation and hybrid versions of those, combining them. This study combines *data saturation* and *inductive thematic saturation*. Data saturation appears solely on the level of data collection with no necessary reference to theory (e.g. an interviewer hears the same statements again in multiple interviews, showing information redundancy). Inductive thematic saturation focusses on new codes or themes generated from new data and is achieved when no further new codes or themes can be identified. For this study, this kind of saturation is achieved by conducting a preliminary analysis after each set of interviews, trying to identify new categories, but also checking for redundancies in all prior interviews.

This approach led to three sets of interviews being conducted. The sets of interviews are given with their respective dates and participants in the following list, while the reasons for conducting further interviews will be discussed afterwards:

- [Set A]: August 2023 – September 2023; 5 teachers from secondary schools. 3 of the 5 teachers qualify for the exploration, 2 of those are willing to participate.
- [Set B]: February 2024 – March 2024; 5 teachers from primary schools. All teachers qualify for the exploration, 4 of the 5 are willing to participate.
- [Set C]: May 2024 – June 2024; 4 teachers from primary schools. All qualify for the exploration and all are willing to participate.

After the release of ChatGPT in November of 2022, this study was initially planned in summer of 2023, with the first interviews being conducted after the development of the interview guide in August and September. For these and all further participants, a convenience sampling approach was chosen, relying on the willingness of teachers in schools in the south-eastern part of North Rhine-Westphalia to participate. Nevertheless, participants from different schools and across a broad range of teaching experience were approached to represent a variety of different teacher backgrounds. This was successful, leading to the total group of participants from multiple schools having a broad range of teaching experience, from teachers who completed their teacher training less than two years ago to those working in school service for over 20 years.

After the initial analysis of the material, the category system appeared not yet fully developed. Furthermore, since the participants up to that point were exclusively secondary school teachers, an additional set of interviews with primary school teachers was planned. The second set of interviews was conducted in February and March of 2024 with primary school teachers. The preliminary analysis after this set showed that multiple further topics, especially those only relevant to primary education, were added by the interviewers. It also became evident that the interviews in this set were significantly shorter and yielded less information compared to those in the initial set. Therefore, the decision for another set of interviews with primary education teachers was made. During the process of asking teachers to participate and selecting further potential candidates, the new version of ChatGPT (4o) was rolled out and also provided in platforms as Fobizz (a user-interface specifically for German schools to implement generative AI) within a few days. As a consequence, the new model was used by some of the participants in the final set of interviews.

With the four interviews conducted in May and June 2024, the preliminary analysis revealed multiple redundancies in the information provided and identified only a few new subtopics related to areas already addressed in earlier interviews. For this reason, we considered the data to be saturated enough to conduct the full analysis without requiring further interviews. This analysis was done independently of the results of the preliminary analyses and followed a strict methodology. The preliminary analyses also encompassed many aspects of the method described in the following section, but to a less formal extent.

4.5 Data analysis

The data from the interviews consists of audio files containing the verbal statements during the interviews. These were transcribed with respect to transcription rules by Dresing & Pehl (2018) and all statements containing private information were anonymized. After completing this preprocessing of data, the main analysis was performed. As the first research question addresses a broad spectrum of beliefs, a method of data analysis that allows to aggregate the data from multiple participants and allows to identify structures in the data was required. A method that does this is the qualitative content analysis (e.g. Mayring, 2014). Specifically, an inductive qualitative content analysis aimed at identifying categories within the material was deemed the most appropriate methodological choice.

Given that the authors of this article had previously engaged with literature addressing topics potentially relevant to the discussion of generative AI in the teaching and learning of mathematics (see, for example, the theoretical framework), there was a risk of influencing the inductive formation of categories by this prior knowledge. As this risk has to be taken into consideration for most studies, where an inductive formation of categories takes place, the authors of this study followed the recommendations by Glaser & Strauss (1967) considering knowledge about different theories. They state that it generally is not possible to completely remove knowledge about other theories and therefore recommend making this transparent and reflecting upon this influence while forming categories, also trying to generate them solely based from the material, instead of trying to fit the material into existing categories. We followed these guidelines as closely as possible and compared our identified categories with existing literature only after the category development was complete.

To conduct the inductive qualitative content analysis, the software MAXQDA was chosen. Segments of one to multiple sentences were coded for each belief as unit of analysis. A segment of text could be coded for multiple beliefs. Such a segment of text was relevant to our study and coded, if it concerned either the teaching or learning of mathematics **and** AI. Therefore, statements as “So, in mathematics lessons there is not much researching at the end of the day. Maybe for applied word problems and Fermi problems” are not coded as beliefs in this study, as they do not concern the topics of our study.

Each interview was analyzed sentence by sentence, coding all statements that either formulated beliefs explicitly or showcased an underlying belief. Except for a handful of beliefs, most fall into the prior category. For this reason, the definitions for most beliefs are a structured summary of the initial anchor example or the multiple coded segments. Therefore, most coded beliefs are close to the data, limiting room for interpretations and increasing the reliability of coding. To give an example for this, the (translated) quote “ChatGPT can’t handle number symbols neither geometry drawings.” has been coded as the definition “AI cannot output symbolic or graphical representations correctly.” for the belief MS-P-3 (see table in the appendix) and is an example for a statement that formulated a belief explicitly, while the quote “It was total chaos, because it can’t handle fixed proof structures – at least not yet.”, in which the teacher speaks in a broader context about a lesson in grade 7, where he let his students use AI to generate proofs, was coded as “Generative AI performs poorly when generating mathematical proofs or arguments”. This is an example of an underlying belief, where the belief is not explicitly stated, but is given through the broader context in the interview.

Going through the interviews one by one, each new segment was compared to the existing coded beliefs and either coded as one of those or coded as a new belief, appending the current list of beliefs. This led to the coding process taking longer with every new belief appending the list and incurring a risk of coding a segment as a new belief, although this belief might already exist in the list. To mitigate this risk, the full set of coded beliefs was grouped by similar topics and keywords in multiple iterations after the initial coding, comparing definitions of beliefs and anchor examples against one another. In this process, only two codes describing the same belief could be identified and were merged into one code, while for a few codes the definitions were clarified to stress differences between beliefs.

While the primary coding of the beliefs was done by author 1 (MH), the formation of categories was done together with author 2 (FD). All codes were distributed on a map (an empty white plane similar to a whiteboard) inside the software and together both authors started clustering similar beliefs. This clustering process was conducted by placing similar beliefs near one another and therefore gradually removing beliefs from the big pool of unclustered beliefs. The topic for each cluster was identified and formed the title for the sub-level category. Similar clusters were then grouped together to find top-level categories. During this process, the definitions and anchor examples of the beliefs were each checked for their validity, discussed among both authors and adapted when necessary. After all beliefs were grouped in sub- and top-level categories, the full system of categories was translated to English. Generative AI was used to support this process.

To make this abstract process transparent for one category, we will showcase the example of the top-level category “mathematical and content-related suitability of AI”. After placing all codes on a map, it became apparent, that mathematics or areas and processes in mathematics (i.e. geometry, calculus, mathematical proofs) were mentioned in multiple beliefs, so we grouped them all together. This led to a total of 11 beliefs, that were grouped together (see all statements with the prefix “MS-” in the full system in the appendix). Comparing these beliefs with one another, two bigger themes could be identified. These were aimed at different *mathematical content areas* on one side and *mathematical processes* on the other side. These two topics had 6 and 4 beliefs each, forming the sub-level categories for the category. Only one belief, stating AI would be more useful in other subjects than in mathematics did fit the top-level category but neither of the sub-level categories, which is why it was left as a singular belief under the top-level category.

The final system of categories was then discussed among all three authors to check the suitability of the translations, but also to make final modifications, should there be any remaining issues with the categories or beliefs. This approach of coding by one person, forming inductive categories by deliberating among two authors and verifying definitions and anchor examples in turn and having a final step of verification and review of the whole system of categories was chosen deliberately to ensure validity and reliability as quality criteria of a content-analytical approach (Mayring, 2014). Due to the nature of the data, such an approach is more appropriate than other measures of reliability, such as Cohens Kappa, as there are a big number of codes and categories, with a small number of coded segments, instead of a very limited number of codes, where it is important to clearly distinguish, if a segment is coded as one code or another.

For the second research question, a different kind of data analysis was conducted. As the changes in beliefs are an individual process, here an in-case analysis according to case study methodology was used (Yin, 2018). For all the cases, the changes of beliefs were analyzed individually. As the original beliefs of the participants were already very different, unsurprisingly no overarching pattern of belief change could be identified. Therefore, a selection of cases had to be made to present them here. In accordance with recommendations on case-selection techniques (Seawright & Gerring, 2008; Yin, 2018), a selection of *diverse cases* was chosen. This selection of cases tries to illuminate a sense of the full variation that exists between the cases with a limited number of cases. We tried to represent positive and negative experiences during the exploration phase as well as a positive and negative outlook in the end. Furthermore, we tried to show the difference in relevance the teachers attribute to generative AI for teaching and learning mathematics.

5. Results

RQ1: Teachers’ beliefs on the use of AI for teaching and learning mathematics

In the coding process, a total of 268 segments were coded across all interviews. Of these, 119 in group A (secondary school, 1st set of interviews), 85 in group B (primary school, 2nd set of interviews) and 64 in group C (primary school, 3rd set of interviews). From these segments, 115 beliefs were identified and aggregated in 28 sub-level and 11 top-level categories. Most beliefs fall under a sub-level category, in turn being under a top-level category. For some beliefs, either no sub-level category was a valid choice, or they were on a meta-level, more similar to whole sub-level categories. These are sorted directly below top-level-categories and can be distinguished by the code prefix ending in a letter, like the sub-level categories, rather than a number like the beliefs under a sub-level category.

The full system of categories, including all beliefs, their definitions and anchor examples, is given in the appendix (due to its length). In the following paragraphs, the top- and sub-level categories (see table 1) will each be defined by describing what kinds of beliefs are aggregated in them. This description will not necessarily include all beliefs in a category, due to the number of beliefs and the limited space in this article. For the presented beliefs in this article, those who best represent the core of the sub-level category have been chosen. All descriptions of the categories always consist of the beliefs that teachers mentioned. For the readers' convenience, these beliefs are given as statements without repeating this disclaimer, that these are not facts or opinions the authors of this article share.

As an example, for the sub-level category of *educational objectives* (under the top-level category of ***fundamental changes to the education system through AI***), it might be described that AI will change what will be learned in school in the future and in what way (FC-E-4). This statement is to be interpreted as one or more of the participants describing the belief that AI will change what will be learned in the future in what way, which is found in the system of categories as the belief FC-E-4 and which has been categorized under the sub-level category considering educational objectives. Looking at the full system of categories, the definition is given as: "AI will change what students have to learn in school in the future and how they will learn it" and the anchor example from the interviews is the following:

> The transmission of knowledge will no longer be the main focus, but rather higher-level skills, perhaps also reflective skills. Also how one acquires knowledge and which knowledge is actually necessary to truly store it in the brain.

Top-level categories are written in bold and italic, sub-level categories in italic and all described beliefs are given with their code-prefix, making it easier to identify them in the full system of categories. The prefix is given as a combination of 1-2 two letters for the top-level category, 1-2 letters for the sub-level category and a number for each belief. The number was given to each belief in the formation process of categories and helps to identify them in the full system of categories. The beliefs are not necessarily numbered in the order in which they are mentioned in the text. Beliefs which fall under a top-level category, but do not fit the sub-level categories have a prefix of 1-2 letters for the top-level category and a unique letter for each belief, to distinguish them from those beliefs that fall under sub-level categories.

The first top-level category is ***fundamental changes to the education system through AI***. It describes how generative AI will reshape the educational system in the future by changing fundamental aspects of teaching and learning in school. The sub-level categories *educational objectives*, *teacher role* and *instructional design* specify the aspects that might be affected. For the *educational objectives*, AI changes what will be learned in school in the future and how (FC-E-4). Beliefs in this category include how evaluating and critically judging information will be more important in the future than knowledge (FC-E-5), how AI will make it necessary to rejustify why certain mathematical skills are still important (FC-E-6) and how some foundational competencies cannot be replaced by AI (FC-E-3). Considering the *teacher role*, there will be a role shift through AI (FC-T-3). While teachers cannot be replaced by AI (FC-T-2), their role might change towards moderating and contextualizing AI output to support students learning (FC-T-1). Looking at *instructional design*, AI can replace other traditional learning resources (FC-I-3) and might lead to more open lessons (FC-I-2), but teachers consider it unlikely to have an effect on early primary education (FC-I-1). A last belief directly below the top-level category shows discomfort or criticism considering a total paradigm shift of learning through AI (FC-P).

The second top-level category ***teacher competencies*** describes the competencies of teachers with generative AI and how generative AI influences their professional competencies. It is split into the two sub-level categories *general statements* and *self-assessments*. While the first of these consists of beliefs about teachers' competencies in general, the second includes self-assessments of the participants considering their very own skills and competencies with generative AI in education.

Most beliefs in *general statements* consider the competencies teachers see as a requirement to use it in a meaningful way for teaching and learning and the relationship between AI usage and teacher competencies. These include prompting skills (TC-G-1), understanding how AI works (TC-G-2) or dealing with unpredictability of AI output (TC-G-4). Surprisingly, not only AI-related competencies are seen as a requirement to use AI effectively, but also content knowledge about the mathematical topics (TC-G-3). The relationship between AI usage and teacher competencies ranges from AI usage prompting more reflection of their own goals and instructional choices (TC-G-7), to teachers with less experience (TC-G-8) or teaching out of field (TC-G-6) especially benefitting from AI usage.

The *self-assessments* of the participants are mostly negatively connotated, expressing a lack of skills in using generative AI (TC-S-1) or technology in general (TC-S-2), finding it hard to evaluate in which daily-life technologies AI is involved (TC-S-3), for what professional activities (TC-S-5) which kind of AI application (TC-S-6) might be useful.

The top-level category ***pedagogical suitability of AI*** encompasses beliefs around the question to what extent generative AI is suitable for the use in school from a pedagogical standpoint. The sub-level categories address the suitability for *primary school*, how *AI and other resources* are related and how the *complexity of AI answe*rs proves challenging for its pedagogical suitability. In *primary school*, according to the beliefs, the use in grade 1 is considered inappropriate (PS-P-3), there seems to be little value in using AI tools at all (PS-P-1) or teachers generally consider AI as a manifestation of digitalization in primary school not pedagogically meaningful (PS-P-2).

Tab. 1: Top-level and sub-level categories in this study

Top-level category	**Sub-level category**
Fundamental changes to the education system through AI	Educational objectives
	Teacher role
	Instructional design
Teacher Competencies	General statements
	Self-assessments
Pedagogical suitability of AI	Primary school
	AI and other resources
	Complexity of AI answers
Risks of AI usage	Outsourcing of learning opportunities
	Cheating with AI
	Harmful content
	Unreliable output
Learners' AI competencies	Competency development through broad usage
	Reflection competence
	Relevance of competence development
Mathematical and content-related suitability of AI	Mathematical content areas
	Mathematical processes
Significance of AI	General presence of AI
	Cross-curricular relevance
Prerequisites and support	Educational objectives
	Technical conditions
	Support
Areas of application: Lesson preparation	Workload reduction
	Resource creation with AI
	Lesson planning with AI
Areas of application: In-class use	Language support
	Diagnosis and support
	Personalized learning
AI characteristics	

Considering the relationship of *AI and other resources*, teachers mention the limited added value of AI in comparison to other resources for learners on one side (PS-R-1) and teachers on the other (PS-R-2). Looking at the *complexity of AI answers*, they seem to be too complex for students to use them in a way that benefits their learning (PS-C-3). This also applies to AI applications, whose complexity presents a major barrier to student usage (PS-C-2), which is why students (especially in primary education) need AI applications that are simple and easy to use (PS-C-1). Beyond the sub-level categories, teachers argued that AI provides greater benefits for high-performing students (PS-A) and that it is unsuitable in all contexts, that require personal student-teacher interaction (PS-U).

The ***risks of AI usage*** mentioned by the teachers are the *outsourcing of learning opportunities*, *cheating with AI*, *harmful content* and *unreliable output*. For the outsourcing of learning opportunities, participants expressed the beliefs, that students may use AI in a copy-paste manner (R-O-1) without reflecting upon the results, missing out on learning opportunities. Furthermore, students may use AI to do their homework for them, instead of doing them on their own (R-O-2). Beyond missing out on learning opportunities, the risk of cheating with AI is also considered by the teachers. They mention that students might use AI to cheat during exams or different forms of assessment (R-C-1), which gets an even bigger problem as teachers consider it difficult to identify, whether a response was written by a student or generated by AI (R-C-2).

Another risk is seen in *harmful content*, as the AI has no moral judgement and may present content that is unethical (R-H-1) to students. Going beyond that, teachers see the general risk of exposing children to an AI system that is not specifically adapted for them (R-H-2). Similarly concerned about AI responses are those participants mentioning the *unreliable output*. They consider AI output as unreliable and requiring critical verification of the responses (R-U-2), which is made more complicated by AI presenting incorrect information in a highly plausible and convincing way (R-U-1). Finally, three beliefs fall directly under the top-level category. Two of these address the risks AI incurs for data privacy (R-D) and the risk of exacerbating educational inequality due to different levels of AI access at home (R-I). The last belief addresses consequences of the many associated risks and demands to not ban AI in schools despite these risks (R-B).

Learners' AI competencies is a top-level category that addresses the competencies of students in the use of generative AI. The sub-level categories address how competencies with AI can be build by the idea of *competency development through broad usage*, what aspects are important considering reflection competence and the *relevance of competence development*. *Competency development through broad usage* encompasses the idea that broad use across different subjects could help students to develop the necessary skills using AI (LC-B-2) and that only this kind of active use gives a user the ability to evaluate the capabilities of an AI system (LC-B-1). For *reflection competence*, teachers deem it necessary to teach students critical thinking, i.e. critically questioning AI-generated content (LC-R-1). They stress this necessity as students – according to them – lack the necessary reflection skills (LC-R-3) and tend to overly trust AI-generated responses (LC-R-4). Teachers express a general *relevance of competence development* considering AI-related competencies, deeming it important to teach students a responsible and critical use (LC-D-2). This relevance is further emphasized by these AI-related competencies becoming relevant for all professional fields later in their students' lives (LC-D-1). Beyond the sub-level categories, teachers also express the belief, that learners are generally unable to distinguish between AI-based tools and conventional digital technologies (LC-T).

In a similar manner to pedagogical suitability, the top-level category of ***mathematical and content-related suitability*** encompasses beliefs that discuss the suitability of generative AI for the teaching and learning of mathematics. In contrast to the prior category, which focusses on general pedagogical aspects, this category is limited to beliefs that are focused on mathematics specifically. In it, there are the two sub-level categories of *mathematical content areas*, which encompasses beliefs discussing the suitability of generative AI for different content areas as geometry and *mathematical processes*, which contains beliefs about the suitability of generative AI for different mathematical processes as problem solving.

Regarding *mathematical content areas*, participants are in disagreement whether AI is useful for them or not. Whether AI is suitable for geometry (MS-C-2) or not (MS-C-5) and whether AI is suitable for arithmetics (MS-C-6) or not (MS-C-1) are two such cases where the teachers disagree. Some offer the alternative, that there is no difference between mathematical topics, considering the suitability of generative AI (MS-C-4). Looking at *mathematical processes*, according to the teachers, generative AI performs poorly at mathematical proofs (MS-P-1) and symbolic notations (MS-P-3) but may be able to assist students for problem solving (MS-P-4). Beyond both sub-level categories, teachers considered generative AI more suitable for their other subjects in contrast to mathematics (MS-S).

Beliefs considering the ***significance of AI*** as a technology for teaching and learning in school form this top-level category. The significance subdivides into the *general presence of AI*, which describes that AI already has arrived in everyday life, so schools have to deal with it (S-G-1) and that it will be permanently present in the school system (S-G-2), as well as the *cross-curricular relevance*, which contains the two beliefs, that AI is relevant to all subjects (S-R-2) and that not only computer-science is responsible for the development of learners' AI competencies (S-R-1).

Prerequisites and support as a top-level category contains beliefs about the requirements (i.e. *educational objectives* and *technical conditions*) to use generative AI in school and the *support* teachers deem necessary for this. The sub-level category of *educational objectives* concerns the political and systemic conditions, especially the objectives of the school system and how they influence the use of generative AI in school. Participants consider the current political conditions as not favorable for using AI in schools (P-E-5), with especially data-protection rules preventing the students from using AI tools (P-E-1). Therefore, a change of the political and legal conditions to use it is seen as a prerequisite (P-E-2) and even a curricular mandate for AI use (P-E-4) is considered an option for this.

Looking at the *technical conditions* at schools, teachers agree that poor school equipment hinders AI use in class (P-T-1) and technical hurdles (accounts, phone verification, …) eat up lesson time and slow down AI use (P-T-2). It seems to them, the digitalization of schools must progress further before AI can be used meaningfully (P-T-3). For support, the participants would consider teacher training (P-S-2) and collaboration (P-S-3) as helpful and important. More knowledge about legal conditions and data protection (P-S-1) as well as best practice examples (P-S-4) are considered especially helpful.

Looking at the *areas of application* of generative AI for teaching and learning, these are split into the two top-level categories of *lesson preparation*, looking at the use outside the classroom for the teacher and *in-class use*, where AI is used in the classroom by teachers or students. For the ***areas of application: Lesson preparation***, the sub-level categories of *workload reduction*, *resource creation with AI* and *lesson planning with AI* were identified. Beliefs about *workload reduction* focus on teachers mentioning that AI can reduce their workload (LP-W-2) and save time, be it in creating and grading exams (LP-W-1) or administrative tasks and general task development (LP-W-3). The *resource creation with AI* is seen as a useful application of the technology, as generative AI seems well suited to create worksheets (LP-R-3) and even novel resources, that were impossible to create with other technologies (LP-R-2). Nevertheless, these created resources must be adapted to the specific learning group (LP-R-4) as AI-generated resources often don't follow an overarching pedagogical logic (LP-R-1), which the teachers see as necessary.

Looking at the applications in *lesson planning with AI*, teachers mention using generative AI for an initial overview of a topic (LP-L-2) and as an idea generator (LP-L-4). For some tasks in lesson preparation that were previously done with a search engine, teachers now use AI (LP-L-1). Considering the impact it has on their teaching, some teachers believe that using AI leads to more creative and varied lessons (LP-L-3), whereas others see this more critically, thinking that AI is not yet capable of designing good lessons (LP-L7). Additionally, understanding students' learning difficulties better (LP-L-5) and being able to adapt the lesson spontaneously (LP-L-8) are seen as advantages of planning lessons without the use of AI.

For the ***areas of application: In-class use***, the sub-level categories of *language support*, *diagnosis and support* and personalized learning were identified. *Language support* concerns two different use-cases, where AI can be helpful for students. On one hand AI can act as support for non-native speakers by translations (CU-L-1) and on the other it can help students in general to express ideas they can't put into words (CU-L-2). For *diagnosis and support*, teachers mentioned that AI can help with evaluating student work and identifying needs for support (CU-S-3). This support can provide enrichment for high-achieving students (CU-S-1) and help low-achieving students (CU-S-2). Similar use-cases are also identified considering *personalized learning*. Here teachers think that AI enables more individualized learning paths (CU-P-3), acting as a digital tutor supporting the students (CU-P-1).

The last top-level category of ***AI characteristics*** aggregates 5 beliefs that consider characteristics of the technology and that did not fit into the other categories. They represent a broad spectrum of ideas, so no sub-level categories could be identified, presenting all beliefs directly. Some teachers believe that good AI results require technically correct and well-formulated prompts (C-P), while others mention that the same prompt can yield a spectrum of different AI answers (C-V). The purpose of AI is seen as an information dispenser (C-I) and one participant mentioned their surprise, as how well AI already solves certain tasks (C-S). Lastly, the belief that each newer AI model leads to a better performance is expressed (C-B).

RQ2: Belief change after the exploratory phase

Three illustrative teacher cases are presented below to exemplify belief changes following the exploratory phase. The cases were selected to capture a range of initial stances and contrasting experiences during the exploration process. Their names have been replaced by a number and the indicator for which interview set (A, B or C) they were in.

Teacher #B1: Relevance of AI in primary education

The first teacher (#B1), who was interviewed in February and March of 2024, is a primary education teacher, whose only prior experience with AI was hearing of it in the media without having ever used it. Directly in the beginning of the pre-interview she is asked about opportunities of generative AI in school. The transcript illustrates her opinion about this:

> **I:** Okay, what opportunities does AI offer for school from your point of view?
>
> **B1:** (5 seconds).
>
> **I:** It might also be that you don't see any opportunities. If you have an idea/
>
> **B1:** Yes, I don't know. Well, I think you can already use it, for example, in higher-level schools for text editing or text processing. Or for creating texts. (…) In elementary school, I don't know. So, I think it's more like that you can check texts to see if the content is correct or so, so that you can work on such things.

As it seems, she does not see a value in adding generative AI to her lessons in primary education. She only mentions applications for secondary education and sees the primary value of AI in generating texts which can then be used as material for a lesson for the students to assess their correctness. The underlying challenge of the reliability of the LLM-output is further emphasized by her, adding that she has to check all material she wants to use in her lessons beforehand as well as the answers it might give to students using it in a learning process. She can't really come up with applications of AI and speaks about her lack of knowledge about AI and how she would appreciate some teacher training to help with this before the end of the pre-interview.

She initially planned to use AI during the exploration phase primarily for generating mathematical word problems involving quantities such as length, money, and time. Accordingly, she employed it to create personalized word problems for her students, using familiar contexts like Pokémon. Reflecting on these experiences from the exploratory phase, she first notes the complexity of digging into AI. As she describes, she would need way more time to develop skills in using it. Her experience with AI was of mixed success: While the tasks made sense in general, she noticed that some key details were wrong in some of the tasks, from which she inferred that critically verifying all of the given information was important:

> Exactly, I think if you really get into it, you could use a lot, but I think you always have to check it. So, the way I see it, I thought, well, for one, 22-centimeter Pokéballs are just factually wrong. That has to be checked. There were also a few good tasks, but not all of them.

As a final statement, she completes the post-interview with her most important takeaway:

> **B1:** Yes, well, I had thought before that this is a topic that doesn't concern us in elementary school, and I would see that differently now. I do think that you can get a lot out of it or also use it.
>
> **I:** So now you do see more opportunities than before?
>
> **B1:** Yes, well, before I would have thought, really, we don't need that at all. That might be something for higher-level schools, and if at all, also for German [lessons], and that has changed. #00:08:18#

As a conclusion for this first case, the teacher was able to identify some general opportunities provided by AI and to see the value even for primary education. Furthermore, she noticed the problem of AI output, that looks convincing but is factually incorrect in some ways and thus deepens her belief on the unreliability of the technology and the necessity to check every answer individually.

Teacher #B4: Negative experience and positive outlook

Teacher #B4 is also a primary education teacher interviewed in February and March of 2024, who states to have never used AI on his own volition before, neither for work nor in his private life. He just mentions a workshop some time ago, where the use of AI in educational settings was a side topic and they were given a tool to create individual learning plans for their students with it:

> That was from some tool. Which one was it? No, I don't remember anymore. He had prepared that for us so we could play around with it for ten minutes. Exactly which one, I don't know anymore; you definitely had to pay for it.

As it seems, the teacher did not see enough value in the tool to pay for it, as he says afterwards. Although the output of the tool was good in his opinion, he does not mention it any further and goes on to describe the chances of AI primarily in the generation of word problems or ideas for lesson planning. He does not see it as a tool to be used by students. In his opinion, students lack the necessary ability to reflect on its output and use it responsibly. For the future, he sees AI as a supporting tool for teachers, but states, that an effective use of the technology will only be possible with enough time invested to understand the technology and its use.

The test phase of AI for teacher #B4 was quite stressful for him. He describes feeling lost, trying to become familiar with the technology without any instructions or help as an additional task on top of his normal workload for his teaching. He intended to use generative AI to plan a series of lessons on the topic of doing calculations with money for his third grade. During his testing, he first tested a special LLM-based AI tool for teachers to help with lesson planning. After taking quite some time to give it all the required information and select fitting settings, it seems like the tool had some kind of error and did not produce any kind of result. After this, he decided to use a generic LLM (ChatGPT) instead. The results were equally disappointing for the teacher. He mentions having started a conversation on planning a lesson series, which seemed promising initially, but when he wanted to continue the planning a day later, the conversation had been deleted by the tool to adhere data privacy regulations.

> With Teachino, it already didn't work, and then with ChatGPT, the chat wasn't saved, and honestly, I just didn't have the nerve anymore to keep dealing with it. So, I continued planning as I usually do, independently, without AI.

As it seems, the frustration of the difficulties occurring during the testing process made #B4 quit his attempts to use AI and keep going without it. With this in mind, it is quite surprising how he reflects his belief change and the future role of AI in school:

> So, I still see AI as an opportunity to make everyday school life easier. Even though my testing phase didn't go so positively, I think that with more time and the right support, one can definitely benefit from it. […] I think with the right guidance, I can definitely imagine using AI regularly. But it shouldn't add extra work; instead, it should really be a relief. Unfortunately, I wasn't able to experience that during my testing phase yet.

Despite the negative experience during the testing phase, the teacher thinks that AI can be beneficial in saving time and making life easier. He blames his missing skills during the test or the current state of technology but is convinced that it will be beneficial in the future. His belief considering the efficiency AI could provide for teaching seems to be important for this.

Teacher #C4: Less reliable than expected

The third case, teacher #C4 is a teacher of primary education who was interviewed in May and June of 2024. Before the exploratory phase, she had only used ChatGPT once in private and seen other teachers use it at a different school to generate lesson plans:

> Lo and behold, it quickly spit out a meaningful concept. Incredible. Yes, of course, I didn't use it for any practical application afterward, […] but I do find it fascinating that it really works so well.

She seemed fascinated by the success of the use and goes on to describe using it for individualized learning, generating material for students or letting students interact with it during the development of own tasks or as a support for creative thinking. In problem solving and mathematical modeling, she sees a special strength of AI in students' hands, allowing for support during the process. Contrasting these chances of the technology, she describes the challenge of students blindly trusting the answers and lacking the ability to critically reflect on them. She sees this issue with existing sources like the internet and considers it likely to get worse with AI.

In her test of the AI, she first asked for some mathematical games to implement in her lessons. She was familiar with all of the games and indicated them to be the standard games used in primary mathematics education. Therefore, she expressed slight disappointment as she had expected more creative and new ideas of the AI. As a second task she tried to generate a set of inverse tasks which continuously failed, as the AI generated different types of tasks. Although trying it in different ways, she could not solve this problem. The AI generated good tasks in her opinion, so the quality of the output was good, just not the type of task what she had asked for:

> Even there, the AI gave me a wonderful addition-subtraction task, but not the inverse task. So, there's still room for improvement, quite a lot of room actually, as it was simply mathematically incorrect. I found that astonishing since the topic isn't particularly difficult and the term is very mathematically precise. I found and still find that quite remarkable.

This inability of the AI to solve a task, that was easy in her opinion, stands in stark contrast to the expectations she formulated beforehand. Considering the statements after this exploratory phase, a strong impact on her beliefs can be postulated as well. After the exploration she argues that AI may be more helpful for career beginners or for spontaneous plannings of replacement lessons, as for experienced professionals a task like the one she gave to AI (see above) would have worked out better just using other resources like a didactical textbook. Furthermore, the belief of the risks associated with the unreliability of AI answers, which she originally stated only in relation to students working with it, is now expanded to teachers and lesson planning:

> And if I have ChatGPT explain something to me, I don't have the assurance that it has to be factually correct down to the smallest detail. Of course, it's correct on a larger scale, but that's exactly where I see a big risk, right? And that's why lesson preparation might still be difficult at the moment, because there are simply still things missing.

In conclusion, this teacher was disillusioned about the abilities of generative AI through the exploration. Her concerns regarding the complexity of assessing the truth in AI output were expanded, now seeing the usage of teachers more critical.

6. Discussion

Comparing the results for the first research question from this study to topics in the research on generative AI use for mathematics education obtained from the literature, multiple categories of the found beliefs are well-aligned with areas in the literature. The *areas of application* mentioned by the participants of this study correspond to many of the benefits generative AI may provide (ElSayary, 2024; Pepin et al., 2025). The applications in the classroom as *personalized learning* and *language support* are examples of such categories, where the categories of teacher beliefs are aligned with benefits mentioned in the literature. Besides these and further benefits, the *risks of AI usage* are also well-aligned with theoretically formulated risks and challenges (Kasneci et al., 2023; Pepin et al., 2025; Busuttil & Calleja, 2025). The perceived *significance of AI* in both of its sub-categories is also described for the participants in the study of Busuttil & Calleja (2025).

The sub-level categories considering the *pedagogical suitability of AI* touch upon themes discussed in the literature as well as the political discourse on AI, such as the usage in primary school and the challenges associated with the complexity of AI answers (Kasneci et al., 2023; Buchholtz et al., 2024; Pepin et al., 2025). Looking at the beliefs in these categories being connotated mostly negative, seeing no use of AI in primary school or no added value of AI in contrast to other resources, the need for research on the role of AI in these cases and teacher training based on this research become apparent. The *suitability regarding mathematical content* is only partially in alignment with results from the literature. While the limitations of large language models for accurate mathematical drawings and their potential negative impact on geometry are reported as a challenge in literature (Wardat et al., 2023), the teachers in this study are in disagreement whether it is suitable not only for geometry, but also other mathematical topics, which shows some differences to the results obtained by Busuttil & Calleja (2025). Especially the shared belief of several teachers, that AI is more suitable for other subjects in comparison to mathematics is an interesting result, as it could not be identified as a common challenge mentioned in the literature. This hints at a need to showcase potential usage scenarios in mathematics education during teacher training.

A general observation regarding the categories and the beliefs they contain is their limited mathematics specificity. Most categories and beliefs are not uniquely related to mathematics and could also be reported for other school subjects. This may be due, on the one hand, to the interview questions, which were not explicitly focused on mathematics. On the other hand, it may reflect that most participating teachers were novices in their use of AI. With regard to teachers' professional competencies in using technology for teaching and learning, the MPC model (Dilling et al., 2024c) suggests that situated experiences with technology use in teaching and learning constitute a foundation for teachers' professional competencies. As the participants in our study were mostly AI novices, it is plausible that they lack a broad base of experience in using generative AI for teaching and learning. If beliefs are considered part of teachers' professional competencies, it follows that the teachers may not yet have developed (many) mathematics-specific beliefs about the use of generative AI for teaching and learning.

Learners' AI competencies are an issue discussed broadly in the literature, with the *relevance of competence development* and the need for *reflection competence* / critical thinking mentioned by the teachers being supported widely (Buchholtz et al., 2024; Pepin et al., 2025). *Teacher competencies* are also a relevant issue in the literature (Buchholtz et al., 2024; Pepin et al., 2025) with multiple of the beliefs mentioned in the *general statements* being represented. The negatively connotated *self-assessments* are not surprising considering the limited experience the participants had with generative AI for their professional activities but point at a need for professional development addressing these, nevertheless. The category of *prerequisites and support* also points at the need for such activities, as the teachers describe the necessity for professional development and teacher training. *Educational objectives* and *technical conditions* as mentioned by the teachers are also discussed in the literature as limiting factors of AI integration into the classroom (Kasneci et al., 2023).

Considering the *fundamental changes to the education system through AI*, these are mostly speculations considering the future, so the comparison to the literature is of a different nature. Some potential changes envisioned in the literature also address changes to *educational objectives*, the *teacher role* and *instructional design* (Buchholtz et al., 2024), which shows alignment between the themes identified in literature and this study. Even though there is alignment of the categories, this does not necessarily apply to the beliefs in the categories. A shift of the teacher role and changes of the competencies students have to acquire are supported by theoretical considerations. Whether generative AI will have no impact on early primary education or will replace resources like textbooks in the near future, are debatable. Lastly, the *AI characteristics* mentioned by the teachers are (in part) characteristics mentioned in the literature, such as the relevance of good prompting and the variety of possible AI answers for a single prompt. Some part of these beliefs are of limited validity, such as that AI gets better with every generation.

Regarding the second research question, the belief changes observed suggest that even a short exploration phase can exert a strong impact, capable of altering teachers' beliefs. Considering the kind of belief changes we observed it is a limitation of this article, that only three cases can be presented due to the length limitations. It became apparent that no general pattern of belief change could be observed in the teachers, such as that all teachers would view AI more positively after the exploration phase. For this reason, three different cases were chosen to represent how different the changes could be. The belief changes in them shed light on the process of belief development and the role of active use of generative AI in this process.

As the first case shows, a first contact with AI and exploring its possibilities might be able to convince teachers of the usefulness and relevance of this resource. The second and third case show how negative experiences might be able to assist teachers in realizing limitations of AI on one hand and understanding the relevance of developing competencies in using AI (i.e. prompting and knowledge about the limitations of the user-interface) as part of their professional development on the other hand. The cases also show the risks associated with letting teachers explore generative AI without guidance and support, as their lack in competencies in using it might lead to unsatisfactory AI output, which might in turn lead to negative beliefs and discourage teachers from using this resource in the future.

The underlying framework of instrumental genesis, in particular the process of instrumentation, proved to be a suitable lens to analyze the belief changes of the teachers. The observed negative experiences in the second and third case show how the teachers formed utilization schemes that incorporated the inabilities of the artifact to perform certain tasks on one side and meta-level reflections about own missing skills in using the artifact on the other. Especially the latter aspect points at the difference between forming utilization schemes alone by exploring the artifact and forming them in social situations by adapting schemes of other users, such as in professional development activities.

Looking at further limitations of this study, first the limited number of participants has to be mentioned. This study is based on data from 14 teachers, that were not randomly sampled from the general population, which might influence the stated beliefs and categories of beliefs. Furthermore, the participants were not equally split among primary and secondary school teachers (9 from primary and 5 from secondary education), which likely influenced the stated beliefs. This is partially mitigated by a higher number of stated beliefs in the interviews of the secondary school teachers (also see the section on data saturation), but some effect might be observed, nevertheless. Social desirability could also have affected the results, as participants might have answered in a certain way to please the interviewers, regardless of their true opinion on certain questions.

Additionally, some of the categories of beliefs that were identified are related to the interview questions asked. To gain enough data to illustrate the broad spectrum of beliefs, specific questions are necessary to touch upon different possible areas for these beliefs. This relation is visible in the question for the suitability of AI for different mathematical topics and the top-level category about *mathematical and content-related suitability*. While this relation of top-level categories and questions applies to multiple of the top-level categories, the interesting results here are the sub-level categories untouched by this. For example, the focus on mathematical processes in contrast to mathematical content-areas in the sub-level category was not initiated by the interviews.

Furthermore, many of the top- and sub-level categories are only vaguely related to the interview questions. As an example, the categories about teachers' and learners' competencies might be related to multiple of the questions in some distant way, but it was the participants' choice to share these specific areas of beliefs in response to the questions. Another limitation of the interview questions lies within their limited focus on mathematics as a subject. Especially for a study in mathematics education, a stronger focus on mathematics is desirable, but as we wanted to keep the questions open enough to not influence the stated beliefs, the choice for more general questions was made.

7. Conclusion & Outlook

The aim of this study was the identification of teachers' initial beliefs about generative AI in mathematics education. In order to obtain a broad picture, teachers from different types of schools were interviewed. Semi-structured guided interviews provided a deep insight into the teachers' beliefs.

Through an inductive approach, top- and sub-level categories could be identified for a broad number of beliefs. These categories touch upon competencies of teachers and students, the suitability of generative AI from pedagogical and mathematical perspectives, prerequisites of the usage in school, which areas of application of AI inside and outside the classroom might be beneficial, which risks are associated with AI, what general characteristics AI has according to teachers and how it will change the education system in the future. Comparing these topics to the themes discussed in the literature, it could be established that many of the found categories are well-aligned with them. While the general topics are well-aligned, some of the beliefs inside the categories are not represented in the literature, such as teachers considering other subjects more suitable for AI usage than mathematics.

The appearance of some of the categories points towards a necessity for professional development activities and the relevance of showcasing meaningful applications of generative AI in mathematics education. Although most participants considered their competencies in the usage of generative AI for their professional activities very limited, there is a sense of basic awareness of the opportunities and risks of AI, which is likely based on superficial assessments of information from social media or other people, due to the lack of own experiences. The quantity of beliefs the teachers described, despite their lack in experience, also show the possibilities for professional development to influence these beliefs. Even though it is unclear, with how much conviction the beliefs are held, this study shows, that even with no professional experience using generative AI, beliefs which may influence teaching decisions are already held.

As many of the beliefs described by the teachers in our study are not related to mathematics as a subject in particular but are on a more general pedagogical level, the need for professional development that addresses the opportunities and challenges by generative AI for teaching and learning mathematics is emphasized. As this study shows, teachers need situated experiences in their mathematics lessons to be able to develop beliefs that are specific to mathematics. While this slightly reduces the relevance of the finding in this study, it points at the necessity of further research in this area.

Our study also showed that these beliefs can be changed by acquiring experience – even in a very short exploration phase. Particularly when it comes to AI, a topic on which many teachers are still novices, initial concrete experiences seem to have a major influence. This result can be seen as a double-edged sword: On the one hand, a successful exploration with a sense of accomplishment is not unlikely, as AI is highly accessible and can produce good results easily without good prompting compared to other mathematics resources. On the other hand, without good prompting and knowledge about this technology, unsatisfactory AI output may be generated easily as well, leading to disappointment and discouragement against trying to integrate this resource into teachers' professional activities. This finding can be constructively taken up in the context of pre-service and in-service teacher training and demonstrates the relevance of professional development in the field of AI.

Future research will examine changes in teachers' beliefs about generative AI in greater depth. The authors are currently conducting empirical studies to assess how extended, more structured interventions—such as AI-focused seminars for pre-service teachers and multi-year professional-development programs for in-service teachers—shape beliefs about AI (see https://kimadu.de/).

Annotations

[1] Please note that our brief literature overview focuses specifically on in-service teachers and their beliefs about generative AI and large language models: Studies focusing on AI in general like from Hwang (2023) were excluded, as beliefs about AI in general and generative AI may greatly differ. Similarly, Studies focusing on pre-service teachers (e.g. Hoya et al., 2024) were excluded, as it is yet unclear how the beliefs of pre-service and in-service teachers might be related considering this topic and whether they are comparable.

Acknowledgements

We would like to thank the reviewers for their helpful and constructive feedback. Furthermore, our thanks go to Judith von Spee and Sandra Horster for their support during the interviews.

Literature

Beege, M., Hug, C., & Nerb, J. (2024). AI in STEM education: The relationship between teacher perceptions and ChatGPT use. *Computers in Human Behavior Reports*, *16*, 100494. https://doi.org/10.1016/j.chbr.2024.100494

Buchholtz, N., Baumanns, L., Huget, J., Noster, N., Rott, B., Schorcht, S., Siller, H.-S. & Sommerhoff, D. (2024). Damit rechnet niemand! Sechs Leitgedanken zu Implikationen und Forschungsbedarfen zu KI-Technologien im Mathematikunterricht. *Mitteilungen der GDM, 117*, 15-24. https://ojs.didaktik-der-mathematik.de/index.php/mgdm/article/view/1249

Buchholtz, N., & Huget, J. (2024). ChatGPT as a reflection tool to promote the lesson planning competencies of pre-service teachers. In E. Faggiano, A. Clark-Wilson, M. Tabach, & H.-G. Weigand (Eds.), *Proceedings of the 17th ERME Topic Conference MEDA 4* (pp. 129–136). University of Bari Aldo Moro

Busuttil, L., & Calleja, J. (2025). Teachers' Beliefs and Practices about the potential of ChatGPT in teaching mathematics in secondary schools. *Digital Experiences in Mathematics Education, 11*, 140-166. https://doi.org/10.1007/s40751-024-00168-3

Clark-Wilson, A., Robutti, O., & Thomas, M. (2020). Teaching with digital technology. *ZDM – Mathematics Education, 52*, 1223–1242. https://doi.org/10.1007/s11858-020-01196-0

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*, 1460337. https://doi.org/10.3389/frai.2024.1460337

Dilling, F., Herrmann, M., Müller, J., Pielsticker, F., & Witzke, I. (2024b). Initiating interaction with and about ChatGPT – an exploratory study on the angle sum in triangles. In E. Faggiano, A. Clark-Wilson, M. Tabach, & H.-G. Weigand (Eds.), *Proceedings of the 17th ERME Topic Conference MEDA 4* (pp. 145–152). University of Bari Aldo Moro.

Dilling, F., Köster, J., & Vogler, A. (2024a). Beliefs of Undergraduate Mathematics Education Students in a Teacher Education Program about Visual Programming in Mathematics Classes. *International Journal for Research in Undergraduate Mathematics Education*. https://doi.org/10.1007/s40753-024-00248-0

Dilling, F., Schneider, R., Weigand, Hans-Georg (2024c). Describing the digital competencies of mathematics teachers: theoretical and empirical considerations on the importance of experience and reflection. *ZDM – Mathematics Education, 56*, 639–650. https://doi.org/10.1007/s11858-024-01560-4

Dilling, F., & Vogler, A. (2022). Pre-service Teachers' Reflections on Attitudes Towards Teaching and Learning Mathematics with Online Platforms at School: A Case Study in the Context of a University Online Training. *Technology, Knowledge and Learning, 28*, 1401–1424. https://doi.org/10.1007/s10758-022-09602-0

Dresing, T., & Pehl, T. (2018). Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende. https://www.audiotranskription.de/ressourcen/praxisbuch/

Eichler, A., Erens, R., & Törner, G. (2023). Measuring changes in mathematics teachers' belief systems. *International Journal of Mathematical Education in Science and Technology, 54*(8), 1490–1508. https://doi.org/10.1080/0020739X.2023.2170835

ElSayary, A. (2024). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, *40*(3), 931–945. https://doi.org/10.1111/jcal.12926

Glaser, B., & Strauss, A. (1967). Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publishing Company.

Grigutsch, S., Raatz, U., & Törner, G. (1998). Einstellungen gegenüber Mathematik bei Mathematiklehrern. *Journal für Mathematik-Didaktik,* *19*(1), 3–45. https://doi.org/10.1007/BF03338859

Grootenboer, P. (2008). Mathematical belief change in prospective primary teachers. *Journal of Mathematics Teacher Education*, *11*(6), 479–497. https://doi.org/10.1007/s10857-008-9084-x

Hoya, F., Mah, D., Prilop, C., Jacobsen, L., & Weber, K. (2024). Pre-service teachers' AI usage: The effects of perceived usefulness, subjective norm, behavioral intention, and self-efficacy. OSF: https://doi.org/10.31219/osf.io/284gk

Hwang, S. (2023). Preservice teachers' beliefs about integrating artificial intelligence in mathematics education: A scale development study. *Research in Mathematical Education, 26*(4), 333–349. https://doi.org/10.7468/jksmed.2023.26.4.333

Kamodi, T., & Garegae, K. (2019). Teachers' perceptions on use of Microsoft excel in teaching and learning of selected concepts in junior secondary school mathematics syllabus in Botswana. *Lonaka Journal of Learning and Teaching, 10*(1). https://journals.ub.bw/index.php/jolt/article/view/1550

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences, 103*, 102274.

Liljedahl, P., Oesterle, S., & Bernèche, C. (2012). Stability of beliefs in mathematics education: a critical analysis. *Nordic Studies in Mathematics Education*, 17 (3-4).

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. https://www.ssoar.info/ssoar/bitstream/handle/document/39517/ssoar-2014-mayring-Qualitative_content_analysis_theoretical_foundation.pdf

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

Pehkonen, E., & Pietilä, A. (2003). On relationships between beliefs and knowledge in mathematics education. In M. Mariotti (Ed.), *Proceedings of the third conference of the European Society for Research in Mathematics Education* Department of Mathematics of the University of Pisa. http://www.dm.unipi.it/~didattica/CERME3/proceedings/Groups/TG2/TG2_pehkonen_cerme3.pdf

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in Mathematics Education. *Digital Experiences in Mathematics Education*, *11*, 9–41. https://doi.org/10.1007/s40751-025-00172-1

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Eds.) *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Information Age.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity, 52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Saralar-Aras, I. (2022). An exploration of middle school mathematics teachers' beliefs and goals regarding a Dynamic Tool in Mathematics Lessons: Case of Geogebra. *Journal of Research in Science, Mathematics and Technology Education, 5*(SI), 41-63. https://doi.org/10.31756/jrsmte.113SI

Scantamburlo, T., Cortés, A., Foffano, F., Barrué, C., Distefano, V., Pham, L., & Fabris, A. (2025). Artificial Intelligence Across Europe: A Study on Awareness, Attitude and Trust. *IEEE Transactions on Artificial Intelligence, 6(2), 477-490.* https://doi.org/10.1109/TAI.2024.3461633

Schoenfeld, A. (1985). *Mathematical Problem Solving.* Academic Press.

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/feduc.2024.1386075

Schorcht, S., Peters, F., & Kriegel, J. (2025). Communicative AI Agents in Mathematical Task Design: A Qualitative Study of GPT Network Acting as a Multi-professional Team. *Digital Experiences in Mathematics Education,* 11, 77–113. https://doi.org/10.1007/s40751-024-00161-w

Seawright, J., & Gerring, J. (2008). Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly, 61*(2), 294–308. https://doi.org/10.1177/1065912907313077

Sjuts, J. (2024). Sinn oder Unsinn? ChatGPT beim Lösen von Aufgaben und Problemen in Mathematik. *MNU-Journal 77*(2), 128–136

Speer, N.M. (2005). Issues of Methods and Theory in the Study of Mathematics Teachers' Professed and Attributed Beliefs. *Educational Studies in Mathematics 58,* 361–391. https://doi.org/10.1007/s10649-005-2745-0

Thurm, D. (2019). Digitale Werkzeuge im Mathematikunterricht integrieren - Zur Rolle von Lehrerüberzeugungen und der Wirksamkeit von Fortbildungen. Springer. https://doi.org/10.1007/978-3-658-28695-8

Thurm, D., & Barzel, B. (2020). Effects of a professional development program for teaching mathematics with technology on teachers' beliefs, self-efficacy and practices. *ZDM – Mathematics Education*, *52*, 1411–1422. https://doi.org/10.1007/s11858-020-01158-6

Thurm, D., & Barzel, B. (2022). Teaching mathematics with technology: A multidimensional analysis of teacher beliefs. *Educational Studies in Mathematics, 109*(1), 41–63. https://doi.org/10.1007/s10649-021-10072-x

Trouche, L. (2004). Managing the Complexity of Human/Machine Interactions in Computerized Learning Environments: Guiding Students' Command Process through Instrumental Orchestrations. *International Journal of Computers for Mathematical Learning*, *9*(3), 281–307. https://doi.org/10.1007/s10758-004-3468-5

Trouche, L, Gueudet, G., & Pepin, B. (2019). *The 'resource' approach to mathematics education.* Springer. https://doi.org/10.1007/978-3-030-20393-1

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need (Version 7). arXiv. https://doi.org/10.48550/ARXIV.1706.03762

Vérillon, P., & Rabardel, P. (1995). Cognition and artefacts: A contribution to the study of thought in relation to instrumented activity. *European Journal of Psychology of Education, 10*(1), 77–101. https://doi.org/10.1007/BF03172796

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7). https://doi.org/10.29333/ejmste/13272

Weygandt, B. (2021). Beliefs – Definitionsversuche eines ›messy construct‹. In: Mathematische Weltbilder weiter denken. Springer Spektrum. https://doi.org/10.1007/978-3-658-34662-1_5

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods. Sage.

Yoon, H., Hwang, J., Lee, K., Roh, K. H., & Kwon, O. N. (2024). Students' use of generative artificial intelligence for proving mathematical statements. *ZDM – Mathematics Education, 56*, 1–21. https://doi.org/10.1007/s11858-024-01629-0

Anschrift der Verfasser

Frederik Dilling
Universität Siegen
Didaktik der Mathematik
Adolf-Reichwein-Straße 2
57076 Siegen
frederik.dilling@uni-siegen.de

Marc Herrmann
Universität Siegen
Didaktik der Mathematik
Adolf-Reichwein-Straße 2
57076 Siegen
marc.herrmann@uni-siegen.de

Ingo Witzke
Universität Siegen
Didaktik der Mathematik
Adolf-Reichwein-Straße 2
57076 Siegen
witzke@mathematik.uni-siegen.de