Educational Studies in Mathematics (2022) 109:263–285
https://doi.org/10.1007/s10649-021-10102-8

Measuring pre-service teachers' noticing competencies within a mathematical modeling context – an analysis of an instrument

Alina Alwast[1] · **Katrin Vorhölter**[1]

Accepted: 16 August 2021 / Published online: 13 October 2021

Abstract
Teaching mathematical modeling is a demanding task. Thus, fostering teachers' competencies in this regard is an essential component of teacher education. Recent conceptualizations of teachers' competencies include situation-specific skills based on the concept of noticing, which is of particular interest for the spontaneous reactions needed when teaching mathematical modeling. The study described in this paper aims to analyze the development of a video-based instrument for measuring teachers' noticing competencies within a mathematical modeling context and obtain evidence for the validity of the instrument. Three kinds of validity are examined in three different studies: content validity, elemental validity and construct validity. Indicators for content validity could be found through different expert ratings and implementation with the target group, where participants were able to perceive all relevant aspects. The qualitative analysis of participants' reasoning, which is consistent with the coded level, indicates elemental validity. Moreover, the results of the confirmatory factor analysis suggest construct validity with one overall factor of noticing competence within a mathematical modeling context. Taken together, these studies imply a satisfactory validity of the video-based instrument.

Keywords Mathematical modeling · Modeling competencies · Noticing · Teachers' competencies · Validity

1 Introduction

Mathematical modeling is included in many curricula and academic standards, such as the Common Core Standards in the United States, and thus is an essential part of mathematics education worldwide. The integration of real-world problems into school lessons can

✉ Alina Alwast
alina.alwast@uni-hamburg.de

Katrin Vorhölter
katrin.vorhoelter@uni-hamburg.de

[1] University of Hamburg, Hamburg, Germany

support students in working autonomously, thinking critically, and developing solutions to real-world problems, thus preparing them to be responsible citizens (Kaiser, 2017).

However, mathematical modeling is demanding for both students and teachers (Blum, 2015). Therefore, preparing teachers to adequately teach mathematical modeling is an important issue. The difficulty of teaching mathematical modeling is, among other reasons, due to the variety of approaches that can be chosen by the students to solve a modelling problem, which often cannot be anticipated. Furthermore, teachers have to spontaneously react to various obstacles that students might encounter, and offer minimal help (Leiß, 2007; Stender, 2016) to support students' autonomous work in a successful and goal-oriented way. To act adaptively and thus support students in the best possible way, teachers must be capable of acting in a situation-specific manner – that is, teachers must perceive students' problems, even if the students do not express them directly, interpret these problems correctly using their professional knowledge, and, finally, make a well-founded decision on how to act. Thus, the concept of noticing (van Es & Sherin, 2002) offers a framework for analyzing teachers' situation-specific competencies and allows researchers to understand and explore – and, consequently, promote – teachers' in-the-moment decisions and interventions. To do so, an instrument to measure those noticing competencies in this specific modelling context is needed.

In this article, we therefore conceptualize noticing competencies within a mathematical modeling context, present a video-based instrument for measuring those competencies of pre-service teachers and argue for the instrument`s validity in terms of content validity (Study 1), elemental validity (Study 2), and construct validity (Study 3).

2 Theoretical framework

2.1 Teaching mathematical modeling

Working on mathematical modeling problems is challenging for both students and teachers (Blum, 2015). In an independent and goal-oriented working process, students not only have to work mathematically but must also make sense of the real-world context of a given problem and go back and forth between the real world and the mathematical procedures. This process is usually illustrated in an idealized way by a modeling cycle. Several subcompetencies for working on modeling problems related to the various phases of the modeling process can be distinguished (Kaiser, 2007, p. 111):

- competencies to understand real-world problems and to construct a reality model;
- competencies to create a mathematical model out of a real-world model;
- competencies to solve mathematical problems within a mathematical model;
- competency to interpret mathematical results in a real-world model or a real situation;
- competency to challenge solutions and, if necessary, to carry out another modeling process.

Global competencies, such as metacognitive competencies (Maaß, 2006; Stillman, 2011; Vorhölter, 2019) and social competencies (Kaiser, 2007), also play an important role. An overview of the different strands and foci in the discussion on modeling competencies can be found in Kaiser and Brand (2015).

To support students in developing modeling competencies as well as working independently and in a goal-oriented manner, teachers need to possess a variety of competencies. In addition to their own modeling competencies, teachers must also possess domain-specific pedagogical content knowledge. In their model, Borromeo Ferri and Blum (2010) categorized this modeling-specific knowledge (and skills) into four dimensions: "(1) a theoretical dimension (incl. modeling cycles or aims and perspectives of modeling as background knowledge), (2) a task dimension (incl. multiple solutions or cognitive analyses of modeling tasks), (3) an instructional dimension (incl. interventions, support and feedback), and (4) a diagnostic dimension (incl. recognising students' difficulties and mistakes)" (Blum, 2015, p. 89). Moreover, Klock and Wess (2018) used this model to conceptualize modeling-specific knowledge and add modeling-specific beliefs and self-efficacy as components of teachers' competence for teaching mathematical modeling based on the COACTIV[1] model.

The complexity of teaching mathematical modeling is made evident by Borromeo Ferri and Blum's aforementioned categorization. However, it becomes even more pronounced when considering the fact that students can solve a problem in multiple ways and therefore a variety of problems, which are sometimes hard to anticipate, can arise. Therefore, teachers have to consider a multitude of aspects to adaptively support their students in the best possible way.

First, it is important for teachers to know the *phase of the modeling process* to diagnose the students' current state and assist them (Stender, 2016). Each step of the modeling process features potential *difficulties* that students must overcome to successfully continue working on the modeling problem (Goos, 2002; Stillman, 2011). Knowing and recognizing the variety of difficulties that can occur during the modeling process in general and regarding specific tasks (see Blum, 2015; Maaß, 2006; for a detailed survey of the literature, see Niss & Blum, 2020) is essential to performing appropriate interventions (Leiß, 2007). Second, an open problem can have a *variety of possible solutions*, which teachers need to recognize and analyze mathematically to support students in their individual approaches (Schukajlow & Krug, 2014). Third, research has shown that *metacognitive strategies*, such as monitoring and regulation, are vital for students to be able to identify a problem, overcome difficulties, and solve the modelling problem in a productive and independent manner (Stillman, 2011; Vorhölter, 2018).

The aspects described above are necessary for teaching mathematical modeling, but are not sufficient for the optimal promotion of students' modeling competencies.

Consequently, to support students in the best possible manner, teachers not only need to possess modeling-specific knowledge but must also perceive and interpret important problem situations based on their knowledge to quickly make a well-founded decision. As these are part of the concept of noticing, this concept is described in the following.

2.2 Teachers' noticing competencies

Research on teacher competencies has tended to focus either on teachers' underlying knowledge or on teachers' behavior for a long time. To overcome the dichotomy between "a behavioral assessment in real-life situations versus an analytical assessment of dispositions underlying such behavior" (Blömeke et al., 2015, p. 5), Blömeke et al.

[1] The Cognitive Activation in the Classroom Project.

(2015) developed a theoretical model that represents competence as a continuum. In their model, competence is defined as a process consisting of three parts: (1) *dispositions*, which involve cognitive and affective-motivational components; (2) the *situation-specific skills* needed to apply dispositional aspects in a specific context; and (3) *performance*, which is the observable behavior. Situation-specific skills can be divided into *perception* of noteworthy aspects, *interpretation* of these aspects based on theoretical concepts, and *decision-making* based on this interpretation. An empirical study by Santagata and Yeh (2016) confirmed that these parts are connected in a circular way and that "situation-specific skills function as the processes through which knowledge and beliefs become relevant in practice" (p. 163) and create new dispositions.

Although the idea of situation-specific skills as component of professional competence is not completely new in teacher education (for early discussions, see, e.g., Erickson et al., 1986; Erickson, 2011), technological developments in the last decade have offered novel opportunities to further investigate this concept, leading to more research being performed in the field of noticing (Sherin et al., 2011b; Schack et al., 2017). Today, many studies on teachers' noticing have been carried out using different conceptualizations of noticing and the related terminology, including *noticing* (van Es & Sherin, 2002), *teacher noticing* (Sherin et al., 2011a), *professional noticing of children's mathematical thinking* (Jacobs et al., 2010), and the *discipline of noticing* (Mason, 2002). Studies have also employed pre-existing theories, such as *professional vision* (Goodwin, 1994), to enrich their conceptualizations. In spite of methodological differences, scholars agree that the effectiveness of teaching should be improved by focusing on situational processes.

However, different studies adopt different conceptualizations, specifications, and foci. Usually, scholars distinguish two or three sub-facets of noticing competencies, which occasionally have overlapping definitions:

1. *Attending to, perceiving, identifying, or paying (selective) attention to noteworthy events in a complex classroom setting, that are essential for student learning*: This sub-facet generally describes the perception of noteworthy aspects when confronted with an overwhelming amount of information in classroom settings. Thus, it is important to perceive relevant information and/or identify essential information based on prior experience, the individual knowledge base, and the learning objective. In addition, theoretical constructs may be drawn upon to classify the perceived event (see Jacobs et al., 2010). Depending on teachers' disposition, their perception may be selective and discriminative (Ball, 2011).
2. *Making sense of, reasoning about, or interpreting students' behavior and thinking*: This sub-facet illustrates the process of analyzing and categorizing perceived information according to one's own domain-specific knowledge and making sense of it by drawing upon abstract theories, former experiences, and personal orientations (see Schoenfeld, 2011). This sub-facet can also be divided into the following two components: (1) making connections between specific events and the broader principles of teaching and learning and (2) using what one knows about the context to reason about a situation (see van Es & Sherin, 2002).
3. *Decision-making, (intended) responses, or additional thinking about alternative actions*: As a third sub-facet, conceptualizations may include teachers' decisions about actions based on their interpretation. This may also be included in the second sub-facet. Erickson (2011) suggested that noticing is an active process that is oriented toward the

goal of an action. Therefore, in some conceptualizations, decision-making is emphasized as a crucial component of noticing. Van Es and Sherin (2021) proposed *shaping* as an alternative third sub-facet, which "involves teachers constructing interactions, in the midst of noticing, to gain access to additional information that further supports their noticing" (p. 23).

In general, while perception is present in all conceptualizations, researchers have considered different combinations of sub-facets with different emphases and, more recently, have gone beyond the aforementioned three components (Dindyal et al., 2021), by proposing, for example, *productive noticing* (Choy, 2016), which involves lesson planning, and *curricular noticing*, which relates to curricular material (Dietiker et al., 2018).

There is no consensus whether these sub-facets can empirically be separated from one another or if they are interrelated (for a discussion on this question, see Thomas, 2017). Researchers have used different methods to investigate (1) different conceptualizations of noticing, (2) manifestations of competence in relation to noticing, and (3) the development and promotion of these competencies, which makes it hard to compare data (Amador, 2019).

To sum up, "noticing is a natural part of human sense making. In our daily lives, we see and interpret based on our own orientations and goals. However, the noticing entailed by teaching is specialized to its purposes" (Ball, 2011, p. xx). Not only is it important to notice with regard to the teaching profession but also to notice specifically embedded in certain contexts (see Dindyal et al., 2021). For example, with a focus on mathematics education, in TEDS-FU[2] items were used concerning mathematics-related classroom demands and pedagogy-related classroom demands (Kaiser et al., 2017). In terms of a more specific research focus, Moreno et al. (2021), for example, examined the development of teachers' noticing competence with regard to the topic length and measurement.

2.3 Conceptualizing noticing within a mathematical modeling context

To notice with regard to a profession entails using a certain lens through which content can be perceived and interpreted. Therefore, noticing within a mathematical modeling context involves looking through a specific lens, while focusing on noteworthy modeling-specific aspects.

As described above, modeling processes cannot be fully anticipated, and teachers have to respond quickly to complex ideas. Below, we elaborate on the three modeling-specific facets of noticing (see Fig. 1; highlighted in grey) based on the modeling-specific aspects described in Section 2.1.

1) *Perception* includes recognizing students' difficulties and their cognitive and affective barriers when working on a modeling problem, considering the mathematical content needed for students' individual approaches to solving the problem, identifying specific procedures when dealing with modeling problems, and paying attention to students' collaborations. Knowledge of typical problems and sensitivity to challenging situations help to identify these events. For example, without being sensitive to the specificities of mathematical modeling, one might perceive students as being too slow in starting to solve a problem mathematically when, in fact, the students are structuring the informa-

[2] Follow-up study of the Teacher Education and Development Study (TEDS-M).

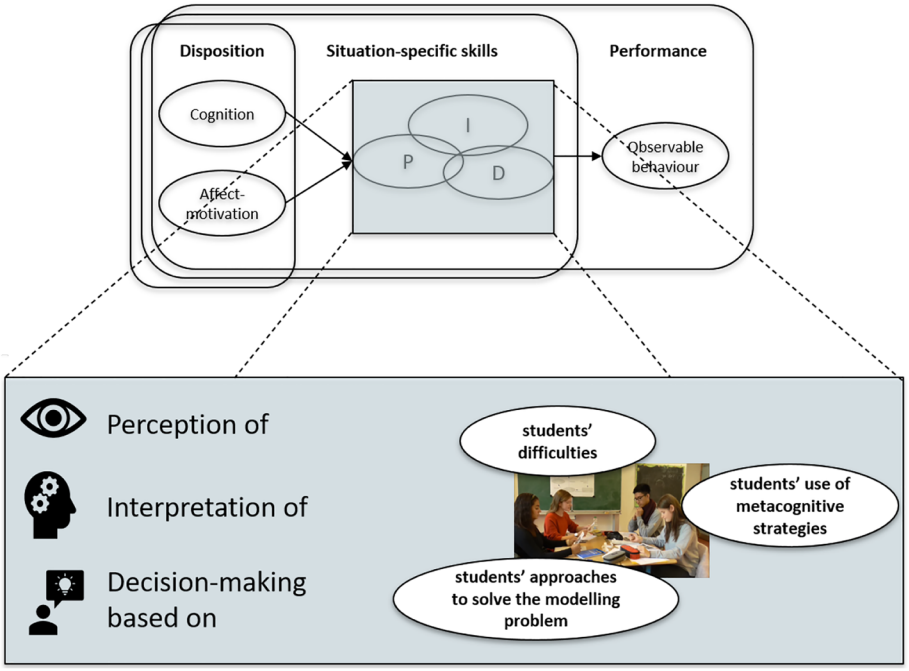

Fig. 1 Conceptualization of noticing within a mathematical modeling context based on the model of "competence as a continuum" by Blömeke et al. (2015)

tion and simplifying the problem, which is an important phase of the modeling process and should not be skipped for the sake of working mathematically only. Furthermore, if teachers are not aware of how important the exchange of ideas is for successful collaboration, they may miss related situations. Therefore, information is perceived and selected as noteworthy through a mathematical modeling lens based on knowledge about mathematical modeling.

2) To *interpret* students' behaviors and problems, teachers have to possess knowledge as conceptualized by Borromeo Ferri and Blum (2010) and Klock and Wess (2018), which influences teachers' interpretations of the modeling process. To interpret a perceived event, a great deal of theoretical and empirical knowledge needs to be considered. For example, a teacher may interpret a student's model as being incorrect if the teacher only knows one way of solving the problem and believes this to be the only correct solution. By contrast, a teacher who is sensitive to multiple possible solutions, which are typical for mathematical modeling problems, would interpret the situation differently and see the opportunity to support the students in finding their individual solution. In addition, a situation in which students in a small group do not work together could be misinterpreted as a lack of collaboration, whereas students have engaged in a meaningful distribution of work.
3) Instant *decision-making* is exceptionally difficult due to the complexity of modeling problems. To support students in the best possible manner, adaptive interventions are needed, which foster students' independent working processes (Stender, 2016) and are based on the interpretation of the situation at hand. For example, teachers have to decide

how to act in case of too complex or misleading approaches. Furthermore, decisions on how to proceed and which sub-competencies of modeling to foster in the moment or in the future must be made.

All in all, noticing within a mathematical modeling context (as illustrated in Fig. 1) requires not only perceiving and interpreting the noteworthy aspects that are specific for the teaching profession and making decisions accordingly but, moreover, being sensitive to content that is specific to and essential for mathematical modeling, such as students' modeling-specific difficulties, use of metacognitive strategies, and diverse approaches to solving modeling problems (see Section 2.1).

2.4 Measuring (prospective) teachers' noticing competencies

Assessing teachers' competencies has always been important for further development of the teaching profession. As noticing is a key competence for teachers, it is important to analyze the structure and development of their noticing competencies to create effective learning environments for teacher education. Recently, many video-based programs for fostering noticing competencies have taken place. Videos are used as a common stimulus to elicit noticing, often accompanied by questions, which vary in their degree of specificity. Van Es et al. (2019) analyzed video-based learning environments for teachers and identified five intended goals for the use of videos, which are not distinct: "developing specialized content knowledge for mathematics teaching, learning to systematically reflect on instructional practice, improving both the quality of mathematics instruction and teachers' noticing practices for teaching, and developing a professional vision of ambitious teaching" (p. 26). In a meta-study, which was not limited to the noticing context, Santagata et al. (2021) identified two types of activities, which involve watching videos, from a noticing perspective: "selective attention and knowledge-based reasoning" (p. 120). For an overview of video-based studies that foster noticing competencies, see Santagata et al. (2021) or Llinares and Chapman (2020). Amador (2019) reports that if noticing is used as a pedagogical tool, it is often also used as an analytical tool in the same project. Thus, for fostering noticing competencies in a goal-oriented way, measuring these competencies is highly relevant.

Video-based assessment instruments have proven to be useful for simulating classroom situations without the disadvantages of analyzing real-life teaching, which include disruptions of the teaching process and the low comparability of data. In 2001, van Es and Sherin (2002) organized so-called "video clubs", with in-service teachers attending professional development meetings to jointly analyze footage of teaching episodes in order to acquire noticing competencies. In their study, data were collected via interviews that used recorded video clips as prompts and were qualitatively analyzed. Since then, scholars have carried out various video-based studies using different stimuli (recordings of one's own teaching, recordings of someone else, or staged videos), participant groups (pre-service or in-service teachers), data collection methods (e.g., interviews, paper-and-pencil tests), and types of data analysis (qualitative or quantitative) (Dindyal et al., 2021). Moreover, different underlying conceptualizations of noticing were used, which determined the lens through which data were viewed and interpreted.

In a meta-study, Stahnke et al. (2016) concluded that most studies in the area of noticing focused on perception, many on interpretation, and fewer on decision-making. Most studies (reviewed in a meta-study by Santagata et al., 2021) used videos

as prompts, especially videos of other teachers' classrooms. However, combinations of different prompts are also possible. For example, a qualitative study by Jacobs et al. (2010) included students' written work as well as video clips and used the highest ranking of both assessments as the overall score. Kersting (2008) asked participants open questions based on video clips but quantified the data, which resulted in a set of three levels of interpretation ranging from descriptive comments to a coherent analysis. Bragelman et al. (2021) compared common methods for analyzing the development of the noticing competence (such as the method implemented by Jacobs et al., 2010) and then used a "micro-analysis of noticing" to identify developments at a smaller, more fine-grained scale. The framework by Jacobs et al. (2010) and the Learning to Notice framework (van Es, 2011) are often used but are also regularly modified to match the specific needs, underlying conceptualizations, and research goals, which makes a comparison of data difficult (Amador, 2019). In general, the majority of studies used qualitative approaches and examined how and to what extent noticing competencies develop (Santagata et al., 2021). By contrast, as a quantitative large-scale study, TEDS-FU used video vignettes and cognitively oriented items to integrate both approaches (Kaiser et al., 2017). Another quantitative instrument is the Observer Research Tool, which uses recorded classroom situations and rating items (Stürmer & Seidel, 2017).

3 Aim of the following studies

As discussed earlier, it is essential that teachers react quickly when confronted with unforeseen obstacles in classrooms. Thus, pre-service teachers should be prepared to make informed decisions based on their knowledge and the requirements of the specific situation at hand. This is especially important when supporting students who are confronted with a complex problem, such as a mathematical modeling problem. To evaluate teacher training programs in terms of improving noticing competencies, it is essential to develop an instrument for measuring teachers' noticing competencies within a mathematical modeling context.

In the following, we present a video-based instrument and analyze it regarding its capacity to measure noticing competencies within a mathematical modeling context. Thus, the aim of the following studies is to discuss different indicators for validity: content validity (Study 1) to ensure all relevant content is covered, elemental validity (Study 2) to show whether the coded levels of interpretation match participants' underlying reasoning, and construct validity (Study 3) to check the assumed structure of noticing competencies.

4 Study 1: Content validity

4.1 Research question

The aim of the first study was to examine the possibility of reconstructing indicators that could justify the instrument's content validity. In line with the argument-based approach to validity proposed by Kane (2016) as well as Kaiser et al.'s (2015) approach to validating video-vignettes, content validity depends on testing goals. In particular, content validity as defined by Moosbrugger and Kelava (2012) indicates

to what extent the theoretical construct is represented by the instrument, in the sense that all (and only) relevant aspects are covered. Content validity is usually examined by theoretically and logically reasoning about the fit of the instrument, which should be complemented by expert judgements. Therefore, we posed the following research questions: (1a) Do the developed videos adequately display a variety of noteworthy aspects of mathematical modeling according to an expert perspective and as identified in Section 2.3? (1b) Are the developed items adequate for generating answers that are in line with the intended purpose according to an expert perspective? (1c) Is the developed instrument feasible for pre-service teachers, i.e., are the generated answers in line with the intended purpose and does it evoke the expected range of responses?

4.2 Methods

As our aim was to assess individual competence and competence development, content-based methods to evaluate the instrument's validity seemed appropriate. Kane (2016) approved of Lissitz and Samuelsen's (2007) suggestion that, if the aim is indeed as mentioned above, a test should be evaluated based on the processes used to develop the test and the content that it is supposed to reflect.

4.2.1 Considerations of the first version of the instrument

First, it was necessary to choose a modeling problem, which students should work on in to be worked on in the staged videos. We selected the problem "Uwe Seeler's Foot, which requires students to check the statement of a newspaper article and compare the volume of two solid figures (Vorhölter and Kaiser, 2016). As this problem has been successfully used in several projects (e.g., Vorhölter, 2018), many task-specific insights regarding a variety of possible misconceptions and difficulties that students must overcome were available. By watching video recordings (recorded during the MEMO[3] project, Vorhölter, 2018, 2019) of students working on this specific modeling problem, we reconstructed typical modeling-specific difficulties and behaviors, which are not specific for this task but for modeling problems in general. Table 1 shows selected common difficulties that may arise when students work on a modeling problem and which have been reported widely in the literature (Blum, 2015; Maaß, 2006; Stillman, 2011; Vorhölter & Kaiser, 2016; for an overview, see Niss & Blum, 2020).

Similarly, to the approach used to identify typical *difficulties* and implement them in a specific situation in a video clip, the aspects *use of metacognitive strategies* and *different approaches to solving the modeling problem* were selected and adapted. Considering these aspects, two scripts were developed. Each script depicted a different phase of the modeling process. The scripts were discussed by an expert team and modified accordingly.

Similar to the approach chosen in the TEDS-FU study (Kaiser et al., 2015), we decided to use longer videos with a length of approximately three minutes each. In contrast to shorter vignettes, longer videos show students' working processes while

[3] Promotion of students ' metacognitive modeling competencies.

Table 1 Difficulties students face during modeling

Difficulty	Implementation in the staged videos	Phase of the modeling cycle (as seen in the staged videos)
Meta-knowledge about how to solve under-determined tasks is missing	Students are clueless about how to get started with the problem and which strategies to use	Understanding and simplifying the real-world problem
Unsuitable use of routines to solve a problem without considering the context	Students randomly combine numbers mentioned in the problem and mathematical operations without understanding their meaning	Understanding and simplifying the real-world problem
Extra-mathematical knowledge is not available or investigated	Students do not know the relation of shoe sizes to centimeters and hesitate to examine this practically	Understanding and simplifying the real-world problem
Interpretation is missing	An interim result is seen as the complete solution to the problem, because the working process stops before it is interpreted	Working mathematically
Mathematical difficulties when dealing with units	Students calculate the relation of the two objects without considering the difference in units (centimeters and meters)	Working mathematically
Lack of motivation	Part of the group lacks motivation and stops working and reflecting upon the solution	Working mathematically

the students are in a certain phase of the modeling cycle in detail. Regarding the choice of a qualitative or quantitative approach, we considered that closed multiple-choice items offer good testing properties, whereas open questions capture competences in a more holistic and comprehensive way. Furthermore, closed items direct the participants' attention toward certain aspects, whereas open questions allow researchers to examine participants' subjective views and investigate their emphases and range of noticed events. Therefore, we chose to use open questions, because they examine the capacity and extent to which an individual person can attend to and interpret a particular event in a classroom while remaining sensitive to individual differences.

Based on the considerations above, we developed open questions that asked the participants about *students' difficulties* in relation to the phase of the modeling cycle, their *use of metacognitive strategies, approaches to solving the modeling problem* and students' *role in the group*. These four questions per video required participants to perceive relevant events on the one hand and give sensible interpretations of these events on the other. To address decision making, participants were asked to state in direct speech how they would react to the problematic situation with which the video ended and to provide a reason for their *intervention*. Therefore, we added a question to address the second video vignette with regard to decision making. In total, we developed nine questions for the first version of the instrument.

This led to the development of the first version of the instrument, including the first version of the staged videos (see Table 2) and items.

Table 2 Description of video vignette 1

Video vignette 1 ("Understanding the real-world situation")
Group of students: Four students (3 females and 1 male or 4 females) in year 9 (age 14–16) of a German school for higher-achieving students (in the so-called "gymnasium" academic track), some of whom had a migrant background
Mathematical content: The students are working on the modeling problem "Uwe Seeler's Foot." They just began the working process and are encountering many difficulties while trying to understand and simplify the real-world situation. They have different ideas about how to approach the problem

4.2.2 Sample and procedure

When developing and analyzing staged videos as authentic prompts to measure pre-service teachers' noticing competencies within a mathematical modeling context, it was essential to include several stages of development, expert ratings, and revisions (see Table 3). This helped ensure that the content covered a broad range of important aspects related to the quality of mathematical modeling teaching based on the conceptualization described in Section 2.3.

The instrument was discussed and rated by different groups of experts as well as implemented with and reflected on by the target group (see Table 3):

- Fourteen experts who were knowledgeable in the domain of noticing and were experienced in using video-based assessments or who were specialized in the field of mathematical modeling discussed the first version of the instrument (sample 1).
- Four PhD students whose doctoral thesis focused on metacognition and mathematical modeling or noticing performed the test and discussed it afterwards (sample 2).
- We used the second version of the instrument in a pilot study with a pretest-posttest design with 15 pre-service teachers at the end of their master's programs who participated in a modeling course (sample 3a).
- The third version of the instrument was used in the main study with two groups at different competence levels at a German university. Group 1 ($n = 72$) had limited knowledge of mathematical modeling, while Group 2 ($n = 36$) had undergone a modeling seminar. Thus, the study design was quasi-longitudinal, although a small number of the participants (25%) completed the test twice (sample 3b).

Table 3 The instrument's development and content validation process in relation to the research questions (see Section 2.3)

	Sample	Instrument version	Research question
Stage 1: Validation of staged videos	1: 14 experts	1	1a
Stage 2: Validation of items	2: 4 PhD students	1	1b
Stage 3: Validation of the complete instrument	3a: 15 pre-service teachers	2	1c
	3b: group 1 ($n = 72$, without modeling experience), group 2 ($n = 36$, with modeling experience)	3	

In line with the conceptualization of competencies for teaching mathematical modeling (Borromeo Ferri & Blum, 2010), the modeling course included knowledge of modeling, modeling competencies, and teaching modeling as well as the concepts implemented in the staged videos. During the modeling course, recorded and staged videos and other student artefacts were regularly used to support pre-service teachers' noticing competencies within a mathematical modeling context (for more details, see Vorhölter, 2018; Alwast & Vorhölter, 2019, 2019).

Data were collected in the form of audio recordings of discussions (stage 1 and 2) and written forms (answers to the open questions of the instrument, stage 3). Regarding the discussion on mathematical modeling and noticing as well as our conceptualization of noticing within a mathematical modeling context, conclusions on the adaptions of the instrument were drawn. Moreover, in stage 3b, the differences between groups were examined using the Mann–Whitney U test for nonparametric data.

4.3 Results

Indicators for content-validity are examined at different stages of the instrument's development.

4.3.1 Stage 1

The first version of the staged videos was critically discussed by experts until consensus was reached. In general, the experts agreed that all noteworthy aspects were covered as intended. However, they mentioned that certain aspects should be emphasized more. Therefore, we implemented a number of adaptations discussed below.

Regarding *technical issues*, the most appropriate position and movement of the camera and the actors were identified to capture important activities and focus on the aforementioned noteworthy aspects, as the position of the camera influences what can be noticed. Close-ups of students' worksheets were added to clarify students' approaches to solving the problem. The focus of the camera, which directs the viewers' attention and is especially important when dealing with noticing, encompassed all four students.

Concerning the *script*, the noteworthy aspects that were hard to detect were highlighted through acting and language to underscore important utterances. Regarding *students' difficulties*, the experts noted that there were too many incidents and suggested reducing their number. By contrast, *metacognitive aspects* were hard to spot and had to be emphasized. Moreover, students' *approaches to solving the problem* were mentioned too quickly and were not clear. Thus, students' explanations were extended. Furthermore, it was noted that the individual student characters should be more consistent, with each role representing a particular trait, such as motivating the group to structure their approach.

Regarding the *implementation of the instrument*, we concluded that to simulate a real classroom situation, participants should be able to read the questions before watching the videos. Moreover, the participants should be allowed to watch each video vignette only once to approximate real classroom teaching. Contextual information regarding the age group, modeling problem, and the solution should be provided beforehand as teachers would have this information before teaching a class.

Fig. 2 Students working on a modeling problem in a staged video

The final two videos were shot with student actors to incorporate the discussed changes. As mathematical modeling is often done in groups, a group of four students who were meant to reflect the diversity of student groups in German classrooms was chosen (see Fig. 2).

4.3.2 Stage 2

Together with the discussion on the videos, the 14 experts talked about first ideas for items as the items are closely connected to the content of the videos, which shaped the development of items. Based on this discussion, the first version of the items was discussed with PhD students who were experienced in the field of metacognition and mathematical modeling or noticing. This resulted in refinements of the open questions to clarify the focus of the questions and changes in wording to improve comprehension. It was noted that an interpretation should request references to abstract theories more clearly. Furthermore, the concept of metacognition should be explained beforehand to separate noticing metacognitive strategies from knowing this concept.

4.3.3 Stage 3a

In this stage, we first implemented the adapted instrument in a pilot study. The results of the pilot study showed that the number of questions was too high, as participants could not answer all questions in depth and focus on a few aspects but were answering superficially and took a long time to do so. Thus, the number of open questions was reduced by erasing two questions about the role of the students in the group. The intended responses to the remaining seven questions could be achieved.

Thus, the final instrument included seven open questions for assessing pre-service teachers' noticing competence within a mathematical modeling context. Several noteworthy events with different topics could be identified, leading to 14 notable aspects in relation to perception and interpretation and one aspect related to decision-making. Due to the design of the instrument, perception and interpretation could not be separated (see Fig. 3) as the

Question 1: Using relevant didactic concepts and terminology, analyze the students' difficulties in relation to the modeling cycle and reason about it.

Fig. 3 An example from the instrument regarding perceiving and interpreting students' difficulties

focus was on interpretation, which depends on perceiving the event first. Therefore, perception was assessed retrospectively.

4.3.4 Stage 3b

As outlined above, it is highly important to recognize problem situations and to perceive them as such. To continue analyzing the instrument's feasibility for pre-service teachers and to examine whether the expected range of answers was covered (research question 1c), we used the third version of the instrument in the main study and examined the breadth of perception with regard to the previously defined noteworthy aspects. These aspects were chosen because they focus on detecting students' *difficulties*, *metacognitive strategies*, or *mathematical approaches* for solving the problem. Regarding each of these topics, the range of perceived aspects was analyzed by comparing groups (in percentage) at different competence levels (with or without intervention).

The number of perceived aspects (see Table 1 for a list of all perceivable aspects related to students' difficulties) in relation to the two staged videos varied across participants (see Fig. 4). It was possible to perceive the maximum number of aspects for all topics, although all possible metacognitive strategies were only perceived by a small number of participants in both groups. A comparison of the two groups revealed slight differences in the number of perceived incidents, although there is still a potential for improvement. The Mann–Whitney U test revealed no significant differences in the participants' abilities to perceive a variety of aspects regarding students' difficulties ($p = 0.32$) and students' approaches to solving the problem ($p = 0.66$). However, significant differences in the participants' abilities to perceive a variety of aspects regarding students' use of metacognitive strategies were found ($p = 0.0107$) with a small effect size (0.22). Overall, the data revealed no ceiling or floor effects in relation to the participants' perception of

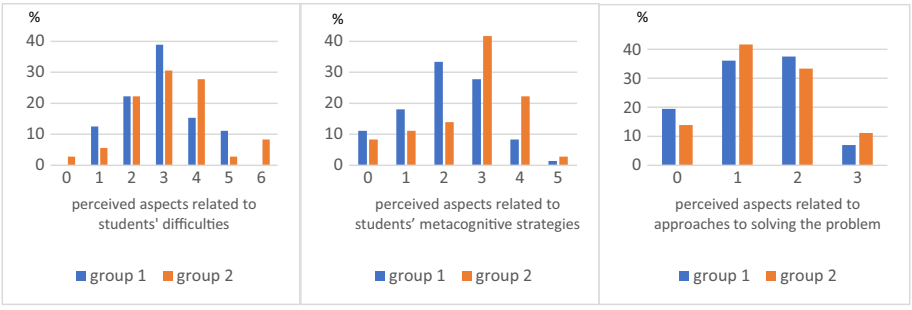

Fig. 4 The number of perceived aspects regarding different topics; Group 1: unexperienced in modeling; Group 2: experienced in modeling

students' difficulties, metacognitive strategies, and *approaches to solving the problem*. Furthermore, the number of perceived aspects varied among the participants, indicating that variances in perception can be measured.

5 Study 2: Elemental validity

5.1 Research questions

Elemental validity indicates whether the provided answers match the participant's reasoning (Schilling & Hill, 2007; Hill et al., 2007). Based on Kane's (2013) argument-based approach and the validity of the scoring inference, the elemental validity argument concerns the individual items in an instrument and indicates how well and consistently these items – or, in our case, the coding of an item – represent the underlying reasoning. Yang et al. (2018) used this approach to analyze closed and constructed response items and showed that it was possible to adapt an instrument for assessing noticing competence that was developed in Germany to another cultural context.

Thus, to address elemental validity, we analyzed the consistency between teachers' thinking as shown by their answers to the open questions and the coding of it. We examined whether the theory-based coding levels, which we adapted from Kersting (2008), are adequate for measuring pre-service teachers' noticing competence by posing the following research questions: 2a) Does the coding of theoretically developed levels of interpretation represent the underlying reasoning of pre-service teachers? 2b) Can all levels of interpretation be reconstructed?

5.2 Method

We take a closer look at the participants' reasoning in relation to deductively coded competence levels. While Group 1 involved pre-service teachers, who had limited knowledge of modeling, participants in Group 2 had taken part in a modeling course and were thus assumed to have a higher competence level.

5.2.1 The coding scheme of the instrument

We used qualitative content analysis (Kuckartz, 2014) to analyze the data. As the items were based on the staged videos, a coding manual was developed deductively based on the incidents included in the videos, with subcodes indicating participants' level of interpretation for each aspect. The participants' levels of interpretation were determined and coded similarly to previous works on the depth of interpretation in the noticing discussion (Kersting, 2008; van Es, 2011). We distinguished three levels according to the depth of interpretation, and quantified the data as described in Table 4.

Table 4 Description of the coding scheme

Code / Level of interpretation	Definition	Example
1	An interpretation is classified as level 1 if it is a description and only mentions a noteworthy aspect, which usually already classified an aspect in a certain way	When answering the question on students' difficulties, a participant states that students mixed up the units
2	A level-2 interpretation contains "some analytic inference" (Kersting, 2008) and includes reasoning about a situation while stating the indicators for selecting a specific aspect	A mathematical mistake based on unit inconsistency is explained in more detail regarding the specific situation: the students used cm^3 and m^3 and forgot to adjust these units when calculating the ratio
3	To achieve level 3, one must reason about a selected aspect using general theoretical concepts related to teaching and learning mathematical modeling or draw conclusions	The aforementioned mistake is made during the phase of the modeling cycle "working mathematically" and leads to difficulties in interpreting the result and results in students' use of the metacognitive strategy "monitoring"

5.2.2 Sample and data analysis

We divided the participants who took part in the main study into two groups according to their competence level (Sample 3b, see Section 4.1). Qualitative data analysis (Kuckartz, 2014) was used to examine whether participants' reasoning was in line with the deductively coded level. Furthermore, we analyzed descriptive statistics to compare different competence levels. Using inferential statistics, the differences in participants' noticing competencies within a mathematical modeling context were analyzed. To check whether significant developments occurred, the Mann–Whitney U test was used because it requires only nonparametric data and does not require a normal distribution. However, it should be kept in mind that a small part of the sample (25%) was not independent, which reduced the explanatory power of the test.

5.3 Results

In this section, we take a closer look at the participants' answers to the open questions to discuss the consistency of reasoning in relation to the different coded levels. Exemplarily, this was done with regard to students' difficulties (see Table 1). One difficulty was related to students' unfamiliarity with modeling problems and became apparent in their confusion on how to start working on the task.

The following statement was classified as a level-1 interpretation:

> *"The students have a problem in understanding the task; they don't know how to work on the task."*

This comment indicates that the participant noticed the difficulty that the students encountered at the beginning of the modeling cycle. The problem, exhibited by the students in the video, is described exactly as it occurs and with references to students' thinking but without any further reasoning as to what may have caused the problem.

The following statement is an example of a level-2 interpretation:

"At the beginning, the students have problems understanding what the actual problem of the task is. They do not succeed in making correct assumptions, which they could use for calculations subsequently."

In contrast to the first quotation, here the participant provides an interpretation that includes reasoning regarding the students' behavior, namely that the students should have made assumptions to deal with the underdetermined modeling problem. Here, an understanding of specific procedures for dealing with modeling problems is evident and is used to interpret the observed difficulty. However, a broader interpretation that provides a reason for this difficulty is missing, and the desired approach to dealing with the modeling problems is not specified.

In contrast, the following statement is given at level 3:

"The students initially seem to have difficulties with the underdetermined nature of the task. At first, they are not aware that they have to identify and determine/research missing data. One indicator of this is that they divide the only given numbers without context."

The above statement is classified as level 3 as a reason for students' thinking or misunderstanding is given, its occurrence is described, and the underlying problem is detected. All in all, an increase in quality in pre-service teachers' interpretations could be recognized, which is illustrated by the coding of three levels of interpretations.

Figure 5 shows the distribution of the different levels of interpretation regarding students' difficulties, use of metacognitive strategies, and approaches to solving the modeling problem. Clearly, only the aspects that were perceived in the first place could be interpreted and subsequently assigned to an interpretation level.

Our analysis (see Fig. 5) showed that level 3 was achieved by only a small number of participants. Group 2 was assumed to perform at a higher level, and we did, indeed, observe differences in favor of Group 2.

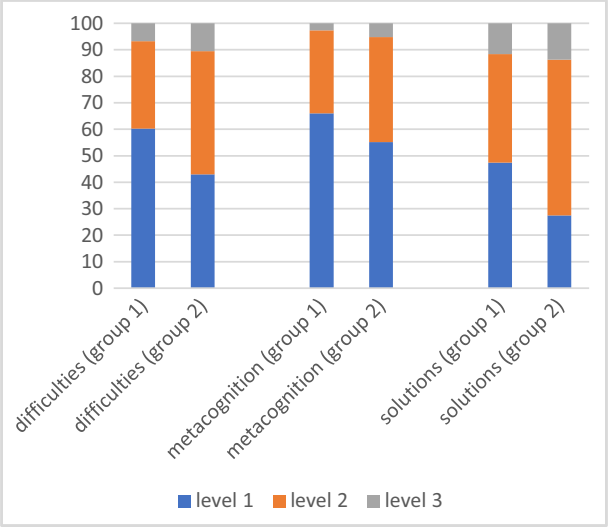

Fig. 5 The level of interpretation for different aspects in relation to different levels of competence

However, there was only a slight difference in the interpretation of students' metacognitive strategies. As metacognition is a difficult topic, it is very challenging to apply knowledge of metacognition to in-the-moment noticing.

The Mann–Whitney U test revealed significant differences in participants' abilities to interpret students' difficulties, which have to do with a lack of motivation ($p = 0.04$), an incomplete modeling process when an interpretation of the mathematical result was missing ($p = 0.04$), and the use of the metacognitive strategy of monitoring ($p = 0.0005$).

Overall, the results demonstrate that all levels of interpretation are achievable and that the differences in participants' ability to interpret can be represented. In addition, the instrument may be applicable to in-service teachers, as no ceiling effects were found.

6 Study 3: Construct validity

6.1 Research question

When it comes to construct validity, a test needs to assess the latent characteristic that it claims to measure. The results of the instrument should allow conclusions to be drawn about participants' characteristics (Moosbrugger & Kelava, 2012). The instrument was designed in a manner that made perception, interpretation, and decision-making closely intertwined. The open questions examined participants' interpretations based on their perceptions of students' actions or their decisions based on their perceptions and interpretations. Thus, it was not possible to merely perceive an event because an interpretation was always asked for. However, it was possible to deduce that an aspect had been perceived if an interpretation was presented. For decision-making, only one item evaluated at the nominal scale level was included in the test, as the instrument's focus was on perception and interpretation. Therefore, we analyzed the structure of the model (research question 3).

6.2 Method

Construct validity as defined above was checked by using a confirmatory factor analysis to check and verify the theoretically assumed structure. In this study, we considered $\chi 2$, df, RMSEA, and SRMR using the fit measures suggested by Schermelleh-Engel et al. (2003). Moreover, we used Cronbach's alpha as another criterion to check the instrument's reliability.

Perception and interpretation were closely intertwined due to the design of our instrument. Therefore, a confirmatory factor analysis was performed using data from the main study (see Sample 3b in Section 4.2) assuming one overall factor, "noticing competence," which included 14 items regarding perception and interpretation.

6.3 Results

The values (see Table 5) clearly indicated a good, or at least acceptable, fit according to the criteria provided by Schermelleh-Engel et al. (2003), with $\text{Chisq/df} < 2$, $\text{RMSEA} < 0.05$, and $\text{SRMR} < 0.08$, indicating a good fit, while the CFI was close to the cut-off value 0.95.

Table 5 Fit measures of the CFA with one factor; Chisq: chi-square goodness of fit; df: degrees of freedom; RMSEA: root mean square error of approximation; SRMR: standardized root mean residual; CFI: comparative fit index

Chisq/df	RMSEA	SRMR	CFI
84.03/77	0.029	0.074	0.93

The reliability of the instrument was thus examined by assuming one scale. Cronbach's alpha was relatively small ($\alpha = 0.63$). However, this value generally increases if the items are highly positively correlated and the number of items is increased (Bühner, 2011). Both influencing factors were disadvantageous in this data set. Some items were also negatively correlated. Therefore, a small value for Cronbach's alpha was a reasonable result. Schmitt (1996) also argues: "When a measure has other desirable properties, such as meaningful content coverage of some domain and reasonable unidimensionality, this low reliability may not be a major impediment to its use." (p. 352).

It was possible to increase Cronbach's alpha to 0.66 by dropping two items (*diff1* and *meta3*). However, these items were meaningful in terms of the content of the instrument and could not be omitted.

7 Summary, limitations, and future perspectives

Current competence measurements regarding the teaching of mathematical modeling mainly focus on teachers' dispositions; to date, researchers have paid little attention to teachers' noticing competencies in the context of mathematical modelling. We aimed to fill this research gap by developing and evaluating an instrument that uses staged videos as prompts for measuring teachers' noticing competencies. The instrument described in this article was developed to enable adequate and standardized measurements of pre-service teachers' noticing competencies within a mathematical modeling context.

Therefore, based on theoretical constructs regarding teachers' competencies for mathematical modeling and noticing, we developed a conceptualization of noticing within a mathematical modeling context (see Section 2.3). Based on empirical and theoretical findings, we identified the following aspects as noteworthy for noticing within a mathematical modeling context: students' modeling-specific difficulties, students' use of metacognitive strategies, and students' diverse approaches to solving a modeling problem. We adapted the model "competence as a continuum" (Blömeke et al., 2015) by including modeling-specific aspects (see Fig. 1). This conceptualization served as the basis for the development of the video-based instrument. The goal of the article was to examine the indicators for different kinds of validity of the instrument. To this end, we carried out three studies.

In Study 1, content validity was evaluated at different stages of the development process by examining whether all relevant content was covered by the instrument. Expert ratings and implementations with target groups resulted in various modifications of the instrument. The experts perceived all intended incidents in the staged videos while proposing changes to make sure that the incidents that were very hard to detect in the first version of the staged videos were emphasized more strongly in the second and final version. Therefore, all theoretically developed aspects were covered adequately (research question 1a).

Furthermore, we examined the focus of the open questions and made the questions more concise (research question 1b). However, this was done with only a small group of experts and more experts should have been involved. Lastly, after reducing the number of open questions, the complete instrument's implementation revealed a satisfactory variance in participants' ability to perceive the intended incidents and in the feasibility to perceive the maximum number of aspects (research question 1c). Significant differences between the two participant groups with different competence levels regarding the breadth of perception were only found in relation to metacognition. This may be due to Group 1 participants' lack of familiarity with this concept, as the participants received only a brief introduction to the concept and may have been incapable of applying this knowledge. Thus, the difference between the two groups was especially noticeable in this regard. Overall, the evidence for content validity was satisfactory.

Regarding elemental validity, Study 2 investigated whether the participants' reasoning, examined via written answers to open questions, was in line with the coded level of interpretation. Exemplarily, participants' interpretation of students' difficulties was analyzed, which showed that the quality of participants' reasoning resulted in different coded levels (research question 2a). The highest level of interpretation was reached by only a small number of participants (research question 2b). Achieving level 3 involved connecting a perceived aspect with an interpretation and referring to broader principles of teaching and learning. Many of the interpretations were classified at level 1. This finding was appropriate because the participants were pre-service teachers and were not expected to attain the maximum level of expertise after a short intervention. We found a few significant differences between the groups, which is in line with the different competence levels of the groups. In general, consensual coding yielded the three levels of interpretation, which was in line with participants' reasoning.

In Study 3, we analyzed the indicators for construct validity. We used one scale due to the design of the instrument. The general model showed an acceptable fit, meaning that the model adequately described the structure of the data (research question 3). The instrument was designed in such a way that perception and interpretation were closely intertwined, and decision-making was only a small part of the instrument. As Stahnke et al. (2016) reported, this is in line with the focus of most video-based instruments. Therefore, if the instrument were to be revised, it would be desirable to separate the different facets of perception, interpretation, and decision-making more clearly, perhaps even adding some closed items. Moreover, decision-making should be a more prominent part of a revised instrument.

In reference to the model of competence proposed by Blömeke et al. (2015), examining the knowledge and affective-motivational aspects of mathematical modeling would complement the results of our study and potentially reveal interesting influences, as shown by the TEDS-M research program (Kaiser & König, 2019; Kaiser et al., 2017) and the associated studies (Yang et al., 2020). Regarding mathematical modeling, Klock and Wess (2018) developed an instrument for measuring pre-service teachers' knowledge, beliefs, and self-efficacy that would complement the measurements of situational skills. Furthermore, as performance is part of the competence model, it would be worthwhile for future studies to examine performance. However, this is a rather difficult aspect to measure adequately.

It was possible to use our developed and validated instrument to analyze teachers' noticing competencies within a mathematical modeling context, which enables assessing

teachers' competencies in this area and measuring the effects of interventions. Thus, our instrument enables investigating trajectories of learning to notice within a mathematical modeling context and evaluating the efficiency of teacher training courses related to noticing competencies for teaching mathematical modeling. In this study, the instrument was applied to pre-service teachers. However, as the highest scores were reached only by a minority of the participants, the instrument may also be useful to assess experienced and in-service teachers. However, one should keep in mind that changes related to noticing are usually slow and require time (Berliner, 2001; Schoenfeld, 2011).

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Alwast, A., & Vorhölter, K. (2019). Enhancing future teachers' situation-specific modelling competencies by using staged videos. In U. T. Jankvist, M. van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education* (pp. 1120–1127). Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Amador, J. (2019). Noticing as a tool to analyze mathematics instruction and learning. In S. Llinares & O. Chapman (Eds.), *International Handbook of Mathematics Teacher Education: Vol. 2,* (pp. 311–336). Brill Sense.

Ball, D. L. (2011). *Foreword*. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), Mathematics teacher noticing: Seeing through teachers' eyes (pp. xx–xxiv). Routledge

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463–482.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies—Competence viewed as a continuum. *Zeitschrift Für Psychologie, 223*(1), 3–13.

Blum, W. (2015). Quality teaching of mathematical modeling: What do we know, what can we do? In S. J. Cho (Ed.), *The Proceedings of the 12th International Congress on Mathematical Education: Intellectual and Attitudinal Challenges* (pp. 73–96). Springer.

Borromeo Ferri, R. & Blum, W. (2010). Mathematical modeling in teacher education: Experiences from a modeling seminar. In V. Durand-Guerrier, S. Soury-Lavergne & F. Arzarello (Eds.), *CERME-6: Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education* (pp. 2046–2055). INRP.

Bragelman, J., Amador, J. M., & Superfine, A. C. (2021). Micro-analysis of noticing: A lens on prospective teachers' trajectories of learning to notice. *ZDM - Mathematics Education, 53*(1), 215–230.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. 3rd edition. Pearson Studium.

Choy, B. H. (2016). Snapshots of mathematics teacher noticing during task design. *Mathematics Education Research Journal, 28*(3), 421–440.

Dietiker, L., Males, L. M., Amador, J. M., & Earnest, D. (2018). Curricular noticing: A framework to describe teachers' interactions with curriculum materials. *Journal for Research in Mathematics Education, 49*(5), 521–532.

Dindyal, J., Schack, E. O., Choy, B. H., & Sherin, M. G. (2021). Exploring the terrains of mathematics teacher noticing. *ZDM - Mathematics Education, 53*, 1–16.

Erickson, F. (2011). On Noticing Teacher Noticing. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Eds.), *Mathematics Teacher Noticing. Seeing Through Teachers` Eyes* (pp. 17–34). Routledge.

Erickson, F., Boersema, D., Brown, M., Kirschner, B., Lazarus, B., Pelissier, C. & Thomas, D. (1986). *Teachers' practical ways of seeing and making sense: A final report*. East Lansing, MI: Institute for Research on Teaching/Washington, DC: Office of Educational Research and Improvement.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633.

Goos, M. (2002). Understanding metacognitive failure. *Journal of Mathematical Behavior, 21*(3), 283–302.

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers non-teachers and mathematicians. *Measurement: Interdisciplinary Research and Perspectives, 5*(2–3), 81–92. https://doi.org/10.1080/15366360701486999

Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional Noticing of Children's Mathematical Thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202.

Kaiser, G., & König, J. (2019). Competence Measurement in (Mathematics) Teacher Education and Beyond: Implications for Policy. *Higher Education Policy, 32*(4), 597–615.

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the complexities of video-based assessments: Theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education, 13*(2), 369–387.

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—Cognitive versus situated approaches. *Educational Studies in Mathematics, 94*(2), 161–182.

Kaiser, G., Brand, S. (2015). Modeling Competencies: Past Development and Further Perspectives. In G. A. Stillman, W. Blum & M.S. Biembengut (Eds.), *Mathematical Modeling in Education Research and Practice (*pp. 129–149). Springer.

Kaiser, G. (2007). Modeling and modeling competencies in school. In C. Haines, P. L. Galbraith, W. Blum & S. Khan (Eds.), *Mathematical modeling (ICTMA 12): Education, engineering and economics* (pp. 110–119). Horwood.

Kaiser, G. (2017). The Teaching and Learning of Mathematical Modeling. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 267–291). National Council of Teachers of Mathematics.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice, 23*(2), 198–211.

Kersting, N. B. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*(5), 845–861.

Klock, H. & Wess, R. (2018). *Lehrerkompetenzen zum mathematischen Modellieren - Test zur Erfassung von Aspekten professioneller Kompetenz zum Lehren mathematischen Modellierens*. MIAMI. Retrieved from https://nbn-resolving.org/urn:nbn:de:hbz:6-36179562140

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. Sage.

Leiß, D. (2007). *„Hilf mir es selbst zu tun". Lehrerinterventionen beim mathematischen Modellieren*. Franzbecker.

Lissitz, R., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*, 437–448.

Llinares, S., & Chapman, O. (2020). *The Handbook of Mathematics Teacher Education: Volume 2: Tools and Processes in Mathematics Teacher Education*. BRILL.

Maaß, K. (2006). What are modeling competencies? *ZDM - Mathematics Education, 38*(2), 113–142.

Mason, J. (2002). *Researching your own practice: The discipline of noticing*. Routledge-Falmer.

Moosbrugger, H., & Kelava, A. (2012). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava, *Testtheorie und Fragebogenkonstruktion* (pp. 7–26). Springer.

Moreno, M., Sánchez-Matamoros, G., Callejo, M. L., Pérez-Tyteca, P., & Llinares, S. (2021). How prospective kindergarten teachers develop their noticing skills: The instrumentation of a learning trajectory. *ZDM - Mathematics Education, 53*(1), 57–72.

Niss, M., & Blum, W. (2020). *The Learning and Teaching of Mathematical Modeling*. Routledge.

Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM - Mathematics Education, 48*(1–2), 153–165.

Santagata, R., König, J., Scheiner, T., Nguyen, H., Adleff, A.-K., Yang, X., & Kaiser, G. (2021). Mathematics teacher learning to notice: A systematic review of studies of video-based programs. *ZDM - Mathematics Education, 53*, 119–134.

Schack, E. O., Fisher, M. H., & Wilhelm, J. A. (2017). *Teacher Noticing: Bridging and Broadening Perspectives, Contexts, and Frameworks*. Springer.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23–74.
Schilling, S. G., & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement: Interdisciplinary Research and Perspectives, 5*(2–3), 70–80. https://doi.org/10.1080/15366360701486965
Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353.
Schoenfeld, A. H. (2011). Noticing Matters. A Lot. Now What? In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Eds.), *Mathematics Teacher Noticing. Seeing Through Teachers` Eyes* (pp. 223–238). Routledge.
Schukajlow, S., & Krug, A. (2014). Do multiple solutions matter? Prompting multiple solutions, interest, competence, and autonomy. *Journal for Research in Mathematics Education, 45*(4), 497–533.
Sherin, M. G., Jacobs, & V. R., Philipp, R. A. (2011a*). Mathematics Teacher Noticing: Seeing Through Teachers' Eyes*. Routledge.
Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (2011b). Situating the Study of Teacher Noticing. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Eds.), *Mathematics Teacher Noticing: Seeing Through Teachers` Eyes* (pp. 3–13). Routledge.
Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: A systematic review of empirical mathematics education research. *ZDM - Mathematics Education, 48*(1–2), 1–27.
Stender, P. (2016). *Wirkungsvolle Lehrerinterventionsformen bei komplexen Modellierungsaufgaben*. Springer.
Stillman, G. A. (2011). Applying Metacognitive Knowledge and Strategies in Applications and Modeling Tasks at Secondary School. In G. Kaiser, W. Blum, R. Borromeo Ferri & G. A. Stillman (Eds.), *Trends in Teaching and Learning of Mathematical Modeling (ICTMA 14)* (pp. 165–180). Springer.
Stürmer, K., & Seidel, T. (2017). A standardized approach for measuring teachers' professional vision: The observer research tool. In E. O. Schack, M. H. Fisher, & J. A. Wilhelm (Eds.), *Teacher noticing: Bridging and broadening perspectives, contexts, and frameworks* (pp. 359–380). Springer.
Thomas, J. N. (2017). The Ascendance of Noticing: Connections, Challenges, and Questions. In E. O. Schack, M.H. Fisher & J.A. Wilhelm (Eds.), *Teacher Noticing: Bridging and Broadening Perspectives, Contexts, and Frameworks* (pp. 507–514). Springer.
Van Es, E. A., & Sherin, M. G. (2021). Extending on prior conceptualizations of teacher noticing. *ZDM - Mathematics Education*, 53, 17–27.
Van Es, E. A., Tekkumru-Kisa, M., & Seago, N. (2019). Leveraging the power of video for teacher learning: A design framework for mathematics teacher educators. In S. Llinares & O. Chapman (Eds.), *International Handbook of Mathematics Teacher Education: Volume 2* (pp. 23–54). Brill Sense.
Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10*(4), 571–596.
Van Es, E. (2011). A framework for learning to notice student thinking. In M. G. Sherin, V. R. Jacobs & R. A. Philipp (Eds.), *Mathematics teacher noticing. Seeing through teachers' eyes* (pp. 134–151). Routledge.
Vorhölter, K. (2018). Conceptualization and measuring of metacognitive modeling competencies: Empirical verification of theoretical assumptions. *ZDM - Mathematics Education, 50*(1–2), 343–354.
Vorhölter, K. (2019). Enhancing metacognitive group strategies for modeling. *ZDM - Mathematics Education, 51*(4), 703–716.
Vorhölter, K., & Kaiser, G. (2016). Theoretical and Pedagogical Considerations in Promoting Students' Metacognitive Modeling Competencies. In C. R. Hirsch & A. Roth McDuffie (Eds.), *Annual Perspectives in Mathematics Education: Mathematical Modeling and Modeling Mathematics* (pp. 273–280). National Council of Teachers of Mathematics.
Yang, X., Kaiser, G., König, J., & Blömeke, S. (2018). Measuring Chinese teacher professional competence: Adapting and validating a German framework in China. *Journal of Curriculum Studies, 50*(5), 638–653.
Yang, X., Kaiser, G., König, J., & Blömeke, S. (2020). Relationship Between Chinese Mathematics Teachers' Knowledge and Their Professional Noticing. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-020-10089-3