

Investigating How German Biology Teachers Use Three-Dimensional Physical Models in Classroom Instruction: a Video Study

Sonja Werner¹ · Christian Förtsch¹ · William Boone² ·
Lena von Kotzebue¹ · Birgit J. Neuhaus¹

Published online: 21 July 2017

© Springer Science+Business Media Dordrecht 2017

Abstract To obtain a general understanding of science, model use as part of National Education Standards is important for instruction. Model use can be characterized by three aspects: (1) the characteristics of the model, (2) the integration of the model into instruction, and (3) the use of models to foster scientific reasoning. However, there were no empirical results describing the implementation of National Education Standards in science instruction concerning the use of models. Therefore, the present study investigated the implementation of different aspects of model use in German biology instruction. Two biology lessons on the topic neurobiology in grade nine of 32 biology teachers were videotaped ($N = 64$ videos). These lessons were analysed using an event-based coding manual according to three aspects of model described above. Rasch analysis of the coded categories was conducted and showed reliable measurement. In the first analysis, we identified 68 lessons where a total of 112 different models were used. The in-depth analysis showed that special aspects of an elaborate model use according to several categories of scientific reasoning were rarely implemented in biology instruction. A critical reflection of the used model ($N = 25$ models; 22.3%) and models to demonstrate scientific reasoning ($N = 26$ models; 23.2%) were seldom observed. Our findings suggest that pre-service biology teacher education and professional development initiatives in Germany have to focus on both aspects.

Keywords Models · Scientific reasoning · Biology instruction · Quantitative video analysis

This research was supported by grants (no. 01JH0904) from the German Federal Ministry of Education and Research (BMBF).

✉ Sonja Werner
s.werner@bio.lmu.de

¹ Biology Education, Department of Biology I, Ludwig-Maximilians University Munich, Winzererstraße 45/II, 80797 Munich, Germany

² Department of Educational Psychology, Miami University, Oxford, OH, USA

Since the publication of the results of the Third International Mathematics and Science Study (TIMSS) 1995 (Baumert et al. 2000) and the Programme for International Student Assessment (PISA) 2000 (Baumert et al. 2001), national education standards in science in several countries have been introduced and have become increasingly important. Scientific reasoning has since been included as one important aspect in these national education standards (e.g. Department for Education and Skills and Qualification and Curriculum Authority [DfEaS&Q] 2004 in England; Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK¹] 2005 in Germany; National Research Council [NRC] 1996 in USA). In this context, models represent an important part of scientific reasoning (Kremer et al. 2012; Mayer 2007). When referring to the model use in instruction, the National Education Standards for England state that students should learn to 'use scientific ideas and models to explain phenomena and events' (DfEaS&Q 2004 p. 72). In the USA, the framework for K–12 Science Education states that students should be able to "make and use a model to test a design, or aspects of a design, and to compare the effectiveness of different design solutions" (NRC 2012 p. 58; NGSS Lead States 2013). Similarly, the German National Education Standards state that students should use models to illustrate relations between structures and functions of objects, to analyse interactions by using models and to evaluate the meaning of models (KMK 2005).

In science education, students' scientific reasoning can be fostered in three different ways: learning science, learning about science and learning how to do science (Hodson 1992). In this regard, models can be used to understand and explain phenomena in nature and to transmit this understanding (Henze and van Driel 2011; Justi and Gilbert 2002b). To get an understanding of scientific reasoning, Justi and Gilbert (2002b) proposed that students should learn to produce and revise models, to learn about major scientific/historical models and to learn about the model used in science and science instruction.

In the following sections, we describe definitions and aspects of model use in science and science instruction, their implementation in science instruction and their importance for scientific reasoning.

Definition and Characteristics of Models in Science and Science Instruction

In the current research literature, there is no common definition of models (Van Driel and Verloop 1999). Gilbert (1994) defined models as "simplified depictions of a reality-as-observed, produced for specific purpose, to which the abstractions of theory are then applied" (Gilbert 1994 p. 116). Giere (2004) pointed out that scientists use "models to represent aspects of the world for specific purposes" (Giere 2004 p. 742). On a more abstract level, models can be seen as "external representations of mental concepts" (Krajcik and Merrit 2012 p. 38). Another definition is that a "model is a simplified representation of a phenomenon, which concentrates attention on specific aspects of it" (Ingham and Gilbert 1991 p. 193). Nevertheless, models can be described in a simple way as "representations of objects, phenomena, processes, ideas and/or their systems" (Gilbert and Boulter 2000 p. 7).

The processes of modelling are shown in the theoretical model of Steinbuch (1977), in which the representational reality as real objects are illustrated first by mental models as awareness of an individual person and subsequently by the physical reality in terms of

¹ Acronym for Kultusministerkonferenz

physical models. Mental models which illustrate only the essential characteristics of the reality (Steinbuch 1977) can be understood as a person's understanding about models and modelling. When a mental model is implemented in the physical reality in terms of a physical model, it includes essential as well as non-essential characteristics of the reality (Steinbuch 1977; Upmeier zu Belzen 2013). Therefore, a model used in science and science instruction can appear in different ways. In the broadest sense, models can be categorized in physical and mental models or mental representations (Coll and Lajium 2011). Mental models are a specific type of mental representations which students "generate during cognitive functioning, and which has the special characteristic that it preserves the structure of the thing it is supposed to represent" (Vosniadou 1994 p. 48). In contrast, physical models are "concrete physical models such as scale models used in engineering design or architecture" (Coll and Lajium 2011 p. 5). Physical models can be further divided into two-dimensional and three-dimensional physical models (Upmeier zu Belzen 2013). Two dimensional models are visual models as diagrams, iconic representations or symbolic representations, whereas three-dimensional physical models are structural or functional models. Therefore, Upmeier zu Belzen (2013) and Collin and Ferguson (1993) subdivided models in three main categories according to their aspect of illustration: structural, functional and dynamic models. Each type of model has its own specific purpose (White et al. 2011). Structural models represent aspects of real objects in a realistic way (Upmeier zu Belzen 2013) and emphasize certain aspects and their relationship in the model (White et al. 2011). Functional models show processes and causal relationships between different elements in the model (Upmeier zu Belzen 2013; White et al. 2011). Dynamic models can be used to test different assumptions about the reality by analysing different processes (White et al. 2011). To decide which model to use in instruction, teachers need to evaluate different models regarding their fitting to the learning goal. It includes a number of characteristics that are described in the literature (Pluta et al. 2011; White et al. 2011). Accordingly, models are as accurate as possible about the real object: they are general and include as many aspects of the real object as possible, they are parsimonious for conceptual coherence and clarity and they are coherent with other scientific evidence of other used models. One approach to use some of these criteria in instruction is to analyse a model regarding its level of complexity and abstraction. In educational research, the level of complexity describes content-related and cognitive structures (Kauertz et al. 2010). Students' cognitive activation increases with the use of higher complexity (Fischer et al. 2007), which can be described by the number and intensity of connections between content elements (Sweller 1994). Kremer et al. (2012), therefore, proposed a model with five levels of complexity, which Wadouh et al. (2014) evaluated for the subject biology. Wadouh et al. also reduced the scale of complexity to three levels: *fact*, *relation* and *concept*. Besides different levels of complexity, models should connect their theory to a phenomenon (Oh and Oh 2011). Therefore, models can be divided into different levels of abstraction concerning their similarity to the real object. They can be an image of the real object and show a high degree of similarity, a focused representation or a theoretical reconstruction of the object, depicting a slight degree of similarity (Grosslight et al. 1991; Grünkorn et al. 2014; Justi and Gilbert 2003; Upmeier zu Belzen and Krüger 2010). In instruction, it is also important to choose the right model that fits the learning goal of the lessons. Students can learn effectively, when the gap—between the learning goal to attain and the provided teaching tools or the used models—is very small (cf. Vygotsky 1978).

Integration of Models in Science Instruction and Fostering Scientific Reasoning Using Models

In science instruction, teachers can use models for different purposes. They mainly use models to show dangerous and complex phenomena and thus to make these available for students (Harrison 2001). Therefore, researchers regard models as teaching tools to illustrate certain aspects (Gilbert et al. 2000; Upmeier zu Belzen 2013) and to transmit content knowledge (Leibold and Klautke 1999). In addition to illustration, models should be used for the process of scientific inquiry (Fleige et al. 2012a; Nowak et al. 2013). Scientific phenomena can be predicted and theories formulated through the use of several models in instruction (Fleige et al. 2012a; Nowak et al. 2013; Schwarz et al. 2009; Treagust et al. 2002; Upmeier zu Belzen and Krüger 2010). These aspects are also included in scientific reasoning of different National Education Standards (KMK 2005; NGSS Lead States 2013).

Besides the integration of models in scientific inquiry, high-quality biology instruction is characterized by critically reflecting on models and introducing them by providing explanations (Wüsten 2010). A focus should be on "critical reflection on the role and nature of models in science" (Henze et al. 2007, p. 105). Additionally, critical reflection on models is also an important part of *nature of science* (Mayer 2007). Therefore, teachers should highlight the importance of models for science education (Justi and Gilbert 2002b). If teachers consider these two aspects, students can develop an elaborate *understanding of models*. In this way of understanding, models are described as a changeable tool to test an idea and represent this idea in the best way. A few studies have already indicated that an understanding of models is fostered by constructing and reflecting on models (e.g. Baek et al. 2011; Schwarz et al. 2009). However, the results of Khan's (2011) study also showed that models are developed but not reflected on or changed in instruction. Nevertheless, students can only develop such an understanding of models when they work with models in the classroom (Grosslight et al. 1991).

Research on Understanding of Models and Model Use in Science

So far, the main focus of recent research has been on investigating students' understanding of models and the model use in science (e.g., Crawford and Cullin 2005; Grosslight et al. 1991; Grünkorn et al. 2014; Justi and Gilbert 2002b; Upmeier zu Belzen and Krüger, 2010). The study of Grosslight et al. (1991) is an important work in this field of research, in which high school students were interviewed about their understanding of models and their use in science. Grosslight et al. focused on five main aspects of models: *kind of models*, *purpose of models*, *designing and creating models*, *multiple models* and *changing models*. After analysing the interview data, Grosslight et al. identified three different levels of model understanding:

- Level 1:
Models are simple copies of reality and useful to illustrate real objects. Students have no idea, why aspects or structures are omitted. Thus, the idea behind the model will not be clarified to students.
- Level 2:
The construction of the model is oriented to a specific purpose or target. The model no longer has to be an exact copy of the real object. Models illustrate their relation to reality and the students begin to identify that the modeller has a specific idea, and why the model

has to be built in a specific way. However, the focus is on the relation between the model and the real object. Testing of models is not in scope.

- Level 3:

The model is developed to test ideas and not to illustrate the reality. The modeller actively constructs the model to evaluate it. Moreover, the model can be changed and improved in a cyclic constructive process. The focus of model construction work is on the idea, which should be tested by the model.

In their study, Grosslight et al. revealed that the majority of students only reached level 1 (67%), only a few reached level 2 (12%) and none had a level 3 understanding of models and their use in science. When teachers provide an elaborate understanding of models and modelling in science (in accordance with level 3) in their instruction, students have the possibility to develop such an understanding. On the basis of Grosslight et al.'s results, Justi and Gilbert (2003) interviewed 39 science teachers from different school types and categorized their understanding of models. Contrary to the five aspects of Grosslight et al., Justi and Gilbert generated seven aspects of understanding models: *nature, use, entities, uniqueness, time, prediction* and *accreditation*. Justi and Gilbert did not find any support for the three levels of Grosslight et al. in their data. However, their results indicated that teachers have "a 'naive' view of models" (Justi and Gilbert 2003, p. 1380) concerning the two aspects of models: their nature and use. Justi and Gilbert also found huge differences with respect to the educational backgrounds of the teachers: teachers with primary teaching certificates had the simplest understanding of models in science. Crawford and Cullin (2005) published a framework regarding the understanding of models based on a literature review. Crawford and Cullin identified five aspects of model understanding (*purpose of models, designing and creating models, changing models, validating/testing models* and *multiple models for the same thing*), for each of which they described three to four levels of understanding. Only three aspects, *kinds of models* (cf. Grosslight et al. 1991) and their *nature and entities* (cf. Justi and Gilbert 2003) were not mentioned in the modelling dimensions of Crawford and Cullin (2005). Based on Upmeier zu Belzen and Krüger (2010), we developed a theoretical framework to assess and investigate students' understanding of models and their use in science. They described the understanding of models in science as 'medial', an illustration of something, as well as 'methodical', an instrument for something (Grünkorn et al. 2014; Mahr 2009; Oh and Oh 2011). Upmeier zu Belzen and Krüger's theoretical framework comprised five aspects: *nature of models, purpose of models, multiple models, testing models* and *changing models*; for every aspect, three levels were defined. The levels represent the epistemological view of models as methods as well as products of science (Gilbert 1991; Mahr 2009). Recent empirical studies confirmed the increasing difficulty of the three levels in Upmeier zu Belzen and Krüger's theoretical framework by using force-choice and multiple-choice items (Krell 2012; Terzer 2013). Another empirical study by Grünkorn et al. also evaluated the framework of Upmeier zu Belzen and Krüger by using open-ended test items. The data of Grünkorn et al. confirmed three levels for the aspects *nature* and *purpose* of models but added an initial level of understanding to the other aspects. However, even in these two aspects, most of the students showed only a limited understanding of models and their use in science. Most of the previous studies indicated that students as well as teachers have a limited elaborate understanding of models and their use in science (e.g. Grosslight et al. 1991; Justi & Gilbert 2003). In contrast to these results, a few studies showed just the opposite (e.g. Chittleborough et al. 2005; Treagust et al. 2002). Chittleborough et al. showed that students can have a "scientifically acceptable understanding of the model concept" (p. 200) when the students connect the role of models to research tools. Different experiences which students had with models in instruction could be the reason for these contradictory results (Ingham and Gilbert 1991;

Treagust et al. 2002). Therefore, the use of models should be an essential part in science education (Grosslight et al. 1991).

In Germany, a main emphasis in science education research is on evaluating National Education Standards with a focus on scientific inquiry and scientific reasoning (e.g. Nowak et al. 2013; Wellnitz et al. 2012). These projects included the use of models. The projects Evaluation of the National Education Standards of Natural Science at the Lower Secondary Level (ESNaS) (Kremer et al. 2012) and Model of Cross-Linking Scientific Inquiry between Biology and Chemistry (VerE-model) (Nehring et al. 2011) both considered models as an important tool in science instruction. Therefore, the use and understanding of models become increasingly important in science education in Germany. In this regard, an elaborate understanding of models represents a key element for fostering students' general understanding of *Scientific Literacy*, including scientific reasoning (Meisert 2008).

Aims of the Recent Study

In summary, we know much about (1) students' and teachers' understanding of models and their use in science (cf. Grosslight et al. 1991), (2) how to operationalize the understanding of models by different frameworks (cf. Upmeier zu Belzen and Krüger 2010) and (3) how elaborate model use in science instruction should be like. To look at several aspects of elaborate model use in science instruction described above, we can summarize these individual aspects as three main aspects as shown in Fig. 1. These also should be considered by teachers. First, they have to select a model according to the aspect of illustration. This model should fit the specific teaching content by taking into consideration the level of abstraction and the level of complexity. The second aspect teachers have to consider is how to integrate the model in science instruction. Therefore, teachers need to use the model for scientific reasoning as its purpose, introduce the model to students and help them work with it. Third, teachers have to

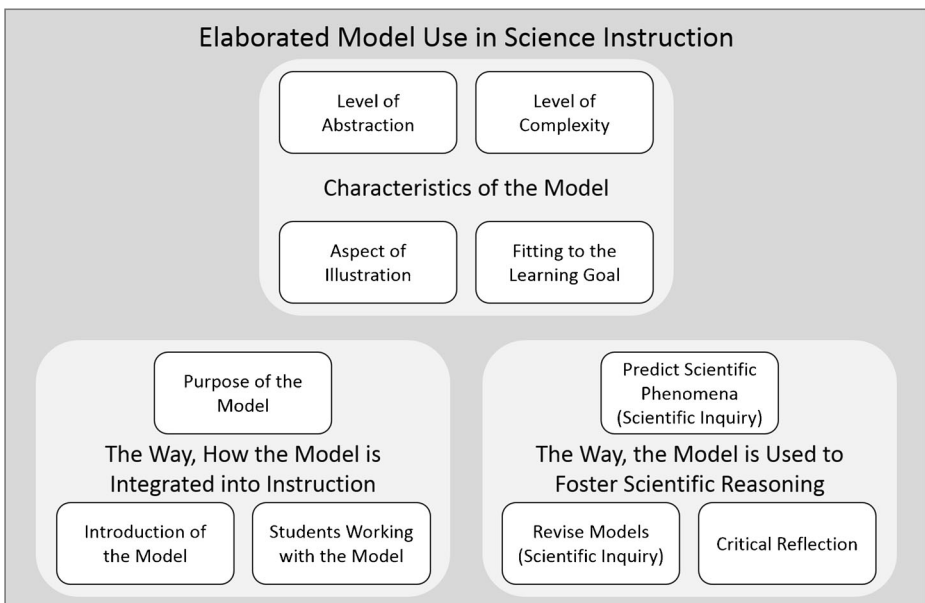


Fig. 1 Three aspects teachers should consider for an elaborate use of models in science instruction

predict scientific phenomena and revise models as part of scientific inquiry and reflect on them critically. Therefore, model use in science instruction is an important key element for acquiring a general understanding of scientific reasoning (Meisert 2008). However, there are no empirical results about how these different aspects of elaborate model use in science instruction are integrated in teaching practice. Previous studies only showed that models are rarely used in classroom practice (Dagher 1995; Justi & Gilbert 2002a, b; Treagust et al. 1992). A first attempt to close this gap was the study of Krell and Krüger (2013). On the basis of the competence model of Upmeyer zu Belzen and Krüger (2010), Krell and Krüger asked 146 German biology teachers how they implemented models in their instruction. Contrary to the results of the studies of Justi and Gilbert (2002a, b) and Khan (2011), Krell and Krüger found that teachers use models more elaborately and that they often reflected on the scientific use of models for scientific inquiry with students. As self-reported data are difficult to interpret and only an indication for the practical implementation, it is important to verify these results by using classroom observations (Krell and Krüger 2013).

Many studies in educational research focus on students' and teachers' understanding of models. However, the implementation of the use of models in German classrooms (cf. Krell and Krüger 2013) in accordance with the National Education Standards and the instruction—for understanding of models provided to students—have rarely been investigated. In order to close this research gap, direct observations of instruction through videos, in contrast to the study of Krell and Krüger (2013), were used in this study with the following aims:

1. To develop an objective, reliable and valid instrument to measure teachers' ability to use models elaborately, and therefore, to provide instruction for understanding of models to students
2. To describe the model use in German biology instruction and the implementation of different aspects of elaborate model use in science instruction
3. To measure which level of understanding of models and modelling will be provided to students in instruction using the three levels of Grosslight et al. (1991) as a prerequisite for the development of an elaborate understanding of models and modelling in science.

Additionally, the focus of the video analysis was on three-dimensional physical models, because the German National Education Standards in the competence area of scientific reasoning describe models as three-dimensional physical models, whereas diagrams as two-dimensional models are included in another competence area called *subject-specific communication* (KMK 2005).

Methods

This study was part of a cooperative project across different universities in Germany. The project aimed to analyse relations between teachers' professional knowledge, their instruction and students' outcomes. The present analysis draws on data from the domain biology and focuses on one feature of instructional quality: elaborate model use in teaching.

Participants in Recent Study

Forty-three biology teachers (60% female) of secondary schools (*Gymnasium* in German) in Bavaria, a federal state of Germany, participated in this study. On average, they were 35 years

old ($SD = 8$; $min = 25$ years, $max = 52$ years) with 6 years of teaching experience after the traineeship ($SD = 5.5$). Each participating teacher was videotaped for two neurobiology lessons in grade nine ($N = 85$ videos). The lessons were on average 42.46 min long ($min = 29.35$ min; $max = 84:34$ min; $SD = 7.34$).

Data Collection

Data collection involved four steps for each class:

Step 1. All participating biology teachers completed a professional knowledge test—with content knowledge (CK) and pedagogical content knowledge (PCK) items based on the study of Jüttner et al. (2013) and pedagogical knowledge (PK) items based on the study of Lenske et al. (2015)—during a professional development initiative.

Step 2. Before videotaping, students completed a pre-achievement test on the topic neurobiology focusing on reflex arc and a questionnaire on estimating their teachers' professional knowledge of different features of instructional quality (elaborate model use among other features).

Step 3. During the teaching unit neurobiology, two lessons were videotaped according to guided standards based on Seidel et al. (2005). The teaching unit neurobiology is part of the German curriculum in grade 9 and includes the topics *nervous system* and *sensory functions*. The first videotaped lesson of each teacher was on the topic *reflex arc*. For the second lesson, teachers could choose a lesson within the teaching unit neurobiology on their own. The selection of one specific content area made a comparison of teachers more valid. After each videotaped lesson, students completed a questionnaire on their situational interest.

Step 4. After the teaching unit neurobiology, post-achievement data and the interest in biology of each student were collected. Additionally, all the teachers completed a questionnaire on their professional knowledge of different features of instructional quality (elaborate model use among other features).

However, this paper refers only to Step 3 of the described data collection.

Description of Category System and Its Different Aspects for Analysing Model Use in Biology Instruction

A theoretically devised event-based category system was developed to reliably, validly and objectively measure elaborate model use in biology instruction. In the first step, we defined models. In our study, models are 'representations of objects, phenomena, processes, ideas and/or their systems' (Gilbert and Boulter 2000, p. 7). This definition comprised thereby structural and functional models as three-dimensional physical models and diagrams and symbols as two-dimensional physical models as well as mental models (Steinbuch 1977; Upmeier zu Belzen 2013). In our study, expert interviews were conducted with nine secondary school teachers about the type of models, which they use in their instruction. All interviewed teachers described the type of models, which they use in instruction, as three-dimensional physical models. This result was also in line with the German National Education Standards, which only refer to three-dimensional physical models in the context of scientific reasoning (KMK 2005). Additionally, on the basis of the model of Steinbuch (1977), it is only possible to capture and analyse validly the physical reality in terms of the physical models in the videos. Therefore, only three-dimensional models were included in our analysis. Examples for three-dimensional physical models in neurobiology are models of a synapse, an eye or a nerve cell.

Before coding the videos, we also defined the event *model use*. A model use event was defined as a phase of instruction where teachers or students use three-dimensional physical models to work out the content of lessons. An additional criterion was the active work with the model of a teacher or students. The term *active* is defined as students being stimulated by instruction to be constructive (Chi 2009), and consequently, cognitively active, in their learning.

The coding process was conducted separately in two steps. In the first step, we did an event-based basic coding using the program *Videograph* (Rimmele 2012) to identify all events of model use. In the second step, all identified events were analysed in depth. Therefore, the category system included three main aspects, which can be derived from theory and empirical studies, on how models should be used in science instruction (Collin and Ferguson 1993; Crawford and Cullin 2005; Fischer et al. 2007; Grosslight et al. 1991; Grünkorn et al. 2014; Justi and Gilbert 2003; Khan 2011; Klahr 2000; Mayer 2007; Nowak et al. 2013; Upmeier zu Belzen 2013; Upmeier zu Belzen and Krüger 2010; Wadouh et al. 2014; White et al. 2011; Wüsten 2010): (1) *characteristics of the model*, (2) *the way the model is integrated into instruction* and (3) *the way the model is used to foster scientific reasoning* (see Fig. 2). Furthermore, each aspect was divided into several categories to simplify the video analysis. In the following subsections, we describe each aspect and its categories and illustrated them with examples.

For analysing the aspect *characteristics of the model*, we identified four important categories: *level of abstraction*, *level of complexity*, *aspect of illustration* and *fitting to the learning goal* (for an overview of the categories and their operationalization, see Table 1). The category

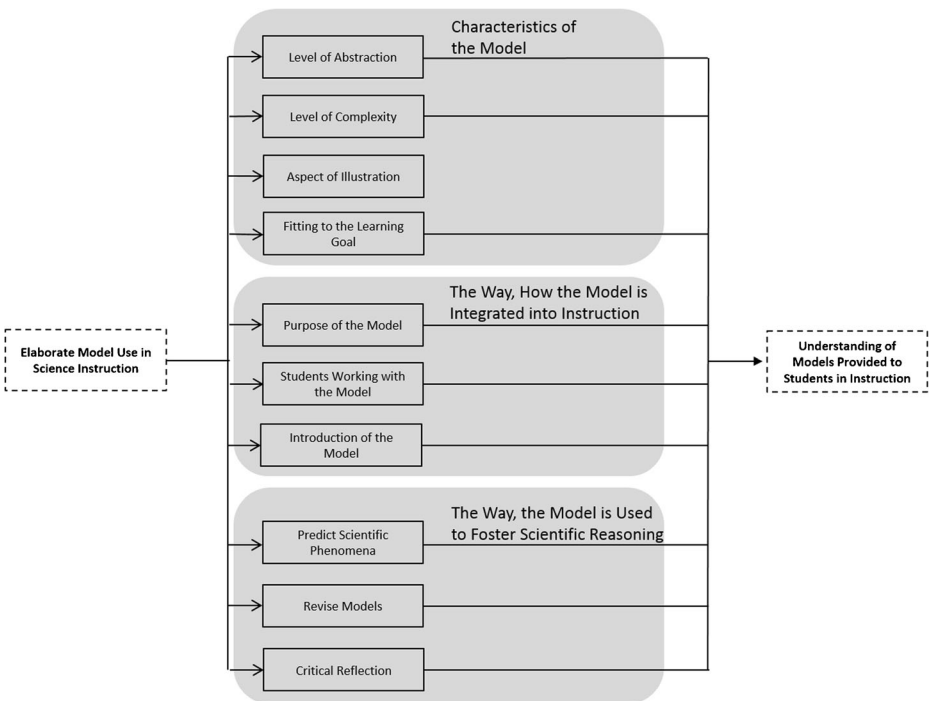


Fig. 2 An overview about the event-based category system including all coded categories. Categories with a *dashed frame* were not directly observed in the videos

Table 1 Description of the coding categories concerning the characteristics of the model

Category	Subcategory	Code	Description
Aspect of illustration	Structural model	1	Structural models illustrate anatomical and morphological aspects. The focus is placed on the structure and structural aspects of the object.
	Functional model	2	Functional models illustrate process and relations. The focus is placed on processes, functions and causal relations.
	Structural and functional model	3	Structural and functional models illustrate the structure of an object, and processes, functions and causal relations.
Level of complexity	Fact	0	The model imparts only facts.
	Relation	1	The model imparts relations. Single facts are connected.
	Concept	2	The model imparts a generic concept and connects different relations. The content of the lesson can be generalized or applied to a new situation.
Level of abstraction	Low abstraction	0	The model resembles the real object. It contains all structures of the real object. It is nearly a copy of the real object.
	Medium abstraction	1	The models resemble the real object only in certain elements. It contains only a part of the structures of the real object.
	High abstraction	2	The model contains only certain elements that are important to the content. The model has no or little similarity to the real object.
Fitting to the learning goal	No fitting to the learning goal	0	The complexity and the level of abstraction of the used model do not fit to the learning goal of the lesson.
	Fitting to the learning goal	1	The complexity and the level of abstraction of the used model fit to the learning goal of the lesson.

aspect of illustration makes a distinction between visualizing different aspects of a model. According to Upmeier zu Belzen (2013) and White et al. (2011), models can be classified into three types regarding the aspect of illustration as subcategories: *structural model*, *functional model*, or a combination of both. Thereby, structural models help to illustrate anatomical and morphological aspects of the real object. For example, the composition of a nerve cell can be illustrated by a structural model. In contrast, functional models focus merely on processes, functions and causal relations. When the model illustrates both aspects, the combined subcategory structural and functional model was used. The category *level of complexity* describes the content of the model according to its complexity. By higher complexity in instruction, students' cognitive activation in the classroom will be increased, and hence, students' learning can be fostered through higher cognitively activating instruction (Fischer et al. 2007; Förtsch et al. 2016a, b). Taking the work of Kauertz et al. (2010) and Wadouh et al. (2014) into account, three different levels (or subcategories) were used and adapted for our analysis: *fact*, *relation* and *concept*. A model of low complexity (or subcategory fact) enables only students to learn individual anatomical structures with no relations between them. By contrast, a high-complexity model can show students overarching scientific concept. The next aspect *level of abstraction* can be found in several frameworks and describes the degree of similarity of a model to the real object. At a low level of abstraction, the model can be seen as a copy of the real object. It contains all structures or aspects of the real object. A high level of abstraction means no or little similarity to the real object (Grosslight et al. 1991; Grünkorn et al. 2014; Justi and Gilbert 2003; Upmeier zu Belzen and Krüger 2010). Structures or aspects, which are irrelevant for learning the lesson content, are not illustrated by the used model. By disregarding several aspects, teachers can draw students' attention to the main aspects of the lesson. The fourth category *fitting to the learning goal* analysed the fitting between the used model and the

learning goal of the lesson. Therefore, the categories *level of abstraction* and *level of complexity* as described above were used. We operationalized the learning goal by analysing the materials used in instruction. Such fitting between the used model as a teaching material and the learning goal of the lesson is important to enable students to learn effectively and successfully (cf. Vygotsky 1978). For example, the learning goal of the lesson is to learn about the neuronal signalling in nerve cells. The teacher has to choose a model which shows on the one hand all structures needed for the signalling (*level of abstraction*) and on the other hand the relation between these structures (*level of complexity*).

The description of science instruction by using models was operationalized through the aspect *the way the model is integrated into instruction* and its categories *introduction of the model*, *students working with the model* and *purpose of the model* (see Table 2). According to Wüsten (2010), introducing models to students in instruction is one aspect of a high-quality biology instruction. Therefore, the category *introduction of the model* was included in our analysis. The category describes whether a teacher introduces a model and if so, with what intensity. This means that the teacher can introduce the model in an incidental or detailed way (Wüsten 2010). In addition to the instruction of a model, it is important that students are confronted with models in instruction and have an opportunity to work with them on their own (Grosslight et al. 1991; Justi and Gilbert 2003). Therefore, the category *students working with the model* includes whether students are working with a model. The work with models in instruction is necessary for students to acquire a comprehensive understanding of models (Grosslight et al. 1991). The category *purpose of the model* includes two applications of models in science instruction. On the one hand, models can be used to simplify and illustrate biological structures and functions. On the other hand, predictions and explanations of the teaching content can be made by models (e.g. Grosslight et al. 1991; Crawford and Cullin 2005; Grünkorn et al. 2014; Nowak et al. 2013; Upmeier zu Belzen and Krüger 2010). The formulation of these predictions and explanations is a main aspect of scientific reasoning and should be fostered through science instruction (cf. KMK 2005; Mayer 2007). Therefore, students' deeper understanding and processing of the content can be enabled (Sins et al. 2009; Smith et al. 2000). As

Table 2 Description of the coding categories concerning the way, how the model is integrated into instruction

Category	Subcategory	Code	Description
Introduction of the model	No introduction	0	There is no introduction of the model in the lesson.
	Short introduction	1	There is a short introduction of the model in the lesson.
	Detailed introduction	2	There is a detailed introduction in the lesson. The teacher highlights the importance of the model for the topic.
Students working with the model	No	0	Only the teacher works with the model.
	Yes	1	Teacher and students together or the students on their own work on a specific topic in the lesson using a model.
Purpose of the model	Illustration	0	The model is mainly used for the illustration of certain aspects of the object.
	Scientific reasoning	1	The model is mainly used for scientific inquiry. A hypothesis can be verified or a prediction can be made. Therefore, students can draw conclusion from the model and transfer them to the real object.

structures, functions or processes are not only illustrated by the model, but also further questions on scientific reasoning are asked, the subcategory *tool for scientific reasoning* was coded. Further questions on scientific reasoning can be questions about the changes of the structures in the model. However, Upmeier zu Belzen and Krüger (2010) stated that models are usually used for illustration.

The last aspect of the category system (*the way the model is used to foster scientific reasoning*) was based on several demands of the German and other national education standards (cf. DfEaS&Q 2004; KMK 2005; NRC 1996, 2012; NGSS Lead States 2013). Therefore, the aspect was divided into three categories: *predict scientific phenomena*, *revise models* and *critical reflection* (see Table 3). Developing scientific research questions or hypotheses and relating back to them are important aspects of scientific reasoning and should be part of scientific instruction (Klahr 2000; Mayer 2007). Thus, the category *predict scientific phenomena* describes whether a teacher refers models to scientific research questions or hypotheses formulated at the beginning of a lesson. Scientific research questions or hypotheses are scientific questions raised which can be solved by working with models. In this context, it is irrelevant whether the teacher or students formulated them. On the basis of this category, we analysed whether the teacher referred to the scientific research question or hypothesis at the end of the lesson (category *Revise Models*). It is important in scientific reasoning that teachers not only formulate hypotheses but also refer to them at the end of the lesson (Mayer 2007). Additionally, we included the category *critical reflection*. A critical reflection is an important part of *nature of science* (Mayer 2007) and a demand of the National Education Standards (cf. KMK 2005). By critical reflection of used models, students' understanding of models and modelling in science can be fostered (Baek et al. 2011; Schwarz et al. 2009). However, several authors are described that critical reflection is rarely done by teachers when using a model in instruction (Khan 2011; Upmeier zu Belzen & Krüger 2010). To analyse this category, we took into account the intensity as well as the number of aspects as subcategories.

Table 3 Description of the categories concerning the way the model is used to foster scientific reasoning

Category	Subcategory	Code	Description
Predict scientific phenomena	No formulation of scientific research question or hypothesis	0	Students or teachers do not formulate a scientific research question or hypothesis during the work with models in instruction.
	Formulation of scientific research question or hypothesis	1	Students or teachers formulate a scientific research question or hypothesis during the work with models in instruction.
Revise models	No regard to scientific research question or hypothesis	0	There is no regard to the scientific research question or hypothesis formulated at the beginning of the lesson.
	Regard to scientific research question or hypothesis	1	There is a regard to the scientific research question or hypothesis formulated at the beginning of the lesson.
Critical reflection	No critical reflection	0	There was no critical reflection during instruction.
	Incidentally	1	Teacher or students mention the limitation incidentally and name one or more aspects.
	Detailed	2	Teacher or students mention limitations in detail and name one or more than one aspect.

In our category system, the optimal critical reflection is detailed and focuses on more than one aspect of the model.

After the coding of the videos, the categories were exported to *SPSS 22* program for further analysis. First, descriptive data were calculated and described with frequency and percentage data.

Psychometric Analysis of the Analysed Category

In addition to the aforementioned data analysis steps, a psychometric analysis of the dataset was also conducted utilizing Rasch measurement techniques. The goal of the analysis was to refer the categories to teacher's ability to use models elaborately in instruction, utilize the data of the video coding which was conducted and to compute an overall teacher measures for further analyses. Additionally, the reliability of the category system could be verified in this way. For the psychometric analysis, we used the coded categories of the analysed videos on the teacher level.

In case we categorized more than one model use event per teacher, the codings for different categories were aggregated by using the highest observed measure. We assumed that the highest observed measure is most suitable for describing teachers' abilities to provide instruction for elaborate model use in instruction. As the category *aspect of illustration* was measured on a nominal level, we did not include this category in the psychometric analysis. Due to the time-consuming work of video coding, the dataset evaluated using Rasch techniques was small. Rasch analysis techniques were used in our study because these techniques are now widely understood to be necessary for the analysis of rating scale data (Wright and Masters 1982). Rating scale data is ordinal, and as a result, the non-linear rating scale data must be converted to a linear scale before statistical tests are conducted. Rasch analysis provides not only linear *person measures* but these person measures are expressed on the same scale as *item measures* (or category measures in our study) facilitating the construction of so-called Wright Maps, which can be used to explain data patterns in a meaningful manner to stakeholders (cf. Bond and Fox 2007; Boone et al. 2014). The analysis conducted in this study was completed utilizing the Rasch-model computer program *Winsteps* (Linacre 2012). The analysis of the dataset was conducted using a Rasch rating scale model, in which each item (category) was viewed as having a potentially different rating set of steps; for example, for the category level of complexity, the meaning of a change from a rating of a 1 to a 2 (for the two subcategories) is not viewed as the same change as a rating from a 1 to a 2 for the scale of level of abstraction).

Rasch analysis allows researchers to combine potentially items with different rating scales in a single analysis for the computation of a person measure as well as an item measure. Therefore, the reliability of the measurement can be verified by conducting a Rasch-analysis (for the technical details of the analysis steps, see the *Winsteps Manual* by Linacre 2012 and references on Rasch analysis in the human sciences, e.g. Boone et al. 2014). The person measure describes the probability to solve the provided task, or in our study, to work with models in a specific way, regardless of his specific person ability (Boone et al. 2014).

The person measures obtained by Rasch analysis for each lesson were averaged for each teacher. Therefore, at the end, we obtained one value per teacher for further analyses.

Description of Coding for the Understanding of Models Provided in Instruction

In addition to the event-based coding described above, we coded the videos for the understanding of models provided in instruction to students in the lesson to order to have a more in-depth analysis of

the model use in instruction. To provide an elaborate understanding of models in classroom, it is a prerequisite for students to develop such an understanding. Therefore, the coding was divided into three levels of understanding according to the classification of Grosslight et al. (1991). At level 1, a teacher only describes models as simple copies of the real object. For example, the teacher only named several morphological structures of the model but did not connect them among each other and/or to their functions. Level 2 was used in the coding if a teacher describes the relationship between the model and the real object with the first indication that the models are constructed for a specific purpose. For example, a process was illustrated by the model and the teacher emphasized that only the relevant structures were shown in the model for students' understanding of the content. The lesson phases at level 3 include the use of models which are built for a certain idea. A teacher emphasizes the idea that the model can be changed and evaluated for a certain purpose. For example, hypothesis raised can be tested by using models and afterwards, the model will be changed, if necessary. Level 3 model use also includes level 1 and level 2 model uses, but goes beyond the latter.

Additionally, the results of video coding using levels of Grosslight et al. (1991) were used for validation of our developed category system. Therefore, the means of our Grosslight et al. codings and the received person measures were calculated for each teacher and afterwards correlated using a Pearson correlation.

Objectivity of the Category System

To verify the objectivity of the results of video coding, 10% of the videos ($N = 9$ videos) were coded by two trained independent raters, and their percentage agreements and the Cohen's kappa coefficients were calculated for each category.

Results

In 85 videotaped lessons, we could identify 68 lessons which teachers or students actively worked with models. In total, 112 different models were used. In the following sections, we describe the quality criteria of our analysis and the model use in German biology instruction according to the developed aspects and their categories.

Quality Criteria of the Coding: Objectivity, Reliability and Validity

One aim of the rating of the lessons according to Grosslight et al.'s (1991) levels was to validate the results of our analysis of the model use in biology instruction by our category system. Each analysed phase of lesson in which teachers worked with a model was correlated with an overall rating of these phases according to levels of Grosslight et al. for the understanding of models and modelling in science using Pearson correlation. The results of the Pearson correlation showed a significant positive correlation between the mean of person measures of our category system and that of the codings according to Grosslight et al. ($r = 0.47$; $p = .002$). Therefore, it can be assumed that a valid measurement was carried out in our study.

To verify the objectivity of our coding, inter-rater agreement of two independent raters were calculated. The percentage agreement ranged between 85.1 and 100%, and the Cohen's κ ranged between 0.79 and 1.00 for all analysed categories (see Table 4). These values showed a satisfactory inter-rater agreement (Wirtz and Caspar 2002).

Table 4 Percentage agreement and Cohen's κ for two independent raters

Category	Percentage agreement	Cohen's κ
Aspect of illustration	93.3	.86
Level of complexity	90.0	.82
Level of abstraction	100.0	1.00
Fitting to the learning goal	95.0	.90
Introduction of the model	93.3	.81
Students working with the model	93.3	.90
Purpose of the model	93.3	.88
Predict scientific phenomena	93.9	.90
Revise models	86.7	.84
Critical models	93.3	.81
Understanding of models (Grosslight et al. 1991)	85.0	.79

Descriptive Results of Model Use in German Biology Instruction

The aspect *characteristics of the model* was operationalized by the categories *aspect of illustration*, *level of abstraction*, *level of complexity* and *fitting to the learning goal*. In our analysis, we categorized structural models (47.3%; $N = 53$) as well as functional models with the same frequency. However, structural-functional models were hardly be observed in our sample (4.5%; $N = 5$). With regard to the *level of abstraction*, models mostly had a low (54.3%; $N = 61$) or high level of abstraction (37.5%; $N = 42$) to the real object (48.2%; $N = 54$). In terms of the *level of complexity*, about half of the used models illustrated relations between single facts (48.2%; $N = 54$), and only 11.6% of models illustrated visualized concepts to student ($N = 13$). Facts were illustrated by 45 of the 112 analysed models (40.2%). In terms of the learning goals of the lessons, more than half of the teachers ($N = 75$; 67.0%) chose models which are fitted to the goals (see Table 5).

The aspect *the way the model is integrated into instruction* used in coding contained the categories *introduction of the model*, *purpose of the model*, and *students working with the model* (see Table 6). In 34% of the cases of model use ($N = 30.4$), teachers did not introduce the model before using it. The model was often introduced by a short sentence (44.6%; $N = 50$). Nearly a quarter of the used models were introduced in detail by the teacher (25.0%; $N = 28$). Furthermore, teachers enabled students to work with models ($N = 90$, 80.4%). For demonstrating phenomena by the teacher, only 19.6% of the models were used ($N = 22$). The main purpose of the analysed cases

Table 5 Descriptive statistics concerning the aspect characteristic of the model ($N = 112$)

Category	Subcategory	Total number (N)	Percentage
Aspect of illustration	Structural model	53	47.3
	Functional model	54	48.2
	Structural and functional model	5	4.5
Level of complexity	Fact	45	40.2
	Relation	54	48.2
	Concept	13	11.6
Level of abstraction	Low abstraction	61	54.5
	Medium abstraction	9	8.0
	High abstraction	42	37.5
Fitting to the learning goal	No fitting to the learning goal	37	33.0
	Fitting to the learning goal	75	37.0
Total number of 'model use' events		112	100.0

Table 6 Descriptive statistics concerning the aspect the way, how the model is integrated into instruction ($N = 112$)

Category	Subcategory	Total number (N)	Percentage
Introduction of the model	No introduction	34	30.4
	Short introduction	50	44.6
	Detailed introduction	28	25.0
Students working with the model	No	22	19.6
	Yes	90	80.4
Purpose of the model	Illustration	86	76.8
	Scientific reasoning	26	23.2
Total number of 'model use' events		112	100.0

of model use was to illustrate structures or functions of the real object (76.8%; $N = 86$). Teachers only asked questions which referred to scientific reasoning in a minority of the cases of model use (23.2%; $N = 26$).

For the aspect *the way the model is used to foster scientific reasoning*, three different categories were analysed: *predict scientific phenomena*, *revise models*, and *critical reflection* (Table 7). In the majority of the analysed cases of model use, scientific research questions or hypotheses was tested (59.8%; $N = 67$). However, in all the cases, the teacher formulated scientific research questions but no hypotheses. The teacher referred to the scientific research questions formulated in 52 of 67 cases (77.7%) during the model use or afterwards. For the 25 analysed cases, the models were just reflected critically in instruction (22.3%). When the models were reflected critically, their limitations were mentioned in detail for one or more aspects in only five cases.

In addition to calculation of percentages or frequencies of the coded categories, the Wright Map calculated by our psychometric analysis was an additional option to describe model use in biology instruction. As shown in Fig. 3, the Wright Map (Linacre 2012) displays the measures of the participating teachers on the left side of the vertical line. The nine rating scale items or categories used to compute a measure for each teacher are presented on the right side of the Wright Map. We computed these measures utilizing the ratings of each teacher for each item or category. Higher teacher measure results represented teachers who received higher raw scores

Table 7 Descriptive statistics concerning the aspect the way the model is used to foster scientific reasoning ($N = 112$)

Category	Subcategory	Total number	Percentage
Predict scientific phenomena ¹	No formulation of scientific research question or hypothesis	45	40.2
	Formulation of scientific research question or hypothesis	67	59.8
Revise models	No regard to scientific research question or hypothesis	15 ^a	22.3
	Regard to scientific research question or hypothesis	52 ^a	77.7
Critical reflection	No critical reflection	87	77.7
	Incidentally	20	17.9
	Detailed	5	4.4
Total number of 'model use' events		112	100.0

^a Total number refers to formulated scientific research questions or hypotheses

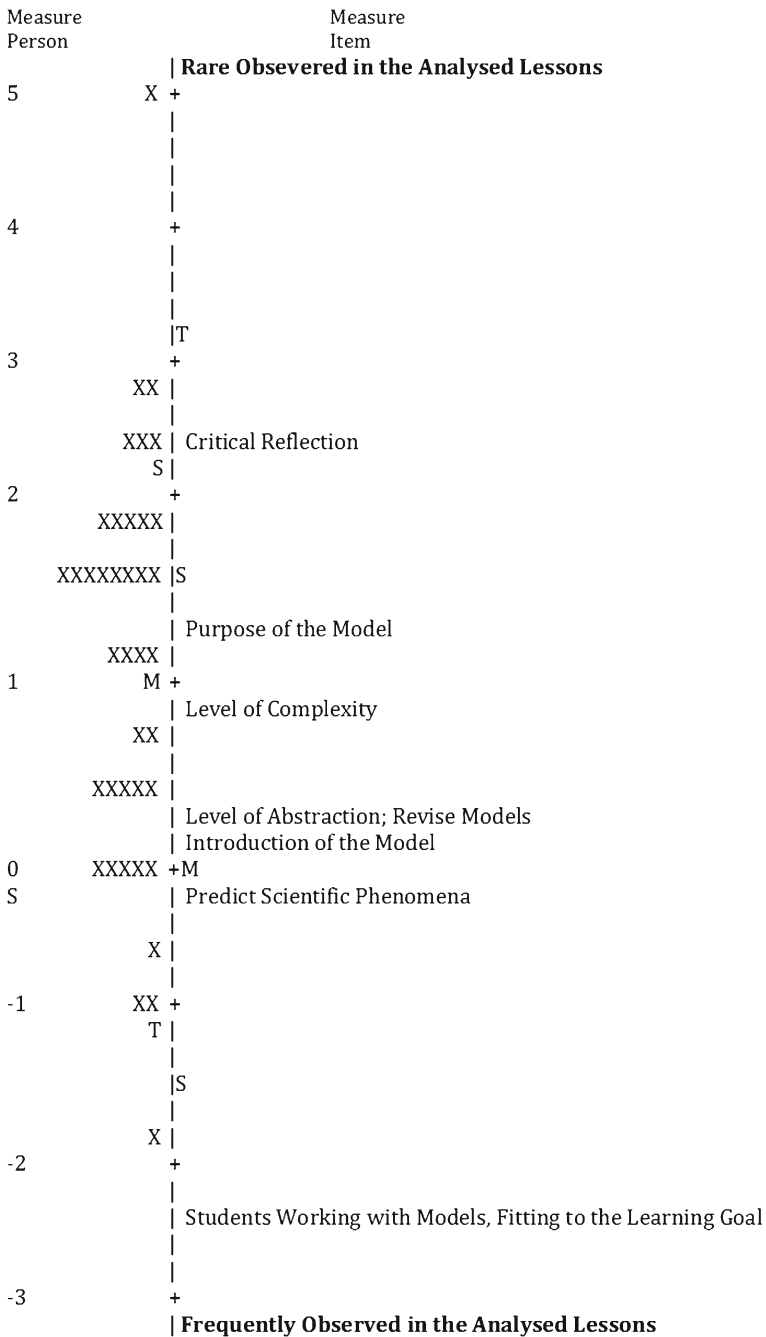


Fig. 3 Wright Map of elaborate model use in biology instruction. Person measures are plotted with an ‘X’ and item measures are plotted using the names of the categories

on the set of nine rating scale categories. This means that these teachers more often showed the specific behaviour in the coded categories in the classroom. Higher item measure results

represented items (categories) for which it was more difficult to be rated for a high raw score or not easy to observe in instruction. *Critical reflection* and *purpose of the model* in the context of scientific reasoning were rarely observed in the video data during model use, whereas *students working with the model* and the used models *fitting to the learning goal* were observed in nearly every lesson (see Fig. 3).

The Wright Map suggested that some coded categories in this study were too easy for the typical respondent (or the participating teacher). The categories *purpose of the model* and *critical reflection* showed the highest difficulty in coding or observing in instruction, whereas the categories *students working with the model* and *fitting to the learning goal* the lowest difficulty. A higher person measure indicates a more elaborate model use in biology instruction; and therefore, a teacher provides instruction for a higher understanding of models to students.

In order to analyse the understanding of models in instruction provided by the teacher, we additionally rated the level of understanding of models and modelling in science, as described by Grosslight et al. (1991). The majority of teachers fostered the understanding of models and modelling at level 2 ($N = 21$; 53.8). This means that their focus of the model use was on the relation between model and real object. Aspects to change or manipulate the used models according to an idea were neglected in instruction. Only four teachers (10.3%) provided an understanding of models as simple copies of the reality to their students. Over one third of the teachers provided an understanding of models in instruction at level 3 ($N = 14$; 35.9%); therefore, the idea behind the modelling process was imparted deliberately to students.

Understanding of Models Provided in Instruction to Students—Psychometric Results

When a Rasch analysis of a dataset is conducted, there are a number of steps, which are commonly taken. Below, we provide an overview of the key Rasch analysis results in this study to verify the reliability of our category system.

For our analysis, we utilized guidance provided by Linacre (2012) and other researchers to evaluate item fit. More specifically, items were evaluated with respect to a targeted MNSQ Outfit less than 1.3. All evaluated categories—considered as items in Rasch analysis—exhibited the MNSQ Outfit and Infit below 1.3 and therefore showed a productive measurement (Bond and Fox 2007). This result provided some evidence that the items or categories can be combined together for the computation of a person measure as well as an item measure.

Reliability is a common index reported in the science education literature. Rasch analyses facilitate the computation of both person reliability as well as item reliability (Boone et al. 2014). These values in a Rasch analysis are corrected for the non-linearity of raw ordinal data. In this study, the item measure reliability was 0.91 and the person measure reliability 0.62. Such values are not unexpected given the small number of items or categories being used to measure the participating teachers. The importance of such values is that the data suggests that one can have a level of confidence in the difficulty ordering of items or the categories, but with respect to person measures, one must be cautious as suggested by Boone et al. (2014). Linacre (1994) also discussed issues associated with sample size in Rasch analysis.

In summary, the results of the Rasch analysis suggested that the data from the teachers and the codings of the elaborate use of models by these teachers using the nine categories in the rating scale fit the Rasch model. Due to the limited number of items or categories used to evaluate each respondent, some caution should be taken when generalizing the results with

respect to the ‘measures’ of each of the teachers. Although a sample of teachers was utilized to compute item (category) measures, Rasch analysis results showed that there is a greater degree of certainty with regard to these measures.

Summary of Results

On the basis of the results, it can be assumed that model use in biology instruction was analysed objectively, validly and reliably using the developed category system. Therefore, conclusions about model use in instruction can be made.

The work with models in the topic neurobiology at secondary schools in our sample can be described as follows. The used models are either structural or functional models. These models give a detailed description of relations between several aspects on a low level of abstraction. Furthermore, the teacher chooses the model appropriate to the learning goal, which he/she wants to achieve in the lesson. In instruction, the teacher shortly introduces the model before working and enables his/her students to involve in the model work phases. The models are also used to illustrate several structural and functional aspects of the real object. Models are not only tools for illustration, but also used to foster students’ scientific reasoning. In working with models, scientific research questions are raised and referred to the models. However, the teacher gives hardly any attention to a critical reflection on models. Additionally, the relations between a model and its related real object are highlighted in instruction.

Unlike what has been assumed, models are not only used for illustration in instruction, but they are also seen as tools to foster students’ understanding of scientific reasoning.

Discussion

First, an objective, reliable and valid instrument to measure teachers’ ability to use models elaborately should be developed. The high value of inter-rater agreement in our study indicated an objective measurement of the model use in biology instruction by using the developed category system (cf. Wirtz and Caspar 2002). Furthermore, the validity of our measurement was verified, as we found a high correlation of our results with the rated levels of understanding of models according to Grosslight et al. (1991). Grosslight et al.’s classification provided the basis for this research in the field of understanding of models in science (e.g. Grünkorn et al. 2014; Justi and Gilbert 2003; Upmeier zu Belzen & Krüger 2010). Therefore, Grosslight et al.’s described levels of model use are a good indication for graduation in understanding of models and modelling in science. The results and fit indices of our Rasch analysis were used to verify the reliability of our measurement. Through Rasch analysis of video data, it was possible for us to use theoretical individual aspects in order to get one measurement for each teacher, and in this case, for the elaborate model use in biology instruction. All fit indices showed satisfactory values (cf. Bond and Fox 2007; Boone et al. 2014). Therefore, the category system in this study could be used for further analyses and to make statements about the model work in German biology instruction.

To foster students’ scientific literacy, the demands of the national educational standards include the work with models in science instruction (DfEaS&Q 2004; KMK 2005; NRC 1996, 2012; NGSS Lead States 2013). Models are not only considered to be copies of the reality but also to be tools for scientific reasoning (Fleige et al. 2012a; Nowak et al. 2013); however, until now, models have rarely been used in science instruction. If teachers use them, structures and

processes in science should be illustrated by means of models (Upmeier zu Belzen and Krüger 2010; Wüsten 2010). One aim of our study was to describe the model use in German biology instruction and how models are used according to scientific reasoning in instruction. The descriptive data of our analysis showed that models are mainly illustrated aspects of the reality; however, more than a quarter of the used models were found to foster students' scientific reasoning. These findings verify the results of Krell and Krüger (2013), which were self-reported data of teachers. Furthermore, formulating scientific research questions and referring back to them are important steps in scientific work (cf. Mayer 2007). Therefore, the understanding of science and its process are clarified. An emphasis lies on the idea behind the model and the possibility to change and manipulate models to test ideas (cf. Grosslight et al. 1991; Upmeier zu Belzen and Krüger 2010). This supports the development of an elaborate understanding of science by students (Meisert 2008). In contrast to formulating scientific research questions, critical reflection could hardly be observed in the analysed lessons in our study. The descriptive results were also verified by the analysis of the Wright Map (see Fig. 3). The Wright Map provides a summary of the ordering and spacing of categories from the most easy to the least easy to observe in the lessons. The ordering of items or categories from *easy* to *difficult* and the spacing of categories provides a picture of which categories or aspects of instruction during model use are easy or difficult to implement in biology instruction. In the Wright Map, critical reflection was also described as the category which was observed least in contrast to predict scientific phenomena and formulating scientific research questions. In the case teachers critically reflected on the model, they did it in an incidental way. The studies of Wüsten (2010) and Khan (2011) showed similar results. The critical reflection is considered as important part to foster students' elaborate understanding of models and modelling in science (Baek et al. 2011; Schwarz et al. 2009). Previous studies already indicated a positive influence of reflecting on models on students' understanding of models (Baek et al., 2011; Schwarz et al. 2009) and consequently on students' achievement. It is necessary that "the learning of scientific models [...] and the act of modelling [...] should go together with a critical reflection on the role and nature of models" (Henze et al. 2007, p. 105). Furthermore, students can develop misconceptions because they lack critical reflection of the used model and this has an adverse impact on their learning. Additionally, critical reflection of models is seen as an important instructional quality feature in biology (Wüsten 2010). Therefore, critical reflection has to be established in model work. Another instructional quality feature is the introduction of the used model (Wüsten 2010). Teachers in this study introduced models to their students; therefore, the importance of the model for learning the content as the learning goal of the lesson could be clarified to them. To facilitate students' learning, teaching material as models which are suitable for the teaching content should be provided to students (cf. Vygotsky 1978). Thus, models fitting to the learning goal of the lessons can support students' learning. Using such a model, students can reach the learning goal more easily (cf. Ergönenc et al. 2014). In the analysed lessons, the majority of the used models were fitted to the learning goal and, therefore, can facilitate the learning process of the students. Models used in the lessons mainly illustrated relations between several aspects of the real object. However, illustrating concepts to students was rarely implemented. Nevertheless, teaching concepts is important for cognitively activating students (Fischer et al. 2007) and increases students' understanding of connections between theory and phenomena (cf. Oh and Oh 2011). The studies of Förtsch et al. (submitted) and Wadouh et al. (2014) showed similar results according to the level of complexity of tasks in biology instruction. The analysed lessons of both studies were characterized by imparting to students single facts or relations between them. Our descriptive results also showed that

students were often involved in working with models. In the Wright Map (see Fig. 3), the aspect *students working with the model* was one of the most observed categories in our results. As a result, students could possibly have deeper processing of the learning content (Craik and Lockhart 1972). Thus, students' cognitive activation can be fostered and their learning improved (cf. Förtsch et al. 2016a, b; Kunter et al. 2013). Additionally, active working with models fosters students' understanding of models and their critical view of scientific phenomena (Grosslight et al. 1991; Lehrer and Schaub 2004; Schwarz and White 2005; Stewart et al. 2005). All these described aspects can contribute to developing an elaborate understanding of science among students.

To analyse the understanding of models provided to students in the analysed lessons by teachers according to the levels of Grosslight et al. (1991) was another aim of our study. Students can only develop an elaborate understanding of models and modelling in science (corresponding to level 3 of Grosslight et al. 1991) when such an understanding is imparted to students in instruction. An elaborate understanding of models and modelling in science also helps students to develop an elaborate understanding of science (cf. Meisert 2008). In contrast to many studies (e.g. Danusso et al. 2008; Harrison 2001; Justi and Gilbert 2002a; Van Driel and Verloop 1999, 2002), most of the teachers in our study provided instruction for an understanding of models at levels 2 or 3 to their students so that the idea behind the models and the relations between the models and reality could be shown to the students. That indicated that teachers can provide an elaborate understanding of models to their students. We also assume that teachers have an elaborate understanding of models, because "there is a strong relationship between what teachers know and think, and the way they teach" (Just and van Driel 2005, p.197). Therefore, our results support the students' perceptions about the role of models in learning science (Chittleborough et al. 2005).

In general, teachers in our study were able to provide an elaborate understanding of models to their students within instruction but had problems in implementing individual aspects of scientific reasoning in their model use (Fig. 4). When using models, teachers also taught generic concepts for a deeper understanding of the content. Considering the results of our descriptive analysis and the Wright map, we can describe which parts of elaborate model use that teachers found difficult in implementing in classroom practice. Teachers show deficits especially in the critical reflection. As a result, teachers have to be aware of the fact that critical reflection is an important aspect of model use. Therefore, we have to include this way of teaching in university education of pre-service teachers and professional development initiatives. We also have to give teachers examples about how they can implement models in scientific reasoning including critical reflection. In Germany, Fleige et al. (2012b) already suggested further approaches about this. They developed a workbook with a focus on using models in the context of scientific reasoning and described examples for different content areas of the curriculum.

There are also some limitations of this study. Only three-dimensional physical models were considered in our analysis according to the classification of the German National Education Standards (cf. KMK 2005). However, two-dimensional models such as diagrams should also be considered in further analysis. In Germany, several studies already dealt with the implementation of diagrams in instruction (e.g. Lachmayer et al. 2007); however, diagrams were not considered in the context of models and modelling. It would be interesting to know if results about two-dimensional model use in instruction are comparable to the results described in this paper. Another limitation was that mental models were not to be taken into account in our analysis. In our video studies, it was not possible to validly analyse the awareness of students or their mental models "which individuals

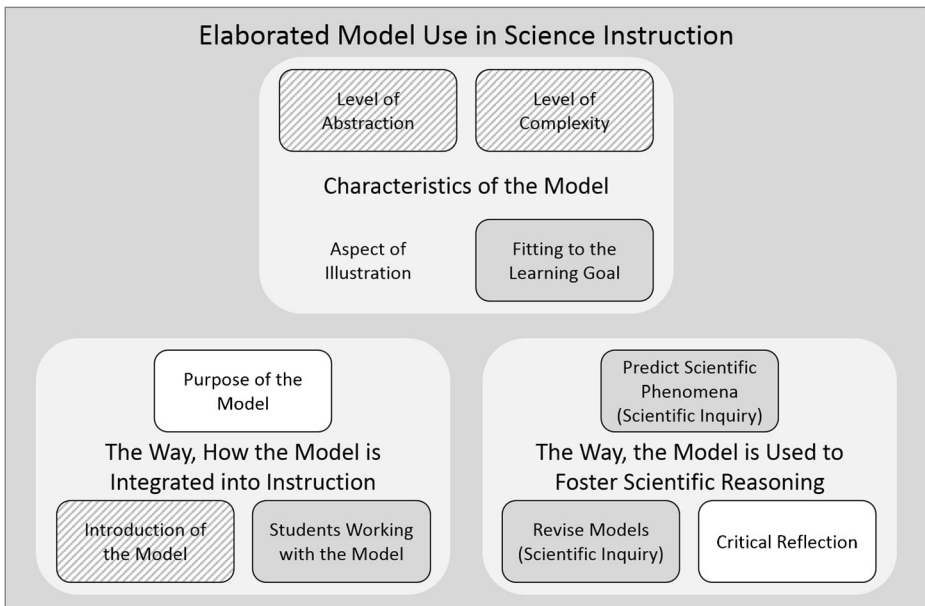


Fig. 4 Model for the elaborate model use in science instruction including the integration of our results. *Note.* Aspects in *grey boxes* are already implemented in biology classroom. Aspects in *grey-white dashed boxes* are to a small extent, whereas aspects in *white boxes* are nearly not putted into practice. The nominal category *aspect of illustration* was not included and therefore not highlighted

generate during cognitive functioning" (Vosniadou 1994, p. 48; cf. Steinbuch 1977). Through instruction, students develop an understanding of models and modelling through mental models. Consequently, teachers have to provide an elaborate understanding of models and modelling in their instruction as a pre-requisite. In this study, the physical reality in terms of physical models (cf. Steinbuch 1977) could only be analysed through the videotaped lessons. In videos, mental models of students or teachers are not observable. To measure mental models of students' awareness, interviews and questionnaires about their mental models have to be conducted in further studies and the results included in the instructional analyses. As a further limitation, all coded categories were only on a descriptive level. Further in-depth analyses of the aspects of model use are necessary to find out problems teachers may have in implementing scientific reasoning or critical reflection on models in their instruction. By using a theory-based coding manual and Rasch analysis in this study, we measured the level of model understanding provided in the teachers' instruction to their students. However, there are other ways to measure the features of instructional quality (Helmke 2003). To get an extensive view on this aspect, perceptions of students and teachers should also be taken into account. Some studies indicated positive correlations between the different perspectives—teachers, students and external observers—but they only focused on general aspects of instructional quality such as classroom management (Clausen 2002; Pauli 2012).

To provide an elaborate understanding of models and modelling in science, teachers also have to have such an understanding. Therefore, further analyses should concentrate on teachers' preconditions for developing of an elaborate understanding of models and modelling, and consequently, using models in an elaborate way in their instruction. These preconditions can also include different

dimensions of professional knowledge, one dimension of which, pedagogical content knowledge, is already linked to several teaching strategies (Shulman 1986). One such teaching strategy in science using models in instruction was described by Van Driel et al. (1998).

The main emphasis of science instruction is to foster students' understanding of science. Instruction has to provide a general understanding of scientific reasoning as part of scientific literacy. Through an elaborate understanding of models and modelling, the understanding of scientific reasoning can be fostered (Meisert 2008). Therefore, the effects of an elaborate understanding of models and modelling as well as elaborate use of models in instruction on student outcome variables, such as achievement or interest in science, should be analysed in further studies.

Acknowledgements We are grateful to the German Federal Ministry of Education and Research for supporting our study.

References

- Baek, H., Schwarz, C., Chen, J., Hokayem, H., & Zhan, L. (2011). Engaging elementary students in scientific modeling: the MoDeLS fifth-grade approach and finding. In M. S. Khine & I. M. Saleh (Eds.), *Models and modeling: cognitive tools for scientific enquiry* (pp. 195–220). Heidelberg: Springer.
- Baumert, J., Bos, W., & Lehmann, R. (eds.). (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* [TIMSS/III. Third International Mathematics and Science Study: mathematics and science education at the end of the school career]. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., et al. (Eds.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* [PISA 2000: basic competences of students in an international comparison]. Opladen: Leske + Budrich.
- Bond, T., & Fox, C. (2007). *Applying the Rasch model: fundamental measurement in the human sciences*. Mahwah, NJ: LEA.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.
- Chi, M. T. H. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73–105.
- Chittleborough, G. D., Treagust, D. F., Mamiala, T. L., & Mocerino, M. (2005). Students' perceptions of the role of models in the process of science and in the process of learning. *Research in Science and Technological Education*, 23(2), 195–212.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? [Instructional quality: a question of perspective?]*. Waxmann: Münster.
- Coll, R. K., & Lajium, D. (2011). Modeling and the future of science learning. In M. S. Khine & I. M. Saleh (Eds.), *Models and modeling: cognitive tools for scientific enquiry* (pp. 3–22). Heidelberg: Springer.
- Collin, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: structures and strategies for guiding inquiry. *Educational Psychologist*, 28(1), 25–42.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. doi:10.1016/S0022-5371(72)80001-X.
- Crawford, B., & Cullin, M. (2005). Dynamic assessments of pre-service teachers' knowledge of models and modelling. In K. Boersma, H. Eijkelhof, M. Goedhart, & O. Jong (Eds.), *Research and the quality of science education* (pp. 309–323). Dordrecht: Springer.
- Dagher, Z. R. (1995). Analysis of analogies used by science teachers. *Journal of Research in Science Teaching*, 32(3), 259–270.
- Danusso, L., Testa, I., Sassi, E., & Vicentini, M. (2008). Teachers' ideas about scientific models and modelling. In E. van den Berg, T. Ellermeijer, & O. Slooten (Eds.), *Modelling in physics and physics education* (pp. 952–957). Amsterdam: University of Amsterdam.

- Department for Education and Skills & Qualification and Curriculum Authority. (2004). *Science—The National Curriculum for England*. London: HMSO.
- Ergönenc, J., Neumann, K., & Fischer, H. E. (2014). The impact of pedagogical content knowledge on cognitive activation and students learning. In H. E. Fischer, P. Labudde, K. Neumann, & J. Viiri (Eds.), *Quality of instruction in physics* (pp. 145–160). Münster: Waxmann.
- Fischer, H. E., Glemnitz, I., Kauertz, A., & Sumfleth, E. (2007). Auf Wissen aufbauen—kumulatives Lernen in Chemie und Physik [To build on knowledge—cumulative learning in chemistry and physics]. In G. Kircher & Häußler (Eds.), *Physikdidaktik, Theorie und praxis [Physics education, theory and teaching practice]* (pp. 657–678). Berlin, Heidelberg, New York: Springer.
- Fleige, J., Seegers, A., Upmeier zu Belzen, A., & Krüger, D. (2012a). Förderung von Modellkompetenz im Biologieunterricht. [Fostering model competence in biology education]. *Der mathematische und naturwissenschaftliche Unterricht*, 65(1), 19–28.
- Fleige, J., Seegers, A., Upmeier zu Belzen, A., & Krüger, D. (2012b). *Modellkompetenz im Biologieunterricht Klasse 7–10: Phänomene begreifbar machen - in 11 komplett ausgearbeiteten Unterrichtseinheiten* [Model competence in biology instruction grade 7–10: making phenomena tangible—11 complete developed teaching units]. Auer Verlag.
- Förtsch, C., Werner, S., Dorfner, T., von Kotzebue, L., & Neuhaus, B. J. (2016a). Effects of cognitive activation in biology lessons on students' situational interest and achievement. *Research in Science Education*. doi:10.1007/s11165-016-9517-y.
- Förtsch, C., Werner, S., von Kotzebue, L., & Neuhaus, B. J. (2016b). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. doi:10.1080/09500693.2016.1257170.
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71, 742–752. doi:10.1086/425063.
- Gilbert, J., & Boulter, C. (2000). *Developing models in science education*. Dordrecht: Kluwer.
- Gilbert, J. (1994). Models and modeling: routes to more authentic science education. *International Journal of Science and Mathematics Education*, 2, 115–130.
- Gilbert, J., Boulter, C., & Elmer, R. (2000). Positioning models in science education and in design and technology education. In J. Gilbert & C. Boulter (Eds.), *Developing models in science education* (pp. 3–17). Dordrecht: Kluwer.
- Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1), 73–78. doi:10.1002/tea.3660280107.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822. doi:10.1002/tea.3660280907.
- Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*. doi:10.1080/09500693.2013.873155.
- Harrison, A. G. (2001). How do teachers and textbook writers model scientific ideas for students? *Research in Science Education*, 31, 401–435.
- Helmke. (2003). Unterrichtsevaluation [Evaluating educational practice]. *Schulmanagement*, 1, 8–11.
- Henze, I., & Van Driel, J. H. (2011). Science teachers' knowledge about learning and teaching models and modeling in public understanding of science. In M. S. Khine & I. M. Saleh (Eds.), *Models and modeling: cognitive tool for scientific inquiry* (pp. 239–261). Heidelberg: Springer.
- Henze, I., van Driel, J. H., & Verloop, N. (2007). Science teachers' knowledge about teaching models and modelling in the context of a new syllabus on public understanding of science. *Research in Science Education*, 37, 99–122.
- Hodson, D. (1992). In search of meaningful relationship: an exploration of some issues relating to integration in science and science education. *International Journal of Science Education*, 14, 541–562.
- Ingham, A., & Gilbert, J. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, 13, 193–202.
- Justi, R. S., & Gilbert, J. K. (2002a). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education*, 24(4), 369–387. doi:10.1080/09500690110110142.
- Justi, R. S., & Gilbert, J. K. (2002b). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education*, 24, 1273–1292.
- Justi, R. S., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education*, 25(11), 1369–1386. doi:10.1080/0950069032000070324.

- Justi, R., & van Driel, J. H. (2005). A Case Study of the Development of a Beginning Chemistry Teacher's Knowledge about Models and Modelling. *Research in Science Education*, 35(2-3), 197–219. doi:10.1007/s11165-004-7583-z.
- Jüttner, M., Boone, W., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, 25(1), 45–67. doi:10.1007/s11092-013-9157-y
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I [Modeling competence according to standards for science education in secondary schools]. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 135–153.
- Khan, S. (2011). What's missing in model-based teaching. *Journal of Science Teacher Education*, 22, 535–560.
- Klahr, (2000). *Exploring science: the cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [Conference of the Ministers of Education in Germany]. (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004 [Educational standards for the subject biology for intermediate-level education. By order of 16 December 2004]*. München: Wolters Kluwer.
- Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: what does constructing and revising models look like in the science classroom? *The Science Teacher*, 79(3), 38–41.
- Krell, M. (2012). Using polytomous IRT models to evaluate theoretical levels of understanding models and modeling in biology education. *Science Education Review Letters, Theoretical Letters*, 2012, 1–5. Retrieved from edoc-server. (urn:nbn:de:kobv:11–100205516).
- Krell, M., & Krüger, D. (2013). Wie werden Modelle im Biologieunterricht eingesetzt?. [How models are used in biology instruction]. *Erkenntnisweg Biologiedidaktik*, 12, 9–26.
- Kremer, K., Fischer, H. E., Kauertz, A., Mayer, J., Sumfleth, E., & Walpuski, M. (2012). Assessment of standards-based learning outcomes in science education: perspectives from the German project ESNaS. In S. Bernholt, K. Neumann, & P. Nentwig (Eds.), *Make it tangible. Learning outcomes in science education*. Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. New York: Springer.
- Lachmayer, S., Nerdel, C., & Prechtel, H. (2007). Modellierung kognitiver Fähigkeiten beim Umgang mit Diagrammen im naturwissenschaftlichen Unterricht [Modelling of cognitive abilities regarding the handling of graphs in science education]. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 145–160.
- Leibold, K., & Klautke, S. (1999). Lerneffektivität des Einsatzes gegenständlicher Modelle in Biologieleistungskursen des gymnasiums [Learning effectiveness of using representational models in biological advanced courses at secondary school]. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(1), 3–23.
- Lenske, G., Thillmann, H., Wirth, J., Dicke, T., & Leutner, D. (2015). Evaluation eines Tests zur Erfassung des pädagogisch-psychologischen Professionswissens von Lehrkräften [Evaluation of an test capturing pedagogical-psychological knowledge of teachers]. *Zeitschrift für Erziehungswissenschaft*, 18(2), 225–245. doi:10.1007/s11618-015-0627-5
- Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 635–679.
- Linacre, J. M. (2012). *A user's guide to Winsteps/Ministep: Rasch-model computer programs*. Retrieved from <http://www.winsteps.com/a/winsteps.pdf>.
- Linacre, M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328.
- Mahr, B. (2009). Information science and the logic of models. *Software & Systems Modeling*, 8, 365–383.
- Mayer, M. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen [Inquiry as Scientific Problem Solving]. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung* (pp. 177–186). Heidelberg: Springer.
- Meisert, A. (2008). Vom Modelwissen zum Modelverständnis [From knowing models to an understanding of models]. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 243–261.
- National Research Council. (1996). *National Science Education Standards*. Washington, DC:
- National Academies Research Council (2012). *A framework for K–12 science education: practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- Nehring, A., K. H. Nowak, R. Tiemann, & Upmeyer zu Belzen, A. (2011). VerE-Studie Vernetzung der Erkenntnisgewinnung zwischen Chemie und Biologieunterricht [Crosslinking scientific inquiry between biology and chemistry lessons]. In D. Höttercke. (Ed.), *Gesellschaft für Didaktik der Chemie und Physik: Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie* [Association of

- chemistry and physics education: scientific education as contribution to create a participative democracy], (pp. 510–512). Berlin: LIT.
- Lead States, N. G. S. S. (2013). *Next generation science standards: for states, by states*. Washington, D.C.: National Academies Press.
- Nowak, K. H., Nehring, A., Tiemann, R., & Upmeyer zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, *47*(3), 182–188.
- Oh, P. S., & Oh, S. J. (2011). What teachers of science need to know about models: an overview. *International Journal of Science Education*, *33*(8), 1109–1130.
- Pauli, C. (2012). Merkmale guter Unterrichtsqualität im mathematisch-naturwissenschaftlichen Unterricht aus der Perspektive von Lernenden und Lehrpersonen [Features of good instructional quality in mathematics and science instruction from the perspective of learners and teachers]. In *Lazarides & Ittel. Differenzierung im mathematisch-naturwissenschaftlichen Unterricht* [Differentiation in mathematics and science instruction], pp.13–28. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching*, *48*(5), 486–511.
- Rimmele, R. (2012). *Videograph* (version 4.2.1.22.X3). [Computer software]
- Schwarz, C. V., & White, B. Y. (2005). Metamodeling Knowledge: Developing Students' Understanding of Scientific Modeling. *Cognition and Instruction*, *23*(2), 165–205. doi:10.1207/s1532690xci2302_1.
- Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. doi:10.1002/tea.20311.
- Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). *How to run a video study: technical report of the IPN video study*. Münster: Waxmann.
- Shulman, L. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.
- Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R., & van Hout-Wolters, B. H. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, *31*(9), 1205–1229. doi:10.1080/09500690802192181.
- Smith, C., Houghton, C., & Hennessy, G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction*, *18*(3), 349–422.
- Steinbuch, K. (1977). Denken in Modellen [Thinking in models]. In G. Schaefer, G. Trommer, & K. Wenk (Eds.), *Denken in Modellen* [Thinking in models] (pp. 10–17). Braunschweig: Westermann.
- Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn* (pp. 515–565). Washington, DC: National Research Council.
- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, *4*, 295–312.
- Terzer, E. (2013). *Modellkompetenz im Kontext Biologieunterricht—Empirische Beschreibung von Modellkompetenz mithilfe von Multiple-Choice Items* [Model competence in the context of biology education—empirical description of model competence using multiple-choice items]. Doctoral dissertation. Retrieved from edoc-server. (urn:nbn:de:kobv:11–100206767).
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24*(4), 357–368. doi:10.1080/09500690110066485.
- Treagust, D. F., Duit, R., Joslin, P., & Lindauer, I. (1992). Science teachers' use of analogies: Observations from classroom practice. *International Journal of Science Education*, *14*(4), 413–422.
- Upmeyer zu Belzen, A. (2013). Modelle [Models]. In H. Großengieser, U. Harns, & U. Kattmann (Eds.), *Fachdidaktik Biologie* [Biology education] (pp. 325–334). Freising: Aulis Verlag.
- Upmeyer zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht [Model competence in biology education]. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 41–57.
- Van Driel, J. H., & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, *21*(11), 1141–1153. doi:10.1080/095006999290110.
- Van Driel, J. H., & Verloop, N. (2002). Experienced teachers' knowledge of teaching and learning of models and modelling in science education. *International Journal of Science Education*, *24*(12), 1255–1272.
- van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *35*(6), 673–695.
- Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*, 45–69.
- Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.

- Wadouh, J., Liu, N., Sandmann, A., & Neuhaus, B. J. (2014). The effect of knowledge linking levels in biology lessons upon students' knowledge structure. *International Journal of Science and Mathematics Education*, 12(1). doi:10.1007/s10763-012-9390-8.
- Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A., et al. (2012). Evaluation der Bildungsstandards – Eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung [Evaluation of the National Educational Standards: an interdisciplinary test design for the competence area acquirement of knowledge]. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 262–291.
- White, B. Y., Collins, A., & Frederiksen, J. R. (2011). The nature of scientific meta-knowledge. In M. S. Khine & I. M. Saleh (Eds.), *Models and modeling: cognitive tools for scientific enquiry* (pp. 3–22). Heidelberg: Springer.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität [Interrater agreement and interrater reliability]*. Göttingen: Hogrefe.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis. Rasch measurement*. Chicago: MESA Press.
- Wüsten, S. (2010). *Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie: Eine video- und Interventionsstudie [General and content-specific features of instructional quality in the subject biology: a video and intervention study]*. Berlin: Logos.