Grünkorn, Juliane; Upmeier zu Belzen, Annette; Krüger, Dirk

Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework.

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
International journal of science education 36 (2014) 10, S. 1651-1684

This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Science Education on 13/01/2014, available online:
http://www.tandfonline.com/10.1080/09500693.2013.873155

Assessing Students' Understandings of Biological Models and their Use in Science to Evaluate a Theoretical Framework

Juliane Grünkorn[a*], Annette Upmeier zu Belzen[b] and Dirk Krüger[c]

[a]Department of Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany; [b]Biology Education, Humboldt-Universität zu Berlin, Berlin, Germany; [c]Biology Education, Freie Universität Berlin, Berlin, Germany

Research in the field of students' understandings of models and their use in science describes different frameworks concerning these understandings. Currently, there is no conjoint framework that combines these structures and so far, no investigation has focused on whether it reflects students' understandings sufficiently (empirical evaluation). Therefore, the purpose of this article is to present the results of an empirical evaluation of a conjoint theoretical framework. The theoretical framework integrates relevant research findings and comprises five aspects which are subdivided into three levels each: nature of models, multiple models, purpose of models, testing, and changing models. The study was conducted with a sample of 1,177 seventh to tenth graders (aged 11–19 years) using open-ended items. The data were analysed by identifying students' understandings of models (nature of models and multiple models) and their use in science (purpose of models, testing, and changing models), and comparing as well as assigning them to the content of the theoretical framework. A comprehensive category system of students' understandings was thus developed. Regarding the empirical evaluation, the students' understandings of the nature and the purpose of models were sufficiently described by the theoretical framework. Concerning the understandings of multiple, testing, and changing models, additional initial understandings (only one model possible, no testing of models, and no change of models) need to be considered. This conjoint and now empirically tested framework for students' understandings can provide a common basis for future science education research. Furthermore, evidence-based indications can be provided for teachers and their instructional practice.

*Corresponding author. Department of Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany. Email: gruenkorn@dipf.de

Keywords: Students' understandings of models and their use in science; Theoretical framework; Competence; Assessment; Empirical evaluation

Introduction

Models and the process of modelling are considered key elements for the work of scientists but also for citizens' participation in social discourses and in decision-making processes in their everyday life (Odenbaugh, 2005; Oh & Oh, 2011). To understand and evaluate the work of scientists as well as their way of conceptualising phenomena and to participate in scientific discourses, it is necessary to learn and know about models and their use in science (Grosslight, Unger, Jay, & Smith, 1991). Thus, science educational standards place a considerable emphasis on models and their use in science and require students to be knowledgeable in these aspects (American Association for the Advancement of Science, 1993; National Research Council, 2000; Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005).

Based on the relevance of models as well as of the process of modelling and its reflection, research has focused on investigating students' (Grosslight et al., 1991; Schwarz et al., 2009; Schwarz & White, 2005; Treagust, Chittleborough, & Mamiala, 2002; Trier & Upmeier zu Belzen, 2009), teachers' (Crawford & Cullin, 2004, 2005; Justi & Gilbert, 2002, 2003; Van Driel & Verloop, 1999), and scientists' (Grosslight et al., 1991; Van Der Valk, Van Driel, & De Vos, 2007) understandings of models and their use in science. In addition to investigating learners' understandings, other scholars have drawn attention to students' handling of models (Louca, Zacharia, Michael, & Constantinou, 2011; Schwarz et al., 2009). As a result, various frameworks were developed. Some of these frameworks are used for the analysis of learners' understandings of models, yet they describe and structure aspects of models differently (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Treagust et al., 2002). Others provide analytical frameworks that combine learners' understandings and their handling of models (Schwarz et al., 2009). The critical issue of all these different frameworks lies in the absence of an overall empirical evaluation, i.e. whether these frameworks describe learners' understandings sufficiently (cf. Schwarz et al., 2009, p. 637).

Consequently, two research gaps can be identified and need to be addressed. First, research needs to provide a framework for analysing learners' understandings of models and their use in science. This framework ought to focus solely on learners' understandings, and bring together as well as integrate the different approaches to learners' understandings. Second, this developed framework needs to be empirically tested with a large sample to evaluate if learners' understandings are sufficiently reflected by the framework (cf. Klieme et al., 2007). Such a conjoint and empirically tested framework can be profoundly beneficial for science education research as well as for science education. For science education research, it could be used as a basis for future research studies. Studies that refer to the same framework allow for valid comparisons between their results (cf. Hartig, 2008). In addition, this framework creates a central precondition for addressing the demand raised by Louca and Zacharia (2012) to investigate the relationships between learners' understandings of models, practical skills, and other components influencing students' processes of modelling. This demand could be realised by using frameworks for each component and by assessing these components separately. Furthermore, teachers can gain infor- mation about evidence-based indications for their instructional practices.

Regarding the first research gap, Upmeier zu Belzen and Kruger (2010) contribute to this area of research by combining and integrating different empirical studies (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) and by consulting science-theoretical approaches (Giere, 2004; Mahr, 2009). This theory-based combination of empirical studies results in a theoretical framework for students' understandings of models and their use in science (Upmeier zu Belzen & Kruger, 2010). The purpose in this article is to target the second research gap and, therefore, show the results of an overall empirical evaluation of the theoretical framework. This issue is addressed by using open-ended items that are situated in different biological model contexts and by assessing a large sample representing students of different ages. The learners' understandings are investigated in the domain of biology as the understandings of models and their use in science in general can be understood as cognitive dispositions which are acquired by learning in a specific domain (cf. Klieme, Hartig, & Rauch, 2008).

Theoretical Background

Different Frameworks for Learners' Understandings of Models and their Use in Science

So far, all efforts concerning learners' understandings of models and their use in science have been focused on generating and describing different frameworks and investigating learners' understandings (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Louca et al., 2011; Schwarz et al., 2009). Grosslight et al.'s (1991) study is considered as fundamental to research on students' understandings of models and their use in science. They elicited five aspects (kinds of models, multiple models for the same thing, purpose of models, designing and creating models, and changing a model) and described three general levels of models and their use in science from the analysis of their interview study with high school students. Students at level 1 understand models as simple copies of reality with the purpose of matching the real thing in colour, shape, dimension, or material. At level 2, they realise that the construction of a model is linked to a specific purpose. Therefore, models are not seen as exact duplicates of reality. Students at level 3 are aware that the original (target, phenomenon, or object) is explained through examination and/ or manipulation of the model. Here, the model is used as a method to test ideas and to draw conclusions on the original. Based on this characterisation, Grosslight et al. (1991) stated that the majority (67%) of students were at level 1, followed by level 2 (12%). None of the students achieved level 3. Similar results have been obtained in other studies showing that students reflect little on their thinking in models and that they are not aware of the role models play in epistemological pro- cesses (Schwarz & White, 2005; Trier & Upmeier zu Belzen, 2009).

Drawing on Grosslight et al.'s (1991) three general levels, Justi and Gilbert (2003) investigated the response patterns regarding models of 39 science teachers. However, they could not provide any support for these general levels. Moreover, they generated seven aspects (nature, use, entities, uniqueness, time, predictions, and accreditation) from the analysis of interviews. Contrary to Grosslight et al.'s (1991) structure, they subdi- vided the aspect of kinds of models into nature and entities and the aspect of purpose of models into use, predictions, and accreditation. Furthermore, Justi and Gilbert (2003) explicitly described the teacher understanding of only one model being possible for a phenomenon within the aspect of uniqueness (p. 1375) in their study. With regard to students' understandings of models, empirical evidence is needed to find out whether students endorse this kind of understanding as well.

A different framework is provided by Crawford and Cullin (2005) which they deduced from a review of relevant literature. They distinguished between five aspects (multiple models for the same thing, purpose of models, designing and creating models, validating/testing models, and changing a model) and described three to four levels (limited, pre-scientific, emerging scientific, and scientific) for each aspect. In contrast to Grosslight et al.'s (1991) aspect of kinds of models and Justi and Gilbert's (2003) aspects of nature and entities, they did not specify an aspect whereby the relationship between the model and the original is characterised. Besides, without an overall empirical evaluation with a large sample regarding structure and content, Crawford and Cullin (2005) used their framework as an analytical basis for tracking the understandings of models and modelling of eight teachers.

Other than the frameworks described above which refer to learners' understandings of models and their use in science, Louca et al. (2011) as well as Schwarz et al. (2009) rather focus on the handling of models and students' practical skills. For instance, Schwarz et al. (2009) developed a learning progression that integrated practical skills (called elements of the practice), learners' understandings of models (called meta-modelling knowledge), and communication aspects. These are further divided into four levels of performance. Schwarz et al. (2009) stated that an 'overall validation of the progression' (p. 637) is still missing. However, in their recent publication (Schwarz, Reiser, Acher, Kenyon, & Fortus, 2012) they suggest refinements of their learning progression after applying it in classroom since it was not possible to clearly assign students to one of the levels. Findings of this study (Schwarz et al., 2012) support an approach to define and assess these different components (understanding, practical skills, and social components of models and modelling) separately. It is thus possible to address the issue claimed by Louca and Zacharia (2012) to investigate and analyse the relationships between the understanding, practical skills, and other components influencing the process of modelling.

In summary, the described approaches show different frameworks for learners' understandings of models and their use in science (for a detailed discussion see Krell, Upmeier zu Belzen, & Kruger, 2013; Upmeier zu Belzen & Krüger, 2010), while no conjoint framework has been developed to integrate the relevant results of these approaches. Furthermore, empirical evidence is needed to evaluate whether the learners' understandings of models and their use in science are sufficiently described. This article meets this demand by portraying a theoretical framework for students' understandings of models and their use in science and by evaluating this framework with a large sample and students of different ages.

Theoretical Framework for Learners' Understandings of Models and their Use in Science

The described frameworks (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) and science-theoretical approaches (Giere, 2004; Mahr, 2009) are integrated into a theoretical framework (Upmeier zu Belzen & Kruger, 2010). This theoretical framework can be used as an analytical framework for assessing and investigating learners' understandings of models and their use in science. Unlike the approach pursued by Schwarz et al. (2009), this framework solely emphasises the cognitive component of models (cf. Klieme et al., 2008; Weinert, 2001). Such a focused approach is strongly in line with a concept of competence underlying international large-scale assessments (e.g., Programme for International Student Assessment). The concept of competencies relates to Klieme and Leutner's (2006) definition of competencies 'as context-specific cognitive dispositions that are acquired by learning and needed to successfully cope with certain situations or tasks in specific domains' (Klieme et al., 2008, p. 9). In this regard, the cognitive component is a basis for successfully solving demanding problems in various situations (cf. Fleischer, Koeppen, Kenk, Klieme, & Leutner, 2013; Klieme et al., 2008; Weinert, 2001). Notably, the authors support the idea that the understanding of models, the handling of models (Louca et al., 2011; Schwarz et al., 2009), as well as social components (Oh & Oh, 2011; Schwarz et al., 2009) are intertwined and thus cannot exist in isolation. Nevertheless, for the purpose of investigating the relationships between these components, a separate definition, framework, and assessment is highly beneficial.

In general, the theoretical framework refers to all models collectively without focusing on specific types of models (Boulter & Buckley, 2000). An essential basis for the theoretical framework is an understanding of models from a medial perspective as an illustration of something, for instance, an idea of the original (target, phenomenon) and from a methodical perspective as an instrument for something, i.e. for testing ideas and drawing conclusions on the original (Gilbert, 1991; Mahr, 2009; Oh & Oh, 2011).

Deduction of the Levels of the Theoretical Framework

Grosslight et al. (1991) generated three general levels of models and their use that are closely tied to a person's epistemological views on science ranging from a naive-realistic view to constructivist understandings of the epistemology of science (Hofer & Pintrich, 1997). Justi and Gilbert's (2003) study, however, provided no support for Grosslight et al.'s (1991) levels but supported the idea of defining three levels for each aspect. In addition, Crawford and Cullin (2005) described at least three levels for each aspect. As a result, Upmeier zu Belzen and Kruger (2010) defined three levels for each aspect (cf. Deduction of the aspects of the theoretical framework) of the theoretical framework. These three levels reflect the epistemological viewpoint of models as products as well as methods of science (Gilbert, 1991; Mahr, 2009): On a basic level I, only the representational model is being considered without focusing on the original (called model object, Mahr, 2009). Perceiving the model as a medium of something and focusing on the creation process describes level II (called model of something, Mahr, 2009). Both perspectives are complemented by a methodological view of models (level III) in their application comprehending their use in science to test and draw conclusions on the original (called model for something, Mahr, 2009). Recently, studies investigating the three levels of this theoretical framework using mul- tiple choice and forced-choice items provide empirical evidence that these levels reflect an increasing degree of difficulty (Krell, 2012; Terzer, 2013).

Deduction of the Aspects of the Theoretical Framework

The theoretical framework comprises five aspects: nature of models, multiple models, purpose of models, testing models, and changing models. Whereas the aspects of nature of models and multiple models describe individual ontological and epistemological con- cepts of models, the aspects of purpose of models, testing models, and changing models refer to cognitive processes while reflecting the act of their use in science (Grosslight et al., 1991; Justi & Gilbert, 2002, 2006).

The aspect of nature of models integrates the aspects of entities and nature (Justi & Gilbert, 2003) as well as the aspect of kinds of models (Grosslight et al., 1991), whereby different views on the nature of science are considered (Hofer & Pintrich, 1997). Here, students compare the model with the original and comment on the extent to which the model is comparable with the original. Three positions are differentiated (Table 1): A model is understood as a replication (level I), as an idealised representation (level II), or as a theoretical reconstruction (level III). Similar aspects were described by Schwarz et al. (2009) as part of metamodelling knowledge and by Oh and Oh (2011) in their review referring to the meanings of a model.

The aspect of multiple models refers to one and the same original being represented by different models (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003). This aspect was deduced from the aspect of multiple models for the same thing (Crawford & Cullin, 2005; Grosslight et al., 1991) and the aspect of uniqueness (Justi & Gilbert, 2003). With the exception of the category of only one model being poss- ible for a phenomenon (Justi & Gilbert, 2003, p. 1375) all described categories were considered. The reason for this exception derived from the fact that studies investigating students' understandings had (so far) provided no clear evidence for this category (Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009). Regarding the three levels, various explanations can be given for these different models (Table 1): Students justify the existence of several models representing one original by describing differences between the shown model objects such as different materials or

Table 1. The theoretical framework for students' understandings of models and their use in science (Upmeier zu Belzen & Kruger, 2010)

	Complexity		
Aspect	Level I	Level II	Level III
Nature of models	Replication of the original	Idealised representation of the original	Theoretical reconstruction of the original
Multiple models	Differences between different model objects	The original allows the creation of different models	Different hypotheses about the original
Purpose of models	Describing the original	Explaining investigated relationships	Predicting connections between variables
Testing models	Testing the model object itself	Comparing the model with the original	Testing hypotheses about the original with the model
Changing models	Correcting errors in the model object	Revising the model due to new findings about the original	Revising the model due to falsification of hypotheses about the original with the model

dimensions (level I), by arguing that the original allows for building different models in order to present all features of the original (structure and function, level II), or by noticing different hypotheses (level III). A corresponding aspect called multiplicity of scientific models was also noted by Oh and Oh (2011).

In the process of modelling, purpose of models, as well as testing, and changing models need to be reflected upon (Justi & Gilbert, 2002, 2006). Therefore, the process of designing and creating a model is an essential part of all three aspects which is why the aspect of designing and creating models (Crawford & Cullin, 2005; Grosslight et al., 1991) is not distinctly definable. Therefore, this aspect is integrated into the aspects of purpose of models, testing, and changing models. For the aspect of purpose of models, Grosslight et al.'s (1991) and Crawford and Cullin's (2005) aspect of purpose of models as well as the aspects of use and prediction described by Justi and Gilbert (2003) were combined. Three diverse purposes are differentiated: models can serve to show facts of the original (level I), to describe and explain a known relationship in the original (level II), and as an instrument to predict the behaviour of the original (level III, Table 1). Oh and Oh (2011) generated the aspect of purpose of modelling and Schwarz et al. (2009) generated the aspect of using models to explain and to predict what can be assigned to the purpose of models.

The aspect of testing models describes and structures the process of testing models. Here, the category of testing within the aspect of purpose of models described by Grosslight et al. (1991), the aspect of accreditation by Justi and Gilbert (2003), and the aspect of validating/testing by Crawford and Cullin (2005) were conjoined. At level I, the model object itself is tested. At level II, the model is compared with the original and at level III, hypotheses about the original are tested with the model (Table 1).

Changes to the model can be made on the basis of test results. Therefore, models are changeable and temporary by definition (Grosslight et al., 1991; Mahr, 2009; Oh & Oh, 2011). Reasons for changing models are errors in the model object itself (level I), new information about the original (level II), or the falsification of a hypothesis about the original with the model (level III, Table 1). The aspect of changing models is derived from the aspects of changing models (Crawford & Cullin, 2005; Grosslight et al., 1991) and time (Justi & Gilbert, 2003). With regard to other studies, the aspects of testing and changing models have also been elicited by other scholars (Oh & Oh, 2011; Schwarz et al., 2009).

Objective, Research Questions, and Hypotheses

Since the theoretical framework (Upmeier zu Belzen & Krüger, 2010) is a theory-based combination of empirical studies, this study meets the demand of empirically evaluating the structure of this framework. Pursuant to this objective, we aimed to address the following main research question.

To what extent are students' understandings of models and their use in science sufficiently and adequately described by the theoretical framework?

- Which understandings of models and their use in science can be identified and described for each aspect and level of the theoretical framework?
- Which understandings of models and their use in science can be identified in addition to the theoretical framework?
- How frequent are the described understandings of models and their use in science regarding the examined sample?

Since the theoretical framework (Upmeier zu Belzen & Kruger, 2010) brings together empirical studies of Crawford and Cullin (2005), Grosslight et al. (1991), and of Justi and Gilbert (2003), identified categories relating to different understandings of models and their use in science were expected to be sufficiently and adequately described by the theoretical framework. In addition, it was likely that categories at level I and II were more frequently used than categories at level III because international and national studies (Grosslight et al., 1991; Trier & Upmeier zu Belzen, 2009) indicated that students' understandings of models and their use in science differed from a scientific understanding.

Research Design and Method

Instruments

For this study, 15 open-ended test items (three items for each aspect of the theoretical framework) were used to elicit students' understandings of models and their use in science. These 15 open-ended test items result from an item evaluation process involving researchers of biology education and a total of 1,231 German students (seventh to tenth grade students, 12–18 years old) in two preceding studies (Grünkorn & Krüger, 2012; Grünkorn, Upmeier zu Belzen, & Krüger, 2011). Considering the objective of the research study described here (the empirical evaluation of the theoretical framework) test items are required that can be interpreted as indicators for students' understanding of models and their use in science (cf. unified framework for validity, Kane, 2001). To this end, the open-ended items were tested in these two preceding studies for understandability of the items and whether the students' answers represent at least the three levels of the theoretical framework and could be assigned to a certain aspect. Focus was less placed on a selection of certain concrete problem contexts or specific model types (Boulter & Buckley, 2000), but rather on meeting the aforementioned requirements.

The open-ended item format was chosen since the answers can be formulated by the respondents themselves and are not determined by written response options

Table 2. Standardised stimuli for each aspect of the theoretical framework which have been used in the open-ended test items

Aspect	Standardised stimuli	Concrete context used in the items
Nature of models	Describe the extent to which this model of the [original] looks like the [original]	– Theoretical reconstruction of a biomembrane structure
	Give reasons for your opinion	– Theoretical reconstruction of a Tyrannosaurus rex – Theoretical reconstruction of a Neanderthal man
Multiple models	Explain why there are multiple models for one [original]	– Different biomembrane structures
	Give reasons for your opinion	– Different human gullet structures – Different taste maps of the human tongue
Purpose of models	Describe what purpose this model of the [original] serves	– Forest (different plants in a pot of soil) – Ocean (different forms of life in a water bowl) – Gut (pig gut filled with fluid lying in a bowl)
Testing models	Explain in detail how people can test if the model of the [original] serves its purpose	– Orientation skills of a beetle – Flying skills of a seed – Flying skills of a dragonfly
Changing models	Name reasons why this model of the [original] could be changed	– Flying skills of a dragonfly – Orientation skills of a crab – Food digestion process in the human mouth

Note: Corresponding to the context the term [original] is substituted (cf. Figure 1).

(Rost, 2004). Compared to interviews, which had been conducted by Trier and Upmeier zu Belzen (2009) to analyse students' conceptions of models based on this theoretical framework on a deeper level, open-ended items allow for identifying different understandings of models as well but they relate to a larger sample size. Thus, an overall empirical evaluation of the theoretical framework is possible.

The theoretical framework served as a basis for the item development to the extent that the items were designed for each aspect and that they combined both understandings of models and their use in science: illustration of something in which the model is seen as a medium and the application for something in which the model is used as a method (Gilbert, 1991; Mahr, 2009). The works of Grosslight et al. (1991) as well as an interview study conducted by Trier and Upmeier zu Belzen (2009) served as primary sources for the development of the initial stimuli. In this presented study, the items are situated in different concrete problem contexts (Table 2) since the concept of competence is defined as context-specific for a domain (cf. Klieme et al., 2008). Furthermore, the validity of the context-independent approach by Grosslight et al. (1991) has been discussed in the field of models and their use in science (Krell, Upmeier zu Belzen, Krüger, 2012; Sins, Savelsbergh, van Joolingen,

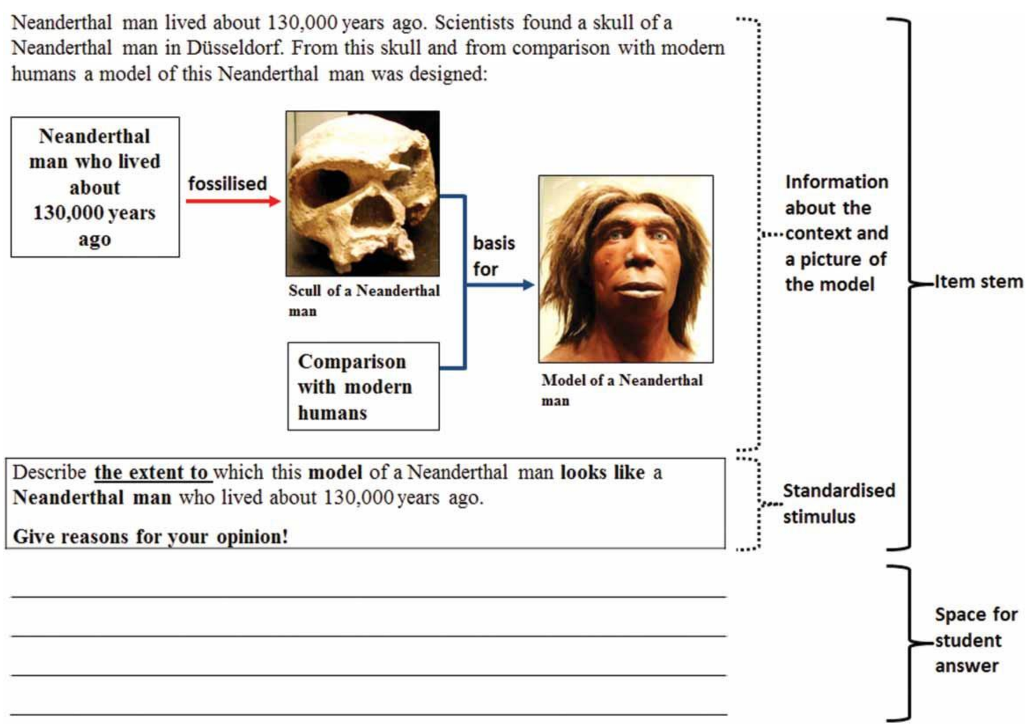

Figure 1. The structure of the open-ended test items exemplified by an item for the aspect of nature of models. The shown item is situated in the context of a theoretical reconstruction of a Neanderthal man (# Pictures: Museum für Naturkunde, Berlin)

& van Hout-Wolters, 2009). Here, the term context relates to situations that are relevant for models in the domain biology (cf. Klieme et al., 2008).

Regarding the structure of the items as exemplified in Figure 1, each item consists of an item stem and the response format—in this case, a blank space for a student's answer (Rost, 2004). The item stem gives essential information about the context, shows pictures of one or several (concerning the aspect of multiple models) model(s) and ends with a standardised stimulus (Table 2). The stimulus for each aspect only differs according to the context in which the item is situated. Regarding the model type, mostly real objects, technical models (e.g. a flying dragonfly-model), or theoretical reconstructions (e.g. Neanderthal man model; Grünkorn et al., 2011) are used.

Sample and Testing Procedure

The study was conducted with a total of 1,177 seventh to tenth grade students (11–19 years old; 48 different school classes) in Germany (Gymnasium[1], Table 3). The sample of different ages and grades was chosen to identify a wide range of different students' understandings of models and their use in science. Only thus, it is possible to evaluate whether the theoretical framework reflects learners' understandings sufficiently. Beyond grade 10, no upper grades were selected for the study because the educational standards in Germany refer to students at the end of grade 10 (KMK, 2005). No further selection criteria were used.

Since one student could not answer all 15 test items and to control order effects, a balanced incomplete block design was developed (Giesbrecht & Gumpertz, 2004). Consequently, the item pool was distributed across 35 test booklets providing each student with three open-ended items and the researchers with about 235 student answers per item (a total of 3,531 student answers). In addition, each booklet included auxiliary variables (such as age, sex, and grade). This block design is appropriate for identifying learners' understandings of models and their use in science on an aggregate level rather than on an individual level. The objectivity of application is ensured by written instructions in a test manual containing information about the aim of the research project, the item format, standardised answers to frequently asked questions, and the testing procedure.

Table 3. Demographic data about the sample of the study

		Sex distribution			Age				
Grade	n	♀	♂	n. i.	Min.	Max.	M	SD	n. i.
7	313	160	150	3	11	14	12.7	0.57	5
8	293	143	149	1	12	16	13.83	0.62	4
9	283	166	113	4	13	17	14.67	0.62	2
10	288	157	130	1	14	19	15.95	0.69	1

n. i. = no information.

Development and Evaluation of the Category System

The data were analysed by qualitative content analysis according to Mayring (2010), assisted by the qualitative data analysis software MAXQDA (Version 10). This software is commonly used in many academic fields such as sociology, psychology, and educational science and it permits easy scoring, structuring, and restructuring of large amounts of text. Furthermore, it allows for coding a given text passage in any number of ways, highlighting interesting text segments, making notes for reference and ideas directly in the text, and exporting all coding in an excel file. Only the above described functions of the software were used in this study.

As a first step of the qualitative content analysis according to Mayring (2010), the students' answers were manually digitised and colloquial expressions were smoothed while the style of the answers was left unaltered. The changes made within a student answer such as the integration of additional words were marked with square brackets to maintain the integrity of students' responses (cf. Table 4, student Qb417). For the purpose of providing evidence in this article, prominent student answers were translated from German into English by a professional translator (J. P.) and checked by two German researchers who regarded them as accurately translated.

As a second step, similar students' answers given to the respective aspect of the theoretical framework were grouped in categories (called inductive approach, Mayring, 2010). For instance, student Qb64 responds to the standardised stimulus question why there are different models for one biomembrane concerning multiple models with 'Models B and C vary in their solidity. Model B is stiffer while Model C is very flexible'. This student compared the model objects and argued with different characteristics or material properties of the shown models. Another student replied to this question by saying 'Model A is not 3D and models B and C are 3D' (student Qb44). In this case, the student also compared the model objects with each other but argued with different methods of presentation. Those and similar answers were summarised in the category different model object properties (cf. Table 5). Answers that showed content-related incorrectness and/or inadequate planning of experiments were equally analysed and not excluded because the study focused on students' understanding of models and their use in science. The whole category system had already been developed in preceding studies and was revised, discussed, checked, and refined ahead of the study presented in this article.

As a third step, the identified categories were compared with the content and definition of the respective aspect of the theoretical framework (cf. Table 1, Upmeier zu Belzen & Kruger, 2010). This was done to evaluate whether the students' understandings (categories) of models and their use in science are sufficiently and adequately described by the theoretical framework. If the framework covered the student's understanding, the category was assigned—otherwise it was added to the structure of the framework (called deductive approach, Mayring, 2010). With regard to the given example of the aspect of multiple models, the category different model object properties (Table 5) focuses only on differences between the model objects and refers to models as a medium (Table 1). Therefore, the described category could be assigned to level I within the aspect of multiple models. The assignments and additions of the categories to the theoretical framework were discussed several times until a consensus was reached with eight researchers of biology education who are experts on the theoretical framework.

As a fourth step, the assignments of students' answers to the categories were also evaluated. For this purpose, three additional trained raters who were familiar with the category system were consulted. To ensure objectivity of the analysis and interpretation, the raters used a coding manual containing information about the items, coding rules for the respective aspect, as well as a detailed description of the category system with prominent student answers from preceding studies. Fifty per cent of the students' answers were double coded and discrepancies in coding were resolved by discussion. Cohen's kappa (k) was used to measure the level of agreement between raters (Fleiss & Cohen, 1973). An agreement between the raters was achieved if all raters assigned the identical text passage within a student answer to the corresponding category. The assignment of an identical text passage to a different category or an additional assignment of text pas- sages to categories by only one rater were noted as discrepancies in coding. The interrater reliability for the assignment of the students' answers to the categories ranged from k = .81–.90 which is characterised as excellent according to Fleiss and Cohen (1973).

Descriptive Analysis

The percentage of each category was calculated based on the number of students who dealt with the respective aspect and the number of coding for each category. The frequencies are presented in Tables 4–10. The sum of frequencies of all categories within each aspect may exceed 100% (cf. Table 8) because the open-ended items allowed for multiple responses in all three levels of the respective aspect. That means one student could present different understandings (categories) in his/her response. As some students did not respond to the open-ended items or their answer did not correspond to the stimulus, the sum of frequencies of all categories within each aspect may not reach 100% (cf. Table 4).

Furthermore, distributions of students across the three levels of each aspect of the theoretical framework are presented to clarify the discussion. For this purpose, all achieved and served levels of a student within each aspect were noted. A level that was mentioned several times by one student was counted only once. The calculation of the percentage of each level was based on the number of students who dealt with the respective aspect and the number of students who showed this level in their response.

Analysis of Student Answers

The descriptive analysis only showed minimal differences between the grades seven to ten, therefore, the different grades were not further considered. Based on 3,531 student answers, 41 categories could be identified in this study describing different students' understandings of models and their use in science.

Categories of Models and their Use in Science and their Frequencies within the Levels of Each Aspect

Twenty-nine out of 41 identified categories in students' responses could be assigned to the respective aspects and the three levels of the theoretical framework. Tables 4–8 summarise the range of categories described for each aspect and their frequencies in the database. The last column of Tables 4–8 presents student answers for each category to provide evidence for each identified category and to substantiate our claims.

Table 4: For the aspect of nature of models, students compared the model with the original and commented on the extent to which the model was comparable with the original. Regarding the frequency distribution across the three levels, the majority of students (69%) could be assigned to level I. They characterised the model as a copy (e.g. student Qb366, Table 4), as a model with great similarity to the original (e.g. student Qd1073, Table 4), or as a model representing their own (non-) subjective conception of the original (e.g. student Qd915, Table 4). In this context, some students showed a high level of confidence in scientists and scientific work (e.g. student Qa485, Table 4). Furthermore, prior knowledge or personal experience with the original might play an essential role when judging a model. For instance, student Qa317 responded:

> Yes, the model looks like a real dinosaur because the dinosaurs back then had also teeth, a head, and a tail like the model shows. I know that because I went to a museum where they had dinosaurs and they look the same.

About 17% of the students (level II) mentioned that only certain features resembled the original of the model (e.g. student Qa97, Table 4), or understood the model as a possible variant among many (e.g. student Qb417, Table 4), or as a focused representation of something (e.g. student Qb6, Table 4). Only 4% of the students answered at level III by referring to models as theoretical reconstructions or ideas (e.g. student Qb24, Table 4).

Table 5: The aspect of multiple models related to the question why one original is represented by different models. Approximately half of the students (44%) compared the model objects describing different model object properties such as different methods of presentation (e.g. student Qc62, Table 5), model features (e.g. student Qb64, Table 5), and construction options (e.g. student Qb311, Table 5). An explicit reference to the original was not made at level I. Almost one-third of the students (31%, level II) focused on the complexity of the original, and mentioned the necessity of having

Table 4. Described categories, their frequencies (%), and student answers within the three levels of the aspect of nature of models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 692)
Level I		
Model as copy – Matches the original – Enlarged/reduced scale copy of original – Accepted as scale model of the original, because there is great confidence in science, in the scientific method, or in the scientists	'The Tyrannosaurus rex looked like the model' (student Qb366) 'The model of a biomembrane is an enlarged copy of a biomembrane' (student Qb502) 'The Neanderthal man looked as it does in the model, because many biologists have certainly worked on this model. These people know what they are talking about, so he looked this way' (student Qa485)	27
Model with great similarity – Resembles the original – Nearly scale model of the original due to dissatisfaction with the modelling process	'The model of the biomembrane is very similar to the real biomembrane. Both have a surface layer that holds it all together and both have tissue in the centre' (student Qd1073) 'The model resembles the Neanderthal man. Only the place on the model where the eyebrows are has to be pushed forward a little, because that's how it is in the skeletal findings. That has to happen; otherwise, the model would be incorrect' (student Qd567)	36
Model represents a (non-) subjective conception of the original – Compares and judges the model based on prior knowledge of, personal experience of, or subjective conceptions about the original	'I don't think the model is correct. The real Neanderthal man looked more like an ape. That's how I imagine one' (student Qd915)	6
Level II		
Parts of the model are a copy – Only certain features resemble the original; other features cannot be judged due to paucity of information or knowledge about the original	'The skeletal findings and the model have the same head shape. It is unknown whether the hair back then was the way it is now. Nothing can be stated with certainty about the eyes, either. On the whole, one can only comment on the shape. Colour and such remains "unknown"' (student Qa97)	9

(Continued)

Table 4. Continued

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 692)
Model as a possible variant – Might resemble the original (or not); abstract statements about similar properties – One conceivable version among many, but less well founded	'The model is comparable in terms of its shape. Nonetheless, one cannot assume that the Neanderthal man really looked like this' (student Qa164) 'Yes, it [the biomembrane] might look like that, but it might also look like this [picture drawn by student]' (student Qb417)	7
Model as focused representation – Focused on one element of the original, highlights certain traits/ properties	'The model only shows the essential parts of a real biomembrane. The main traits, structures and colours are shown here' (student Qb6)	0.3
Level III		
Model as hypothetical representation – Presents a justified hypothesis about the original, possible similarity between original and model is discussed	'No one can know for certain what a living Tyrannosaurus rex looked like back then. Scientists can only make assumptions about how it looked. They analyse the skeleton and use that to calculate how its body might have been constructed' (student Qb24)	4

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

different models in order to show every aspect of the original as exemplified by student Qb50: ' . . . Model A shows the structure of the biomembrane precisely and Model C shows it roughly. Model B shows the idealised function'. Here, an explicit reference to the original was made. For students it was rather rare (8%, level III) to justify different models of one and the same original with various assumptions or ideas about the original (e.g. student Qb5, Table 5). The scientific purposes of these different hypotheses about the original such as starting points for discussions, comparison of different assumptions, and testing assumptions with the models were named by only a few students (e.g. student Qd353, Table 5).

Table 6: In response to the purpose of models, a good proportion of the students gave answers at level I (52%) and II (50%). At level I, the purpose of models was seen as showing facts without using the model to recognise a certain relationship between different aspects in the original (e.g. student Qa403, Table 6). The latter was characteristic for level II in which the model served to identify (e.g. student Qa217, Table 6) or explain relationships (e.g. student Qb175, Table 6) in order to understand known facts. With regard to level III, 24% of the students stated that models were instru-

ments to examine ideas by testing hypotheses about the original. Here, the facet of gaining new information and learning something new about a specific phenomenon

Table 5. Described categories, their frequencies (%), and student answers within the three levels of the aspect of multiple models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 705)
Level I		
Different model object properties – Differing methods of presentation (2D or 3D, different colours, etc.) – Differing model features (moveable or immoveable, soft or hard, large or small, etc.) – Differing construction options (thin or thick materials, separated elements or one piece, etc.)	'The tongue can be shown a little differently. No colours are used in Model C, for example, whereas they are in Models A and B' (student Qc62) 'Models B and C vary in their solidity. Model B is stiffer while Model C is very flexible' (student Qb64) 'The different elements are easy to recognise in Model A. The rods on Model C are thinner than in B' (student Qb311)	44
Level II		
Focus on different aspects – The complexity of the original allows diverse perspectives or ways of focusing on the original (interior or exterior, profile or cross-section, structure or function, diverse sections or states of the original, etc.)	'Since each of these models highlights something different, there are different models. Model A focuses on the different elements and the structure, while Model B and C look more at the construction of a biomembrane' (student Qd744)	31
Level III		
Different assumptions – There can be various assumptions and ideas about the original; different models are valid at the same time – Differing interpretations of the data	'Since there are various theories/ideas about the human oesophagus, there will also be alternative models. Scientists might have other opinions' (student Qb5) 'The persons have drawn different conclusions from their observations, which is why there are different models of this biomembrane' (student Qd468)	8
Different assumptions with prospects of application – Differing assumptions about the original are named after scientific purposes (basis of discussion, comparison of different assumptions, testing assumptions with the models, etc.)	'Probably because there are many people who have researched this and so different ideas developed which still have to be examined in studies' (student Qd353)	1

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

Table 6. Described categories, their frequencies (%), and student answers within the three levels of the aspect of purpose of models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 706)
Level I		
Model for showing the facts – Presenting the facts	'The model shows the different plants that grow in a forest' (student Qa403)	52
Level II		
Model to identify relationships – Describing relationships between different aspects in the original and serving to understand known facts	'The model shows that it is possible to observe how the leaves and blossoms develop and spread' (student Qa217)	45
Model to explain relationships – Describing and explaining relationships between different aspects in the original and serving to understand known facts	'It is meant to demonstrate that the sea is a good habitat for animals e.g. fish and plants. It also shows and explains how the sea constitutes a "circulatory chain". The plants could not survive without the oxygen and the water; the fish could not live without the plants' (student Qb175)	6
Level III		
Model to examine abstract ideas – Serving as an instrument to test hypotheses about the original; general ideas are mentioned	'Perhaps so that certain tests can be viewed and carried out so that the effects of certain differences such as temperature can be studied' (student Qd1168)	19
Model to examine concrete ideas – Both testing hypotheses about the original and serving to draw conclusions about the original; concrete ideas are mentioned – Serving to transfer findings about the original to other phenomena	'One could also test which plants grow best and most quickly in which types of soil and compare these results. It might be that a certain plant draws so many nutrients out of the soil that there are less available for another. If the model were to prove this, we would know that these types of plant should not be planted too closely together' (student Qd263) 'Such a model is probably there to test whether plants can "multiply". This information helps to say something about the forest and to capture this situation in figures or charts. Then we could relate this to other ecosystems, like the sea and such' (student Qb291)	5

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

Table 7. Described categories, their frequencies (%), and student answers within the three levels of the aspect of testing models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 711)
Level I		
Testing of material – Testing the resistance of the material (flexibility, stability, elasticity, weight, etc.)	'One should test if the material of the model is strong enough to remain unharmed by something such as wind' (student Qd793)	6
Testing of basic requirements – Naming fundamental requirements for that model	'For starters, the model should be able to fly in any case. Otherwise, I don't think the model would be very good' (student Qa4)	28
Level II		
Comparison between original and model – Comparing the properties (structure and/or function) of the original with those of the model	'The model has to be compared to a real beetle' (student Qc69)	33
Comparison and matching of original and model – Both comparing properties and describing the necessary adjustments for congruity between the model and the original; naming criteria for a good model	'The model can be tested for its dimensions, its weight. The structure of the model must match the original or it isn't suitable' (student Qb206)	34
Level III		
Testing hypotheses – Testing hypotheses about the original using the model and listing general ideas for studies	'This model could simulate the flight of such a seed. Such simulations would show where the seed flies to and how it gets implanted into the soil. The model could also be used to test the effects the impact has on the soil, on the flight, and on the seed' (student Qb278)	6
Testing of hypotheses with research designs – Describing a concrete application for the model (research design) to test a hypothesis about the original	'One has to try to obtain videos of the original flight manoeuvres and attempt to recreate and compare these with the model in a wind tunnel to see if the model behaves as the original. If so, one has to change the environmental influences in the wind tunnel to determine what the dragonfly needs to fly' (student Qb200)	2

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

Table 8. Described categories, their frequencies (%), and student answers within the three levels of the aspect of changing models

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 712)
Level I		
Alterations to improve the model object – Optimising the functioning/aesthetics of the model object – Optimising the technology of model creation	'The only reason why most models are changed is because their movement and functionality can be improved' (student Qa378) 'To change the model of the dragonfly, a new technology is needed that allows the model to stay up without needing to attach a booster for uplift on the long back legs' (student Qb521)	22
Alterations when there are errors in the model object – Fundamental considerations for fixing errors in the model – Referencing concrete, incorrect properties of the model (e.g. defective materials)	'I think it's because errors are always being found which need to be corrected' (student Qd629) 'Perhaps the wings have to be made out of harder materials; otherwise they cannot resist the pressure during flight' (student Qb508)	5
Alterations when basic requirements are not met – Reviewing the basic requirements of each model and correcting defects if necessary	'If the model is meant to fly and it doesn't, then the scientists have to work on it' (student Qb496)	9
Level II		
Alterations when model does not match the original – Optimising how the (structure and/or function of the) model matches the original with consideration of the necessary congruity between the original and the model	'The model doesn't look exactly like a crab. The legs of a real crab are longer. The body of a crab is somewhat narrower. This is not the shape of a crab. That should definitely be changed, because it has to match the real crab' (student Qd1010)	61
Alterations due to new findings about the original – Integrating new findings about the original into the model; improved technology leads to new findings about the original	'In a few years, we will have better technology, so we can learn more about the dragonfly. The model could be changed when something new about the dragonfly is discovered' (student Qd1199)	4
Alterations due to changes in the original – Reflecting changes (e.g. individual developments) or advancements (e.g. evolutionary adaptation) in the original as new information in the model	'There are always changes in biology and in history. The same is true of the crab. Evolution changes the environment and animals have to adapt again. Changes to the environment force animals to change as well. That's why the model can be changed' (student Qd1165)	5

(Continued)

Table 8. Continued

Name and description of category	Evidence from student answers (student ID no.)	%* ($n_{students}$ = 712)
Level III		
Alterations due to findings from model experiments – Adjusting the model to reflect findings about the original based on a model experiment or falsification of the hypothesis behind the model	'If tests of a flying object show that the model flies completely differently than thought or than a real dragonfly does, then something could be changed on the gliding surfaces. The scientists may have had a different assumption' (student Qd352)	1

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

played a decisive role, as exemplified by student Qb291 (Table 6). At level III, students noticeably more often expressed general ideas (category model to examine abstract ideas, 19 %, Table 6) than concrete ideas (category model to examine concrete ideas, 5 %, Table 6).

Table 7: For the aspect of testing models, students explained how people could test whether the model served its purpose. The most popular responses (level II, 68 %) to this question were to test a model by comparing the original to the model (e.g. student Qc69, Table 7) or by comparing properties and describing necessary adjustments for congruity between the original and the model (e.g. student Qb206, Table 7). Considering the frequencies within the different levels, level II (68 %) was followed by level I (30 %). At level I, students tested mainly the material of the model object for robustness (e.g. student Qd793, Table 7) or tested if the model objects fulfilled basic requirements (e.g. student Qa4, Table 7). A few students' responses assigned to level III (8 %) concerned testing hypotheses about the original by describing general (e.g. student Qb278, Table 7) or concrete (e.g. student Qb200, Table 7) ideas for studies with the aim of gaining purposeful new insights into biological topics. Similar to the aspect of purpose of models (Table 7), students mentioned general ideas (category testing hypotheses, 6 %, Table 7) more often than concrete ideas (category testing hypotheses with research designs, 2%, Table 7).

Table 8: The aspect of changing models pursued the question of what conditions could lead to changing a model. Similar to testing models, most students (68 %) described understandings that could be assigned to level II of the theoretical frame- work paying attention to the original: a majority of students (61 %, category alteration when model does not match the original, Table 8) at level II argued that a model is changed when it does not match the original in terms of structure and/or function (e.g. student Qd1010, Table 8). Only a few students named reasons such as alteration due to new findings about the original (4 %, e.g. student Qd1199, Table 8) or alteration due to changes in the original (5 %, e.g. student Qd1165, Table 8). Approximately one-third of the students (31 %, level I) referred to alterations that improve (e.g. student Qa378, Table 8) or fix the model object when defects were found in

the model object itself (e.g. student Qb508, Table 8) or the basic requirements were not met (e.g. student Qb496, Table 8). Students at this level did not mention the original as a reason for a change; they rather focused on general technical issues in the model object, as exemplified by student Qb521: ‘The model of the dragonfly is changed when a new technology allows the model to stay up without needing to attach a booster for uplift on the long back legs’. Rarely (1%) did the students change a model due to findings from model experiments in which their hypothesis, implicitly stated by students in their answers, was rejected (level III, e.g. student Qd352, Table 8).

Additional Categories of Models and their Use in Science to Each Aspect

For the aspect of nature of models, no additional categories could be identified within the data. However, concerning the aspects of multiple models, purpose of models, testing models, and changing models, additional categories could be generated (Tables 9 and 10).

Table 9: The aspect of multiple models and its three levels refer to the understanding of one and the same original being represented by different models. Therefore, students were confronted with three different models of one original in the open- ended items. Special attention was paid in the development of the open-ended items to ensure that distinct references to one original were made in the item stems and obviously different models were shown (such as different foci, colours, material, and dimensions). Still, 14% of the students gave one of the following responses when asked about the presence of multiple models (Table 9): (a) all models were the same (e.g. student Qd240, Table 9), (b) various models representing different originals were shown (e.g. student Qb331, Table 9), or (c) only one model was the final and correct one (e.g. student Qd643, Table 9). Those students rejected the existence of multiple models as representations of one original and conceived only one model as representation of an original. By giving one of the listed responses (Table 9), they stuck to their understanding. As the aspect of multiple models demands the acceptance and understanding of multiple models referring to one original, all three facets can be seen as an initial understanding of this aspect.

For the aspect of testing models, 1% of the students (e.g. student Qc71, Table 9) thought it was unnecessary to test a model and expressed their perplexity about why and how a model is tested. Students kept this understanding although the purpose of the presented models could plausibly be used for testing. The concept of testing models includes the acceptance and necessity for testing, which is why this kind of response can be interpreted as an initial understanding and level of this aspect.

Similarly, an initial understanding and level could be identified for the aspect of changing models reflecting the opinion that models should not be changed. Students (10%) who showed this understanding responded by either rejecting changes of a model (e.g. student Qd341, Table 9), or by changing the presented model to represent another original (e.g. student Qd954, Table 9). Since the concept of this aspect

Table 9. Initial understandings of the aspects of multiple models, testing models, and changing models exemplified by student answers and their frequencies (%)

Name and description of category	Evidence from student answers (student ID no.)	%*
Multiple models ($n_{students} = 705$)		
All models are the same – All models are or show the same; no description of differences between models	'All three models show basically the same. I don't know why there should be different models at all. That makes no sense' (student Qd240)	4
Various models of different originals – Each model represents a different original	'One might also make three different models to show the biomembranes of different life forms, e.g. a human being, a bird and a cow' (student Qb331)	13
Only one final and correct model – Only one of the various models is final and correct; the others are incorrect – Only one model is the final model; they are not valid contemporaneously	'Perhaps there is only model which is the best model. All the others are wrong. There is just one correct model. How else should this work?' (student Qd643) 'I think that two models are old models. At that time, one did not have all information like we have today. However, one model is the final model which happens to be true. I've seen it in my textbook' (student Qd321)	1.4
Testing models ($n_{students} = 711$)		
No testing of models – Rejecting model testing in general or of this model	'Why should this beetle model even be tested? I don't think it's necessary' (student Qc71)	1
Changing models ($n_{students} = 712$)		
No reason for alterations – Rejecting changes to a model	'I don't think the model should be changed' (student Qd341)	3
Alteration of how different originals are represented – Creating different models for different originals; each original is represented by its own model	'Because not all dragonflies are alike and models can be made for different dragonflies' (student Qd954)	7

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

demands the acceptance of changing a model and that the creation of a model as well as the process of changing a model are always linked to a specific purpose and a specific original, these categories were not assigned to the three levels of the aspect of changing models.

Table 10: For the aspects of purpose of models and changing models, further categories could be generated that were not considered as initial understandings of these aspects. As the question regarding purpose of models was formulated in a rather open way

Table 10. Further categories, their frequencies (%), and student answers for the aspect of purpose of models presenting everyday understandings, and a category of the aspect of changing models showing an alternative understanding

Name and description of category	Evidence from student answers (student ID no.)	%*
Purpose of models ($n_{students}$ = 706)		
Model as toy – Serving for pleasure, interest, and/or enjoyment	'There might be children who are interested in it and will have fun with it, so this model was developed' (student Qd580)	0.4
Model for decorative purposes – Serving to beautify a room or building – Contributing to the well-being of people and other life forms	'It might be that the forest model can be put to good purpose decorating the home' (student Qd1137) 'Some people simply find it [the model] relaxing when they aren't on holiday or simply want to daydream' (student Qd1035)	2
Model as replacement object – Serving as prosthesis or replacement organ	'Its purpose might be that it was developed as a replacement organ. If someone has a defective intestine, it could be replaced with an artificially manufactured, fully functional one' (student Qd31)	1
Model as blueprint – Serving as a small-scale blueprint for presenting the future object	'This model is meant to be used like a map. One could say that a hotel miniature is to be placed on the model and then one could build the real hotel exactly on the spot represented in the model. It serves for orientation, so to speak' (student Qd144)	0.1
Model for advertising purposes – Serving as advertisement to reach political, social, and/or environmental goals	'They want to use it to make the world aware that forests need more water' (student Qd613)	1
Changing models ($n_{students}$=712)		
Alterations of the model conditions – Changes to the model to reflect changes in how it is used or handled (movements of the model, etc.)	'The model of a mouth could change during rotation' (student Qd1208)	2

*Each frequency is based on all noted students' understandings and is calculated with reference to the number of students who dealt with this aspect.

(Table 2), the student responses touched a wide range of purposes. For example, students thought of models as toys (e.g. student Qd580, Table 10), as something for decorative purposes (e.g. students Qd1137 and Qd1035, Table 10), as replacement objects (e.g. student Qd31, Table 10), as blueprints (e.g. student Qd144, Table 10), or as something for advertising purposes (e.g. student Qd613, Table 10). With regard to changing models, a few students had an alternative understanding of the term change by thinking that a change referred to the alteration of conditions underlying the model (e.g. student Qd1208, Table 10).

Discussion and Conclusion

The purpose of this study was to empirically evaluate the theoretical framework. Before contrasting the findings with the results of other scholars in this field, methodological constraints need to be discussed: the preceding studies have already shown that the 15 open-ended test items can be interpreted as indicators for the understanding of models and their use in science (Grünkorn & Krüger, 2012; Grünkorn et al., 2011). Therefore, the items are suitable for evaluating the structure of the theoretical framework concerning learners' (grades 7–10; 11–19 years old) understandings of models and their use in science. Concerning the target group (grades 7–10) for whom these instruments are designed, the results only allow for reliable statements concerning these particular learners.

For the presented study, a context-dependent approach was chosen since the validity of a context-independent approach by Grosslight et al. (1991) as well as Treagust et al. (2002) have been discussed and questioned in the field of models (Krell et al., 2012; Sins et al., 2009). All items were situated in different concrete contexts (Table 2). Thus, different results concerning the frequencies may be achieved when choosing a context-independent approach or when choosing other contexts (Krell et al., 2012). Since the focus of this study was rather to evaluate the theoretical framework on an aggregate level than to diagnose students' understandings of models and their use in science on an individual level, this issue can be neglected (cf. Leutner, Fleischer, Spoden, & Wirth, 2007).

Bearing in mind that the items are situated in biological contexts, findings cannot—without further investigations—be generalised and therefore only hypothetically be transferred to other science domains such as physics or chemistry. This precaution is based on the fact that we define learners' understandings as competencies. Competencies are acquired and applied in a specific domain (cf. Fleischer et al., 2013; Klieme et al., 2008). Furthermore, in scientific subjects such as biology, chemistry, and physics models are used in different ways. Whereas in biology structural, functional, or dynamic models are often used, abstract and mathematical models predominate in chemistry and physics (cf. Beerenwinkel & Parchmann, 2008; Mikelskis-Seifert & Fischler, 2003).

Empirical Evaluation of the Theoretical Framework

With regard to the main research question whether the theoretical framework (Upmeier zu Belzen & Krüger, 2010) accurately reflects the different aspects and levels of learners' understandings concerning models and their use in science, the hypothesis was only partly confirmed: the students' understandings of the aspect of nature of models and purpose of models are sufficiently and adequately described by the theoretical framework. However, the aspects of multiple models, testing models, and changing models need to be discussed and, therefore, revised since initial understandings for these aspects were identified. Table 11 summarises the findings of the study and provides a revised theoretical framework, that contributes to research in this field.

Table 11. Revised framework for students' understandings of models and their use in science including levels of complexity and their categories

	Complexity			
Aspect	Initial level	Level I	Level II	Level III
Nature of models	—	Model as copy	Parts of the model are a copy	Model as hypothetical representation
		Model with great similarity Model represents a (non-) subjective conception of the original	Model as a possible variant Model as focused representation	
Multiple models	All models are the same Various models of different originals Only one final and correct model	Different model object properties	Focus on different aspects	Different assumptions Different assumptions with prospects of application
Purpose of models	—	Model for showing the facts	Model to identify relationships Model to explain relationships	Model to examine abstract ideas Model to examine concrete ideas
Testin g models	No testing of models	Testing of material	Comparison between original and model	Testing hypotheses
		Testing of basic requirements	Comparison and matching of original and model	Testing of hypotheses with research designs
Changing models	No reason for alteration	Alterations to improve the model object	Alterations when model does not match the original	Alterations due to findings from model experiments
	Alteration of how different originals are represented	Alterations when there are errors in the model object	Alterations due to new findings about the original	
		Alterations when basic requirements are not met	Alterations due to changes in the original	

The initial understanding of the aspect of multiple models could be described by the conception only one model of an original. Grosslight et al. (1991) indicated that some students did not believe that it was possible to have multiple models representing one original. Justi and Gilbert (2003), however, explicitly described the conception uniqueness (p. 1375) in a study with teachers. The students' responses concerning the uniqueness of models named in this article (categories: all models are the same and only one final and correct model) were similar to findings reported by Justi and Gilbert (2003). Students giving the latter explanation probably understood models as final form devices and evaluated multiple models for correctness (Crawford & Cullin, 2004). Those students' understandings differ from scientific understandings which ' . . . do not ask whether it [the model] is right or not' (Crawford & Cullin, 2004, p. 1382). Contrary to Crawford and Cullin (2004), Grosslight et al. (1991), and Justi and Gilbert (2003), the category of various models of different originals was eli- cited among students' responses and played an important role when explaining the presence of multiple models representing one original. These results demonstrate that this understanding is also prominent among students in different variations and therefore needs to be considered.

Based on the analysis of the students' responses, a new initial understanding could be identified for the aspect of testing models. Thus, a few students rejected the necessity of testing a model and, therefore, showed no clear awareness of the work of scientists which is predominated by the creation as well as the testing of models (Giere, 2004; Odenbaugh, 2005; Oh & Oh, 2011). The studies by Crawford and Cullin (2004, 2005), Grosslight et al. (1991), and Justi and Gilbert (2003) did not give any indi- cation as to this category. This newly described category is a result from this study and it needs to be considered and focused on in future research.

Regarding the aspect of changing models, some students mentioned the understand- ing of models being unchangeable (Table 9). This understanding has previously been reported by other studies (Crawford & Cullin, 2004, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003; Treagust et al., 2002), and it confirms the idea that some stu- dents think of models as final form devices (Crawford & Cullin, 2004). Besides clearly rejecting changes to the model, several students explained that a model could be changed in order to present another original. Similar to the category of various models of different originals in the aspect of multiple models, students again referred to different originals instead of focusing on one original. The category also emerges from this study and has not been described by other studies (Grosslight et al., 1991; Justi & Gilbert, 2003).

Based on these results, for the aspects of multiple models, testing, and changing models a fourth level of students' understandings of models and their use in science needs to be added (Table 11). Therefore, the results rather support the approach by Crawford and Cullin (2005) who defined three to four levels for each aspect than the approach of Grosslight et al. (1991), of Justi & Gilbert (2003), or of Upmeier zu Belzen and Krüger (2010) who differentiate between three levels.

Besides describing initial understandings, newly described categories within the three levels of the aspects of purpose of models and testing models could be identified

(Table 11). Unlike the categorisation for the aspect of purpose of models shown in the studies by Crawford and Cullin (2005), Grosslight et al. (1991), and Justi and Gilbert (2003), the categories of models to examine abstract ideas (19 %) and models to examine concrete ideas (5 %) could be described and distinguished within the data (Table 6, level III). Similar categories (Table 7, level III) were described in the aspect of testing models in which hypotheses about the original are tested by either describing general (6 %) or concrete ideas (2 %). This analysis allowed for the conclusion that students have an abstract and/or concrete understanding and that it is probably more difficult and complex to establish a concrete understanding than an abstract one. This argument is supported by the fact that a qualitative study using thinking aloud protocols to elicit students' understandings (Terzer, 2013) showed that students did have an abstract understanding of models but could not use it to solve the problem in concrete situations with models. Therefore, future research needs to focus on these different understandings.

In summary, the purpose of this study was to evaluate a theoretical framework that brings together different approaches (Crawford & Cullin, 2005; Grosslight et al., 1991; Justi & Gilbert, 2003) and is empirically evaluated with a large sample and students of different ages. This overall empirical evaluation did not only provide evidence for students having an initial understanding concerning the 'uniqueness' (Justi & Gilbert, 2003) of models, but newly described categories were identified that need to be considered in future research (Table 11). This conjoint and empirically tested framework can now serve as a basis for future investigations concerning students' understandings of models and their use in science.

Further Categories within the Data

The analysis of students' responses to the aspect of changing models showed that a few students had an alternative understanding of the term changing in this context. They perceived a change as an alteration of the model condition (Table 10). This understanding does not agree with the understanding that is fundamental in the theoretical framework (Upmeier zu Belzen & Krüger, 2010). Although this category is not being considered in the theoretical framework, there is a need to gain access to this understanding by asking students in a separate questionnaire to prevent misunderstandings.

Frequencies of the Generated Categories

We were also interested in how often the described understandings of models and their use in science could be determined in the examined sample on an aggregate level. Regarding the percentage distribution within each aspect of the theoretical framework (Upmeier zu Belzen & Krüger, 2010), students responded more frequently at level I and II than at level III. These findings were as expected and corresponded largely to results of other studies (Grosslight et al., 1991; Schwarz & White, 2005; Trier & Upmeier zu Belzen, 2009). Differences in the percentage distribution were noted in the aspects of multiple models and changing models.

In contrast to the studies of Grosslight et al. (1991), in which students often explained the presence of multiple models with different foci (level II), students in the presented study rather focused on differences between the model objects in terms of colour, shape, dimension, and material (level I). Concerning changing models, a popular response did not concern alterations due to new findings about the original (4 %) but alterations when the model does not match the original (59 %). The comparison between the original and the model played an important role and showed that students were firmly anchored in a medial perspective on their use in science (Oh & Oh, 2011). The differences concerning the percentage distribution for the aspects of multiple models and changing models might have been caused by the fact that this presented study was conducted with a larger sample than the previous studies (Grosslight et al., 1991). Another possible explanation might be that in the presented study, the items were situated in different concrete problem contexts. Sins et al. (2009) noted and a study using forced-choice items (Krell et al., 2012) showed empirically that the context-dependent and the context-independent approaches lead to different results. However, further investigations concerning the impact of different concrete problem contexts and model types on students' performances need to be conducted.

Students in this sample see models less as a method of science (Mahr, 2009; Oh & Oh, 2011) and more from a medial perspective (Mahr, 2009; Oh & Oh, 2011). Possible reasons might be a more frequent use of models as a substitute for the original or as a medium for transmitting information in biology lessons (Crawford & Cullin, 2004; Van Driel & Verloop, 1999). This and the role of teachers in 'distinguishing the positive and negative analogies as clearly as possible' (Hardwicke, 1995, p. 64) might lead to the fact that students primarily focus on the relationship between the model and the original. Students draw comparison to achieve the best possible match without considering the purpose of the model.

Final Implications

With regard to science education research the following contributions and implications can be derived from the results of this study: first, to assess and diagnose the current state or development of students' understandings of models and their use in science, more sensitive instruments might be helpful. For this purpose, the comprehensive category system can provide student-based under- standings and could serve as a basis for the development of those instruments. Second, a conjoint and empirically tested framework for students' understandings of models and their use in science is now available. This framework is applicable to future research in this field, for instance, by having created a precondition to address the demand raised by Louca and Zacharia (2012). In their paper, they requested that research needs to focus on investigating the relationships between students' understandings of models, their practical skills, and other components such as the social facet that influence students' processes of modelling. This research gap can be addressed by developing frameworks for each component that allow for assessing these components separately and consequently for investigating the relationships between them. In terms of providing an empirically tested framework for assessing students' understandings, this was accomplished by this study. Regarding the other facets, conjoint and consensual frameworks need to be developed. For this purpose, the works of Louca et al. (2011) as well as Schwarz et al. (2009, 2012) already provide valuable information on students' practical skills. Third, the conjoint and empirically tested framework and the comprehensive category system presented in this article can be used to assess and investigate the cognitive facet of models, i.e. students' competencies of models and their use in science. As the concept of competencies is defined as a basis for successfully solving problems in various situations (cf. Klieme et al., 2008; Weinert, 2001), the students' achieved level gives valuable information to teachers and researchers concerning the extent to which they are able to solve problems with models in certain modelling situations.

Concerning science education, several studies already dealt with learning progressions for scientific modelling and showed that constructing models and reflecting about models help to facilitate students' understandings of models and modelling (Baek, Schwarz, Chen, Hokayem, & Zhan, 2011; Schwarz et al., 2009). This study contributes to the field by giving further implications regarding students' understandings of models and their use in science: first, we suggest being aware not only of the initial students' understandings of models and their use in science but also of the other understandings. Thus, the category system might be used in teacher training to assist the teaching staff to keep abreast of their students' different understandings of models and their use in science (cf. Fleige, Seegers, Upmeier zu Belzen, & Krüger, 2012). Second, in this context, a reflective use of historical models in school needs to be discussed as already demanded by Justi and Gilbert (2002). The findings of this study suggest that epistemological processes in science and the hypothetical character of models could lead to an elaborated under- standing of models and their use in science (Justi & Gilbert, 2003) as exemplified by student Qd468 ' . . . Models could represent different ideas of scientist which changed over time'. However, an unreflective use of historical as well as current models might contribute to the understanding why the presence of multiple models representing one original is rejected. According to student Qd321 (Table 9), the historical processes lead to a final and correct model which does not change. Therefore, it is important to carefully reflect current models and to consider and discuss the conclusions students draw from historical processes. Third, as giving feedback is an integral part of fostering students adequately, the category system might be used by teachers as an instrument to provide students with detailed feedback.

Note

1. Gymnasium is a type of secondary school in Germany in which students finish school after grade 12 or 13 with A-levels (Abitur). The Abitur qualifies them for admission to university.

References

American Association for the Advancement of Science. (1993). Benchmarks for science literacy: Project 2061. New York, NY: Oxford University Press.

Baek, H., Schwarz, C., Chen, J., Hokayem, H., & Zhan, L. (2011). Engaging elementary students in scientific modeling: The MoDeLS fifth-grade approach and findings. In M. Khine & I. Saleh (Eds.), Models and modeling (pp. 195–220). Dordrecht: Springer.

Beerenwinkel, A., & Parchmann, I. (2008). Metadiskussion über Modelle. Historische Aspekte als Impuls [Meta-discussion about models. Historical aspects as impulses]. Praxis der Naturwissenschaften Chemie [Practice of science chemistry], 4(57), 13–16.

Boulter, C., & Buckley, B. (2000). Constructing a typology of models for science education. In J. Gilbert & C. Boulter (Eds.), Developing models in science education (pp. 41–57). Dordrecht: Kluwer Academic.

Crawford, B. A., & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modelling in science. International Journal of Science Education, 26(11), 1379–1401. doi:10.1080/0950069041000167377 5

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.

Fleige, J., Seegers, A., Upmeier zu Belzen, A., & Kruger, D. (2012). Förderung von Modellkompetenz im Biologieunterricht. [Fostering model competence in biology education]. Der mathematische und naturwissenschaftliche Unterricht [Mathematical and science teaching], 65(1), 19–28.

Fleischer, J., Koeppen, K., Kenk, M., Klieme, E., & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms [Modelling of competencies: Structure, concepts and research approaches of the DFG priority program]. Zeitschrift für Erziehungswissenschaften [Journal for educational science], 16(1), 5–22. doi:10.1007/s11618-013-0379-z

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement, 33, 613–619. doi:10.1177/001316447303300309

Giere, R. N. (2004). How models are used to represent reality. Philosophy of Science, 71, 742–752. doi:10.1086/425063

Giesbrecht, F., & Gumpertz, M. (2004). Planning, construction, and statistical analysis of comparative experiments. Hoboken, NJ: John Wiley & Sons.

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., & Krüger, D. (2012). Entwicklung und Evaluierung von Aufgaben im offenen Antwortformat zur empirischen Überprüfung eines Kompetenzmodells zur Modellkompetenz [Development and test of open-ended test items to empirically assess a model for model competency]. In U. Harms & F. X. Bogner (Eds.), Lehr- und Lernforschung in der Biologiedidaktik [Teaching/Learning research in biology education] (pp. 9–27). Innsbruck: Studienverlag.

Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2011). Design and test of open-ended tasks to evaluate a theoretical structure of model competence. In A. Yarden & G. Carvalho (Eds.), Authenticity in biology education (pp. 53–65). Braga: CIEC, Universidade do Minho.

Hardwicke, A. J. (1995). Using molecular models to teach chemistry: Part 1 using models. School Science Review, 77(278), 59–64.

Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen [Competencies as results of educational processes]. In N. Jude, J. Hartig, & E. Klieme (Eds.), Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden [Assessing competencies in pedagogical contexts. Theories, concepts, and methods] (pp. 17–25). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67, 88– 140. doi:10.3102/00346543067001088
Justi, R. S., & Gilbert, J. K. (2002). 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. (2003). Teachers' views on the nature of models. International Journal of Science Education, 25(11), 1369–1386. doi:10.1080/0950069032000070324
Justi, R. S., & Gilbert, J. K. (2006). The role of analog models in the understanding of the nature of models in chemistry. In P. J. Aubusson, A. G. Harrison, & S. M. Ritchie (Eds.), Metaphor and analogy in science education (pp. 119–130). Dordrecht: Springer.
Kane, M. T. (2001). Current concerns in validity theory. Journal of Educational Measurement, 38, 319 – 342. doi:10.1111/j.1745-3984.2001.tb01130.x
Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., . . . Vollmer, H. J. (2007). The development of national educational standards: An expertise. Bonn: BMBF.
Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), Assessment of competencies in educational contexts (pp. 3 –22). Cambridge, MA: Hogrefe & Huber.
Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG [Competence models for assessing individual learning outcomes and evaluating educational processes: Description of a newly established GRF priority program]. Zeitschrift für Pädagogik [Journal of pedagogy], 52(6), 876–903.
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., Upmeier zu Belzen, A., & Kruger, D. (2012). Students' understanding of the purpose of models in different biological contexts. International Journal of Biology Education, 2, 1–34. Retrieved from http://www.ijobed.com/2_2/Moritz-2012.pdf
Krell, M., Upmeier zu Belzen, A., & Kruger, D. (2013). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? Research in science education, 1–24. doi:10.1007/s11165-013-9365-y
Leutner, D., Fleischer, J., Spoden, C., & Wirth, J. (2007). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik [State-wide standardised assessments of learning between educational monitoring and individual diagnostics] [special issue]. Zeitschrift für Erziehungswissenschaften [Journal for educational science], 8, 149–167.
Louca, T. L., & Zacharia, C. Z. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. Educational Review, 64(4), 471–492. doi:10.1080/00131911.2011.628748
Louca, T. L., Zacharia, Z., Michael, M., & Constantinou, P. C. (2011). Objects, entities, behaviors and interactions: A typology of student-constructed computer-based models of physical phenomena. Journal of Educational Computing Research, 44(2), 173–201. doi:10.2190/EC.44.
Mahr, B. (2009). Information science and the logic of models. Software & Systems Modeling, 8, 365– 383.

MAXQDA (Version 10) [Computer Software]. Berlin: VERBI Software. Retrieved from http://www.maxqda.de/

Mayring, P. (2010). Qualitative Inhaltsanalyse: Grundlagen und Techniken [Qualitative analysis: Basics and techniques] (11th, updated and revised ed.). Weinheim: Beltz.

Mikelskis-Seifert, S., & Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen—Stand der Forschung und Entwurf einer Unterrichtskonzeption [The relevance of thinking in models with regard to conceptions of particles—state of research and draft of a teaching conception]. Zeitschrift für die Didaktik der Naturwissenschaften [Journal of science education], 9, 75–88.

National Research Council. (2000). Inquiry and the national science education standards. Washington, DC: National Academy Press.

Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. Biology & Philosophy, 20(2–3), 231–255. doi:10.1007/s10539-004-0478-6

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. doi:10.1080/09500693.2010.502191

Rost, J. (2004). Lehrbuch Testtheorie—Testkonstruktion [Textbook test theory—test construction]. Bern: Huber.

Schwarz, C., Reiser, B., Acher, A., Kenyon, L., & Fortus, D. (2012). MoDeLS: Challenges in defining a learning progression for scientific modeling. In A. Alonzo & A. Gotwals (Eds.), Learning progressions in science. Current challenges and future directions (pp. 101 – 137). Rotterdam: Sense. Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., …, Krajcik, J. (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

Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. Cognition and Instruction, 23(2), 165–203. doi:10.1207/s1532690xci2302_1

Sins, P. H. M., Savelsbergh, E. R., Joolingen, W. R. van, & Hout-Wolters, B. H. A. M. van (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

Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (Eds.). (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.

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

Trier, U., & Upmeier zu Belzen, A. (2009). 'Die Wissenschaftler nutzen Modelle, um etwas Neues zu entdecken, und in der Schule lernt man einfach nur, dass es so ist.': Schülervorstellungen zu Modellen ['Scientists use models to discover something new and in school, you only learn that this is the case': Students' conceptions of models]. In D. Krüger, A. Upmeier zu Belzen, S. Hof, K. Kremer, & J. Mayer (Eds.), Erkenntnisweg Biologiedidaktik 8 [Path of knowledge biology education 8] (pp. 23–37). Kassel: Universität Kassel.

Upmeier zu Belzen, A., & Kruger, D. (2010). Modellkompetenz im Biologieunterricht [Model competence in biology education]. Zeitschrift für Didaktik der Naturwissenschaften [Journal of science education], 16, 41–57.
Van Der Valk, T., Van Driel, J., & De Vos, W. (2007). Common characteristics of models in present-day scientific practice. Research in Science Education, 37(4), 469–488. doi: 10.1007/s11165-006-9036-3
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
Weinert, F. E. (2001). Concepts of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), Defining and selecting key competencies (pp. 45–65). Seattle, WA: Hogrefe & Huber.