

# Students' Understanding of the Purpose of Models in Different Biological Contexts

**Moritz Krell\***

**Freie Universität Berlin**, Biology Education, Schwendenerstraße 1, 14195 Berlin, Germany,  
0049 30 838 51 171, [Moritz.Krell@fu-berlin.de](mailto:Moritz.Krell@fu-berlin.de)

**Annette Upmeier zu Belzen**

**Humboldt Universität zu Berlin**, Biology Education, Invalidenstraße 42, 10115 Berlin,  
Germany, 0049 30 2093 8457, [Annette.Upmeier@biologie.hu-berlin.de](mailto:Annette.Upmeier@biologie.hu-berlin.de).

**Dirk Krüger**

**Freie Universität Berlin**, Biology Education, Schwendenerstraße 1, 14195 Berlin, Germany,  
0049 – 30 – 838 56 714, [Dirk.Krueger@fu-berlin.de](mailto:Dirk.Krueger@fu-berlin.de)

---

## *Abstract*

The present article analyses context dependencies in students' ranking of three perspectives on the purpose of biological models, i.e. to show, to explain, or to predict. German students (N = 1,207; 11 to 18 years old; secondary schools) have been assessed using one decontextualized forced choice task (i.e. without referring to a specific model) as well as six contextualized forced choice tasks (each presenting a different biological model in the task stem). Students' responses have been compared using the Wilcoxon test as well as within an IRT approach. The findings show that the respondents systematically preferred more elaborated perspectives concerning the purpose of models in biology in the contextualized tasks than in the decontextualized task. Further, students' answers were slightly inconsistent even within the contextualized tasks. Based on these findings, implications for assessment in science education and science teaching are discussed.

**Key words:** Purpose of models, biology teaching, science education

## Introduction

Models are an indispensable part of scientific inquiry and communication (e.g. Giere, Bickle, & Mauldin, 2006; Laubichler & Müller, 2007; Magnani & Nersessian, 2002; Morgan & Morrison, 1999). Consequently, the understanding of the nature of models is conceptualized as an integral part of the understanding of the *Nature of Science* (NOS) (Gobert *et al.*, 2011) and models are said to be effective means for teaching scientific literacy (Gilbert, 1991; Halloun, 2007). Above that, models seem to be effective tools for teaching biological content knowledge (Chabalengula & Mumba, 2012). Thus, it is suggested to highlight the role of models as research tools in science curricula and teaching practice (Gilbert, 2004; Prins, Bulte, Van Driel, & Pilot, 2009).

### *Models and Modeling in Science Education*

Contemporary research about models and modeling in science education focuses on scientists' (Van Der Valk, Van Driel, & De Vos, 2007), experienced or prospective teachers' (Crawford & Cullin, 2005; Justi & Gilbert, 2003; Van Driel & Verloop, 1999, 2002), and students' (Grosslight, Unger, Jay, & Smith, 1991; Grünkorn, Upmeier zu Belzen, & Krüger, 2011; Treagust, Chittleborough, & Mamiala, 2002) understanding of models and modeling in science. Most studies indicate that both teachers and students do not have an elaborated understanding of the role of models as research tools in science but primarily associate models with descriptive entities or teaching tools (e.g. Grosslight *et al.*, 1991; Justi & Gilbert, 2003). Biology teachers seem to follow this understanding to a greater extent than other science teachers (Justi & Gilbert, 2003). However, in contrast to these findings, a few studies indicate that students may have a 'scientifically acceptable understanding of the model concept', since they recognize the role of models as research tools (Chittleborough, Treagust, Mamiala, & Mocerino, 2005, p. 200; Treagust *et al.*, 2002). Because the empirical findings are found to be

ambiguous, it can be assumed that students' responses to questions about models may depend on the experience students have made with models in school (Ingham & Gilbert, 1991; Treagust et al., 2002). This way of thinking is reasonable, because students who are asked about the purpose of a model, which they have experienced as a teaching tool, may recall this when they are asked about the same or a similar model in a questionnaire or an interview. Furthermore, students may adopt a general concept of models as teaching tools but consider them research tools when an appropriate model is shown in an assessment task (e.g. a scientific model; Gilbert, Boulter, & Elmer, 2000).

### *Context Dependencies*

In other areas of science assessment (e.g. NOS), the impact of different task stems on students' responses is analyzed and the task stems are sometimes called item features (Nehm & Ha, 2011) or contexts (Urhahne et al., 2011). However, the term 'context' is not used consistently in the field of science education (Gilbert, 2006). From the perspective of science learning, 'context' is most widely understood as a learning situation or a learning activity (Van Oers, 1998). In the field of science assessment, on the other hand, the term is commonly used in the sense of task stem or item feature, because different 'contextualizing elements such as activities, personal perspectives, and concrete examples' (Son & Goldstone, 2009, p. 75) may be described in a task stem. By doing so, the task stem defines and introduces the background of a given task and contributes to how respondents perceive and internally process the task (Nehm & Ha, 2011). In the following, 'context' is used in the sense of task stem or item feature and the term 'task context' will be used to highlight this. More precisely, in the present study task contexts are said to be different as long as the task stems are constituted of the illustration and explanation of different biological models.

As mentioned above, there are studies in which context effects in students' or teachers'

understanding of NOS are analyzed (e.g. Clough & Driver, 1986; Leach, Millar, Ryder, & Séré, 2000; Murcia & Schibeci, 1999; Urhahne, Kremer, & Mayer, 2011). For example, Leach et al. (2000) assessed students' epistemological understanding in science and found that students use a range of epistemological reasoning across different task contexts. The authors conclude that this may be explained with situated perspectives on learning ('situated learning'; e.g. MacLellan, 1996). Similarly, Clough and Driver (1986) show that students' science concepts vary across many task contexts. The authors conclude that students' perception of tasks may be different to researchers' perception. Therefore, also tasks contexts which refer to identical scientific concepts (e.g. force and motion, Clough & Driver, 1986; evolution, Nehm & Ha, 2011) and are therefore said to be quite similar, may be seen differently by students (Clough & Driver, 1986; Song & Black, 1991). Guerra-Ramos (2012) emphasizes that the assumption of stable concepts about NOS in different task contexts is questionable but that 'different ideas can be applied in different situations, and therefore that the context matters' (p. 642). Nehm and Ha (2011) underline that findings concerning context effects may be used to improve assessment designs in such a way that those effects are balanced as well as to give researches hints regarding respondents' cognitive coherence.

### *Context Dependencies in Students' Understanding of Models and Modeling*

Models and modeling can be seen as being a 'subset' of NOS (Gobert et al., 2011) since the process of science can be understood as developing, testing, comparing, and changing hypothetical models for natural phenomena (Giere et al., 2006). Considering both the importance of context dependencies in students' understanding of NOS which is sketched out above and the role of models in science, it is reasonable that students' understanding of models and modeling may also vary across different task contexts. Consequently, it is argued that systematic research should be done in order to analyze students' understanding of models across

different task contexts (Al-Balushi, 2011). However, there are several studies in which students' understanding of models and their use in science is assessed either decontextualized, i.e. without referring to an example or a situation (e.g. Gilbert, 1991; Treagust et al., 2002), or contextualized by referring to concrete examples, i.e. to single models in an assessment task or an interview question (e.g. Grünkorn et al., 2011; Justi & Gilbert, 2003). These different methods of assessment may be the cause of the ambiguous empirical findings concerning students' understanding of models and modeling sketched out above. Furthermore, both approaches have limitations: When assessing students' understanding decontextualized, it is not clear which instances the respondents have in mind when answering a task (Guerra-Ramos, 2012). Tasks which refer to concrete examples or situations (i.e. contextualized tasks) are likely to assess students' understanding of these examples but general conclusions are problematic as long as context effects are not considered and appropriately balanced. This is highly important for the assessment of students' understanding of the purpose of models (in biology) since there is no purpose of models per se but different models may have different purposes. The present study investigates consistencies in students' answers across different tasks concerning the purpose of biological models.

Danusso, Testa, and Vicentini (2010) point out that it is 'intrinsically problematic' (p. 872) to develop an agreed definition of the term 'model'. This is mainly due to the reason that models may be discussed from different points of view, like psychology (e.g. Gentner & Stevens, 1983), philosophy of science (e.g. Bailer-Jones, 2003), and science education (e.g. Halloun, 2007), but also because there is a wide range of different kinds of models (e.g. Boulter & Buckley, 2000). Consequently, due to economic and argumentative reasons, the focus of the present article shall not be on finding a distinct definition but on the purpose of biological models. Other perspectives such as the multiplicity or the changing nature of models (cf. Oh & Oh, 2011; Upmeier zu Belzen & Krüger, 2010) will not be considered here.

*Theoretical Background: About the Purpose of Models in Biology*

There are a great number of publications in which the purpose of scientific models is elucidated (e.g. Bailer-Jones, 2003; Braithwaite, 1962; Giere et al., 2006; Hestenes, 1992; Magnani & Nersessian, 2002; Morgan & Morrison, 1999; Suckling, Suckling, & Suckling, 1978; Suppes, 1962). In most publications on the topic, one purpose or a number of purposes of models in science are highlighted. For example, Morgan and Morrison (1999) point out the purpose of scientific models to mediate between theory and reality ('models as mediators'). Suppes (1962) emphasizes the use of models to describe and process data ('models of data'). Schwartz and Lederman (2005) found that scientists are aware of the diversity of model-purposes in science. However, the authors state that '[n]ot all models explain empirical observations and not all models take an abstract concept and make it more concrete' (p. 14). In the following, three theoretical descriptions about the purpose of models in science are outlined which focus on the fact that there is no purpose of models per se but that different kinds of models may have different purposes. These theoretical approaches by Harré (1970), Leonelli (2007), and Odenbaugh (2005) are explained to stress the necessity to systematically consider the context which may be used when assessing students' understanding of models and their use in science.

Harré (1970) argues that there are no objects which are models in themselves, but that something can function as a model when seen in a certain relationship to, for example, a biological phenomenon. This relationship is referred to as a 'projective convention' (p. 46). To explain this approach in detail, Harré distinguishes between the subject and the source of a model. While the subject is said to be the corresponding (biological) phenomenon which is represented by the model (elsewhere called the 'original'; e.g. Black, 1962), the source is 'whatever it is the model is based upon' (Harré, 1970, p. 38). For example, taking a corded ladder as a model for the DNA's structure, the DNA is the subject and the corded ladder is the

source of this model. Based upon the distinction between a model's subject and source, Harré mainly distinguishes between two kinds of models, homeomorphs and paramorphs. Sometimes the object or process which is to be modeled is already known or understood. In such cases it is possible to use the same mechanisms, molecules, principles which exist in the subject to construct the model (homeomorph). In other cases, when a process is unknown, one may use a source different from the subject. For example, modeling the interaction between two molecules as a lock-and-key-interaction uses a known mechanism in order to explain why, e.g., a given molecule only interacts with one of various other molecules in a certain way (paramorph). So, homeomorphs are models with the subject being the same as the source, paramorphs are models with the subject and the source being different from each other. Harré (1970) argues that the development of scientific knowledge is mainly based on the construction of models, i.e. especially paramorphs. In many cases, a paramorph is constructed which suggests: 'whatever is in the black box [...] could be like this' (p. 39), therefore, using hypothetical explanations based on information from another domain. Harré calls this modeling-strategy 'making models for unknown mechanisms' while the construction of homeomorphs is called 'making models of known things and processes' (p. 40), since the subject has to be known to use it as the source. It is important in the present context that Harré clearly differentiates between the purpose of models for something (to provide hypotheses) and the purpose of models of something (to show or explain phenomena already understood). Consequently, the author concludes that there is no purpose of models per se, but that the purpose of a model depends on whether it is a paramorph or a homeomorph.

Similarly to Harré (1970), but with focus on theory development in science, Leonelli (2007) emphasizes the need of what she calls a multimodel approach. Despite of elaborating a single definition of the modeling process and of highlighting one prior purpose of models (single-model approach; Leonelli, 2007) one may capture and understand the role of models in

theory development within a pluralistic account. Leonelli develops such an account by distinguishing between material models and theoretical models. She points out that the initial choice of a researcher's epistemic goal determines the respective modeling-strategy. Her argumentation is based on Levins (1966) who underlines that the 'contradictory desiderata of generality, realism, and precision' (p. 431), the complexity of nature, as well as the limitation of both the human mind and computational power forces researchers to choose the kind of modeling-strategy depending on one's objective. With her focus on theory development in biology, Leonelli (2007) argues that models which are manipulated materially (material models) have other epistemological functions than models that are manipulated conceptually (theoretical models). She argues that only the complementary use of both leads to 'intelligible theories' (p. 16) about biological phenomena. Intelligible theories are said to have empirical content and explanatory power. Theoretical models ground on an already developed theory and are used to test or illustrate this theory (models with 'predictive accuracy'; p. 30). Following Leonelli, such models strictly rely on the respective theory and it is tested to what extent it is possible to, for instance, elaborate this theory using the model. Unlike theoretical models, material models are tangible objects, e.g. scale models, diagrams, robots, or model organisms, based directly on biological phenomena (models with 'empirical accuracy'; p. 30). As long as biologists do not have a theoretical explanation for a phenomenon, material manipulation is necessary to provide 'epistemic access' (p. 27) to the phenomenon. Following Leonelli, both modeling-strategies are important since material models are manipulated materially in order to secure the empirical content of the theory which is developed by using the model, and theoretical models are manipulated conceptually to secure the explanatory power. Similar to Harré (1970), Leonelli concludes that there are two different kinds of models, one largely used as representatives of a given phenomenon (material models) and the other as representatives for a phenomenon (theoretical models). By emphasizing the importance of material manipulation, Leonelli



underlines the role of physical interaction and perception for scientific inquiry.

With focus on theoretical ecology, Odenbaugh (2005) describes five main purposes of models: They are used (1) to explore possibilities, (2) to investigate more complex systems, (3) to provide conceptual frameworks, (4) to generate accurate predictions, and (5) to generate explanations. Svoboda and Passmore (2011) summarize Odenbaugh's (2005) approach with a focus on science education. Since the purposes (4) and (5) are well described in the literature, Odenbaugh emphasizes the first three purposes: For instance, (1), a model may represent a certain relation between various variables of a biological system and therefore allows proposing hypotheses about how the system might be or might work under these relations. This helps scientists to organize their ideas about how a system might work and explore possibilities about the system (Odenbaugh, 2005; Svoboda & Passmore, 2011). Hence, this modeling strategy is highly creative and may be called 'abductive' since new inferences may be made by analogy (Magnani, 1999). Secondly, simple models may be used to investigate more complex systems (2): Different models may be arranged on a continuum of simplicity, from extremely simplified models to highly complex ones. Using simple models, the focus is on relatively few parameters. Therefore, it is possible to locate the source of model misspecification better than in complex models (Odenbaugh, 2005; Wimsatt, 1987). As a result, relatively simple models can be used to find out the reasons why a more complex model represents something inaccurately. Concerning (3), Odenbaugh (2005) claims that models provide conceptual frameworks, i.e. 'new ideas that have the potential to transcend the model' (Svoboda & Passmore, 2011, p. 6). For example, a scientist has to decide which parameter of a biological phenomenon to include in a model and how it should be represented. Thus, 'positive analogies' (Hesse, 1966) have to be chosen in accordance with the purpose of the respective research agenda. By doing so, a researcher may create new conceptual frameworks by highlighting single parameter (Svoboda & Passmore, 2011). Summarizing, Odenbaugh (2005) provides a more differentiated analysis of the possible

roles of models as research tools than Harré (1970) and Leonelli (2007). Odenbaugh concludes that ‘models must be evaluated according to their functional roles not against jobs that they are not designed to carry out’ (p. 253).

Mahr (2008) emphasizes that once something is used as a model, there are always some similarities between the model and the original. These ‘positive analogies’ (Hesse, 1966) may be structural or functional but they reflect, in general, the creation of the model as a model of the respective phenomenon (Mahr, 2008). Certainly, this creation may be only mentally (Gentner & Stevens, 1983). Since every model is constructed for a purpose, it is likely that there will be an application of the model. If the model is a homeomorph, its purpose may be to show or to explain the corresponding phenomenon while a paramorph may be used to investigate unknown mechanisms (Harré, 1970). Mahr (2008) argues that in its application, a model is a model for something. Consequently, both features, being a model of something and being a model for something, are constructive relations of the model and the corresponding phenomenon which every model fulfills: ‘All modelers [...] agree that a model is always of some things and for a specific purpose’ (Halloun, 2006, p. 22). However, the focus of a given model may be on one of both constructive relations. Similarly, Leonelli (2007) stresses that the distinction between material models (representatives of sth.) and theoretical models (representatives for sth.) is not rigorous but that material models are said to be largely manipulated material and theoretical models are largely manipulated conceptually (p. 25). Thus, a model may be developed primarily as a model of something, having more empirical accuracy than predictive power, or, it may be developed primarily as a model for something, having more predictive power than empirical accuracy. In the former case, the focus is on the retrospective relation between model and phenomenon whereas in the second case the focus is on their prospective relation.

#### *What Should Students Learn About the Purpose of Models in Biology?*

It is argued that models and modeling should be integrated in science education

curriculums more prominently (Gilbert, 2004). Consequently, it has to be discussed which aspects of models and the modeling process are important for science education. Recently, Upmeier zu Belzen and Krüger (2010) have developed a theoretical structure of model competence in biology education with five aspects of models and modeling: Nature of models, multiple models, purpose of models, testing models, and changing models. Similar structures have been developed by Crawford and Cullin (2005; who describe the five aspects purpose of models, designing and creating models, changing a model, multiple models for the same thing, and validating/testing models), Grosslight et al. (1991; again five aspects: kinds of models, purpose of models, designing and creating models, multiple models for the same thing, changing a model), or Justi and Gilbert (2003; who name seven aspects: nature, use, entities, uniqueness, time, prediction, accreditation). Consistently, the purpose (or the use; Justi and Gilbert, 2003) of models is seen as an important aspect of the understanding of models and modeling in science. Since the focus of the current article is on the purpose of models (in biology) this aspect will be illustrated in the following.

According to Upmeier zu Belzen and Krüger (2010), three different purposes of models in biology can be distinguished: (I) Models can be used to show or to describe the corresponding phenomenon, (II) models can be used to explain relations of variables/parameters of the corresponding phenomenon, and (III) models can be used to test or to generate hypotheses about the corresponding phenomenon. This differentiation between three main purposes of models is also made by Van Driel and Verloop (1999) as well as Oh and Oh (2011). However, Oh and Oh additionally mention the use of models to communicate. Similarly, Schwarz et al. (2009) distinguish between the usage of models in ‘sensemaking’ (e.g. to describe, to predict) on the one hand and ‘communication’ on the other hand. The authors explain the difference between both with the audience of the respective model: While using models as sense-making entities means to make models for oneself, using models as entities to communicate means to make

models for others (Schwarz et al., 2009). However, in communication models may be used to describe or to explain something (to others). Therefore, the purpose of models to communicate may be included in purposes (I) and (II) by Upmeier zu Belzen and Krüger (2010). Both purposes are related to the retrospective relation between a model and its original (Upmeier zu Belzen & Krüger, 2010). Hence, these purposes may be summarized as the ‘descriptive nature of models’ (Treagust et al., 2004). However, while (I) refers to the purpose of models to show what a phenomenon may look like and tries to give an answer to the question of what actually exists, (II) refers to the purpose of a model to show how the phenomenon may behave and gives answers to the causal question of why something happens (cf. Treagust et al., 2004; Halloun, 2007; Oh & Oh, 2011). In contrast to the descriptive purposes of models, (III) is concerned with the ‘predictive nature of models’ (Treagust et al., 2004). Table 1 summarizes the purposes of models as developed by a philosopher (Mahr, 2008), theoreticians of science (Harré, 1970; Leonelli, 2007; Odenbaugh, 2005) as well as researchers in science education (Van Driel & Verloop, 1999; Treagust et al., 2004; Halloun, 2006; Upmeier zu Belzen & Krüger, 2010; Oh & Oh, 2011). However, this list is neither exclusive nor exhaustive.

Table 1. The retrospective and the prospective relation between a model and the corresponding biological phenomenon as developed by Harré (1970), Odenbaugh (2005), Leonelli (2007), and Mahr (2008). Additionally, approaches which have been developed for science education are added (Van Driel & Verloop, 1999; Treagust *et al.*, 2004; Halloun, 2006; Upmeier zu Belzen & Krüger, 2010; Oh & Oh, 2011).

	<b>retrospective relation</b>		<b>prospective relation</b>
Harré (1970)	models <i>of</i> known things and processes		models <i>for</i> unknown mechanisms
Odenbaugh (2005)	generate explanations		explore possibilities, investigate more complex systems, provide conceptual frameworks, generate accurate predictions
Leonelli (2007)	representatives <i>of</i> sth.		representatives <i>for</i> sth.
Mahr (2008)	creation (model <i>of</i> sth.)		application (model <i>for</i> sth.)
Treagust <i>et al.</i> (2004)	descriptive nature of models		predictive nature of models
Van Driel & Verloop (1999), Halloun (2006), Oh & Oh (2011)	describing	explaining	predicting
Upmeier zu Belzen & Krüger (2010)	describing the original	explaining the original	predicting something about the original

In line with other conceptualizations about students' and teachers' understanding of models and modeling in science (e.g. Crawford & Cullin, 2005; Grosslight *et al.*, 1991) the three perspectives on the purposes of models (I, II, III) are ordered by complexity and called 'levels' of understanding (Upmeier zu Belzen & Krüger, 2010). There are several studies which suggest that the descriptive nature of models is dominant in the science classroom and that modeling activities with a focus on the predictive nature of models are sparsely implemented in science education (cf. Danusso *et al.*, 2010). Consequently, most findings support the conclusion that students seem to have more difficulties in understanding the predictive nature of models than the descriptive nature of models (e.g. Grosslight *et al.*, 1991).

#### *Importance of the Study and Research Question*

As outlined above, there is no purpose of biological models per se. In agreement, there

is research evidence that students' understanding of models and modeling in biology may vary across different task stems (e.g. Grünkorn et al., 2011; Ingham & Gilbert, 1991) and across different science domains (Gobert et al., 2011). Task contextuality is also reported in other areas of science education (e.g. Clough & Driver, 1986; Nehm & Ha, 2011). Consequently, it is reasonable that students' understanding of biological models may vary across different task contexts. The following research question has been addressed:

*To what extent does students' understanding of the purpose of models in biology vary across different task contexts?*

## **Method**

### *Data Collection and Sample*

#### *Instrument*

Forced choice tasks (Hicks, 1970) have been deductively developed based on the three levels of understanding concerning the purpose of models in science (cf. Upmeyer zu Belzen & Krüger, 2010; Table 1). In these tasks the respondents have to rank operationalizations of the three levels, i.e. they have to decide which of the alternatives they agree most with: The purpose of a model is (I) to describe the original, (II) to explain the original, or (III) to predict something about the original. The tasks present abstract operationalizations of the theoretical perspectives (Figure 1). Therefore the forced choice tasks are likely to assess 'a type of nature of science understanding' (Schwarz et al., 2009, p. 634) about models which is related to, but not identical with, students' modeling ability (Schwarz et al., 2009; Treagust et al., 2004).

Forced choice tasks have been used since they have several advantages. For example, individually different interpretations of category labels in rating tasks and tied judgments are avoided (Böckenholt, 2004). Due to the advantages, this task-format has already been used by

others to assess students' understanding of science (e.g. Chittleborough et al., 2005; Kleickmann et al., 2010).

<b>What do you think about the purpose of a model?</b>	
Decide: Which is most like you think and which is least like you think? Write the letters A ( <b>most</b> ), B ( <b>median</b> ), and C ( <b>least</b> ) next to the respective statements.	
The purpose of a model is ...	
... to show the original as precise as possible.	<input type="checkbox"/>
... to explain relationships between parts of the original.	<input type="checkbox"/>
... to acquire assumptions about the original.	<input type="checkbox"/>

Figure 1. The *general* forced choice-task for the aspect purpose of models. For the *model-specific* tasks, individual task stems have been developed which describe and picture the models and the corresponding biological phenomena

Seven tasks have been developed in order to discuss the research question:

Task 0: In this task, the respondents have to rank the three levels decontextualized (i.e. without a concrete model in the task stem; Figure 1).

In the other six tasks, different models are described and pictured in the task stem (Figure 2). These models have been chosen for the present study since findings of a preliminary study (N = 725, students) suggested that they represent a diverse spectrum of biological models from students' point of view. These six models are quite diverse, for example they have different modes of representations (e.g. a three-dimensional object, a drawing, or a computer simulation), and they are different kinds of models (e.g. a functional model, a model organism). This ensures to capture an appropriate part of the broad range of different models which are used in biology.

Context of task 1: A mathematical curve which represents the relation of predators and preys (more precisely: birds and earthworms) diagrammatically ('diagrams are typically regarded as

models'; Giere, 2002).

Context of task 2: A technical model of the human mouth which may be described as a '3D model that moves' (Boulter & Buckley, 2000) or a functional model (cf. Penner, Giles, Lehrer, & Schauble, 1997). A drawing of the model is presented in the task stem (for a closer description of this model cf. Arvisenet et al., 2008).

Context of task 3: A model organism which is used in neurobiological research (*Aplysia californica*; e.g. Kandel, 1983).

Context of task 4: A computer simulation which models crossbreeding of mice (simulations can be seen as 'dynamic models'; Guala, 2002).

Context of task 5: A functional model (cf. Penner et al., 1997) of a palm leaf which is also a '3D model that moves' (Boulter & Buckley, 2000). However, not a drawing but a photography of the model is presented in the task stem (Figure 2).

Context of task 6: A theoretical reconstruction of the Neanderthal man. However, the model is best described as a scale model (Boulter & Buckley, 2000).

Fulfilling these six tasks, the respondents have to rank the three levels with respect to the particular model. Hence, the tasks 1 to 6 are called contextualized (Figure 2).



The properties of a palm leaf (e.g. the stability) are influenced by its shape and its structure.  
Figure 1 shows a **model of a palm leaf**. Figure 2 shows a **folded palm leaf**.


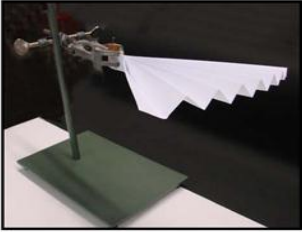


Figure 1: Model of a palm leaf      Figure 2: Palm leaf

**What do you think about the purpose of the model of the palm leaf?**

Decide: Which is most like you think and which is least like you think?  
Write the letters A (**most**), B (**median**), and C (**least**) next to the respective statements.

The purpose of the model of the palm leaf is ...	
... to show the palm leaf as precise as possible.	<input type="checkbox"/>
... to explain relationships between parts of the palm leaf.	<input type="checkbox"/>
... to acquire assumptions about the palm leaf.	<input type="checkbox"/>

Figure 2. The *model-specific* task which refers to the model of a palm leaf (model 4).

### Test Booklets

The presented findings are a part of a bigger study concerning the empirical dimensionality of model competence (cf. Terzer, Krell, Krüger, and Upmeier zu Belzen, 2011). In this study tasks for all five aspects of model competence have been included. Therefore a booklet-design with systematic gaps has been used to keep the number of tasks for each student small (mainly an incomplete latin square design; Cochran & Cox, 1957).

First, the respondents answered the decontextualized task concerning the aspect purpose of models as one among three decontextualized tasks (about 60 % of the students got a test-booklet which included a decontextualized task concerning the purpose of models in biology). Thereafter, they got between one and four contextualized tasks concerning the aspect purpose of models in between a set of ten contextualized tasks (the six contextualized tasks were answered each by about 30 % of the students). However, only the aspect purpose of models is analyzed in the following. Consequently, the present study aims to provide information on the group level rather than investigating students' individual concepts.

*Sample*

7th to 10th graders answered the tasks (N = 1,209), currently 11 to 18 years old (ms = 14.08; sd = 1.24) and attending secondary public schools in Berlin, Germany.

*Data analysis*

First of all, ‘partially ipsative data’ (Hicks, 1970) has been generated by scoring only the level which was ranked as ‘most like me’ (cf. Figures 1, 2). Hence, students who ranked level I first, have been scored with 0, students who ranked level II first have been scored with 1, and those who ranked level III first have been scored with 2. The relative count of students who answered on level I, II, and III was analyzed for each task stem and compared using the Wilcoxon signed rank test. The test statistics indicate if there are significant differences between the students’ responses in the case of different task stems. Following Fritz, Morris, and Richler (2012), Cohen’s  $r$  should be used as a measure of the effect size for the Wilcoxon test (with  $r = .1$ ,  $r = .3$ , and  $r = .5$  indicating small, medium, or large effects). Since an incomplete booklet design was developed, these analyses are each based on a subsample of  $N$ . To include the whole sample, the software ConQuest (Wu, Adams, & Wilson, 2007) has been used to analyze the data using the partial credit-model (PCM) – an IRT measurement model which assumes more than two ordered response categories (Masters, 1982, 2010). This corresponds to the theoretical assumptions of model competence by Upmeier zu Belzen and Krüger (2010). Furthermore, the data of forced choice tasks had been fitted to the PCM in a previous study (Krell, Upmeier zu Belzen, & Krüger, 2012). In the PCM the probability  $p$  of person  $v$  answering in response category  $k$  on item  $i$  is estimated as follows (Masters, 2010):

$$p(X_{vi} = k) = \frac{\exp \sum_{s=0}^k (\theta_v - \delta_{is})}{\sum_{r=0}^k [\exp \sum_{s=0}^r (\theta_v - \delta_{is})]} , \text{ where } \delta_{i0} \equiv 0, \text{ so that } \sum_{s=0}^0 (\theta_v - \delta_{is}) = 0 .$$

Using ConQuest, item parameters and person parameters are computed based on the marginal maximum likelihood estimator (Wu *et al.*, 2007). Person parameters ( $\theta_v$ ) indicate the ‘ability’ of a person. In the PCM, item thresholds ( $\delta_{is}$ ) indicate the relative difficulty of answering in a specific response category, i.e. in a specific level of understanding (Masters, 2010). Item parameters are computed as the mean of the item thresholds to indicate the general ‘difficulty’ of each item (Wu *et al.*, 2007). However, since the PCM is an IRT measurement model, it has to be analyzed if the estimated parameters are appropriate before interpreting the relevant values (Smith, 2000).

## **Results**

The research question addresses differences between students’ understanding of the purpose of biological models across different task contexts. Table 2 shows the relative count (%) of responses in level I, II, and III for the six *contextualized* tasks as well as the *decontextualized* task. Looking at the results of the *contextualized* tasks, students primarily understand models as something to show or to explain the corresponding original: For four models, level II has the highest relative count (task contexts 1, 2, 3, 5), whereas the purpose of the model of crossbreeding (task context 4, 39.4 %) and the model of the Neanderthal man (task context 6, 41.3 %) is primarily understood to be level I. In all cases, only a small group of students understood the purpose of the models primarily in predicting something (level III). However, concerning the model of the palm leaf (task context 5), level I (28.7 %) has approximately the same count as level III (27.1 %).

According to the Wilcoxon test, only the comparison of task context 1 and task context 5 (cf. Table 2) results in significant differences: Students answered on a significantly higher level of understanding when a representation of the relation between predators and preys (task

context 1) is presented in the task stem ( $mode = 2$ ) than when the model of a palm leaf (task context 5) is shown ( $mode = 2$ ,  $Z = 2.46$ ,  $p < .05$ ). In this case, the effect size is small to medium ( $r = .23$ ).

Table 2. The absolute (n) and the relative count (%) of students' responses in level I, II, and III.

Task context	n	Level I	Level II	Level III
0 general	712	48.17	36.66	15.17
1 predators and preys	368	32.07	51.36	16.58
2 mouth	368	28.80	52.99	18.21
3 organism	371	35.58	40.97	23.54
4 crossbreeding	373	39.41	35.39	25.20
5 palm leaf	387	28.68	44.19	27.13
6 Neanderthal	383	41.25	31.59	27.15

Table 2 also shows the response pattern concerning the *decontextualized* task. Students primarily answered on level I, followed by level II and III. Compared with the response patterns of the six *contextualized* tasks (Table 2), more students answered on low levels in the *decontextualized* task. A closer analysis shows significant differences between the response patterns of the *decontextualized* task and all *contextualized* tasks (Table 3). For example, students answered on a significantly lower level of understanding when no model was presented in the task stem ( $mode = 1$ ) than when the model of the palm leaf was shown (task context 5,  $mode = 2$ ,  $Z = 4.18$ ,  $p < .01$ ,  $r = .28$ , pair e). The smallest effect size ( $r = .13$ ) occurs when the *decontextualized* task is compared with the model of the Neanderthal man (task context 6, pair f). But also in this case, the *decontextualized* task is answered on a significantly lower level of understanding ( $p < .05$ ).

Table 3. The results of the Wilcoxon test for the *decontextualized* task compared with the *contextualized* tasks; \*:  $p < .05$ , \*\*:  $p < .01$ , \*\*\*:  $p < .001$ .

	<b>Pair</b>	<b>n</b>	<b>mode</b>	<b>Z</b>	<b>p</b>	<b>r</b>
(a)	decontextualized. vs. prey and pred.	221	1 2	2.88	**	.19
(b)	decontextualized. vs. mouth	255	1 2	4.37	***	.27
(c)	decontextualized vs. organism	143	1 2	3.38	**	.28
(d)	decontextualized vs. crossbreeding	263	1 1	2.24	*	.14
(e)	decontextualized vs. palm leaf	224	1 2	4.18	***	.28
(f)	decontextualized vs. Neanderthal	267	1 1	2.19	*	.13

The PCM shows a good fit to the data: For the general item parameters, the MNSQ-values range from 0.98 to 1.01 (unweighted) and from 0.98 to 1.01 (weighted), the  $t$ -values range from -0.20 to 0.20 and from -0.30 to 0.20. For the step parameters, the MNSQ-values range from 0.96 to 1.02 and from 0.97 to 1.01, the  $t$ -values range from -0.50 to 0.30 and from -0.70 to 0.20. These values indicate a good fit between the assumptions of the PCM and the data (Smith, 2000).

Table 4 shows the item parameters and the thresholds. Looking at the item parameters, the *decontextualized* task is the ‘hardest’ and the *contextualized* task with the model of a palm leaf is the ‘easiest’ one. However, the thresholds allow a more differentiated analysis. Since  $\delta_{i1} < \delta_{i2}$  for all items, the thresholds indicate that it is comparatively ‘harder’ to take the step from level II to III than to take the step from level I to II. But the distances ( $\delta_2 - \delta_1$ ) depend on the task stems. For instance, it is only 0.10 when the model of the Neanderthal man (task context 6) is presented in the task but 1.91 with respect to the model of the human mouth (task context 2). For instance, the relatively high value of  $\delta_1$  indicates that it is rather challenging to answer on level II when the model of the Neanderthal man is presented in the task stem. In this case,  $\delta_1$  is

quite close to the threshold  $\delta_1$  of the *decontextualized* task. Above that, there are two *contextualized* tasks which have higher thresholds  $\delta_2$  than the *decontextualized* task (mouth:  $\delta_2 = 1.20$ , predators and preys:  $\delta_2 = 1.28$ , *decontextualized*:  $\delta_2 = 1.07$ ). So, it is even more challenging to answer on level III when these models are shown than when no model is presented in the task stem.

Table 4. The item parameters (item) and the thresholds ( $\delta_1$ ,  $\delta_2$ ) of all tasks.

	Task context	Item	$\delta_1$	$\delta_2$
0	general	0.66	0.24	1.07
1	predators and preys	0.37	-0.54	1.28
2	mouth	0.24	-0.71	1.20
3	organism	0.23	-0.23	0.69
4	crossbreeding	0.25	0.02	0.48
5	palm leaf	0.04	-0.53	0.62
6	Neanderthal	0.25	0.20	0.30

Summarizing, the current data shows mainly two things. Firstly, students' responses to the *decontextualized* task are different to their *contextualized* responses. Basically, students seem to understand the *general* purpose of models primarily in showing something (level I) but the *model-specific* purpose of models especially in explaining something (level II). However, secondly, students' understanding of the purpose of biological models seems to vary across different task contexts.

## Discussion

The research question concerns students' understanding of the purpose of biological models across different task contexts. Especially, both the comparison across different *contextualized* tasks and the analysis of students' responses to a *decontextualized* task is addressed. Students' answers to the *decontextualized* task are seen as indicators for students'

*general* concepts or understanding of the purpose of biological models whereas students' answers to the *contextualized* tasks indicate their understanding of a specific model.

The results of the Wilcoxon test show that there are mainly no significant differences between students' response patterns when referring to different instances. Only the comparison of the model of a palm leaf and the model of the relation of predators and preys results in significant differences. This finding may be explained with the different modes of representation of these two models. While the model of the palm leaf is a three-dimensional functional model, the relation of predators and preys is represented in a diagram (i.e. a curve). There are other studies indicating that students' understanding may depend on the model's dimensionality (e.g. Grünkorn *et al.*, 2011; Ingham & Gilbert, 1991; Schwarz & White, 1998). For example, Grünkorn *et al.* (2011) describe that students set concrete models apart from drawings. In an open-ended task one student wrote: 'A is a real model, however, B is only a drawing' (p. 9) – despite A and B were both labeled as models. Similarly, Schwarz and White (1998) implemented an intervention into their studies and asked students in a pre-post-design whether they see different entities as models. Referring to a causal rule, only 14 % of the students thought of it as a model before the intervention (48 % thereafter). The comparison between students' understanding of the purpose of the model of the palm-leaf and the model of predators and preys supports these findings: Comparatively more students seem to understand a functional model primarily as a predictive entity than a diagrammatical model. Although the Wilcoxon test does not show significant differences in the other cases, a closer analysis of the relative counts (Table 2) indicates differences in students' understanding of the models. For instance, the model of crossbreeding of mice (39.4 %) and the model of the Neanderthal man (41.3 %) are understood to have primarily the purpose of showing the original. In contrast, only about 28 % of the students see this purpose for the model of the palm leaf or the model of the human mouth – which are both '3D models that move' (Boulter & Buckley, 2000). Looking at the model of

the human mouth, a majority (53.0 %) answered that its purpose is to explain the original, where this count is relatively small for the model of the Neanderthal man (31.6 %). Finally, the relative count of level III ranges from 16.6 % (model of predators and preys) to more than 25 % (27.1 % with respect to the model of a palm leaf and 27.2 % with respect to the model of the Neanderthal man). This indicates that students understand the predictive nature of models fairly good when an appropriate model is presented. The thresholds additionally support this conclusion. They underline, for instance, that it is relatively ‘easy’ to answer on level III when the model of the Neanderthal man ( $\delta_2 = 0.30$ ) or the model of crossbreeding of mice is shown in the task stem ( $\delta_2 = 0.48$ ).

Summarizing, the present findings clearly indicate that students’ understanding of the purpose of models varies across different task context. Results from other studies seem to support these findings: Next to general questions about models, Grosslight *et al.* (1991) also asked their respondents about four different models. The interviewees referred to different criteria to decide if these items are models or not ‘depending on what item they were shown’ (p. 817). Similarly, Justi and Gilbert (2003) asked teachers about their views of models and modeling in science and worked out different aspects (e.g. *use of models*; cf. above), each with several (sub-)categories. Justi and Gilbert (2003) used different models in their research and pointed out that the teachers expressed the subcategories for each aspect while referring to different models. However, neither Grosslight *et al.* (1991) nor Justi and Gilbert (2003) describe the effect of different models on the respondents’ understanding in detail.

With respect to students’ responses to the *decontextualized* task, the results indicate that students’ *general* understanding of biological models is primarily on level I and II. In school, models are primarily used as teaching tools to show or to explain something but sparsely as a method to explore new phenomena (cf. Danusso *et al.*, 2010). There are several authors who argue that students’ understanding of models is determined by the experience students have



made with models in school (e.g. Gobert *et al.*, 2011; Ingham & Gilbert, 1991; Treagust *et al.*, 2002). Furthermore, studies suggest that students' *general* understanding of the purpose of models in science is primarily connected with the descriptive nature of models (e.g. Grosslight *et al.*, 1991). Consequently, the usage of *decontextualized* questions is likely to reveal primarily the role of models in students' science classes. Similarly, Guerra-Ramos (2012) argues that *decontextualized* questions about NOS are problematic since '[t]he lack of context in questions makes [it] difficult to gain insights on the instances that respondents may have in mind when answering a particular item' (p. 640).

The results of the present study are in line with those reported by Grosslight *et al.* (1991) since they, next to other things, asked *decontextualized* questions about models and modeling and approximately half of the respondents saw the purpose of models in showing something. However, there are other studies which conclude that students' understanding of models may be more elaborated (e.g. Chittleborough *et al.*, 2005). Furthermore, Treagust *et al.* (2002) found inconsistencies in students' understanding of models: For example, Australian year 8 and year 9 students think of models as exact replicas but are simultaneously aware of the multiplicity of scientific models. The authors conclude that these inconsistencies occur because the students are aware of different purposes regarding different models (e.g. school models vs. scientific models) and argue: 'By highlighting these subtle differences between different types of models, they may be used more effectively in teaching and learning science' (p. 366). The present findings support this insistence. It has been shown that students' understanding of the purpose of models varies across different task contexts. It is also emphasized in theory that there is no purpose of biological models *per se* but that the development of a model depends on the modeler's objective (e.g. Harré, 1970; Leonelli, 2007; Odenbaugh, 2005).

The present findings may be used to develop more accurate assessment instruments related to models and modeling in biology. Clough and Driver (1986) argue that the

interpretation of a given task depends on both formal concepts and experienced-based intuition concerning the respective task context. Guerra-Ramos (2012) questions the use of *decontextualized* tasks. Hence, having the aim of assessing students' understanding of models and modeling in biology, the effect of the model which is presented in the task stem should be considered carefully. Using only some kinds of models in assessment instruments (e.g. three-dimensional functional models) may give researchers insight in students' understanding of this kind of models. The present findings suggest that more than 25 % of the students have a fairly elaborate understanding of the purpose of such models, but only less than 20 % of the students show this understanding in the context of models with diagrams, for example. In contrast to such an approach, a broad range of models may be implemented in assessment instruments to analyze the consistency of students' understanding across different kinds of models.

Since both, being a model *of* something and being a model *for* something, are constructive features of all models (Halloun, 2006; Mahr, 2008), context-related analyses may give practitioners fruitful insights in students' understanding of biological models. If, for example, students think of a given model as a descriptive entity, teachers could explicitly point out and discuss the predictive nature of the given model – and *vice versa*. The present findings show that students understand level III relatively good in the context of the model of the palm leaf. Hence, such models may be used to introduce this concept (cf. Treagust *et al.*, 2002). However, Treagust *et al.* (2004) distinguish between students' theoretical and practical understanding of models and it is said that the forced choice tasks assess the former. Therefore, the questionnaire can only give some initial clues concerning students' practical understanding of the purpose of biological models. Qualitative (quasi-)experimental research would be necessary to expose the psychological and epistemological processes when learning about models and about using models in biology lessons. Furthermore, students were 'forced' to choose one preferred purpose of models in each forced choice task. This helps to uncover even

slight differences in students ranking of the different purposes (cf. Böckenholt, 2004). However, as a result of the task format, the differences between the task contexts might be overestimated. Therefore, additional investigations using rating scales or similar task formats might be helpful to evaluate the effect of the task format in the present study.

In the present study, the students had to choose between the purpose of a model to show (level I), to explain (level II), or to predict something (level III; Upmeier zu Belzen & Krüger, 2010). Similar theoretical differentiations between purposes of models have also been made by others (Table 1). Within a normative frame it is suggested that students should be fostered towards an elaborated understanding of level III. With a focus on NOS, Guerra-Ramos (2012) criticizes such normative approaches since they do not consider the experience of their respondents. For example, teachers may not have enough opportunities to learn about NOS (Guerra-Ramos, 2012) and students may not have opportunities to experience the scientific nature of models – the latter is suggested by studies about the dominant use of models in school (cf. Danusso *et al.*, 2010). Hence, by contrasting the respondents' (e.g. students') understanding with a normative (e.g. scientific) expectation one might get to know that students do not have a scientific understanding of models and modeling. Instead of contrasting students with scientists or philosophers, the present study conducted a more differentiated analysis of students' responses across different task contexts.

Svoboda and Passmore (2011) criticize the theoretical restriction to only three main model-purposes (cf. Table 1). Referring to Odenbaugh (2005), they claim to consider the diversity of models and model-purposes even in science education to achieve adequate assessment and teaching. Consequently, the present findings may be extended in future research based on more differentiated theoretical descriptions of model-purposes. For instance, level III may be differentiated based on Odenbaugh's (2005) description of four different prospective purposes of models in science (Table 1).

Seven tasks have been used in the present study: One *decontextualized* and six *contextualized* task contexts. The models which are described in the *contextualized* tasks are quite diverse, for example they have different modes of representations, different dimensionalities, and they are from different biological disciplines. This ensures to capture an appropriate part of the broad range of different models which are used in biology. Nevertheless, a constraint had to be made due to economic reasons. Using even more different models may reveal more differentiated insights in students' understanding of biological models. For example, only one model organism has been used in the task stems but a great number of these can be found in biological research (Harré, 2009). However, even with these constraints, the present study has shown that students' understanding of biological models varies, as suggested by theoretical literature (e.g. Leonelli, 2007), across different models. Hence, descriptions of students' understanding of models and modeling should take these differences into account. This could be done by either emphasizing which kind of model students understand in a specific way or by using a broad range of different models to describe students' understanding of specific models.

## References

- Al-Balushi, S. (2011). Students' evaluation of the credibility of scientific models that represent natural entities and phenomena. *International Journal of Science and Mathematics Education, 9*, 571–601.
- Arvisenet, G., Billy, L., Poinot, P., Vigneau, E., Bertrand, D., & Prost, C. (2008). Effect of apple particle state on the release of volatile compounds in a new artificial mouth device. *Journal of Agricultural and Food Chemistry, 56*, 3245–3253.
- Bailer-Jones, D. (2003). When scientific models represent. *International Studies in the Philosophy of Science, 17*, 59–74.
- Black, M. (1962). *Models and metaphors*. Ithaca, NY: Cornell U.P.

- Böckenholt, U. (2004). Comparative judgments as an alternative to ratings: Identifying the scale origin. *Psychological Methods*, 9, 453–465.
- Braithwaite, R. (1962). Models in the empirical science. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science. Proceedings of the 1960 international congress* (pp. 224–231). Stanford: Stanford University Press.
- 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.
- Chabalengula, V. & Mumba, F. (2012). Promoting biological knowledge generation using model-based inquiry instruction. *International Journal of Biology Education*, 2, 1-24.
- Chittleborough, G., Treagust, D., Mamiala, T., & Mocerino, M. (2005). Students' perceptions of the role of models in the process of science and in the process of learning. *Research in Science and Technological Education*, 23, 195–212.
- Clough, E. & Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education*, 70, 473–496.
- Cochran, W. & Cox, G. (1957). *Experimental designs*. New York, NY: John Wiley & Sons.
- Crawford, B. & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 309–323). Dordrecht: Springer.
- Danusso, L., Testa, I., & Vicentini, M. (2010). Improving prospective teachers' knowledge about scientific models and modelling: Design and evaluation of a teacher education intervention. *International Journal of Science Education*, 32, 871–905.
- Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18.
- Gentner, D. & Stevens, A. (Eds.). (1983). *Mental models*. Hillsdale & London: Erlbaum.
- Giere, R. (2002). Models as parts of distributed cognitive systems. In L. Magnani & N. Nersessian (Eds.), *Model-based reasoning. Science, technology, values* (pp. 227–241). New York, NY: Kluwer Academic.
- Giere, R., Bickle, J., & Mauldin, R. (2006). *Understanding scientific reasoning*. London: Thomson Learning.
- Gilbert, J., Boulter, C., & Elmer, R. (2000). Positioning models in science education and in

- design and technology education. In J. Gilbert & C. Boulter (Eds.), *Developing models in science education* (pp. 3–18). Dordrecht: Kluwer Academic.
- Gilbert, J. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education*, 2, 115–130.
- Gilbert, J. (2006). On the nature of “context” in chemical education. *International Journal of Science Education*, 28, 957–976.
- Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28, 73–79.
- Gobert, J., O'Dwyer, L., Horwitz, P., Buckley, B., Levy, S., & Wilensky, U. (2011). Examining the relationship between students' understanding of the nature of models and conceptual learning in biology, physics, and chemistry. *International Journal of Science Education*, 33, 653–684.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28, 799–822.
- 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: Benefits and challenges* (pp. 53–65). Braga: CIEC, Universidade do Minho.
- Guala, F. (2002). Models, simulations, and experiments. In L. Magnani & N. Nersessian (Eds.), *Model-based reasoning. Science, technology, values* (pp. 59–74). New York, NY: Kluwer Academic.
- Guerra-Ramos, M. (2012). Teachers' ideas about the Nature of Science: A critical analysis of research approaches and their contribution to pedagogical practice. *Science & Education*, 21, 631–655.
- Halloun, I. (2006). *Modeling theory in science education*. Dordrecht: Springer.
- Halloun, I. (2007). Mediated modeling in science education. *Science & Education*, 16, 653–697.
- Harré, R. (1970). *The principles of scientific thinking*. London & Basingstoke: Macmillan and Co Ltd.
- Harré, R. (2009). *Pavlov's dogs and Schrödinger's cat: Tales from the living laboratory*. Oxford: Oxford University Press.

- Hesse, M. (1966). *Models and analogies in science*. Notre Dame, IN: Notre Dame University Press.
- Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics*, 60, 732–748.
- Hicks, L. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, 74, 167–184.
- Ingham, A. & Gilbert, J. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, 13, 193–202.
- Justi, R. & Gilbert, J. (2003). Teacher's views on the nature of models. *International Journal of Science Education*, 25, 1369–1386.
- Kandel, E. (1983). From metapsychology to molecular biology: Explorations into the nature of anxiety. *American Journal of Psychiatry*, 140, 1277–1293.
- Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion. [The modeling of scientific competence in primary school age: Theoretical conception and test construction.] *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 265–283.
- Krell, M., Upmeyer zu Belzen, A., & Krüger, D. (2012). Assessment of students' concepts of models and modeling: Empirical evaluation of a model of model competence. In C. Bruguière, A. Tiberghien, & P. Clément (Eds.), *Ebook proceedings of the ESERA 2011 conference*. (pp. 68-74). Retrieved from: [http://lsg.ucy.ac.cy/esera/e\\_book/base/ebook/ebook-esera2011.pdf](http://lsg.ucy.ac.cy/esera/e_book/base/ebook/ebook-esera2011.pdf).
- Laubichler, M. & Müller, G. (Eds.). (2007). *Modeling biology*. Cambridge, MA: MIT.
- Leach, J., Millar, R., Ryder, J., & Séré, M.-G. (2000). Epistemological understanding in science learning: The consistency of representations across contexts. *Learning and Instruction*, 10, 497–527.
- Leonelli, S. (2007). What is in a model? Combining theoretical and material models to develop intelligible theories. In M. Laubichler & G. Müller (Eds.), *Modeling biology* (pp. 15–35). Cambridge, MA: MIT.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54, 421–431.



- MacLellan, H. (1996). *Situated learning perspectives*. Englewood Cliffs, NJ: Educational Technology.
- Magnani, L. (1999). Model-based creative abduction. In L. Magnani, N. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (p. 219–238). New York, NY: Kluwer Academic.
- Magnani, L. & Nersessian, N. (Eds.). (2002). *Model-based reasoning*. New York, NY: Kluwer Academic.
- Mahr, B. (2008). Ein Modell des Modellseins. [A model of model-being.] In U. Dirks & E. Knobloch (Eds.), *Modelle* (pp. 187–218). Frankfurt am Main: Peter Lang.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.
- Masters, G. (2010). The partial credit model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 109–122). New York, NY: Routledge.
- Morgan, M. & Morrison, M. (Eds.). (1999). *Models as mediators*. Cambridge: Cambridge University Press.
- Murcia, K. & Schibeci, R. (1999). Primary student teachers' conceptions of the nature of science. *International Journal of Science Education*, *21*, 1123–1140.
- Nehm, R. & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*, 237–256.
- Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology & Philosophy*, *20*, 231–255.
- Oh, P. & Oh, S. (2011). What teachers of science need to know about models: An overview. *International Journal of Science Education*, *33*, 1109–1130.
- Penner, D., Giles, N., Lehrer, R., & Schauble, L. (1997). Building functional models: Designing an elbow. *Journal of Research in Science Teaching*, *34*, 125–143.
- Prins, G., Bulte, A., Van Driel, J., & Pilot, A. (2009). Selection of authentic modelling practices as contexts for chemistry education. *International Journal of Science Education*, *30*, 1867–1890.
- Schwartz, R. & Lederman, N. (2005). What scientists say: Scientists' views of models. Paper presented as part of the symposium, "Learning about models and modeling in science: International views of research issues" at the annual meeting of the American Educational Research Association, April 12, 2005. Montreal, Canada.



- Schwarz, C. & White, B. (1998). Fostering middle school students' understanding of scientific modeling. *Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.*
- Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B., & 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*, 632–654.
- Smith, R. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, *1*, 199–218.
- Son, J. & Goldstone, R. (2009). Contextualization in perspective. *Cognition and Instruction*, *27*, 51–89.
- Song, J. & Black, P. (1991). The effects of task contexts on pupils' performance in science process skills. *International Journal of Science Education*, *13*, 49–58.
- Suckling, C., Suckling, K., & Suckling, C. (1978). *Chemistry through models: Concepts and applications of modelling in chemical science and industry*. Cambridge, MA: Cambridge University Press.
- Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and the philosophy of science: Proceedings of the 1960 international congress* (pp. 252–261). Stanford: Stanford University Press.
- Svoboda, J. & Passmore, C. (2011). The strategies of modeling in biology education. *Science & Education*, Advance online publication. doi: 10.1007/s11191-011-9425-5
- Terzer, E., Krell, M., Krüger, D. & Upmeier zu Belzen, A. (2011). Assessment of students' concepts of models and modelling using multiple- and forced-choice items [Abstract]. In C. Bruguière (Ed.), *Science learning & citizenship. 9th international conference ESERA 2011* (p. 220). Lyon: Lyon Ingenierie Projets.
- Treagust, D., Chittleborough, G., & Mamiala, T. (2002). Student's understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24*, 357–368.
- Treagust, D., Chittleborough, G., & Mamiala, T. (2004). Students' understanding of the descriptive and predictive nature of teaching models in organic chemistry. *Research in Science Education*, *34*, 1–20.

- Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57. Retrieved from: [http://www.ipn.uni-kiel.de/zfdn/pdf/16\\_Upmeier.pdf](http://www.ipn.uni-kiel.de/zfdn/pdf/16_Upmeier.pdf).
- Urhahne, D., Kremer, K., & Mayer, J. (2011). Conceptions of the nature of science - Are they general or context specific? *International Journal of Science and Mathematics Education*, 9, 707–730.
- 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, 469–488.
- Van Driel, J. & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *International Journal of Science Education*, 21, 1141–1153.
- Van Driel, J. & Verloop, N. (2002). Experienced teacher's knowledge of teaching and learning of models and modeling in science education. *International Journal of Science Education*, 24, 1255–1277.
- Van Oers, B. (1998). From context to contextualization. *Learning and Instruction*, 8, 473-488.
- Wimsatt, W. (1987). False models as means to truer theories. In M. Nitecki & A. Hoffman (Eds.), *Neutral models in biology* (pp. 23–55). New York, NY: Oxford University Press.
- Wu, M., Adams, R., & Wilson, M. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, Vic: ACER Press.