Res Sci Educ (2014) 44:109–132
DOI 10.1007/s11165-013-9365-y

Students' Levels of Understanding Models and Modelling in Biology: Global or Aspect-Dependent?

Moritz Krell · Annette Upmeier zu Belzen · Dirk Krüger

Published online: 11 August 2013
© Springer Science+Business Media Dordrecht 2013

Abstract It is argued that knowledge about models is an important part of a profound understanding of Nature of Science. Consequently, researchers have developed different 'levels of understanding' to analyse students', teachers', or experts' comprehension of this topic. In some approaches, *global* levels of understanding have been developed which mirror the idea of an understanding of models and modelling as a whole. Opposed to this, some authors have developed levels of understanding for distinct aspects concerning models and modelling in science (i.e. *aspect-dependent* levels). This points to an important issue for science education research since *global* conceptualisations might lead to less differentiated assessments and interventions than *aspect-dependent* ones. To contribute to this issue, the article summarises conceptualisations of both *global* and *aspect-dependent* levels of understanding models and modelling that have been developed in science education. Further, students' understanding of the aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* has been assessed (N=1,180; 11 to 19 years old; secondary schools; Berlin, Germany). It is discussed to what extent the data support the notion of *global* or *aspect-dependent* levels of understanding models and modelling in science. The results suggest that students seem to have a complex and at least partly inconsistent pattern of understanding models. Furthermore, students with high nonverbal intelligence and good marks seem to have a comparatively more consistent and more elaborated understanding of models and modelling than weaker students. Recommendations for assessment in science education research and teaching practice are made.

Keywords Models and modelling · Levels of understanding · Latent class analysis · Biology education

M. Krell (✉) · D. Krüger
Biology Education, Freie Universität Berlin, Schwendenerstraße 1, 14195 Berlin, Germany
e-mail: moritz.krell@fu-berlin.de

A. Upmeier zu Belzen
Biology Education, Humboldt Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany

Introduction

The importance of models and the modelling process for scientific communication and for scientific reasoning is described extensively (e.g. Adúriz-Bravo 2012; Coll and Lajium 2011; Giere 1988, 2006; Harré 1970; Suckling et al. 1978). According to philosophical authors, the purpose of models in science is multifaceted. For example, models support the development of scientific theories, they facilitate the testing and revision of scientific theories, they mediate between theory and phenomenon, they substitute theories, they allow the description and interpretation of data, they are part of scientific explanations, they guide experimentation, and they promote creative insight and imagination (e.g. Bailer-Jones 2003; Frigg and Hartmann 2006; Magnani and Nersessian 2002; Morgan and Morrison 1999; Odenbaugh 2005).

Modelling is not only important for scientific communication and reasoning but is also seen as 'a key process in teaching and learning science' (Acher et al. 2007, p. 399). Therefore, models are 'effective pedagogical tools' for teaching scientific literacy (Halloun 2007, p. 653). Ever since the importance of models for science education has been recognised (e.g. Clement and Rea-Ramirez 2008; Gilbert 1991, 2004; Gilbert and Boulter 1998, 2000; Justi and Gilbert 2000; Khine and Saleh 2011; Koponen 2007; Windschitl et al. 2008), many studies have been conducted concerning this issue. For instance, Gericke and colleagues analyse the way models are described in science textbooks (Gericke and Hagberg 2010) and its relation with students' concepts of models and modelling in science (Gericke et al. 2012). Wang and Barrow (2011) investigate students' ability to think with mental models. Furthermore, several studies analyse students' (Chittleborough et al. 2005; Grosslight et al. 1991; Grünkorn et al. 2011, 2012; Treagust et al. 2002), prospective teachers' (Crawford and Cullin 2004, 2005), experienced teachers' (Van Driel and Verloop 1999, 2002; Justi and Gilbert 2003), or scientists' (Van Der Valk et al. 2007) understanding of models and modelling in science. In some of these approaches, the respondents' understanding of models is described in terms of different 'levels of understanding' (e.g. Crawford and Cullin 2005; Grosslight et al. 1991; Grünkorn et al. 2011). For example, Grosslight et al. (1991) developed three *global* levels 'of thinking about models, reflecting different epistemological views about models and their use in science' (p. 817). In contrast to the development of such *global* levels, *aspect-dependent* levels of understanding models and modelling in science have been developed, too. For example, Crawford and Cullin (2005) differentiate between five aspects of models and modelling in science and propose levels of understanding for each aspect. In sum, it does not seem clearly investigated whether the assumption of *global* levels of understanding models and modelling is appropriate for science education or not.

The importance of a theoretical conceptualisation of the understanding of models and modelling in science arises from the perspective of both science teaching and assessment in science education. Concerning teaching about models and modelling in science, Oh and Oh (2011) recently developed five aspects (called 'subtopics') of models and modelling which they argue should be known by science teachers. Further, if a deep understanding of models and modelling means to be aware of different aspects, students should study and discuss such aspects in lessons about models and modelling. Regarding assessment, the theoretical conceptualisation affects the outcome of a measurement—especially in quantitative research. Therefore, the appreciation of different aspects of models and modelling allows a differentiated insight into respondents' understanding of this topic. However, there is always a trade-off between a highly differentiated theoretical conceptualisation and economic tenets (the 'principle of parsimony'; cf. Burnham and Anderson 2000). Therefore, empirical evidence is needed to get insights into what extent the most parsimonious conceptualisation of the understanding of models and modelling in science, the *global* approach, is appropriate to be used in science

education research and practice. The present study contributes to this issue by analysing students' level of understanding of different aspects of models and modelling in biology by analysing students' patterns of understanding. In such an approach, the notion of *global* levels of understanding would be supported if a consistent pattern of understanding across different aspects was found (Justi and Gilbert 2003). However, inconsistencies in students' understanding between different aspects would support the idea of levels of understanding which are dependent on different aspects of models and modelling (i.e. *aspect-dependent* levels of understanding models and modelling).

In the following section, we summarise studies in which students' or teachers' understanding of models and modelling in science is either conceptualised as being *global* or *aspect-dependent*. It is concluded that there is still no consensus in science education research about how to conceptualise respondents' levels of understanding models and modelling. To analyse consistencies in students' levels of understanding across several aspects, the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* are distinguished in the present study. It is pointed out to what extent these five aspects are also developed by other researchers in science education and thus might be seen as important aspects of the understanding of models and modelling in science.

Theoretical Background

Levels of Understanding Models and Modelling

In an exploratory study, Grosslight et al. (1991) assessed students' understanding of models and modelling in science. The authors interviewed thirty-three 7th formers and twenty-two 11th formers and asked them both general questions about models and modelling in science and questions about concrete instances (for example a toy airplane or a schematic diagram of the water cycle). Based on the interviews, Grosslight et al. (1991) developed three *global* levels of understanding models and modelling in science (cf. Carey and Smith 1993). The levels have been developed based on Carey et al. (1989) who analysed students' epistemological understanding of science. They can be called *global* since 'they summarise the kind of understanding and conceptions of models that emerged from the interview as a whole' (p. 817) and each level includes views about six different aspects of models and modelling, such as the *role of the modeller*, *testing models*, or *multiplicity in model building* (Grosslight et al. 1991). In the understanding of level I, students think of models as being simple copies of reality or toys. Consequently, the purpose of a model is seen to be providing a simplified version of a real object or process. If some parts of reality are not included in a model, students do not reflect on this. In Grosslight et al.'s (1991) level II, students appreciate the construction of a model for a concrete purpose. Hence, the modeller's subjective ideas gain importance, which is why a model does not have to match reality. However, the focus is still on the relation between the model and the corresponding phenomenon and not on the model and underlying ideas or hypotheses. Therefore, the testing of a model is seen as a testing of the workability of the model object itself. Finally, a level III understanding is characterised by the view that the construction of models has the purpose to develop and test ideas and not to provide replicas of reality. Therefore, the active and selective role of the modeller is recognised and models are seen as tools for testing and revising hypotheses about the corresponding phenomenon (Carey and Smith 1993; Grosslight et al. 1991). Since each level summarises views about six different aspects of models and their use in science, students were scored as 'pure' or 'mixed'—'pure' describing students who showed the same level of understanding in at least five (out of six)

aspects and 'mixed' for the other cases. In sum, Grosslight et al. (1991) assigned 67 % of the 7th formers as pure level I, 18 % as mixed level I/II, and 12 % as pure level II. For the 11th formers, 23 % were assigned as pure level I, 36 % as mixed level I/II, and 36 % as pure level II. None of the students reached a level III understanding. Although Grosslight et al. (1991) had to score some students as 'mixed', they do not discuss in which aspects students reached a higher or lower level of understanding, leading to a 'mixed' overall score. Furthermore, the authors do not describe the different aspects in each level of understanding in detail.

The three levels developed by Grosslight et al. (1991) were used by Crawford and Cullin (2005) to analyse the success of teaching units with dynamic computer modelling. However, the authors expanded and refined Grosslight et al.'s three levels in order to describe *aspect-dependent* levels of understanding models and modelling in science. Based on a review of relevant literature (e.g. Van Driel and Verloop 1999; Justi and Gilbert 2003), Crawford and Cullin (2005) distinguished the five aspects (called 'dimensions'): *Purpose of models*, *designing and creating models*, *changing a model*, *multiple models for the same thing*, and *validating/testing models*. Based on the differentiation between a *limited*, a *pre-scientific*, an *emerging scientific*, and a *scientific* understanding, Crawford and Cullin (2005) developed three to four levels of understanding within each aspect: Three levels are described in the aspects *designing and creating models* and *validating/testing models* since no *limited* understanding is developed by the authors in these aspects, whereas four levels of understanding are developed for the other aspects. Crawford and Cullin (2005) were able to describe positive changes in their respondents' understanding of models and modelling after dynamic computer modelling. However, the subjects 'tenaciously held on to some scientifically uninformed views' (p. 318). The authors conclude that the description of *aspect-dependent* levels of understanding helps to uncover teachers' understanding in detailed way.

Justi and Gilbert (2003) asked teachers about models and worked out seven aspects which are similar but not identical to the ones described by Grosslight et al. (1991) and Crawford and Cullin (2005): *Nature of models*, *use of models*, *entities of which the models are composed*, *uniqueness of models*, *stability over time*, *making predictions*, and *accreditation of models*. Justi and Gilbert (2003) explicitly address the question of *global* levels in their analysis and conclude that 'data provide no support for the notion of "level" in the teachers' understanding of the notion of "model"' (p. 1381). In fact, most of the teachers showed a complex pattern of understanding which varied across the different aspects and therefore did not match the idea of *global* levels of understanding developed by Grosslight et al. (1991) (Justi and Gilbert 2003).

In the work by Schwarz and colleagues, a learning progression about models and modelling in science with the two major dimensions *models as generative tools for predicting and explaining* and *models change as our understanding improves* has been developed (e.g. Baek et al. 2011; Schwarz et al. 2009, 2012; Schwarz and White 2005). The first dimension reflects that models are important generative tools which can be used to explain and predict something about the corresponding phenomenon, whereas the second dimension focuses on the changeable nature of models (Schwarz et al. 2009). In addition, four levels of understanding are developed in each dimension which may be seen as *global* levels of understanding since they include different aspects of models and modelling. For example, in the dimension *models as generative tools for predicting and explaining*, level III refers to students who think of models as supportive entities when thinking about scientific phenomena and who analyse different possible models for one phenomenon (Schwarz et al. 2009). Hence, an understanding of the aspects *purpose of models*, *multiple models*, and *validating/testing models* proposed by Crawford and Cullin (2005) is included in this level of understanding. Recently, Schwarz et al. (2012) described strengths and weaknesses of their theoretical approach. Among other

things, the authors stress that in some cases it was not possible to assign students clearly to one level of understanding, since indicators of different levels were found for individual students. Therefore, Schwarz et al. (2012) developed sub-dimensions for each level, which describe students' understanding and acting more closely. For example, the dimension *models as generative tools for predicting and explaining* has been differentiated into four sub-dimensions (*the model's level of abstractness*, *audience and clarity of communication*, *evidence or authority*, *relationship between model and phenomenon*) each with *aspect-dependent* levels of understanding. However, the authors do not describe the four levels of understanding for each sub-dimension in detail.

Danusso et al. (2010) asked teachers three open-ended questions about models (*What is a scientific model?; What are its main components?; What are its main functions?*) and developed a coding scheme based on a combined analysis of the teachers' answers. In this coding scheme, five levels of understanding (called 'macro clusters') are described. At each level, ideas concerning the three questions are subsumed which is why these levels may be regarded as *global* levels. For example, the highest level of understanding (macro cluster one) describes teachers who understand a model as an 'abstract representation of a phenomenon with the aim of characterising and studying it' (p. 886). Danusso et al. (2010) were able to assign each respondent to one of the five levels of understanding. However, the authors state that a 'non-negligible percentage of prospective teachers gave incoherent responses' (p. 885). The authors do not describe this incoherency in detail, but it may hint at similar difficulties in the application of *global* levels of understanding reported by Schwarz et al. (2012).

In sum, the studies sketched out above (Crawford and Cullin 2005; Danusso et al. 2010; Grosslight et al. 1991; Justi and Gilbert 2003; Schwarz et al. 2009, 2012) provide ambiguous answers to the question whether levels of understanding models and modelling should be regarded as *global* or *aspect-dependent*. Whereas Danusso et al. (2010) and Grosslight et al. (1991) developed *global* levels, both Justi and Gilbert (2003) and Crawford and Cullin (2005) were not able to identify *global* levels of understanding across the different aspects. Schwarz et al. (2009, 2012) developed *global* levels for each dimension and reported difficulties in clearly assigning students to one level of understanding. Hence, the more recent studies seem to support the view of Justi and Gilbert (2003) who suggest considering different aspects concerning models and modelling in science and to make allowance for complex patterns of understanding. However, consensus has not yet been reached because many researchers base their investigation on the *global* levels developed by Grosslight et al. (1991) or use them as a reference for the analysis of results (e.g. Chittleborough and Treagust 2007; Harrison 2001; Harrison and Treagust 1996). Hence, there is no consensus in science education research about whether there is an understanding of models and modelling as a whole (i.e. *global* levels) or if the level of understanding varies between different distinct perspectives (i.e. *aspect-dependent* levels). The present article addresses this issue related to a theoretical structure of understanding models and modelling which is called the 'model of model competence' (Upmeier zu Belzen and Krüger 2010).

'Model of Model Competence'

Based on a review of relevant literature (e.g. Crawford and Cullin 2005; Grosslight et al. 1991; Justi and Gilbert 2003) and theoretical reflections about the term 'model' (e.g. Mahr 2008) Upmeier zu Belzen and Krüger (2010) developed the 'model of model competence'. In Germany, this structure is used to provide researchers with an analytical framework to investigate in detail both students' comprehension of models and modelling and the success of teaching units (Fleige et al. 2012; Grünkorn et al. 2011). In addition, practitioners can focus on

the aspects described in the 'model of model competence' while teaching about models and modelling (cf. Oh and Oh 2011). Finally, for the development of test instruments, it is essential to know which aspects are relevant to understanding a given topic (Klieme et al. 2008).

The 'model of model competence' distinguishes the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* (Table 1). Three (*aspect-dependent*) levels of understanding have been proposed for each aspect (Upmeier zu Belzen and Krüger 2010). Hence, the 'model of model competence' is similar to what Crawford and Cullin (2005) proposed with their 'Matrix of Modelling Dimensions'. In the following sections, the levels of understanding as well as the five aspects are described more closely.

Levels of Understanding

Basically, level I and II refer to the 'descriptive nature of models', but level III refers to the 'predictive nature of models' which reflects the scientific perspective of models as research tools (cf. Treagust et al. 2004; Table 1). Hence, three different views concerning each aspect are described in the 'model of model competence' which represent an increasing ability to reflect on models in biology (Upmeier zu Belzen and Krüger 2010). The ascending order of the three levels in each aspect is based both on theoretical-normative and empirical considerations. Level III in each aspect is related to an understanding of models as theoretical reconstructions and research tools in science. *Normatively*, this is seen as a goal for science education (Bybee 2011) and is also established in science education standards documents (e.g. AAAS 1993; KMK 2005; NRC 2000). Therefore, the three levels describe a normative frame about what students (and teachers; cf. Justi and Gilbert 2003; Oh and Oh 2011) should know about models in biology (or science in general). Therefore, students who are able to understand the perspectives described in level III in each aspect are seen to have a more elaborated (i.e. more scientific) understanding of models. Such a normative framework is also proposed by other authors (e.g. Crawford and Cullin 2005; Grosslight et al. 1991). Furthermore, *empirical* findings suggest that the predictive nature of models is understood only by a minority of students (e.g. Grosslight et al. 1991; Grünkorn et al. 2011), and there is research evidence that models are seldomly used as predictive entities in science classes (cf. Danusso et al. 2010). Recently conducted studies based on the 'model of model competence' provide empirical evidence that the three levels in each aspect reflect an increasing degree of difficulty (Krell 2012; Terzer and Upmeier zu Belzen 2011). However, it is not proposed by Upmeier zu Belzen and Krüger (2010) that students (and teachers) should only understand the predictive nature of models (or level III in each aspect). It is assumed that a profound understanding of models and modelling means to understand both

Table 1 The 'model of model competence' by Upmeier zu Belzen and Krüger (2010)

	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	Different model objects	Different foci on the original	Different hypotheses about the original
Purpose of models	Describing the original	Explaining the original	Predicting something about the original
Testing models	Testing the model object	Compare the model and the original	Testing hypotheses about the original
Changing models	Correcting defects of the model object	Revise due to new insights	Revise due to the falsification of hypotheses about the original

their descriptive and their predictive nature. In the following, the five aspects by Upmeier zu Belzen and Krüger (2010) are sketched out and compared with similar approaches of other researchers in science education.

Nature of Models

Under the aspect *nature of models*, different ontological and epistemological beliefs about the relation of a model to its original are described, similar to the description by Grosslight et al. (1991): A model may be seen as an exact copy of the original (level I), or the possibility (level II) or necessity (level III) of differences between the model object and the original may be realised (Upmeier zu Belzen and Krüger 2010; cf. Grünkorn et al. 2011). Similarly, Grosslight et al. (1991) described the aspect *kinds of models*, which refers to the relationship between the model and its original but additionally include views about what can be modelled (objects, abstractions) as well as what a model can be (e.g. objects, visual entities). Treagust et al. (2002) developed tasks with rating scales and analysed the corresponding data using an exploratory factor analysis (cf. Gobert et al. 2011). This approach resulted in five factors, one of them labelled *models as exact replicas* which includes different views about the extent to which a model is an exact replica of its original. Similar to the aspect *nature of models* by Upmeier zu Belzen and Krüger (2010), Justi and Gilbert (2003) developed the aspects *nature* and *entities*. The first describes different ontological beliefs about models (e.g. a model is a reproduction of something, a model is a mental image). The second is related to different possible originals of a model (e.g. an object, a process, an idea). Schwarz et al. (2009) described the aspect *models are generative tools for predicting and explaining* as an important aspect of meta-modelling knowledge. Recently, Oh and Oh (2011) wrote a review article concerning models and modelling in science education and developed five aspects which they argue should be understood by science teachers. One of them is the aspect *meanings of a model* which is similar to the aspect *kinds of models* by Grosslight et al. (1991) since it includes ideas about what can be modelled and how these entities (objects, processes, or ideas) can be modelled (Table 2).

Multiple Models

The aspect *multiple models* focuses on the existence of different models representing the same original (Upmeier zu Belzen and Krüger 2010; cf. Grünkorn et al. 2011). Hence, different arguments against the idea of a model's *uniqueness* (Justi and Gilbert 2003) are summarised in this aspect (cf. Crawford and Cullin 2005; Treagust et al. 2002). In level I, students explain differences between several models of the same phenomenon as differences between the model objects themselves (e.g. literally different views, different angles, different mode/symbols; Grosslight et al. 1991, p. 814). In level II, different models are seen as ways to represent different parts of the original (Grünkorn et al. 2011). In level III, models are seen as predictive tools. Consequently, differences between various model objects occur because each is made for a different application (e.g. testing of another hypothesis; cf. Crawford and Cullin 2005). This aspect was also developed by Oh and Oh (2011) who called it the *multiplicity of scientific models*.

Purpose of Models

The aspect *purpose of models* summarises different views about the purpose of a model. Similar to Treagust et al. (2004), three different purposes of models are distinguished by

Table 2 Different aspects of models and modelling in science

Reference	Aspects of models and modelling				
Upmeier zu Belzen and Krüger (2010)	Nature of models	Multiple models	Purpose of models	Testing models	Changing models
Grosslight et al. (1991)[a]	Kinds of models	Multiple models	Purpose of models	(purpose of models)	Changing a model
Treagust et al. (2002)	Models as exact replicas	Models as multiple representations	Uses of scientific models; models as explanatory tools	–	Changing nature of models
Justi and Gilbert (2003)[a]	Nature; entities	Uniqueness	Use; prediction	–	Time
Crawford and Cullin (2005)[a]	–	Multiple models for the same thing	Purpose of models	Validating/testing models	Changing a model
Schwarz et al. (2009)[a]	Models are generative tools	–	Using models to explain and to predict	Evaluating models	Revising models; models change to capture improved understanding
Oh and Oh (2011)[a]	Meanings of a model	Multiplicity of scientific models	Purposes of modelling	(change in scientific models)	Change in scientific models

The references are arranged by year of publication. The theoretical framework of the present study presented in the first line. Phrases in brackets indicate that the authors discuss the testing of models within another aspect (e.g. within the aspect *purpose of models*; Grosslight et al. 1991).

[a] Additional aspects are developed (cf. the discussion section)

Upmeier zu Belzen and Krüger (2010): Models may be used to show (level I), to explain (level II), or to predict (level III). Hence, the scientific use of models as research tools is seen as the most elaborated understanding of the purpose of models (cf. Crawford and Cullin 2005; Grosslight et al. 1991). The aspect *purpose of models* has also been developed by Grosslight et al. (1991), Treagust et al. (2002), Justi and Gilbert (2003), Crawford and Cullin (2005), Schwarz et al. (2009), as well as Oh and Oh (2011). Specific purposes are highlighted by some authors, for example the use of *models as explanatory tools* (Treagust et al. 2002), *prediction* (Justi and Gilbert 2003), or both (Schwarz et al. 2009) (Table 2).

Testing Models and Changing Models

In the aspects *testing models* and *changing models* as developed by Upmeier zu Belzen and Krüger (2010), level I focuses on the testing or changing of a model object itself, level II on the comparison of a model with its original, and level III on the application of a model. Again, level I and level II refer to the descriptive nature of models, which focuses on 'models *of* known things and processes' (Harré 1970, p. 40; italics in the original). Therefore, at level II, the testing of a model is seen as a comparison between *already known* properties of the original and the model. As a consequence of this the model may be changed, for example, to include additional information, which was not considered when the model had been made (Upmeier zu Belzen and Krüger 2010). In contrast, level III refers to the predictive nature of models or 'models *for* unknown mechanisms' (Harré 1970, p. 40; italics in the original). From this point of view, models are hypothetical entities which are used to make predictions about the original (Odenbaugh 2005; Van Der Valk et al. 2007). Hence, the model is tested by evaluating the generated predictions concerning the corresponding scientific phenomenon and would be altered if these predictions turn out to be inappropriate (Upmeier zu Belzen and Krüger 2010). Both *testing models* and *changing models* are also developed by others (e.g. Crawford and Cullin 2005; Schwarz et al. 2009; Treagust et al. 2002). Grosslight et al. (1991) only described the aspect *changing models* but included different ideas about how a model might be tested in the aspect *purpose of models*. Justi and Gilbert (2003) developed the aspect *time*, which refers to reasons why a model might change over time. Finally, Oh and Oh (2011) did not develop an extra aspect *testing models* but discuss the testing of models within the aspect *change in scientific models* (Table 2).

Research Question and Hypotheses

The research question addresses the issue of *global* levels of understanding models and modelling in biology: To what extent do students have a consistent understanding of the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*?

As outlined above, Grosslight et al. (1991) developed three *global* levels of understanding models and modelling in science (cf. Carey and Smith 1993). However, Grosslight et al. (1991) were not able to assign all respondents clearly to one level of understanding. Furthermore, there are others who developed *aspect-dependent* levels of understanding (e.g. Crawford and Cullin 2005) and Justi and Gilbert (2003) found no support for the notion of global levels in their data. Therefore, one can formulate the hypothesis (*H 1*) as follows:

H 1 The preference of level I, II, and III varies across the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*.

Referring to Grosslight et al. (1991), it is likely that students' understanding of models increases at higher levels of education. Furthermore, studies have shown that the understanding of models and modelling is positively related to students' science learning (Gobert and Pallant 2004; Schwarz and White 2005) and to the depth of students' cognitive processing (Sins et al. 2009). Consequently, students with better marks may have a more elaborated understanding of models and modelling in science. Clough and Driver (1986) argue that students' response behaviour, in general, may be affected by pre-existing, experienced-based, and intuitive ideas as well as by formally acquired stable concepts. Hence, the extent of consistency in students' response behaviour may be seen as an indicator for stable concepts concerning the assessed issue. Hence, hypothesis two (*H 2*) is as follows:

H 2 Students with better marks and better cognitive abilities have a more consistent pattern of understanding models and modelling than other students.

Methods

Instrument of Assessment

Task Format

On the basis of the theoretical structure of model competence by Upmeier zu Belzen and Krüger (2010), forced-choice tasks have been deductively developed and qualitatively content validated (Krell and Krüger 2010; Krell et al. 2012). In forced-choice tasks, a set of alternatives are presented and respondents are asked to rank them in an order of preference, for example 'most like I think', 'median like I think', and 'least like I think'. When the alternatives are ranked directly, the task is called a ranking task. Alternatively, respondents get several pairs, selected from the set of alternatives, and have to select the more preferred alternative from each pair. These tasks are called paired-comparison tasks (Brown and Maydeu-Olivares 2011; Hicks 1970; Maydeu-Olivares and Böckenholt 2005). Because in both kinds of tasks respondents are 'forced' to choose the most preferred alternative, ranking tasks and paired-comparison tasks are called forced-choice tasks (Brown and Maydeu-Olivares 2011). There are several advantages of forced-choice tasks. For example, they avoid effects of individually different interpretations of category labels, which may occur in rating tasks, and also tied judgements are avoided completely (Böckenholt 2004). Due to the advantages, the task format has already been used by others to assess students' understanding of science (e.g. Chittleborough et al. 2005; Kleickmann et al. 2010).

In the present study, ranking tasks for the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* (Table 1) were developed. In each task, statements which represent the three levels of understanding of one of the five aspects are presented as alternatives and respondents are asked to rank them in the order of their preference. Figure 1 shows a task for the aspect *purposes of models* for illustrative purpose. In this task, a possible student answer is exemplified. It reflects an understanding of the purpose of the palm leaf's model as explanatory (cf. Table 1). Due to their deductive development (Krell and Krüger 2010), the alternatives in each ranking task are rather abstract representations of the three levels. Therefore, students' responses are seen as indicators for their 'theoretical understanding' of models and modelling (Treagust et al. 2004) or their 'metamodelling knowledge' (Schwarz et al. 2009, 2012) concerning the five aspects. Both terms are used to describe 'a type of nature of science understanding' (Schwarz et al. 2009, p. 634) about models, which is said to be related to, but not identical to, students' modelling ability (Schwarz et al. 2009; Treagust et al. 2004).

Similar distinctions between the knowledge about models on the one hand and modelling skills on the other hand have also been made by other authors (e.g. Gilbert and Boulter 1998; Henze et al. 2007; Justi and Gilbert 2002).

Task Stems

In each task, a task stem introduces a biological model. Next to a short description of a biological phenomenon (i.e. the original) and a model that represents it, one picture of the model and one of the original is shown. Six different task stems have been developed, which refer to different biological models: In the first task stem, a model of a palm leaf is pictured, which represents the structure of a folded palm leaf (Fig. 1).

The second task stem introduces a model of a bee's leg. This model refers to the hairy surface of the leg and its function in collecting pollen. In the third task stem, a model of a plant seed (*Alsomitra macrocarpa* M. Roem.) is explained and illustrated. It is due to its structure that this seed is able to glide in the air for long distances. The fourth task stem presents a model of the shape of the human foot arch and its relation to the foot's stability.

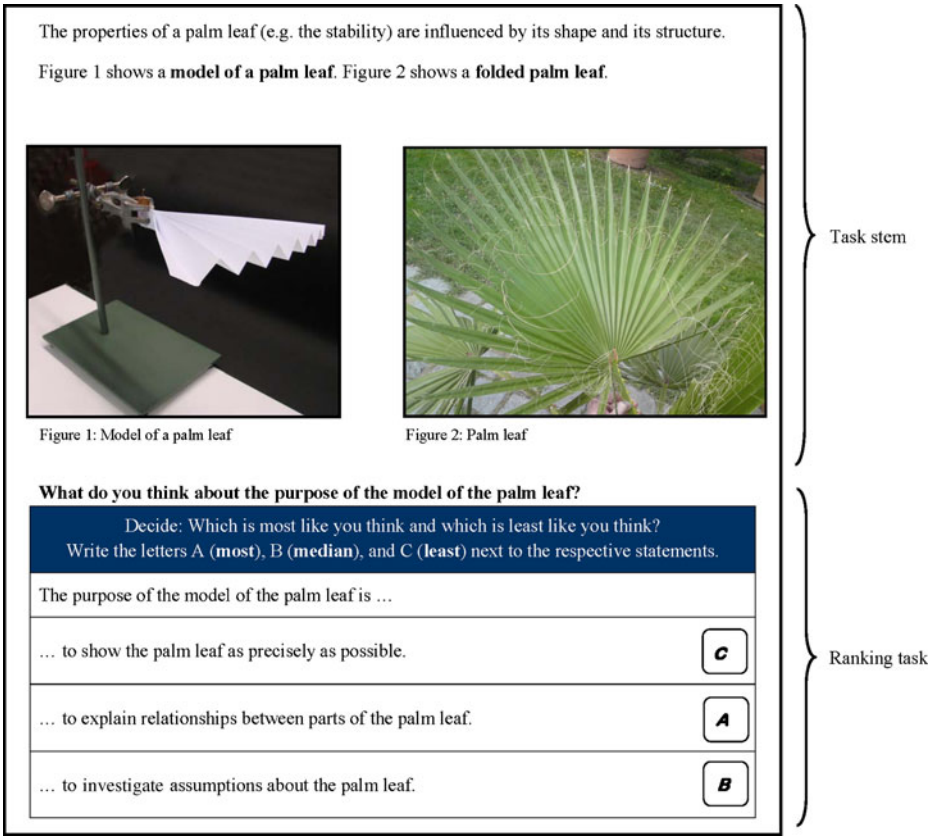

Fig. 1 The ranking task, which refers to the aspect *purpose of models*, and the model of a palm leaf. In the task stem, the model and the biological phenomenon are introduced. In the ranking task, the respondents have to rank the three levels of the respective aspect. For illustrative purposes, one possible student answer is pictured, which reflects the preference of level II (*A*)

The fifth task stem shows and describes a model of a swallow's pinion, which represents the relation between the pinion's shape and its features. Finally, a model of a dragonfly, which especially focuses on the comparatively large and light wings of dragonflies (cf. Eisma 2012), is shown in the sixth task stem. Hence, all six models focus on the relation between the shape or structure and the function of the original (or parts of the original). Therefore, the models may be seen as functional models (Penner et al. 1997). Pictures of the models, which have been used for the development of the task stems two to six, are shown in Fig. 2.

To allow for a non-biased comparison of the five aspects of understanding models and modelling, the six task stems have been used to develop tasks for each aspect. Therefore, six times five stem equivalent ranking tasks have been developed (i.e. tasks with identical task stems; Martinez 1991). For example, there are five tasks (one for each aspect) with a task stem referring to the model of a palm leaf. In sum, this faceted test design (Martinez 1999) keeps the influence of task stems constant but systematically varies the aspect of understanding models and modelling in science.

Sample and Data Collection

Students from secondary public schools currently in 7th to 10th form (*ms*=8.46; *sd*=1.13) and between 11 and 19 years old (*ms*=14.26; *sd*=1.35) in Berlin (Germany) answered the ranking tasks (*N*=1,180), 544 students (44.7 %) were male and 626 students (51.5 %) were female, 10 students (3.8 %) did not specify their gender. As well as students' understanding of models and modelling, data relating to the students' age, sex, form, reading skills, and marks in biology, chemistry, physics, mathematics, German, and the first foreign language were also collected. Regarding reading skills, both reading speed and reading comprehension were assessed using the Reading Speed and Comprehension Test (in German: Lesegeschwindigkeits- und Verständnistest; LGVT) developed by Schneider et al. (2007). In addition, nonverbal intelligence as one aspect of students' cognitive abilities was assessed using a subscale of the Cognitive Abilities Test (in German: Kognitiver Fähigkeitstest; KFT, Skala 'N2').

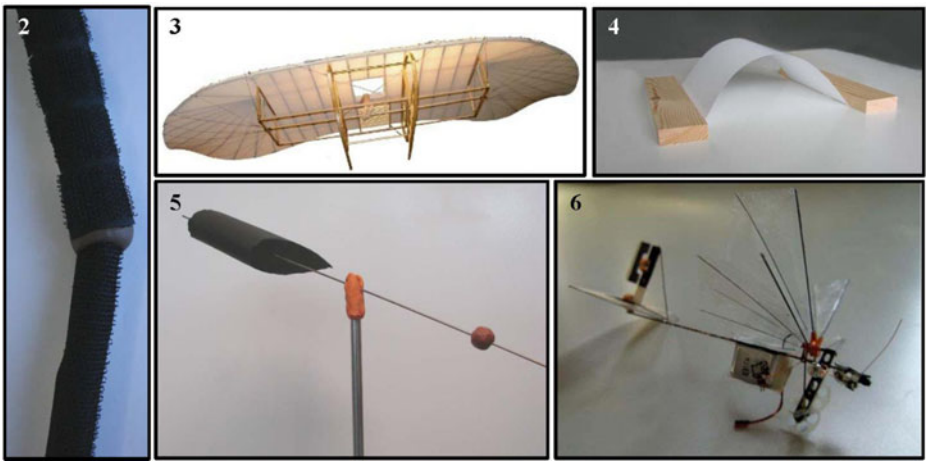

Fig. 2 The biological models which have been shown and explained in the task stems two to six. Note that in these task stems pictures of the corresponding original have also been shown as illustrated in Fig. 1. (© Picture 3: Ökopark Hartberg, Ausstellung Zukunft-Technik lernt von der Natur. © Picture 6: Eisma 2012)

The present study took place within a larger assessment of students' understanding of models and modelling in biology education. In addition to the ranking tasks, the students also answered open-ended (Grünkorn et al. 2011) and multiple-choice tasks (Terzer et al. 2011). These tasks were developed using the same theoretical background as the ranking tasks (i.e. the 'model of model competence'; Upmeier zu Belzen and Krüger 2010). Different task formats were included in the assessment to get a broad insight into students' understanding of models and modelling because different cognitive abilities are needed to solve different task formats (Martinez 1999). However, the present article focuses on the findings from the ranking tasks.

A balanced incomplete block design was developed to keep the total number of tasks for every single student small (Giesbrecht and Gumpertz 2004). For the ranking tasks, 35 different test booklets were developed, each including 6 of 30 ranking tasks (20 %). Consequently, the present study aims to provide information about students' patterns of understanding on the group level rather than investigating students' individual concepts.

Data Analysis

In general, the raw data of a ranking task provides the respondents' individual rank order of the respective alternatives (Brown and Maydeu-Olivares 2011; Coombs 1950; Hicks 1970). Hence, the raw data of the developed ranking tasks provide an insight into which level of each aspect is most, median, and least preferred by the students concerning the six biological functional models. Since no external scale is provided to rate the three levels individually, the data collected by the ranking tasks are called 'ipsative data' (Hicks 1970). However, in a qualitative approach it was found that students' second preference (i.e. 'median like I think') may not be content valid in some cases (Krell et al. 2012). Therefore, the ipsative raw data were transformed into 'partially ipsative data' (Hicks 1970) by scoring only the level that was ranked first (i.e. 'most like I think'). Since the three levels are assumed to be in an ordinal order due to normative and empirical considerations (Upmeier zu Belzen and Krüger 2010), the students were scored with 0 when they ranked level I as 'most like I think', they were scored with 1 when they preferred level II, and with 2 when they preferred level III. Thus, the response pattern exemplified in Fig. 1 would result in a score of 1 for this task. However, as mentioned above, it is argued that students should understand both the descriptive nature of models (levels I and II) and the predictive nature of models (level III). The distinction between level I, level II, and level III modellers is made for analytic purposes in order to be able to describe differences in students' understanding of the five aspects.

Both Hypothesis 1 and Hypothesis 2 address students' pattern of understanding models and modelling related to the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*. To analyse this pattern, a latent class analysis (LCA) was conducted with the software Mplus (Muthén and Muthén 1998–2007). In a LCA, students' responses are analysed on the latent level, all variables are assumed to be (at least) on a nominal level, and there are no restrictions on the kind of relation between the (manifest) variables (Collins and Lanza 2010; Hagenaars and Halman 1989; Langeheine and Rost 1988). Basically, a LCA computes different groups (i.e. latent classes) of students, with each group consisting of students with a response pattern that is as homogenous as possible but different to the response patterns of the other groups. Thus, the aim of a LCA is similar to that of a cluster analysis (Kaufman and Rousseeuw 2005).

Analysing data using a LCA, one has to decide which LCA model fits the data best: 'In LCA a fundamental question involves deciding on the appropriate number of latent classes to use to represent a data set' (Collins and Lanza 2010, p. 87). To compare different LCA

models, Mplus provides the indices AIC, BIC, and ssaBIC. These indices factor in the parsimony, the sample size, and the likelihood of the LCA models—each index in a different manner (Henson et al. 2007):

$$\mathrm{AIC} = -2\log[ML(k)] + 2[m(k)],$$

$$\mathrm{BIC} = -2\log[ML(k)] + \log(n)[m(k)],$$

$$\mathrm{ssaBIC} = -2\log[ML(k)] + \log\Big((n+2)\Big/24\Big)[m(k)].$$

Where -2log[$ML(k)$] represents the likelihood of the LCA model k, n labels the sample size, and $m(k)$ denotes the number of estimated parameter of LCA model k (Henson et al. 2007). The AIC does not include the parameter n. Therefore, this index does not consider the sample size. Both the BIC and the ssaBIC include n, but the ssaBIC, which is a derivate form of the BIC, does not emphasise the parsimony of the LCA model as strongly as the BIC since n is replaced by $(n+2)/24$ (Henson et al. 2007). When comparing different LCA models with these information indices, the smallest value of each index points out the comparatively best LCA model. However, both the BIC and the ssaBIC are 'better indicators of the number of [latent] classes than the AIC' (Nylund et al. 2007, p. 557), which is why these indicators are used in the present study. However, the BIC and the ssaBIC often do not identify the same LCA model as optimal (Collins and Lanza 2010). Therefore, one has to use a combination of different insights to decide how many latent classes represent the data set best (Collins and Lanza 2010).

It is an important characteristic of LCA that the subjects are not assigned to the different latent classes in a deterministic but in a probabilistic sense. For diagnostic purposes, it is common to classify each subject to the latent class with the highest probability of assignment. Therefore, an 'additional indicator [of model-goodness] is the average membership probability within each [latent] class' (Spiel and Glück 2008, p. 52). The higher this probability, the better the LCA model is. Furthermore, one should analyse the item parameters for extreme values, which indicate an estimated probability of 0 or 100 % to solve a task. The fewer extreme values there are, the better the LCA solution is (Spiel and Glück 2008).

Findings

To analyse students' patterns of understanding models and modelling, the appropriate LCA model has to first be selected. Table 3 provides the information indices BIC and ssaBIC for four different LCA models. Both indices indicate the LCA model with two latent classes as the most appropriate to represent the data. The probability of assignment ranges from .82 to .91 for the LCA model with two latent classes and from .78 to .82 (three latent classes), from .76 to .84 (four latent classes), and from .76 to .85 (five latent classes) for the other LCA models. In addition, the number of extreme values increases with the number of latent classes: In the LCA model with two latent classes, there are only four such cases, in the model with three latent classes 36 cases, and even more in the other LCA models (66 and 102).

These findings consistently suggest that the response pattern is best represented using two latent classes. The indices BIC and ssaBIC show that the two latent classes are the comparatively best representation of the data when the likelihood and the parsimony of

Table 3 The information indices BIC and ssaBIC for the four LCA models

	BIC	ssaBIC
Two latent classes	13,947.02	13,562.69
Three latent classes	14,187.26	13,609.17
Four latent classes	14,451.86	13,680.02
Five latent classes	14,746.05	13,780.45

the LCA model as well as the sample size are taken into account. These latent classes have probabilities of assignment of .91 (latent class A) and .82 (latent class B), which indicates that the students, on average, could be assigned to one of the two latent classes with a relatively high certainty. The latent classes consist of about 80 % or 944 students (latent class A) and 20 % or 236 students (latent class B) of the sample.

Analysing Consistencies in Students' Level of Understanding (*H 1*)

H 1 is related to the issue of *global* levels of understanding models and modelling and assumes inconsistencies in students' understanding of the five aspects. The LCA provides two groups (i.e. latent classes) of students. In each latent class, the probability of ranking level I, level II, or level III as 'most like I think' is estimated for the 30 ranking tasks. For further analysis, the mean probabilities for each aspect were computed. Thus, in the following, both latent classes are characterised by their average probabilities of preferring level I, level II, or level III in each aspect. This allows a better comparison between the five aspects than by describing the response behaviour in 30 single tasks. Furthermore, as mentioned above, stem equivalent ranking tasks were developed to allow for a non-biased comparison between the five aspects.

The (average) preference probabilities of level I, II, and III estimated in the LCA vary across the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*. Figure 3 provides the response pattern for the two latent classes.

Students assigned to latent class A most probably prefer level II in the aspects *nature of models*, *multiple models*, and *changing models* but level III in the aspect *testing models*. In the aspect *purpose of models*, the preference probabilities of level II (40 %) and level III (41 %) are quite similar and higher than of level I (19 %). In latent class A, level I has the smallest preference probability in all aspects. The preference probability of level II is relatively high and ranges from 40 to 57 %. Thus, students assigned to latent class A most likely think that the biological functional models are idealised representations (*nature of models*), that there are multiple models of the same phenomenon since they have different foci (*multiple models*), that the functional models have the purpose of explaining the phenomenon or predicting something about it (*purpose of models*), that the models are tested by testing predictions that can be generated with them (*testing models*), and that the models should be changed when there is new information about the original (*changing models*).

In latent class B, the preference probability of level I is higher than in latent class A in all aspects. The students who were assigned to latent class B most probably prefer level I in the aspects *multiple models* and *purpose of models* and least likely prefer level III in any aspect (Fig. 3). Hence, students assigned to latent class B most probably think that biological functional models are idealised representations (*nature of models*), that there are multiple models of the same phenomenon since there are differences in the model objects themselves (*multiple models*), that the functional models are used to show the phenomenon (*purpose of models*), that the models are tested by comparing them with already known information

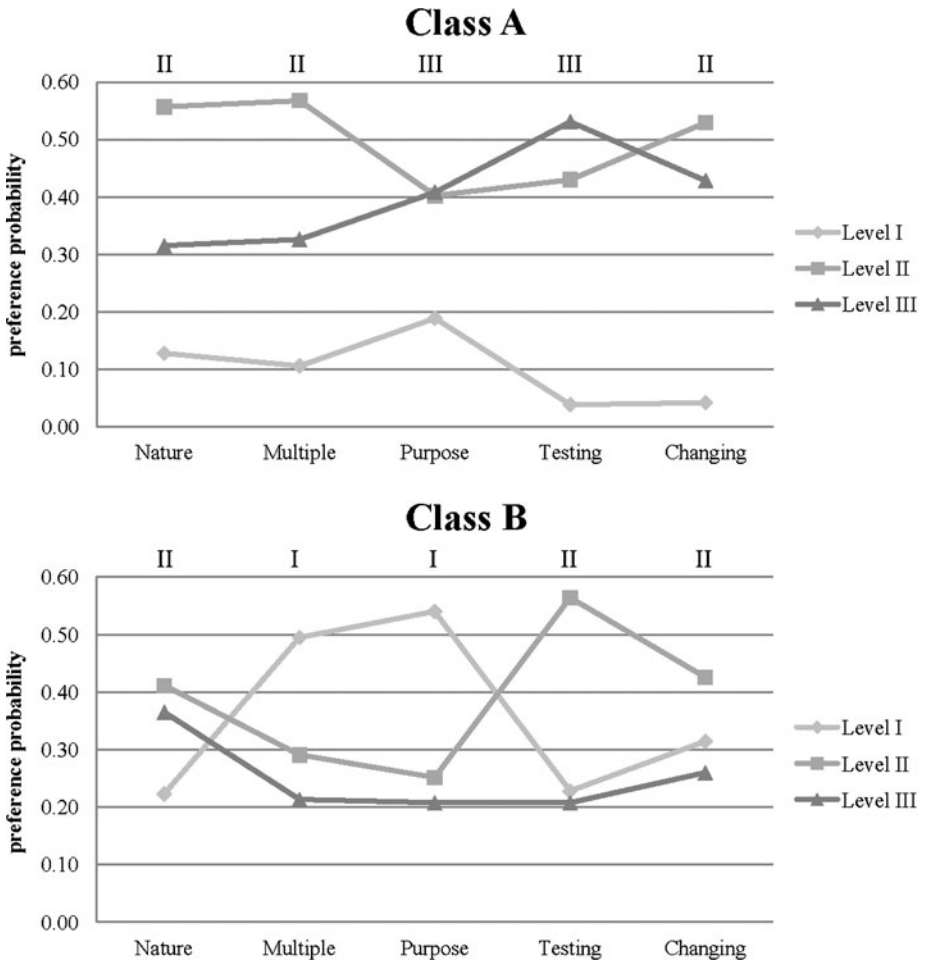

Fig. 3 The results of the latent class analysis. The preference probability of level I, level II, and level III for the aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*. For each aspect, the level with the highest preference probability is highlighted *above the curves*

about the original (*testing models*), and that the models should be changed when there is new information about the original (*changing models*).

Summarising, students in latent class A seem to have a more elaborated understanding of the biological functional models based on the 'model of model competence' (Upmeier zu Belzen and Krüger 2010) than students in latent class B. Furthermore, students in latent class A have a more consistent pattern of understanding.

Analysing Differences Between Latent Classes A and B (*H 2*)

Table 4 provides the mean scores of the auxiliary variables for both latent classes. There are significant differences between them for most of the variables (except reading speed and mark in the first foreign language), and the students of latent class A consistently have a better mean score in all auxiliary variables. For example, latent class A has a significantly better nonverbal

Table 4 Mean score (*ms*) and standard deviation (*sd*) of the auxiliary variables for the two latent classes A and B

Variable	Latent class	ms	sd	*p*	*d*
Reading (speed)	A	47.39	27.88	.162	.10
	B	44.44	29.60		
Mark (first foreign language)	A	2.69	0.90	.084	.13
	B	2.81	0.90		
Age	A	14.31	1.36	.048	.15
	B	14.11	1.35		
Mark (German)	A	2.75	0.83	.035	.16
	B	2.88	0.81		
Mark (biology)	A	2.67	0.87	.030	.17
	B	2.82	0.94		
Mark (chemistry)	A	2.74	0.98	.031	.17
	B	2.91	0.99		
Mark (mathematics)	A	2.86	0.99	.009	.20
	B	3.06	0.97		
Form	A	8.52	1.12	.002	.22
	B	8.27	1.13		
Nonverbal intelligence	A	0.71	0.21	.003	.22
	B	0.66	0.25		
Mark (physics)	A	2.68	0.93	.004	.22
	B	2.88	0.92		
Reading (comprehension)	A	33.83	30.70	.000	.38
	B	22.90	26.74		

Significance (*p*) and effect size (*d*) of the *t* test. The variables are arranged by effect size. In Germany marks range from 1 to 6, with 1 as the best mark. Reading speed and reading comprehension was assessed using the LGVT (Schneider et al. 2007); the scores are percentile ranks. Nonverbal intelligence was assessed with the KFT (Heller and Perleth 2000); the scores indicate the rate of right responses.

intelligence (ms=0.71; sd=0.21) than latent class B (ms=0.66; sd=0.25; p<.01). However, the effect size is small (d<.2) to medium (.2<d<.5; Fritz et al. 2012) in all cases.

Discussion

Before contrasting the findings of the present study with related findings of other authors, some methodological constraints have to be discussed. In the present study, 30 ranking tasks which refer to 6 different biological functional models were developed. In each task, the students had to choose one most preferred ranking option out of three ranking options, which represented three levels of understanding as developed by Upmeier zu Belzen and Krüger (2010; Table 1). Consequently, the present findings represent students' understanding of the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* concerning biological functional models.

Additionally, the forced-choice tasks are likely to assess a theoretical understanding of models and modelling and not students' practical skills (cf. Henze et al. 2007; Schwarz et al. 2009; Treagust et al. 2004). Consequently, different findings may occur when the assessment

tasks refer to other kinds of models or when students' modelling abilities are assessed. Furthermore, students were 'forced' to choose one preferred level of understanding in each ranking task. On this basis, they were scored as level I, level II, or level III modellers. This helps to uncover even slight differences in students' ranking (cf. Böckenholt 2004). However, as a result of the task format, the differences between the five aspects 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. Finally, the 'model of model competence' (Upmeier zu Belzen and Krüger 2010) was used as the starting point for the task development. Since a closed task format was used, it was not possible to investigate whether there are additional important aspects concerning models and modelling. According to international research findings, there are a few such perspectives. For example, some authors developed an aspect which refers to the making of a model: *Designing and creating models* (Crawford and Cullin 2005; Grosslight et al. 1991) or *constructing models* (Schwarz et al. 2009). Justi and Gilbert (2003) describe another aspect which belongs to the social *accreditation* of a scientific model. Since Oh and Oh (2011) focus on what teachers should know about models and modelling in science, these authors additionally developed the aspect *uses of models in the science classroom*. Furthermore, Crawford and Cullin (2005) describe four levels of understanding in some aspects of their 'Matrix of Modelling Dimensions' (pp. 316–317), whereas three levels have been assumed in the present study. Both issues could not be addressed within the methodological frame of the present study, but they are investigated in another study using open-ended tasks as well as hands-on tasks (Grünkorn et al. 2012). Despite of these constraints, the present study suggests that students' understanding of models and modelling is likely to vary across different aspects.

The research question is concerned with the consistency in students' understanding of models and modelling in biology. Hence, this study investigated to what extent there are differences in students' level of understanding across the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*. A LCA was conducted to allow for a more differentiated analysis than would have been possible if doing an overall analysis.

Analysing Consistencies in Students' Level of Understanding (H 1)

Based on research findings about students' and teachers' understanding of models in science (Crawford and Cullin 2005; Grosslight et al. 1991; Justi and Gilbert 2003; Schwarz et al. 2009), it was hypothesised that there would be inconsistent patterns of understanding in students' understanding of models and modelling (*H 1*). The findings support this supposition: In latent class A, the scientific perspective of models as research tools (level III) is most likely preferred in the aspects *purpose of models* and *testing models* but not in the aspects *nature of models*, *multiple models*, and *changing models*. For instance, students' who were assigned to this latent class primarily think that models may be tested by testing conclusions that can be drawn from them (*testing models*, level III) but do not prefer the corresponding ranking alternative that models may be changed in order to acquire further (alternative) predictions about the corresponding phenomenon (*changing models*, level III).

In latent class B, students think that models are idealised representations (*nature of models*, level II) but simultaneously prefer to explain *multiple models* of one original with differences in the model objects themselves (level I) and not with different foci on the respective original (level II), for example. Hence, students in both latent classes show inconsistencies in their understanding of models and modelling based on the levels described in

the 'model of model competence'. These findings support the results of others. For instance, Justi and Gilbert (2003) found complex patterns of understanding models and modelling and 'were not able to identify "profiles of understanding" for individuals that cut completely across the seven aspects' (p. 1381). Crawford and Cullin (2005) used the three general levels of understanding by Grosslight et al. (1991) as a starting point to develop levels of understanding for five different aspects of models and modelling (cf. Table 2). The authors emphasise: 'The expansion and refinement of the Grosslight et al. (1991) levels […] enabled us to fine-tune our assessments and track small changes across participants' (p. 320). Treagust et al. (2002) found inconsistencies in students' understandings of models. For example, they point out that students think of models as exact replicas but simultaneously are aware of the multiplicity of scientific models.

In summary, these studies suggest that students' and teachers' understandings of models and modelling in science is likely to vary across different aspects. Consistently, the present findings suggest that students may have *aspect-dependent* understandings of models and modelling in science. Therefore, the findings support the suggestion of Justi and Gilbert (2003) to question Grosslight et al.'s (1991) description of *global* levels of understanding models and modelling. Whereas Crawford and Cullin (2005), Danusso et al. (2010), and Justi and Gilbert (2003) asked teachers about their understandings of models and modelling, students were assessed by Treagust et al. (2002). However, whereas Treagust et al. (2002) explicitly referred to scientific models, biological functional models were used in the task stems of the present study.

Analysing Differences Between Latent Classes A and B (H 2)

It was hypothesised that students with a more consistent and more elaborated pattern of understanding models and modelling would have better marks and better cognitive abilities than other students (*H 2*). As discussed above, the students of both latent classes do not have a consistent understanding of models and modelling in biology (Fig. 3). However, students who were assigned to latent class A show a *relatively* consistent understanding compared to students who were assigned to latent class B since in latent class A the probability of answering in one specific level of understanding (level II) is relatively high and the probability of answering in another level of understanding (level I) is relatively low across all aspects. So, students who were assigned to latent class A seem to have a more stable concept (Clough and Driver 1986) of models and modelling in biology. Furthermore, the students of latent class A show a more elaborated understanding of models and modelling than students of latent class B (Fig. 3). Therefore, the present findings support *H 2*: Consistently, students who were assigned to latent class A have significantly better marks in the subjects biology, chemistry, physics, mathematics, and German. They are significantly older and in higher forms. In addition, they have a better nonverbal intelligence and better reading skills (Table 4).

The positive relationship between the understanding of models and modelling and science learning (e.g. Schwarz and White 2005) as well as cognitive processing (Sins et al. 2009) is well described. The present findings seem to support the suggestion that the understanding of models and modelling in science is positively related to the understanding of science and mathematics in school: Students with a more consistent and more elaborated understanding of models and modelling (i.e. students assigned to latent class A) achieve better marks in science school subjects as well as in mathematics. However, the effect size *d* is relatively small, which reflects the fact that students in both latent classes have average marks of about 3 in these subjects (Table 4). Therefore, the statistically significant differences between both latent classes concerning the marks indicate slight differences between latent classes A and B, which may seem less important from a practitioner's point of

view. In a similar way, nonverbal intelligence seems to be (only) slightly positive related to an elaborated and consistent understanding of the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models* (Table 4).

The present findings also suggest that students in higher forms may have a more elaborated understanding of models and modelling in science. However, the effect size *d* is small in this case, too (Table 4). The lack of relevant differences between students in different school years may be explained by findings which propose that the scientific role of models is sparsely discussed in science classrooms (cf. Danusso et al. 2010). As a consequence of this, it is reasonable that students' understanding of models and modelling does not increase meaningfully over the years in school. However, the present study is not longitudinal. Therefore, the development of students' understanding of models and modelling and its relation to science learning in school needs to be investigated in more detail (Patzke and Upmeier zu Belzen 2011).

Regarding reading skills, only reading comprehension but not reading speed is significantly better in latent class A than in latent class B. Consequently, some students may have scored poorly because of poor reading comprehension. The significant differences concerning the German mark between latent class A and latent class B seem to support this conclusion. However, with regard to reading comprehension, both latent classes A and B show a relatively poor percentile rank of about 34 (latent class A) and 23 (latent class B). These scores are still in the range of +/− one standard deviation, compared to the LGVT's norm group (Schneider et al. 2007). Hence, the students in both latent classes may be seen as average to poor readers. Consequently, one of the reasons for the more inconsistent response behaviour of the students assigned to latent class B may be their comparatively poor reading comprehension skills. However, it is not likely that the differences between the two latent classes' understanding of models and modelling are influenced strongly by the differences in their reading comprehension.

Implications

Two major implications may be derived from the present findings, one concerning the assessment of students' understanding of models and modelling and one concerning the teaching about this issue.

As discussed above, the theoretical framework of a study may influence both the development of assessment instruments and the interpretation of data. As a matter of principle, a theory which describes different *aspect-dependent* levels of understanding allows a more differentiated analysis than a *global* approach. However, for reasons of parsimony, the degree of complexity should be kept as low as possible but as high as necessary (cf. Burnham and Anderson 2000). Therefore, different aspects of models and modelling (i.e. *aspect-dependent* levels of understanding) should only be taken into account if they empirically improve the analysis of data. On the other hand, the *global* approach would be appropriate to be used in science education research if respondents show a consistent understanding across different aspects of models and modelling (cf. Justi and Gilbert 2003).

The present findings suggest that a differentiated framework with *aspect-dependent* levels of understanding models and modelling may help to analyse respondents' understanding of this issue more precisely. For example, Fleige et al. (2012) assessed students' understanding of the five aspects described in the 'model of model competence' (Table 1) before and after lessons which had a focus on models and modelling in biology. The authors

were able to show that in most cases students' understanding significantly improved in only a few of the five aspects. Such findings offer insights into which aspects of models and modelling may be difficult to understand or which aspects may be discussed in depth within a given teaching unit.

With reference to models and modelling, Justi and Gilbert (2002) proposed three major aims of science education: (1) learning of major scientific models, (2) learning about the nature of models and modelling, and (3) learning to use models (cf. Henze et al. 2007). In a later article, Justi and Gilbert (2003) argue that a profound understanding about the nature of models and modelling is reached when there is an elaborated view 'in respect of each and every one of the aspects' (p. 1382). The present findings indicate that students may understand some aspects of models and modelling in a more elaborated way than others. Hence, in science classes, the nature of models and modelling needs to be discussed in a differentiated way. The 'model of model competence' (Table 1) can be used as a theoretical basis for such a discussion in order to develop understandings in the five aspects *nature of models*, *multiple models*, *purpose of models*, *testing models*, and *changing models*.

Acknowledgments The authors want to thank the two anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

Acher, A., Arcà, M., & Sanmartí, N. (2007). Modeling as a teaching learning process for understanding materials: A case study in primary education. *Science Education, 91*, 398–418.

Adúriz-Bravo, A. (2012). A 'semantic' view of scientific models for science education. *Science & Education*. doi:10.1007/s11191-011-9431-7.

American Association for the Advancement of Science [AAAS]. (1993). *Benchmarks for Science Literacy, Project 2061*. New York: 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.

Bailer-Jones, D. (2003). When scientific models represent. *International Studies in the Philosophy of Science, 17*, 59–74.

Böckenholt, U. (2004). Comparative judgments as an alternative to ratings: Identifying the scale origin. *Psychological Methods, 9*, 453–465.

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*, 460–502.

Burnham, K., & Anderson, D. (2000). *Model selection and inference*. New York: Springer.

Bybee, R. (2011). Scientfic and engineering practices in K-12 classrooms: Understanding a framework for K-12 science education. *The Science Teacher, 78*, 34–40.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*, 235–251.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*, 514–529.

Chittleborough, G., & Treagust, D. (2007). The modelling ability of non-major chemistry students and their understanding of the sub-microscopic level. *Chemistry Education: Research and Practice in Europe, 8*, 274–292.

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.

Clement, J., & Rea-Ramirez, M. (2008). *Model based learning and instruction in science*. Dordrecht: Springer.

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.

Coll, R., & Lajium, D. (2011). Modeling and the future of science learning. In M. Khine & I. Saleh (Eds.), *Models and modeling* (pp. 3–21). Dordrecht: Springer.
Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis*. Hoboken: Wiley.
Coombs, C. (1950). Psychological scaling without a unit of measurement. *Psychological Review, 57*, 145–158.
Crawford, B., & Cullin, M. (2004). Supporting prospective teachers' conceptions of modeling in science. *International Journal of Science Education, 26*, 1379–1401.
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.
Eisma, J. (2012). *Flow visualization and force measurements on a flapping-wing MAV DelFly II in forward flight configuration* (Master of Science Thesis). Delft University of Technology, Delft
Fleige, J., Seegers, A., Upmeier zu Belzen, A., & Krüger, D. (2012). Förderung von Modellkompetenz im Biologieunterricht. [Fostering model competence in biology education.]. *MNU, 65*, 19–28.
Frigg, R., & Hartmann, S. (2006). Scientific models. In S. Sarkar & J. Pfeifer (Eds.), *The philosophy of science. An encyclopedia. Volume 2, N-Z, index* (pp. 740–749). New York: Routledge.
Fritz, C., Morris, P., & Richler, J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General, 141*, 2–18.
Gericke, N., & Hagberg, M. (2010). Conceptual incoherence as a result of the use of multiple historical models in school textbooks. *Research in Science Education, 40*, 605–623.
Gericke, N., Hagberg, M., & Jorde, D. (2012). Upper secondary students' understanding of the use of multiple models in biology textbooks: The importance of conceptual variation and incommensurability. *Research in Science Education*. doi:10.1007/s11165-012-9288-z.
Giere, R. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.
Giere, R. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press.
Giesbrecht, F., & Gumpertz, M. (2004). *Planning, construction, and statistical analysis of comparative experiments*. Hoboken: Wiley.
Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching, 28*, 73–79.
Gilbert, J. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education, 2*, 115–130.
Gilbert, J., & Boulter, C. (1998). Learning science through models and modelling. In B. Fraser & K. Tobin (Eds.), *International handbook of science education. Part one* (pp. 53–66). Dordrecht: Kluwer Academic.
Gilbert, J., & Boulter, C. (Eds.). (2000). *Developing models in science education*. Dordrecht: Kluwer Academic.
Gobert, J., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology, 13*, 7–22.
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* (pp. 53–65). Braga: CIEC, Universidade do Minho.
Grünkorn, J., Hänsch, J., Upmeier zu Belzen, A., & Krüger, D. (2012). Determination of students' model competence using open-ended and hands-on tasks. In C. Bruguière, A. Tiberghien, & P. Clément (Eds.), *Ebook proceedings of the ESERA 2011 conference. Part 5: Nature of Science, History, Philosophy, Sociology of Science* (pp. 38–44). Retrieved from http://www.esera.org/media/ebook/strand5/ebookesera2011_GRUNKORN-05.pdf.
Hagenaars, J., & Halman, L. (1989). Searching for ideal types: The potentialities of latent class analysis. *European Sociological Review, 5*, 81–96.
Halloun, I. (2007). Mediated modeling in science education. *Science Education, 16*, 653–697.
Harré, R. (1970). *The principles of scientific thinking*. London: Macmillan.
Harrison, A. (2001). How do teachers and textbook writers model scientific ideas for students? *Research in Science Education, 31*, 401–435.
Harrison, A., & Treagust, D. (1996). Secondary student's mental models of atoms and molecules: Implications for teaching chemistry. *Science Education, 80*, 509–534.
Heller, K., & Perleth, C. (2000). *KFT 4-12+ R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision. [KFT 4–12 + R: Cognitive abilities test for 4th-12th graders, revision.]*. Göttingen: Beltz Test GmbH.

Henson, J., Reise, S., & Kim, K. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling, 14*, 202–226.

Henze, I., Van Driel, J., & Verloop, N. (2007). Science teachers' knowledge about teaching models and modelling in the context of a new syllabus on public understanding of science. *Research in Science Education, 37*, 99–122.

Hicks, L. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167–184.

Justi, R., & Gilbert, J. (2000). History and philosophy of science through models: Some challenges in the case of 'the atom'. *International Journal of Science Education, 22*, 993–1009.

Justi, R., & Gilbert, J. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education, 24*, 1273–1292.

Justi, R., & Gilbert, J. (2003). Teacher's views on the nature of models. *International Journal of Science Education, 25*, 1369–1386.

Kaufman, L., & Rousseeuw, P. (2005). *Finding groups in data*. Hoboken: Wiley.

Khine, M., & Saleh, I. (Eds.). (2011). *Models and modeling*. Dordrecht: Springer.

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. Retrieved from http://www.ipn.uni-kiel.de/zfdn/pdf/16_Kleickmann.pdf.

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). Göttingen: Hogrefe & Huber.

Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss [Biology education standards for the Mittlere Schulabschluss]*. München: Wolters Kluwer.

Koponen, I. (2007). Models and modelling in physics education: A critical re-analysis of philosophical underpinnings and suggestions for revisions. *Science Education, 16*, 751–773.

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 http://edoc.hu-berlin.de/serl/2012-1/PDF/2012_1.pdf.

Krell, M., & Krüger, D. (2010). Diagnose von Modellkompetenz: Deduktive Konstruktion und Selektion von geschlossenen Items [Diagnosis of model competence: Deductive development and selection of closed items]. *Erkenntnisweg Biologiedidaktik, 9*, 23–37.

Krell, M., Czeskleba, A., & Krüger, D. (2012). Validierung von Forced Choice-Aufgaben durch Lautes Denken [Validation of forced choice-tasks using thinking aloud protocols]. *Erkenntnisweg Biologiedidaktik, 11*, 53–70.

Langeheine, R., & Rost, J. (Eds.). (1988). *Latent trait and latent class models*. New York: Plenum Press.

Magnani, L., & Nersessian, N. (Eds.). (2002). *Model-based reasoning*. New York: 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/Main: Peter Lang.

Martinez, M. (1991). A comparison of multiple-choice and constructed figural response items. *Journal of Educational Measurement, 28*, 131–145.

Martinez, M. (1999). Cognition and the question of test item format. *Educational Psychologist, 34*, 207–218.

Maydeu-Olivares, A., & Böckenholt, U. (2005). Structural equation modeling of paired-comparison and ranking data. *Psychological Methods, 10*, 285–304.

Morgan, M., & Morrison, M. (Eds.). (1999). *Models as mediators*. Cambridge: Cambridge University Press.

Muthén, L. & Muthén, B. (1998–2007). Mplus user's guide: Fifth edition. Los Angeles, CA: Muthén & Muthén.

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

Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569.

Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology and 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.

Patzke, C., & Upmeier zu Belzen, A. (2011). Entwicklung von Modellkompetenz im Biologieunterricht. [Development of model competence in biology education.] [Abstract]. In F. Bogner (Ed.), *Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB) im VBIO* (pp. 168–169). Bayreuth: Difo Druck.

Penner, D., Giles, N., Lehrer, R., & Schauble, L. (1997). Building functional models: Designing an elbow. *Journal of Research in Science Teaching, 34*, 125–143.

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6–12: Lesegeschwindigkeits- und verständnistest für die Klassen 6–12. [LGVT 6–12: Reading speed and reading comprehension test for 6th-12th graders.]*. Göttingen: Hogrefe.

Schwarz, C., & White, B. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction, 23*, 165–205.

Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*, 632–654.

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.

Sins, P., Savelsbergh, E., Van Joolingen, W., & Van Hout–Wolters, B. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education, 31*, 1205–1229.

Spiel, C., & Glück, J. (2008). A model-based test of competence profile and competence level in deductive reasoning. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 45–68). Göttingen: Hogrefe & Huber.

Suckling, C., Suckling, K., & Suckling, C. (1978). *Chemistry through models: Concepts and applications of modelling in chemical science and industry*. Cambridge: Cambridge University Press.

Terzer, E. & Upmeier zu Belzen, A. (2011, September). *Model competence in biology education: Evaluation of a theoretical structure using multiple-choice items*. Paper presented at the 14th EARLI Conference, Exeter.

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).

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 [Model competence in biology education]. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 41–57. Retrieved from http://www.ipn.uni-kiel.de/zfdn/jg16.html.

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.

Wang, C.-Y., & Barrow, L. (2011). Characteristics and levels of sophistication: An analysis of chemistry students' ability to think with mental models. *Research in Science Education, 41*, 561–586.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*, 941–967.