

# Identifying the Item Hierarchy and Charting the Progression across Grade Levels: Surveying Taiwanese Students' Understanding of Scientific Models and Modeling

Silvia Wen-Yu Lee<sup>1</sup>

Received: 15 March 2017 / Accepted: 2 September 2017 / Published online: 16 September 2017  
© Ministry of Science and Technology, Taiwan 2017

**Abstract** The purpose of this study was, first, to understand the item hierarchy regarding students' understanding of scientific models and modeling (USM). Secondly, this study investigated Taiwanese students' USM progression from 7th to 12th grade, and after participating in a model-based curriculum. The questionnaire items were developed based on 6 aspects of USM, namely, *model type*, *model content*, *constructed nature of models*, *multiple models*, *change of models*, and *purpose of models*. Moreover, 10 representations of models were included for surveying *what a model is*. Results show that the *purpose of models* and *model type* items covered a wide range of item difficulties. At the one end, items for the *purpose of models* are most likely to be endorsed by the students, except for the item “models are used to predict.” At the other end, the “model type” items tended to be difficult. The students were least likely to agree that models can be text, mathematical, or dynamic. The items of the *constructed nature of models* were consistently located above the average, while the *change of models* items were consistently located around the mean level of difficulty. In terms of the natural progression of USM, the results show significant differences between 7th grade and all grades above 10th, and between 8th grade and 12th grade. The students in the 7th grade intervention group performed better than the students in the 7th and 8th grades who received no special instruction on models.

**Keywords** High school · Middle school · Modeling · Scientific model

---

✉ Silvia Wen-Yu Lee  
silviawyl@cc.ncue.edu.tw

<sup>1</sup> Graduate Institute of Science Education, National Changhua University of Education, No.1, Jin-Der Rd., Changhua City 500 Taiwan, Republic of China

## Introduction

Promoting students' understanding of models and modeling (USM) is one of the major goals of science teaching (National Research Council, 1996, 2007, 2012). Researchers generally relate USM to the understanding of the nature, purpose, evaluation and process of modeling (Grosslight, Unger, Jay, & Smith, 1991; Pluta, Chinn, & Duncan, 2011; Schwarz et al., 2009; Schwarz & White, 2005). Scientific models are defined as models representing the characteristics of phenomena, providing a mechanism that accounts for how a phenomenon operates, and predicting the observable aspects of the phenomenon (Nicolaou & Constantinou, 2014).

Recently, an increasing number of researchers have emphasized the importance of promoting understanding of scientific models and modeling (Gobert et al., 2011; Prins, Bulte, & Pilot, 2011; Raghavan & Glaser, 1995). The importance of USM is supported both in theory and by empirical evidence. From a theoretical perspective, the nature of models and modeling is a part of the nature of science and is also a specialized aspect of the epistemology of science (Justi & Gilbert, 2002; National Research Council, 2007), which are essential to science education. Nicolaou and Constantinou (2014) further argued that the understanding of the nature of models and modeling is one of the major constructs in the theoretical framework of modeling competence. In their theoretical framework, modeling competence includes two major categories, namely modeling practices and meta-knowledge. Meta-knowledge of modeling is further divided into "meta-modeling knowledge" and "meta-cognitive knowledge of the modeling process." Nicolaou and Constantinou (2014) defined "meta-modeling knowledge" as the "epistemological awareness about the nature and the purpose of models" (p. 53) which is similar to what other researchers have called understanding of models and modeling or views of models and modeling. Based on this theoretical framework, one cannot fully develop one's modeling competence in science unless one advances one's understanding of models and modeling while improving one's performance in modeling practices and developing meta-cognitive knowledge.

Empirical studies have also provided evidence of the advantages of sophisticated understanding of models and modeling in science learning. For example, students with advanced understanding of the nature of models have been found to learn better in model-based tasks when compared to their counterparts with less sophisticated understanding (Gobert & Pallant, 2004). Additionally, positive correlations have been found to exist between higher levels of understanding of models and modeling and deep cognitive processing (Sins, Savelsbergh, van Joolingen, & van Hout-Wolters, 2009), and between advanced understanding of models and modeling and better science performance (Cheng & Lin, 2015; Gobert & Pallant, 2004). Researchers have found that engaging students in model-based activities can improve their understanding of models and modeling (Schwarz et al., 2009). Reciprocally, students' advanced understanding of the different aspects of models and modeling supports their use and creation of models for learning science (Gobert et al., 2011; Sins et al., 2009).

Students' understanding of models and modeling include different aspects, such as model type, model evaluation, constructed nature of models, multiple models, change of models, model revision, and purpose of models (Grosslight et al., 1991; Krell, Reinisch, & Krüger, 2015; Schwarz & White, 2003; Sins et al., 2009; Treagust, Chittleborough, & Mamiala, 2002). Researchers have used interviews, ranking tasks,

and questionnaires to elicit students' USM. Questionnaires have been created in past studies to measure students' USM for different aspects. For instance, the students' understanding of models in science (SUMS) survey include items of five subscales, namely scientific models as multiple representations, models as exact replicas, models as explanatory tools, how scientific models are used, and the changing nature of scientific models (Treagust et al., 2002). Additionally, there have been different views of the same aspect of USM that were interpreted through interview studies. For example, Schwarz and White (2003) categorized students' responses to each aspect of USM into strong, moderate, and weak responses. However, thus far, one cannot conclude based on empirical evidence which aspect or view is more difficult for students to understand and agree with, and which aspect is more likely for students to accept and agree with. Although in theory all aspects and all views under each aspect are valuable, the insights into the item level and aspect level may further inform model-based teaching in the future.

Thus, the first purpose of the current research is to investigate at a finer grained level what it is about USM which appears to be more likely for students to agree with (i.e., easier items), and what appears to be less likely for students to agree with (i.e., more difficult items), in other words, the item hierarchy extend over the range of difficulty within the latent trait of USM. The item hierarchy allows all items across different aspects to be compared on the same scale, which has rarely been discussed in previous studies. For this particular aim, the authors developed the questionnaire items based on the six aspects of USM. Specifically, the items refer to the students' diverse responses to interview questions about their understanding of models and modeling (Lee, Chang, & Wu, 2017; Pluta et al., 2011; Schwarz & White, 2003). The purpose was to include a wider range of views within each aspect. Details of the item design are provided in the methodology section.

Past studies, by using interviews or Likert scale questionnaires, did not make direct comparisons of the different subscales due to the limitations of the research instrument or methodology (e.g., Liu, 2006; Sins et al., 2009). Item response theory (IRT) and Rasch modeling have been used in the development and validation of Likert scale instruments in science education and in other fields of the social sciences (Bond & Fox, 2007; Boone, Townsend, & Staver, 2011; Neumann, Neumann, & Nehm, 2011; Sondergeld & Johnson, 2014). Because a Likert scale is ordinal, researchers have argued that it is mathematically inappropriate to use total item scores for the analysis of Likert scale data (Bond & Fox, 2007). Rasch statistics have the following advantages (Linacre, 2002; Neumann et al., 2011; Wei, Liu, & Jia, 2014; Wu & Chang, 2008). First, it converts ordinal data (i.e., Likert scale data) into a metric of logits (ratio-scale data), by which person and item parameters can be estimated on an equivalent interval scale. That is, person ability and item difficulty for a latent trait are both measured by logits and mapped on a single continuum. Second, based on Rasch statistics, indices for person and for item are independent. In other words, the measures on a person's ability are independent from how difficult the instrument for the same construct is and vice versa. Third, Rasch statistics provide information on category functioning which allows further optimization of the rating scale categories (e.g., 5-point scale).

Secondly, this study also aimed to investigate two kinds of progression the students can make in terms of their understanding of models and modeling. The first progression regards how Taiwanese students' understanding of models and modeling progresses

from seventh grade to 12th grade, and the second progression concerns whether explicitly teaching models and modeling can advance students' understanding of models and modeling. Although some studies have found that students' understanding of models and modeling can be advanced as they gain more experience of learning science (e.g., Krell et al., 2015; Willard & Roseman, 2010), some other studies have shown no significant differences among the students of different grades (e.g., Treagust et al., 2002). Additionally, the studies so far have only compared a narrow range of school years. For instance, studies have compared students in the sixth, seventh, and eighth grades (Willard & Roseman, 2010) or between 7/8 and 9/10 (double grades) (Krell et al., 2015). As there has not been enough evidence collected across a wider range of age groups measured by a consistent instrument, the goal of this study is to present a more comprehensive picture of the extent to which students' understanding of models and modeling progresses over the years. By using a Rasch unidimensional model (Linacre, 2012), the results of all items can be calibrated into one score to measure each student's understanding of models and modeling. The study also included a group of students who received intervention in modeling. The goal was to understand the magnitude of the effect of the modeling instruction in relation to the overall progression.

To summarize, in this study, the following research questions are posed:

1. What is the item hierarchy of USM?
2. To what extent does Taiwanese students' USM progress over grade levels?
3. To what extent does the students' USM progress after participating in a model-based science curriculum?

## Literature Review

### Different Aspects of USM

Through reviewing the literature, six aspects emerged as the most commonly researched aspects of students' understanding of models and modeling, namely, "model type," "model content," "constructed nature of models," "multiple models," "change of models," and "purpose of models" (Grosslight et al., 1991; Pluta et al., 2011; Schwarz et al., 2009; Schwarz & White, 2005). Detailed descriptions of each are provided in the following.

The "model type" aspect concerns the identification of different kinds of models and their attributes. The term "model" refers to a wide range of things, such as diagrams, physical objects, simulations, or mathematics equations (Harrison & Treagust, 2000). These models can be categorized into verbal, visual, mathematical, gestural, concrete, and a mixture of the above depending on the level of abstraction (Boulter & Buckley, 2000). For science teaching and learning, there are expressed models, consensus models, historical models, and curricular models (Gilbert, 2013). "Model content" is concerned with what is represented by a model. A model can represent scientific concepts, mechanisms, theories, or structures and functions, as well as non-visible processes and features (Schwarz et al., 2009). However, a model does not necessarily represent absolute reality, and does not have to be an exact copy of the real thing.

The “constructed nature of models” aspect concerns “How close does a model have to be to the real thing?” (Gobert & Pallant, 2004; Schwarz & White, 2005). Without in-depth understanding of models, students might think that a model has to be as close as it can be to the real thing (Treagust et al., 2002). As students develop more understanding of models and modeling, they may realize that a model can be a simplified and abstract form of the reality or evidence. Although models are built based on evidence or phenomena, they are an idealized representation of the reality, or a theoretical reconstruction of the reality (Krell, Upmeier zu Belzen, & Krüger, 2014). The “constructed nature of models” is also referred to as the “nature of models” by some researchers (Krell et al., 2014).

The question of “Can there be different models for the same object?” concerns the “multiple models” aspect of USM (e.g., Grosslight et al., 1991; Schwarz et al., 2009; Sins et al., 2009). Students with a more naïve understanding of models might believe that “all of the models are the same” or “there is only one final and correct model” (Chittleborough, Treagust, Mamiala, & Mocerino, 2005; Grünkorn, Upmeier zu Belzen, & Krüger, 2014). A more sophisticated understanding of the multiple models is that the different models can represent different ideas or explanations of the same phenomenon, and different models may have different advantages (Grosslight et al., 1991; Schwarz et al., 2009). Krell et al. (2015) further specified that from a lower level to a higher level of understanding, students should understand that different models can exist because of the different model objects, the different foci of the original, or the different hypotheses of the original.

Another important aspect of the USM is the belief of “change of models.” Researchers have been interested in the extent to which students understand the circumstances that would require a model to be changed (Gobert & Pallant, 2004). At the initial level, students might think that models are unchangeable and are in their final form (Grünkorn et al., 2014). As students develop more sophisticated understanding, they come to realize that scientists change models due to many reasons. Models change because errors are found in the model, or because new information or evidence is found, or because the original hypotheses or purposes of modeling change (Grosslight et al., 1991; Krell et al., 2015). Additionally, for scientists and also for students, any change in a model also reflects a growth in understanding of the phenomenon (Schwarz et al., 2009).

As for the “purpose of models,” researchers are concerned about the question of “What are models for?” (Gobert & Pallant, 2004). In earlier research, Grosslight et al. (1991) suggested that models are for communication, observation, learning and understanding, testing, and provision of reference and examples. Sins et al. (2009) suggested three levels of model purposes. From the lower to the higher levels, models are used to measure things; models are used for showing how things work; and models are used to make predictions. In addition to the illustrating, explaining, and predicting purposes, Schwarz et al. (2009) also stated that models are for sense-making and knowledge construction, and for communicating understanding of knowledge. Although there are a variety of different views on the purposes of models, the commonly accepted ideas among scholars are that scientific models are used to describe or illustrate, explain, and predict phenomena (Krell et al., 2014).

The questionnaire items in this study were designed based on the aforementioned aspects. In addition to the 6 aspects, 10 representations of models are also included in

the questionnaire. Sample models have been used in surveying students' understanding of "what a model is" in earlier studies. For example, in the original interviews conducted by Grosslight et al. (1991), specific examples of scientific models, such as a toy airplane and a subway map, were shown to students who were asked if they were models. Researchers in subsequent studies continued to follow this method and present questions or tasks with representations, such as flowcharts, written explanations, causal diagrams, and pictorial models (e.g., Pluta et al., 2011). Students were asked to circle what they considered as models or to discuss "what a model is." More details about the questionnaire items are described in the methodology section.

### **Empirical Evidence of the Differences among Different Aspects of USM**

Students' understanding of models and modeling has been elicited through different research methods including interviews, ranking tasks, and surveys. However, only a few researchers have focused on the difference among the aspects of understanding of models. Two studies using the SUMS survey (Treagust et al., 2002) concluded that the models as exact replicas (ER) subscale appeared to be more difficult for the students. In Liu's (2006) study, the students learned the gas law by computer-based modeling and hands-on experiments. The results showed that the ER subscale resulted in a lower median (i.e., 2.0) and lower mode (i.e., 2.0) than the rest of the factors (i.e., 3.0) than the other subscales. Cheng and Lin (2015) also used the SUMS questionnaire to survey middle school students in Taiwan, but the responses were collected on a 5-point Likert scale. The researchers found that the CNM sub-factor had the highest mean score, while ER had the lowest mean. The authors concluded that the students had a better understanding of the changing nature of scientific models but were not sure whether scientific models are exact replicas of target objects.

By using an open-ended questionnaire, Sins et al. (2009) surveyed 26 11th grade students who completed a computer-based modeling task. Students answered questions probing their understanding of the four dimensions of USM, including the nature of models, purposes of models, design and revision of models, and evaluation of models. Students' responses regarding the four dimensions were categorized from level 1 (lower level) to level 3 (higher level). Sins et al. found that for three of the four dimensions, more than one half of the students reached level 2. For the dimension of purposes of models, 46% of the students reached level 2 and 42% reached level 3. The results imply that it is more likely for students to achieve a higher level of understanding in the dimension of purpose of models than in other dimensions of USM.

In a recent study, a ranking task was designed by Krell et al. (2014) to assess students' different levels of understanding regarding the nature of models, multiple models, purpose of models, testing models, and changing models. Each item included a description of the original phenomenon and a model representing the phenomenon. In each task, the students had to rank the three statements, which represented three levels of understanding of models and modeling in science. The results showed that the students had "partially inconsistent" views of models across the five aspects. The researchers found that students can be categorized into two "latent classes" based on their profiles of understanding of models and modeling. Students in "latent class A" most likely prefer level II in the aspects nature of models, multiple models, and changing models, but level III in the aspect of testing models. Students in "latent class

B” were likely to have a preference for level I in the aspects of multiple models and purpose of models, and were least likely to prefer level III in any aspect. The results indicated that the students did not have the same levels of understanding across the five aspects.

Because the empirical evidence is still limited, and interpretations across different studies are difficult, it is inconclusive regarding which aspect of USM appears to be easier or more difficult to understand. Thus far, the research has suggested that it is possible to achieve better understanding for the “purpose of models” (Sins et al., 2009) and the “changing nature of scientific models” (Cheng & Lin, 2015), but that students seem to have difficulty developing understanding of “models as replicas” (Cheng & Lin, 2015).

### **Progression of Students' USM**

To understand students' natural progression of USM by grade levels, some studies interviewed or surveyed students without implementing intervention regarding models and modeling. Comparisons were made between students at the middle school and high school levels. In an earlier study, Grosslight et al. (1991) found that more 11th grade than 7th grade students had higher levels of understanding of models and modeling. Based on the interview data, Grosslight et al. developed three general levels of understanding of models and modeling. At the first level, students thought that models are exact replicas of real objects. By achieving level 2, students are aware of the modeler's purposes, while at level 3, they should be able to further examine the relationship between model designs and model purposes. The results showed that the majority of the 7th graders were at level 1, while the majority of the 11th graders were categorized into levels 1/2 and level 2. Lee, Chang, and Wu (2017) further compared eighth graders and 11th graders' views of scientific models and modeling. The students' responses for model identification and the utility of multiple models indicated that the 11th grade students may have more sophisticated understanding than the eighth grade students. In terms of the different representations of models, the 11th graders were more likely to recognize textual representations and pictorial representations as models than the eighth graders, while the 11th graders were also more likely to appreciate the differences between 2D and 3D models. Lee et al. (2017) suggested that students' educational level may impact how they interact with different representations.

Most of the studies have researched the differences among grade levels within middle school or within high school. Willard and Roseman (2010) developed 61 items based on five key ideas about models and modeling. In their pilot study, more than 2000 students in sixth, seventh, and eighth grade were surveyed. They found that seventh graders did better than sixth graders, but there existed no difference between students in seventh and eighth grade. For the high school levels, Gobert et al. (2011) analyzed the students' understanding of models and modeling prior to receiving any intervention about modeling. The analysis of the pre-tests showed that the group of students who had statistically significantly less understanding in terms of ET, MR, and USM than the other groups were the youngest. The researchers argued that the younger students in their study had less sophisticated understandings of models, probably because they had taken fewer science courses. In another study that surveyed high school students, Krell et al. (2015) found

significant differences, although with small effect sizes, between the double grades of 7/8 and 9/10 students in Germany. One study also included students at university level. When surveying students from eighth grade to the first year university level, Chittleborough et al. (2005) found a tendency that more students at the university level showed appreciation of scientific models than the younger students due to their age and learning experience.

However, some other studies found no significant differences between students' understanding of models across different age groups who had received the same instruction about modeling. For instance, using the SUMS instrument, Treagust et al. (2002) found that there were no statistically significant differences for any tested subscales between eighth, ninth, and tenth graders who had not received instruction about modeling. None of the students in the aforementioned studies had received additional instruction about models and modeling; thus, their responses were based on their science curriculum. The differences between age groups can, however, be reversed after the students received instruction in modeling. Saari and Viiri (2003) found that seventh graders who received a series of modeling instruction had better understanding of models than a control group at ninth grade who did not have any instruction regarding models.

Although it is the general assumption that the students of higher educational level may have better understanding of models and modeling due to more learning experience, thus far, we can only gain a vague picture of how the students progress based on the sporadic evidence in the literature. The results were inconclusive, mostly because different instruments were used in different studies and the participating students were from different countries. In addition, the number of studies is too limited to conduct a meta-analysis. Thus, a more comprehensive survey using a consistent measurement would be helpful for researchers to understand when the difference occurs and why.

## Methodology

### Participants

The data are from 983 students in Taiwan including 502 male students, 402 middle school students (i.e., 168 seventh grade students, 158 eighth grade students, and 76 ninth grade students), and 581 high school students (i.e., 149 10th grade students, 228 11th grade students, and 204 12th grade students). All of the school teachers and students participated in this study voluntarily. Nine students' responses containing a large portion of missing data were dropped from the data. All of the items were computer-based and the questionnaire was administered on desktop computers. The non-intervention students answered the questionnaire at the beginning of the school year, except for 87 12th grade students who answered the questionnaire at the end of the school year, and were marked as "12E" in the data. Based on our current science curriculum, modeling is not taught explicitly in science classes in middle and high school in Taiwan.

In the data set, two classes of 11th grade students ( $n = 79$ ) of the "social science track" were included as well. In Taiwan, from 3rd grade until 10th grade (first year



in high school), all students must study science. In 11th and 12th grade, students can choose either to enroll in the “social science track” or the “science track.” The students in the “science track” must continue studying advanced physics, chemistry, and biology in the 11th and 12th grades. The “social science track” focuses on social science subjects such as history and geography. One class of 7th grade students ( $n = 23$ ) had received special instruction on models and modeling. The students in the treatment class participated in a 12-h modeling unit about marine ecology and fishery sustainability. Details description of the curriculum is provided in [Appendix 1](#).

## Research Instrument

The USM computer-based questionnaire (USM-CBQ) was created based on the literature about scientific models (Gobert et al., 2011; Grosslight et al., 1991; Schwarz & White, 2005; Treagust et al., 2002). A total of 40 items were originally designed, 4 of which were deleted after a test run of the questionnaire in a pilot study. The final version of the questionnaire consisted of 36 items, including 10 representations of models (MD1-MD10) (see [Table 1](#)) and 26 written statements (see [Table 2](#)). A 5-point Likert scale rating was used for all items in the questionnaire, where 5 represents *strongly agree* while 1 represents *strongly disagree*.

In the current study, items MD1 through MD10 were designed based on the typology of models created by Harrison and Treagust (1998, 2000), including “scale model,” “theoretical model,” “map and diagram,” “concept-process model,” and “simulation.” For each item, the students were asked to which degree they agree that the representation is a scientific model. “Scale models” are designed to represent

**Table 1** Model form, representation, and biological themes of items MD1 to MD10

Item	Theme	Model form	Representation type	Measure	S.E.
MD5	Photosynthesis mechanism	Concept-process model	Verbal (text)	1.49	0.05
MD9	Mendel's laws	Theoretical model	Verbal (text)	1.37	0.05
MD2	Food web	Map/diagram	Visual (symbolic, static)	1.00	0.05
MD8	Inheritance pattern and gene combination	Map/diagram	Visual (symbolic, static)	0.86	0.05
MD7	Heart function and circulation mechanism	Simulation	Visual (2D, animated)	-0.28	0.05
MD10	Enzyme and reactions	Simulation	Visual (symbolic, animated)	-0.53	0.05
MD1	Structure of bones and muscles	Simulation	Visual (3D, animated)	-0.67	0.05
MD6	Breathing mechanism	Simulation	Visual (3D, animated)	-0.76	0.05
MD3	Cell and organelles (plant)	Scale model	Visual (3D, static)	-0.95	0.05
MD4	Human anatomy	Scale model	Concrete (plastic model)	-1.29	0.05

Items are arranged by item difficulty

**Table 2** Items grouped by aspect and item difficulty. The underlined items were interpreted as of more sophisticated level of USM based on the literature

Aspect	Description	Item	Measure (item difficulty)	S.E.
Model content (MC)	A model represents <i>invisible features</i>	MC5	0.01	0.05
	A model represents <i>concepts</i>	MC4	- 0.24	0.05
	A model represents <i>theories</i>	MC2	- 0.41	0.05
	A model represents <i>scientific principles</i>	MC1	- 0.45	0.05
	A model represents <i>structures and functions</i>	MC3	- 1.00	0.05
Model type (MT)	A model can be <i>text</i>	MT2	1.57	0.05
	A model can be <i>mathematical</i>	MT6	1.50	0.05
	A model can be <i>dynamic</i>	MT5	1.08	0.05
	A model can be <i>graph</i>	MT3	0.70	0.05
	A model can be <i>concrete</i>	MT1	- 0.05	0.05
	A model can be a <i>3-dimensional object</i>	MT4	- 0.09	0.05
Constructed nature of models (CM)	A model is abstract	CM1	- 0.02	0.05
	A model is composed of elements	CM2	- 0.08	0.05
	A model is built based on data	CM3	- 0.29	0.05
Multiple models (MM)	Scientists decide a better model based on the evaluation of evidence when multiple models are available	MM3	0.08	0.05
	All models have advantages, therefore there is no need to decide which one is better	MM2	0.06	0.05
	Scientists use multiple models for the same phenomenon	MM1	- 1.07	0.05
Change of models (CN)	Scientists change models when they rethink the purpose of the model	CN1	0.39	0.05
	Scientists change models when new evidence is found	CN3	0.34	0.05
	Scientists change models because there are errors in the existing model	CN2	0.31	0.05
Purpose of model (PM)	Models are used to predict	PM2	0.73	0.05
	Models are used to explain	PM1	- 0.67	0.05
	Models are for sense-making	PM4	- 1.21	0.05
	Models are for observation	PM3	- 1.43	0.05

reality and are used to depict the color, external shape, or structure (Harrison & Treagust, 1998). “Theoretical models” are abstract models that are designed to communicate scientific theories. “Maps and diagrams,” “concept-process models,” and “simulations” are models for describing multiple concepts and processes. As defined by Harrison and Treagust (2000), “maps and diagrams” are models that visualize patterns and relationships, while “simulations” are computer-based multimedia or games that employ animations or real-life situations. Also, based on Buckley and Boulter (2000) categorization, the model representations in this study can be further divided into concrete, verbal, and visual models. The visual

representations can be further described as symbolic, 2-dimensional (2D) drawing versus 3-dimensional (3D) drawing, and animated versus static. Descriptions of the model form, representation type, and biological themes of the 10 models are shown in Table 1. In Appendix 2, the representations for four sample items, MD2, MD4, MD5, and MD6, are shown. The representations were selected to include a variety of different model representations and visual representations explained above. The 10 items included one concept-process model, one theoretical model, two map/diagram models, four simulations, and two scale models. Both the “theoretical model” and the “concept-process model” were presented in the text. Concerning the representation type, the 10 items can be categorized as two verbal representations, seven visual representations, and one concrete representation. Among the four “simulations,” two items are 2D animations, and the other two are 3D animations. All representations used in this questionnaire are based on the common models in the seventh grade biology textbooks in Taiwan.

The other statements concern the six aspects of USM, including “model content” (five items), “model type” (six items), “the constructed nature of models” (five items), “change of models” (three items), “multiple models” (three items), and “the purpose of models” (four items). The students were asked to mark to what extent they agreed with the statement in the item also on a 5-point Likert-scale where 5 represents *strongly agree* while 1 represents *strongly disagree*. The exact wording of the statements can be found in Table 2. The items belonging to the same aspect were not repeating the same idea but were representing various ideas of the same aspect. Based on the literature (Lee et al., 2017; Pluta et al., 2011; Schwarz & White, 2003), we underlined the items that were interpreted as of a higher level of difficulty for the students to agree with. For instance, for the multiple models (MM) aspect, item MM3 “scientists decide a better model based on the evaluation of evidence when multiple models are available” is considered as being of a higher level. The evidence-based criteria for evaluating among multiple models is considered more sophisticated understanding than thinking that all models are available disregards their quality (Pluta et al., 2011; Schwarz & White, 2003).

## Analysis

A unidimensional Rasch model for the rating scale data was used in this study. Iterative Rasch modeling was performed to examine whether the psychometric properties of the items were satisfactory. Important indicators for item quality including item and person fit statistics, person and item reliability and separation, and rating scale functioning (Sondergeld & Johnson, 2014) were examined. Then, the item difficulty and person ability generated from the Rasch modeling were used to compare across different items and different groups of students. Based on the item measures, a Wright map was constructed to visually present the hierarchy of the items and the ability of a person. A unidimensional rather than a multi-dimensional Rasch model was chosen in this study because the focus is more on the items than on the subscales (aspects). All items belong to the overall construct of USM. Winsteps version 3.75.1 (Linacre, 2012) was used to conduct the unidimensional Rasch analysis. The Rasch measurement was then exported to SPSS for ANOVA and post hoc analysis with Bonferroni correction.

## Findings

### Quality of Item Measures

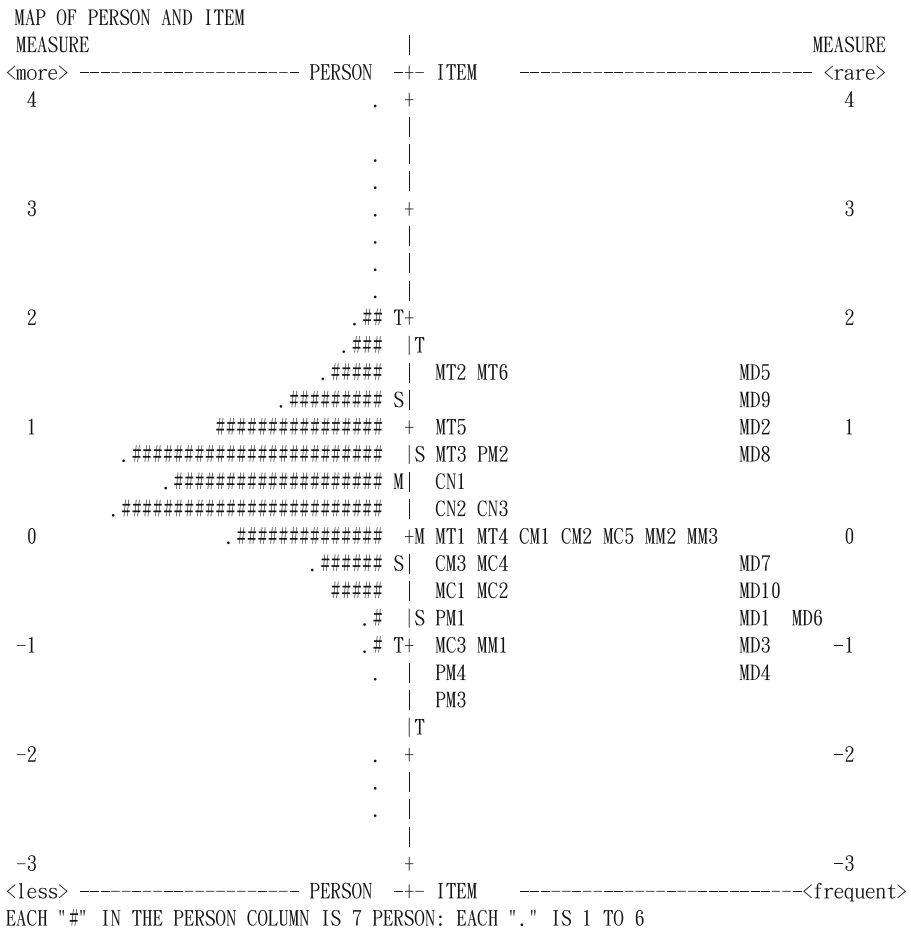
**Fit Indices** Rasch modeling was conducted to verify if the items represented the latent trait of students' understanding of the nature of models. Infit and outfit indices and point-biserial correlation (pbis) are commonly used to determine how well the data fit the Rasch model (Bond & Fox, 2007). There are two kinds of fit statistics—mean square (MNSQ) fit statistics and z-standardized (ZSTD) fit statistics. However, ZSTD is very sensitive to sample size and is suitable for data numbering less than 300 and over 30 (Linacre, 2012). Thus, in this study, infit MNSQ, outfit MNSQ, and pbis were used (see Appendix 3). For rating scales, infit or outfit MNSQ values ranging between 0.6 and 1.4 suggest productive measurement and a reasonable fit of the data to the model (Wright & Linacre, 1994). Items with MNSQ over 1.4 provide more noise than information (i.e., misfit), while items with MNSQ under 0.4 provide no new information to the construct (i.e., overfit). Two reversed items, “models need to be exactly like the real thing” (CM5), and “a model is a copy of the real thing” (CM4) resulted in MNSQ over 1.6 and were deleted from the model. All of the remaining items resulted in infit values ranging between 0.74 and 1.33, and outfit values ranging from 0.74 to 1.38. Additionally, all items' pbis correlations were positive (see Appendix 3) indicating that all items contribute to the measure (Sondergeld & Johnson, 2014). The results showed that all items were within the acceptable range, confirming the unidimensional construct.

**Separation and Reliability** The separation and reliability indices are also indications of measurement quality. Separation represents “statistically distinct groups that can be classified along a variable,” while reliability shows “statistical reproducibility” (Sondergeld & Johnson, 2014). In a Rasch model, separation and reliability are calculated independently for person abilities and item difficulties. According to the literature, separations and reliability of 3.00 and 0.90, respectively, are considered excellent; separations and reliability of 2.00 and 0.80, respectively, are considered good, and separations and reliability of 1.00 and 0.70, respectively, are of acceptable levels (Duncan, Bode, Lai, & Perera, 2003). In the current study, person separation (2.28) and person reliability (0.84) are at good levels, and item separation (16.42) and item reliability (1.00) are at excellent levels.

**Rating Scale Analysis** Scale functioning output from Winsteps was used to examine the rating scale categories. Based on the literature (Sondergeld & Johnson, 2014), the peak of probability curves for each response category should reach approximately 0.6 probability as an indication of the good functioning of the scale. Otherwise, some scale categories should be collapsed or investigated further. The initial output of the analysis showed that the response categories 2 (disagree) and 3 (neutral) did not meet the criteria. After combining these two categories, the probability curves for the four categories meet the criteria.

### Item Difficulties and Item Hierarchy

A Wright map was created and showed both student ability distribution and item difficulty distribution on the same scale for direct comparison (see Fig. 1). The top of the right side of the Wright map is labeled “rare” where the most difficult to endorse item is located, and the bottom of the map is labeled “frequent” where the most likely to be endorsed item (i.e., least difficult item) is located. Items at similar locations provide similar measurement information (Boone et al., 2011). On the left side of the Wright map, a range of students' ability is shown. The top represents students with more understanding of models and modeling, while the bottom represents students with less understanding of models and modeling. The item difficulty ranges from 1.57 to -1.43 logits. All MD items are listed separately on the far-right side of the graph for better illustration.



**Fig. 1** Person measure and item difficulty presented on a Wright map. *Note.* Numbers on the left side of the map are logits. Each “single asterisk” represents seven people and each “dot” represents one to six people on the scale. M = the location of the mean person and item measures; S = one standard deviation away from the mean; T = two standard deviations away from the mean. <more> = most able person; <less> = least able person; <frequent> = least difficult to endorse item; <rare> = most difficult to endorse item

The difficulty of items MD1 to MD10 is shown in Table 1, along with the biological theme, the model form, and the representation type of each item. Models from the most difficult to be endorsed to the least difficult to be endorsed are as follows: concept-process models (MD5), theoretical models (MD9), map/diagrams (MD2, MD8), simulation (MD7, MD10, MD1, MD6), and scale models (MD3, MD4). In terms of the presentation type, the two text models seem to be the most difficult to be endorsed, followed by the symbolic models, the animated models, and then the concrete model. Among the four animated models, the students tended to agree with the 3D animated models more than the 2D animated models. The students considered MD5 and MD9 the most difficult, possibly because they were both represented in text as well as being more abstract and theoretical than the other models.

As for the items of the six aspects, five were located between one and two standard deviations above the mean difficulty (see Fig. 1). The five most difficult to be endorsed items included four “model type” items and one “purpose of models” item. It was difficult for the students to understand that models can be text (MT2), mathematical (MT6), dynamic (MT5), and graph (MT3). Also, it was difficult for the students to agree that “models are used to predict” (PM2). There were also five items located between one and two standard deviations below the average. These easy-to-agree items include “models are used to explain” (PM1), “models represent structures and functions” (MC3), “scientists use multiple models to represent the same phenomena” (MM1), “models are for sense-making” (PM4), and “models are for making observation” (PM3). As shown on the Wright map, the “model type” and “purposes of models” items covered a wide range of item difficulties. At one end of the map, “model type” items appear to be relatively more difficult than items of the other aspects. At the other end, the “purpose of models” items appear to be the most likely to be endorsed by the students, with the exception of item PM2.

The items of “model content” spread out between the mean level and around one standard deviation below. The students seemed to have difficulty grasping the idea that models can include invisible components (MC5), while they tended to agree that models “represent structures and functions” (MC3). The three items of the “change of model” were located above the average, meaning that understanding the nature of changing models is slightly difficult for the students. Nonetheless, the students agreed to a similar extent among the three scenarios why scientists change models. The “constructed nature of models” items were consistently located around the mean level of difficulty. The students were inclined to agree with the ideas that models are “abstract” (CM1), “composed of elements” (CM2), and “deducted from evidence” (CM3).

The results of the six items representing the “model type” aspect also verified some of the aforementioned findings through the MD items. The results show consistency in terms of the text models being the most difficult type of model, and the concrete and three-dimensional models being the least difficult type of models (see Table 2). The students responded to the dynamic type model (MT5) differently to their response to the four animated/simulation items (i.e., MD7, MD1, MD10, and MD6). In other words, some of the MD items appeared to be easier than the MT items even though they all target similar types of models. These findings will be discussed later in the paper.

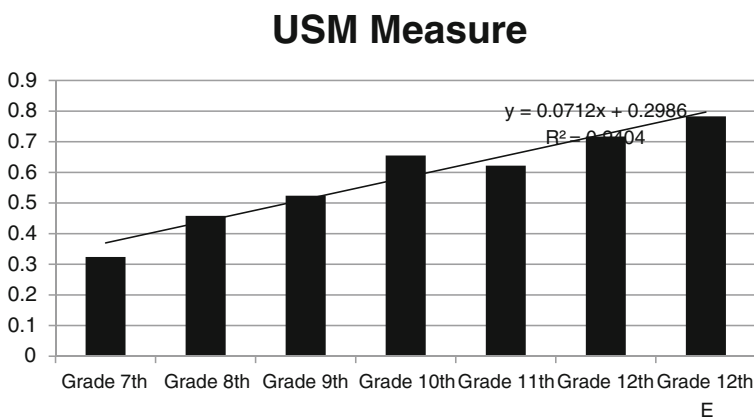
## Progression of Students' USM

The overall trend shows a slow but steady progression from seventh grade through 12th grade (Fig. 2). Furthermore, the results of the ANOVA analyses and Bonferroni tests (post hoc analyses) show significant differences between some grade levels ( $F = 5.630$ ,  $p < .001$ ; see Appendix 4). Tenth, 11th, and 12th graders (both 12th and 12thE) showed statistically significantly better understanding of models and modeling than the 7th grade students. There was also a statistically significant difference between the 8th grade students and 12th grade end of school year students (12thE). The students in the 11th grade social science track had a lower mean score than their science track counterparts; however, the differences were not statistically significant.

One important goal of this study was to know whether students who received formal instruction in models and modeling would advance further in their USM. The post hoc analysis showed that after participating in the modeling curriculum for about 2 weeks, the 7th grade treatment group developed significantly better understanding of USM (mean = 0.1.079, SD = 0 .661) than the non-intervention 7th and 8th grade students.

## Conclusions and Discussion

In summary, based on this questionnaire USM-CBQ, the “model type” and “change of models” items appear to be more difficult for the student to endorse. The items of “multiple models” and “constructed nature of models” are of middle level of difficulty based on the Rasch statistics. All items that were considered of higher level (i.e., MC5, MT2, MT6, CM1, MM3, CN1, PM2, PM1) as defined in the literature were consistently less likely for the students to endorse (i.e., more “difficult”) in all aspects in the current study. One possible explanation of why the “model type” items are more difficult is because they deal with abstract forms of



**Fig. 2** Students' USM measures by grade level. Grade 12thE: 12th grade students surveyed at the end of the school year

model. The “change of models” items seem to be more difficult for the students to endorse because the students might not be familiar with the epistemic reasons of changing models. Therefore, future instruction should place more emphasis on USM that appears to be above the average level of difficulty, including understanding of the “model type” and “change of models” aspects, and the predictive nature of models in the “purpose of models.”

The “purpose of model” items seem to be relatively easy for the students to agree with, except for PM2 (“models are used to predict”). The difference in terms of item difficulty between “models are used to predict” and “models are used to explain” in the current study validated the three levels of understanding proposed by Upmeier zu Belze and Kruger (Krell et al., 2015; Upmeier zu Belzen & Kruger, 2010). In their framework of “model of model competence”, level III represents “predicting something about the original;” level II refers to “explaining the original;” and level I is “describing the original.” Upmeier zu Belzen and Kruger (2010) theorized that the viewing models as research tools for science purposes is of higher level of understanding. Treagust, Chittleborough, & Mamiala, (2004) also found that 11th grade students had limited understanding of the predictive nature of modeling, while the same group of students had good understanding of the descriptive nature of modeling. One possible explanation is that for students in Taiwan, most of their experience with science learning has focused on making sense of and memorizing historical and curricular models. Thus, they have limited understanding or experience that models are built to predict scientific events or phenomena.

Regarding the students’ progression over grade levels, the results show significant differences between the 7th graders and the students in the 10th grade and above, and between 8th grade and 12th grade students who finished the school year. This result suggests that it is not likely to find major progression between students who are close in age. Thus, it possibly takes years of experience of learning science to develop sophisticated views of models and modeling when no explicit instruction on modeling is implemented. Even though historical and curricular models are frequently found in textbooks, students may not be aware of the nature of models and modeling. After experiencing model-based instruction, the students’ understanding of models accelerated.

One subset of the items in the questionnaire allowed us to continue the inquiry of the students’ understanding of “what a model is.” The Wright map and item difficulty in the current study further provide direct comparison among the different representations of models as well as different model types. The findings of the current study synthesized some findings suggested by different researchers. Some researchers found that abstract forms of models, such as text or diagrams, appear to be more difficult for young students (Bean, Sinatra, & Schrader, 2010; Schwarz, 2009), and other researchers have suggested that students are more likely to identify three-dimensional and dynamic models (Lee et al., 2017). Again, the previous studies compared some representations but did not compare all different types of models, such as dynamic or 2D models. This study goes further by providing a finer grained hierarchy of various forms of models in between these two extremes of text-based and 3D models.

More interestingly, the results showed that the items may become easier when representations of models are provided. For instance, item MT5 and items MD7,



MD10, MD1, and MD6 were all targeting *dynamic* types of models. The students were asked if they agreed with the statement that “a model can be dynamic” in MT5, while they were asked if they agreed that dynamic simulations (e.g., items MD7, MD10, MD1, and MD6) are models. The result shows that the item difficulty level for MT5 (1.08) was higher than the item difficulties for the four MD items with representations of models (from  $-0.28$  to  $-0.76$ ). Similar patterns of differences can also be found in comparing the *concrete* types of items and models. These results imply that the provision of sample models or context in the questionnaire rather than only describing the nature of models in text may help the students to answer questions related to “what is a model.” This finding should be taken into account in future questionnaire design.

Finally, the results showed that the USM-CBQ has satisfactory psychometric properties, and the Rasch statistic is suitable for the research goals of the study. Additionally, the current tool covers a wide distribution of item difficulty (i.e., between 1.57 and  $-1.43$ ). One might expect a different pattern of progression after students participate in model-based instruction. Yet, the methodology can still be applied to future studies in which more effects of model-based instruction are examined.

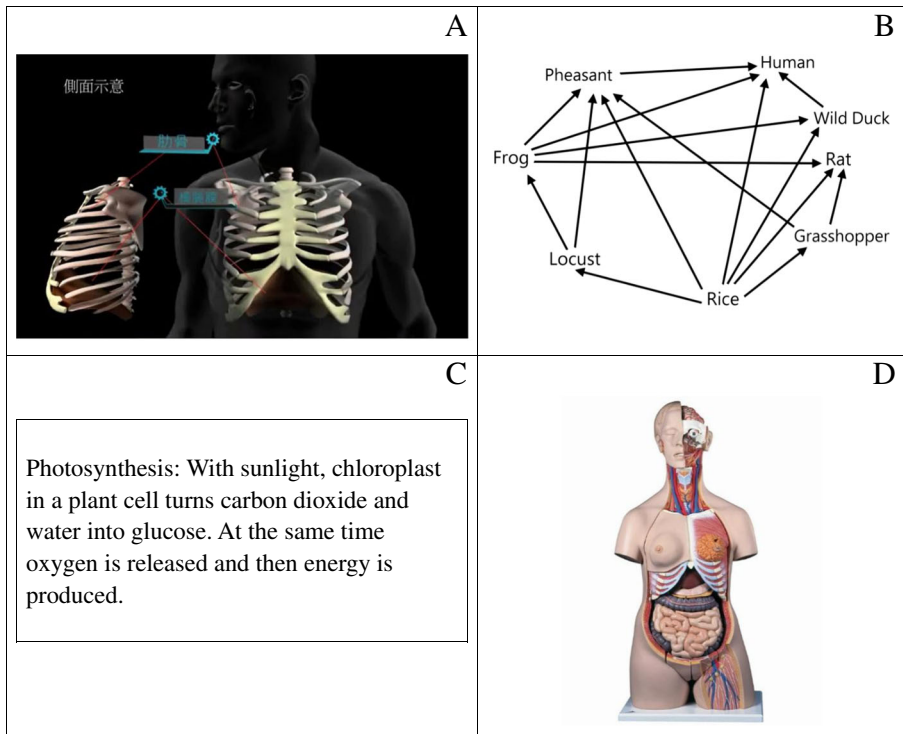
**Acknowledgements** Many thanks to Dr. Hsin-Kai Wu for her comments on an earlier version of the manuscript. This work was supported by the Ministry of Science and Technology, Taiwan [Grant numbers: MOST 104-2511-S-018-013-MY4, and MOST 103-2628-S-018-002-MY3].

## Appendix 1

Description of the model-based curriculum:

One class of seventh grade students ( $n = 23$ ) had received special instruction on models and modeling. The students in the treatment class participated in a 12-h modeling unit about marine ecology and fishery sustainability. This unit modified the Model-centered Instructional Sequence proposed by Baek, Schwarz, Chen, Hokayem, and Zhan (2011). During this unit, the students were engaged in a series of activities that helped them to build and revise models. The unit included five major stages: (1) brief introduction of the nature of models and modeling; (2) anchoring by the driving question of “Is it possible that 1 day we can no longer catch any cod fish in the ocean?” and developing an initial model; (3) investigation of the phenomenon through simulation; (4) introduction of scientific concepts and model revision; and (5) introduction of more concepts and final model revision. The students built concept maps as conceptual models of the relationships between marine ecology and human activities. In the initial stage, the students read a brief introduction about the nature of scientific models and modeling and participated in whole-class discussion led by the teacher. During the investigation stage, the students participated in the Fishbanks game (Meadows, Sterman, & King, 2017), in which they competed for fish in a multiple-player simulation. As fishermen, the students experienced the challenges of managing sustainable marine resources. In the fourth and fifth stages, through analyzing data, reading learning materials, and engaging in small group discussion, the students explored more concepts, including the marine food web, population dynamics, fishing methods, fishing laws, and international negotiation of sustainable resources. Then the students built their final models.

## Appendix 2



**Fig. 3** Translated sample representations (originally in Chinese). **a** A screenshot of item MD6; a 3D dynamic (animated) representation of the breathing mechanism (Wu, Lin & Hsu, 2013). **b** Item MD2; a visual representation of the food web. **c** Item MD5; a verbal (text) model of the photosynthesis mechanism. **d** Item MD4; a concrete scale model of the human body

## Appendix 3

**Table 3** Item difficulty and fit statistics for all items

Item	Infit MNSQ	Outfit MNSQ	pbis
MD1	0.926	0.940	0.381
MD2	1.257	1.345	0.205
MD3	0.907	0.916	0.396
MD4	0.977	1.000	0.371
MD5	1.337	1.313	0.145
MD6	0.935	0.988	0.411
MD7	0.984	1.004	0.405
MD8	1.138	1.126	0.243
MD9	1.134	1.098	0.177
MD10	0.904	0.920	0.408
MC1	0.818	0.836	0.410
MC2	0.767	0.760	0.477
MC3	0.763	0.766	0.437
MC4	0.868	0.860	0.443
MC5	1.301	1.284	0.317
PM1	0.741	0.743	0.463
PM2	1.094	1.062	0.349
PM3	0.829	0.823	0.443
PM4	0.771	0.771	0.478
CM1	0.910	0.899	0.397
CM2	1.046	1.043	0.333
CM3	0.766	0.766	0.501
MT1	1.185	1.179	0.249
MT2	1.002	0.968	0.200
MT3	1.075	1.046	0.375
MT4	1.097	1.096	0.332
MT5	0.954	0.918	0.334
MT6	1.008	0.982	0.279
CN1	1.081	1.069	0.271
CN2	0.964	0.951	0.310
CN3	1.086	1.071	0.258
MM1	0.955	0.959	0.372
MM2	1.334	1.380	0.217
MM3	1.065	1.111	0.271

## Appendix 4

**Table 4** Comparison of USM among different grade levels

Grade level	<i>n</i>	<i>M</i>	<i>SD</i>	LL 95% CI	UL	<i>F</i>	<i>p</i>	Post hoc
7th	145	0.324	0.873	0.180	0.467	5.630	< .001	10, 11, 12, 12E, 7T > 7
8th	158	0.458	0.597	0.365	0.551			12E, 7T > 8
9th	76	0.524	0.980	0.300	0.748			
10th	149	0.655	0.718	0.538	0.771			
11th	149	0.622	0.661	0.515	0.729			
12th	117	0.717	0.679	0.592	0.841			
12thE	87	0.783	0.631	0.648	0.917			
11thS	79	0.496	0.871	0.301	0.692			
7thT	23	1.079	0.661	0.778	1.380			

*12thE* indicates 12th grade students surveyed at the end of the school year; *11thS* indicates 11th grade students (social science track); *7thT* indicates 7th grade treatment class

## References

- Baek, H., Schwarz, C. V., 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, The Netherlands: Springer.
- Bean, T. E., Sinatra, G. M., & Schrader, P. G. (2010). Spore: Spawning evolutionary misconceptions? *Journal of Science Education and Technology*, *19*(5), 409–414. doi:<https://doi.org/10.1007/s10956-010-9211-1>.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (Second ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education*, *95*(2), 258–280. doi:<https://doi.org/10.1002/sce.20413>.
- Boulter, C. J., & Buckley, B. C. (2000). Constructing a typology of models for science education. In J. K. Gilbert & C. J. Boulter (Eds.), *Developing models in science education* (pp. 41–57). Dordrecht, The Netherlands: Springer.
- Buckley, B. C., & Boulter, C. J. (2000). Investigating the role of representations and expressed models in building mental models. In J. K. Gilbert & C. J. Boulter (Eds.), *Developing models in science education* (pp. 119–135). Dordrecht, The Netherlands: Springer.
- Cheng, M.-F., & Lin, J.-L. (2015). Investigating the relationship between students' views of scientific models and their development of models. *International Journal of Science Education*, *37*(15), 2453–2475. doi:<https://doi.org/10.1080/09500693.2015.1082671>.
- Chittleborough, G. D., Treagust, D. F., Mamiala, T. L., & Mocerino, M. (2005). Students' perceptions of the role of models in the process of science and in the process of learning. *Research in Science & Technological Education*, *23*(2), 195–212. doi:<https://doi.org/10.1080/02635140500266484>.
- Duncan, P. W., Bode, R. K., Lai, S. M., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: The stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950–963. doi:[https://doi.org/10.1016/S0003-9993\(03\)00035-2](https://doi.org/10.1016/S0003-9993(03)00035-2).
- Gilbert, J. K. (2013). Representations and models. In R. Tytler, V. Prain, P. Hubber, & B. Waldrip (Eds.), *Constructing representations to learn in science* (pp. 193–198). Rotterdam, The Netherlands: Sense.

- Gobert, J. D., & Pallant, A. (2004). Fostering students' epistemologies of models via authentic model-based tasks. *Journal of Science Education and Technology*, 13(1), 7–22. doi:<https://doi.org/10.1023/B:JOST.0000019635.70068.f6>.
- Gobert, J. D., O' Dwyer, L., Horwitz, P., Buckley, B. C., Levy, S. T., & 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(5), 653–684. doi:<https://doi.org/10.1080/09500691003720671>.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28(9), 799–822.
- Grünkom, J., Upmeier zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, 36(10), 1651–1684. doi:<https://doi.org/10.1080/09500693.2013.873155>.
- Harrison, A. G., & Treagust, D. F. (1998). Modelling in science lessons: Are there better ways to learn with models? *School Science and Mathematics*, 98(8), 420–429.
- Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22(9), 1011–1026.
- Justi, R. S., & Gilbert, J. K. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education*, 24(12), 1273–1292.
- Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2014). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in Science Education*, 44(1), 109–132. doi:<https://doi.org/10.1007/s11165-013-9365-y>.
- Krell, M., Reinisch, B., & Krüger, D. (2015). Analyzing students' understanding of models and modeling referring to the disciplines biology, chemistry, and physics. *Research in Science Education*, 45(3), 367–393. doi:<https://doi.org/10.1007/s11165-014-9427-9>.
- Lee, S. W.-Y., Chang, H.-Y., & Wu, H.-K. (2017). Students' views of scientific models and modeling: Do representational characteristics of models and students' educational levels matter? *Research in Science Education*, 47(2), 305–328. doi:<https://doi.org/10.1007/s11165-015-9502-x>.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- Linacre, J. M. (2012). *Winsteps (Version 3.75.1)*. Beaverton, OR: Winsteps. Retrieved from <http://www.winsteps.com>.
- Liu, X. (2006). Effects of combined hands-on laboratory and computer modeling on student learning of gas laws: A quasi-experimental study. *Journal of Science Education and Technology*, 15(1), 89–100. doi:<https://doi.org/10.1007/s10956-006-0359-7>.
- Meadows, D., Serman, J., & King, A. (2017). *Fishbanks: A Renewable Resource Management Simulation*. Retrieved from <https://mitsloan.mit.edu/LearningEdge/simulations/fishbanks/Pages/fish-banks.aspx>
- National Research Council. (1996). *National Science Education Standards*. Washington DC: National Academy Press.
- National Research Council. (2007). Understanding how scientific knowledge is constructed. In R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.), *Taking science to school: Learning and teaching science in grades K-8* (pp. 168–185). Washington, DC: National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and Core ideas*. Washington, DC: The National Academies Press.
- Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, 33(10), 1373–1405. doi:<https://doi.org/10.1080/09500693.2010.511297>.
- Nicolaou, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13, 52–73. doi:<https://doi.org/10.1016/j.edurev.2014.10.001>.
- Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching*, 48(5), 486–511. doi:<https://doi.org/10.1002/tea.20415>.
- Prins, G. T., Bulte, A. M., & Pilot, A. (2011). Evaluation of a design principle for fostering students' epistemological views on models and modelling using authentic practices as contexts for learning in chemistry education. *International Journal of Science Education*, 33(11), 1539–1569. doi:<https://doi.org/10.1080/09500693.2010.519405>.
- Raghavan, K., & Glaser, R. (1995). Model-based analysis and reasoning in science: The MARS curriculum. *Science Education*, 79, 37–62.

- Saari, H., & Viiri, F. (2003). A research-based teaching sequence for teaching the concept of modelling to seventh-grade students. *International Journal of Science Education*, 25(11), 1333–1352.
- Schwarz, C. V. (2009). Developing preservice elementary teachers' knowledge and practices through modeling-centered scientific inquiry. *Science Education*, 93(4), 720–744. doi:<https://doi.org/10.1002/sc.20324>.
- Schwarz, C. V., & White, B. Y. (2003). Developing a model-centered approach to science education (pp. 1–141). Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165–205.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., & Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R. V., & van Hout-Wolters, B. H. A. M. V. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, 31(9), 1205–1229. doi: <https://doi.org/10.1080/09500690802192181>.
- Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, 98(4), 581–613.
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, 24(4), 357–368.
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2004). Students' understanding of the descriptive and predictive nature of teaching models in organic chemistry. *Research in Science Education*, 34, 1–20.
- Upmeyer zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht [Model competence in biology teaching]. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 41–57.
- Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch measurement to validate the instrument of students' understanding of models in science (SUMS). *International Journal of Science and Mathematics Education*, 12(5), 1067–1082.
- Willard, T., & Roseman, J. E. (2010). *Probing students' ideas about models using standards-based assessment items*. Paper presented at the 83rd NARST Annual International Conference, Philadelphia, PA.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370. Retrieved from <http://www.rasch.org/rmt/rmt83b.htm>.
- Wu, P.-C., & Chang, L. (2008). Psychometric properties of the Chinese version of the Beck Depression Inventory-II using the Rasch model. *Measurement and Evaluation in Counseling and Development*, 41(3), 13–31.
- Wu, H.-K., Lin, Y. F. & Hsu, Y. S. (2013). Effects of representation sequences and spatial ability on students' scientific understandings about the mechanism of breathing. *Instructional Science*, 41(3), 555–573. doi: <https://doi.org/10.1007/s11251-012-9244-3>.