International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

Modeling-Oriented Assessment in K-12 Science Education: A synthesis of research from 1980 to 2013 and new directions

Bahadir Namdar[a] & Ji Shen[b]
[a] Primary Education, Faculty of Education, Recep Tayyip Erdogan University, Cayeli, Rize, Turkey
[b] Department of Teaching and Learning, University of Miami, Coral Gables, FL, USA
Published online: 18 Feb 2015.

International Journal of Science Education, 2015
Vol. 37, No. 7, 993–1023, http://dx.doi.org/10.1080/09500693.2015.1012185

Modeling-Oriented Assessment in K-12 Science Education: A synthesis of research from 1980 to 2013 and new directions

Bahadir Namdar[a] and Ji Shen[b*]
[a]*Primary Education, Faculty of Education, Recep Tayyip Erdogan University, Cayeli, Rize, Turkey;* [b]*Department of Teaching and Learning, University of Miami, Coral Gables, FL, USA*

Scientific modeling has been advocated as one of the core practices in recent science education policy initiatives. In modeling-based instruction (MBI), students use, construct, and revise models to gain scientific knowledge and inquiry skills. Oftentimes, the benefits of MBI have been documented using assessments targeting students' conceptual understanding or affective domains. Fewer studies have used assessments directly built on the ideas of modeling. The purpose of this study is to synthesize and examine modeling-oriented assessments (MOA) in the last three decades and propose new directions for research in this area. The study uses a collection of 30 empirical research articles that report MOA from an initial library of 153 articles focusing on MBI in K-12 science education from 1980 to 2013. The findings include the variety of themes within each of the three MOA dimensions (modeling products, modeling practices, and meta-modeling knowledge) and the areas of MOA still in need of much work. Based on the review, three guiding principles are proposed for future work in MOA: (a) framing MOA in an ecology of assessment, (b) providing authentic modeling contexts for assessment, and (c) spelling out the connections between MOA items and the essential aspects of modeling to be assessed.

Keywords: *Modeling-oriented assessment; Model-based learning; Modeling; Assessment*

Introduction

There is a wide range of approaches that scientists use to investigate and explain the natural world. One such approach is scientific modeling. Many scholars have

*Corresponding author. Department of Teaching and Learning, University of Miami, 1320 S. Dixie Hwy, Coral Gables, FL 33124, USA. Email: ji.shen1221@gmail.com

promoted model-based teaching and learning, or modeling-based instruction (MBI) in science education (Buckley, 2000; Cheng & Brown, 2010; Gilbert & Boutler, 2000; Gobert, 2000; Johnstone & Mahmoud, 1981; Louca & Zacharia, 2012; Roth, Woszczyna, Smith, Universiq, & Va, 1996; Shen, Lei, Chang, & Namdar, 2014; White & Frederiksen, 1998). This approach allows students to actively participate in their own learning (Penner, Lehrer, & Schauble, 1998; Schwarz et al., 2009) and externalize their ideas via multiple representations (Shen & Confrey 2007, 2010). Modeling tasks can also serve as an authentic environment in which students develop and apply various scientific practices similar to what scientists do (Nersessian, 2008, 2009; Penner et al., 1998).

The new *Framework for K-12 Science Education* (National Research Council, 2012) recognizes developing and using models as one of the eight core practices in science education. Accordingly, the *Next Generation Science Standards* (NGSS Leads States, 2013) advocates that 'the practice of modeling will need to be taught throughout the year and indeed throughout the entire K-12 experience' (pp. 2–3).

Although many MBI studies have developed genuine ways to evaluate specific types of modeling approaches (e.g. Hogan & Thomas, 2001; Prins, Bulte, Driel, & Pilot, 2009; Rivet & Kastens, 2012), the majority of the assessments in these studies only targeted student gains in content knowledge and/or affective domains (e.g. Angell, Kind, Henriksen, & Guttersrud, 2008; Fazio, Guastella, Sperandeo-Mineo, & Tarantino, 2008; Klopfer & Um, 2005; Liu, 2006; Pallant & Tinker, 2004). Much less attention has been paid to directly assess students' knowledge, skills, and practices related to models and modeling, or modeling-oriented assessment (MOA) as termed in this paper.

This is problematic. First of all, given the importance of scientific modeling in NGSS (NGSS Lead States, 2013) and the call to align assessment with curriculum (National Research Council, 2001, 2014), it is imperative to employ appropriate assessments in a modeling-based curriculum that truly capture students' learning gains (van Borkulo, van Joolingen, Savelsbergh, & de Jong, 2012). Furthermore, assessments should be designed in a systematic way that informs teachers (and students) about student learning (Binkley et al., 2012). A clear and systematic understanding of MOA is necessary to provide valuable feedback to improve students' modeling practices.

This review paper, therefore, serves as a starting point to build a sound understanding of MOA. The following research questions guided the study: (1) How has MOA been approached by researchers in K-12 settings in the last three decades?, (2) What are the patterns and gaps regarding MOA in the literature?, and (3) What insights can we gain for future work in MOA?

Theoretical Framework

Models, Modeling, and Meta-modeling Knowledge

In science education, models and modeling have been defined in many ways. In this paper, we adopt a broad definition of model: a *model* is a human construct used to describe, explain, predict, and communicate with others a referent such as a natural phenomenon, an event, or an entity (Shen, 2006). Human constructs may include physical representations (e.g. Fretz et al., 2002), computer simulations (e.g. Scalise, Timms, Moorjani, Clark, & Holtermann, 2011; Smetana & Bell, 2012; Stern, Barnea, & Shauli, 2008; Stratford, Krajcik, & Soloway, 1998), mathematical formula and equations (e.g. Hestenes, 1987; Lehrer & Schauble, 2006), and rules and relations (e.g. Bravo, van Joolingen, & de Jong, 2009). We also distinguish between mental models and expressed models of scientific phenomena (Gilbert, Boutler, & Elmer, 2000). Mental models are approached as 'internal, personal, idiosyncratic, incomplete, unstable, and essentially functional' models (National Research Council, 2012, p. 56). Expressed models, on the other hand, are external representations of the phenomena under study (Gilbert & Boutler, 2000). Therefore, they embody to a certain degree corresponding mental models. In this paper, unless being explicitly pointed out, the term model(s) refers to expressed model(s).

Modeling is the ensemble of the processes that generate models. For instance, Schwarz et al. (2009, p. 635) lays out four key processes:

- Students *construct* models consistent with prior evidence and theories to illustrate, explain, or predict phenomena.
- Students *evaluate* the ability of different models to accurately represent and account for patterns in phenomena, and to predict new phenomena.
- Students *use* models to illustrate, explain, and predict phenomena.
- Students *revise* models to increase their explanatory and predictive power, taking into account new evidence or additional aspects of a phenomenon.

Many researchers also recognize the importance of planning where students make decisions before or during the construction of their models (e.g. Prins et al., 2009; Wu, 2010). Another important process is model interpretation, in which students explain, discuss, and reflect on their models (e.g. Chang, Quintana, & Krajcik, 2010; Hogan & Thomas, 2001). In reality, these modeling processes neither have well-defined boundaries nor do they necessarily follow a prescribed sequence. For instance, the planning process may go before model construction or after evaluation (to plan for a revision); model interpretation is often embedded within the other processes.

Executing the modeling processes requires a set of integrated knowledge, skills, and cognitive strategies relevant to modeling (e.g. Stratford et al., 1998), which we call *modeling practices* in this paper (Fretz et al., 2002; National Research Council, 2012; Schwarz et al., 2009). Constructing a model, for instance, involves practices such as identifying a system and its components in the specified knowledge domain, and specifying component behaviors while making meaningful connections among them. Model evaluation includes practices such as testing the model behaviors and deciding whether the model works or not based on domain knowledge, and calibrating and validating models under certain conditions. Model revision includes practices such as changing the model based on new knowledge and/or prior results of model evaluation.

Performing high-quality modeling practices requires learners to 'understand the rationale, norms, and the nature of scientific modeling' (Schwarz et al., 2009, p. 634). *Meta-modeling knowledge* refers to the knowledge about the nature and purpose of scientific models and modeling. Schwarz and White (2005) argued that without this knowledge, students could neither comprehend the nature of science nor perform modeling practices in a sound manner. They further argued that enhancing students' meta-modeling knowledge would enhance their evidence-based reasoning and integration of scientific knowledge. Hence, they advocated for adding a meta-modeling layer in a modeling curriculum and identified the four aspects of meta-modeling knowledge: (a) nature of models, (b) nature or process of modeling, (c) evaluation of models, and (d) purpose or utility of models. We adopt this meta-modeling definition in this paper.

Modeling-based Instruction

Many scientific phenomena and experiments cannot be directly introduced in the classrooms because of time, safety, scale, and budget constraints. Models are used to help overcome these difficulties through visualization, simplification, manipulation, scale transformation, and mathematization (National Research Council, 2012). We use MBI to denote model-based teaching and learning (Gobert & Buckley, 2000) and emphasize the non-separable nature between teaching and learning in a classroom environment (Branch & Kopcha, 2013). MBI includes the cognitive aspect that concerns learners' reasoning and development of mental models (Norman, 1983; Vosniadou, 1994), the social aspect that focuses on students' construction and negotiation of expressed and consensus models (Gilbert & Boutler, 2000), and the pedagogical aspect that recognizes the use of teaching models as a meditational means for sensemaking and social interaction (Harrison & Treagust, 2000). With the fast advancement in the information communication technologies, MBI has moved up to a whole new level from both cognitive and social perspectives (Buckley et al., 2004; Fretz et al., 2002; Shen et al., 2014).

Researchers have proposed MBI frameworks and curricula from different theoretical perspectives. For instance, Hestenes (1987) and Hestenes, Wells, and Swackhamer (1992) established an MBI approach that emphasized procedures moving towards mathematical formulation. Hestenes (1987) asserted that models, especially the ones in physics, are 'physical properties represented by quantitative variables' (p. 441). Therefore, to teach scientific knowledge effectively, mathematical modeling of the physical world should be prioritized. Lehrer and Schauble (2000, 2006) stressed the qualitative resemblance between a base system (model) and a target system (natural world) in modeling. Considering analogies as the basic forms of models, they developed a typology of models, including physical microcosms (e.g. mechanical models of the solar system), representational models (e.g. maps of the world), syntactic models (e.g. models that have little resemblance to target systems but summarize the essential function that is being modeled), and emergent models (e.g. models based on relations between objects that lead to emergent behaviors). They also pointed out that incorporating analogical development is the key to support model-based reasoning. Looking at a longer time scale, Schwarz et al. (2009) proposed a learning progression, which centers around two modeling dimensions, 'models as generative tools for predicting and explaining', and 'models change as our understanding improves incorporating the modeling practices and meta-modeling knowledge' (p. 637). They argued that these dimensions depict the interrelationship between students' modeling practices and their meta-modeling knowledge. They emphasized the importance of creating a learning progression as it would 'capture a coherent incremental trajectory' (p. 652) and the interconnected aspects of meta-modeling knowledge.

Modeling-oriented Assessment

Educational assessment, on the one hand, can be described as 'a formal attempt to determine students' status with respect to educational variables of interest' (Popham, 2010, p. 7); on the other hand, the *assessment for learning* perspective submits that students' dynamic learning progress can be facilitated by the practice of (formative) assessments (e.g. Binkley et al., 2012; Black & William, 2010; National Research Council, 2001). In this paper, we take an integrated perspective and define MOA as both a way to determine students' status with respect to variables of interest from a modeling perspective and a way to enhance student learning through modeling.

We envision that MOA encompasses three interrelated dimensions: (a) assessment of modeling products (e.g. Manlove, Lazonder, & de Jong, 2009; Nuhoglu, 2008), (b) assessment of modeling practices (e.g. Chang et al., 2010; Ergazaki, Zogza, & Komis, 2007), and (c) assessment of meta-modeling knowledge (e.g. Papaevripidou, Constantinou, & Zacharia, 2007). The assessment of modeling practices targets evaluating in real-time how students go about using, constructing, and revising models. The dynamic and iterative nature of modeling processes makes the data collection challenging and the analyses time consuming. It is also context and task specific. In many modeling tasks, students are expected to construct their own models. In these tasks, the assessment of modeling products is made feasible. Even in their premature forms, these concrete products may indicate students' modeling knowledge and abilities. Compared with the assessment of modeling practices, the assessment of modeling products is relatively easy to approach as long as one establishes a good set of criteria to evaluate models (Pluta, Chinn, & Duncan, 2011); but it is also conditioned upon the specific modeling task and resources. The assessment of meta-modeling knowledge goes beyond a specific modeling task, and therefore, provides a relatively robust estimate of a person's modeling disposition over a longer term.

Methods

MBI Literature Collection and Coding

The literature reviewed in this study was drawn from a larger project that aimed to synthesize MBI in the last three decades (Shen, Lei, Namdar, & Chen 2013). In that project, the MBI literature in K-12 science education settings was searched and collected using the electronic database ERIC, as well as the specific science education journals. Various combinations and permutations of the following keywords were used: 'model-based', 'science education', 'model-based teaching', 'model-based learning', 'model-based instruction', and 'model-based science'. The search was repeated after the word 'model' was replaced with 'modeling' and then with the British spelling, 'modelling'.

The search generated a large number of references and the following criteria were used to include or exclude relevant literature in the project. The study should be based on MBI that focuses on expressed models. We excluded literature focusing on mental model research, which greatly overlaps with the misconception and conceptual change literature (e.g. Vosniadou, 1994). Also excluded were papers only focusing on students' understanding of models without any description of an MBI intervention (e.g. Grosslight, Unger, Jay, & Smith, 1991; Hsu, Lin, Wu, Lee, & Hwang, 2012; Rivet & Kastens, 2012; Treagust, Chittleborough, & Mamiala, 2002). Because the focus was on K-12 educational settings, studies conducted with college students or K-12 teachers were excluded (e.g. Crawford & Cullin, 2004). We only included studies published in English from 1980 to 2013.[1]

After the articles were included in our library, we coded the individual articles based on our coding schema that included Study Information (e.g. author, year, source, and article type), MBI Theory (e.g. overall theoretical description, attention to mapping between base and target), MBI Design (e.g. modeling processes, technology, embedded scaffolding, and collaboration type), and MBI Assessment (e.g. outcome variables, assessment format, delivery mode, reliability and validity, and statistical results). For instance, *outcome variables* include the following codes: (a) Affective: students' affective outcomes including interest, attitudes, motivation, etc.; (b) Conceptual: students' content knowledge/conceptual understanding; (c) MOA: students' modeling practices, products, and meta-modeling knowledge; and (d) Other: all other outcomes (e.g. tool familiarity).

The literature collection and coding in the project were completed by two collaborating teams at the University of Georgia and Syracuse University. First, both teams coded the same subset of the articles (20%) to establish inter-rater reliability and refine the coding scheme. Then, each team was assigned to code one half of the rest of the articles. Questions and concerns from each team's coding were brought up to discussion. After this, the teams switched and recoded the other team's articles to double check. All coding inconsistencies were discussed and negotiated until the final agreement was reached between the two teams.

Analyzing MOA Articles

For this study, only research articles that reported MOA empirically and explicitly were included; that is, in order to be included, the article needs to report quantitative methods or qualitative rubrics to code, score, rate, rank, or compare students' models, modeling practices, or meta-modeling knowledge with empirical data. MBI articles only assessing other areas such as students' conceptual understanding and affective domains were excluded. To ensure the minimum quality of the literature gathered, we only included those that clearly reported the following: clear research question(s), reasonable theoretical elaboration, appropriate research methodology with clear description of appropriate sampling, study context, data collection and analysis procedures, and findings backed up with evidence. Several other codes (e.g. theoretical description, study type, and statistical results) also indicated the varying quality of the research articles. Although an *article* may report multiple *studies*, for the purpose of this paper, individual articles were used as our unit of analysis because the related studies within an article typically propose coherent assessment strategies and are framed under the same MBI perspective.

Our MOA synthesis was informed by high-quality synthesis studies in the field (e.g. Minner, Levy, & Century, 2010; Scalise et al., 2011). In order to better depict the landscape of MOA, we followed six steps of thematic analysis proposed by Braun and Clarke (2006). First, we familiarized ourselves with the articles having some prior knowledge about the topic. Based on our theoretical framework, we identified the three dimensions of MOA as assessing modeling products, modeling practices, and meta-modeling knowledge. In the second step, the first author, in consultation with the second author, generated initial codes of MOA within each of the three dimension based on a thorough reading of each article. In the third step, after the articles had been initially coded, we applied different strategies to organize the codes into a smaller number of themes. For the dimension of modeling products, we sorted out similar codes and negotiated how these codes could combine in overarching themes. For the dimension of meta-modeling knowledge, we adopted Schwarz and White's (2005) categorization because they matched perfectly what we found in the articles. Realizing the complexity of modeling practices, we revisited all the articles reporting the assessment of modeling practices and coded them independently. We then compared, discussed, and negotiated all the codes and organized them into themes until we reached complete agreement. In the fourth step, we reviewed and refined all the themes to assure that there were enough data supporting them for the dimensions of modeling practices and products. As the theme creation process was an inductive and dynamic process, we constantly went back and forth between the articles and the themes, and eventually matched all of our codes with the themes. In the fifth step, after refining the themes, we defined what each theme represents and refined them again (e.g. merged related themes into a new one), and determined what aspect of the data each theme captured and named them. Finally, we produced the final report that includes all the themes and specific examples in the three dimensions of MOA, as well as patterns and gaps observed in the literature.

We acknowledge that our taxonomy only provides one way to examine MOA. Our classification of the MOA dimensions and themes may not be consistent with what the authors claimed in their articles as they may have used different terminologies. For instance, Ergazaki et al. (2007) examined students' major modeling *actions* including analysis, synthesis, testing, and interpreting. Each of these actions included several *operations*.

Results

Based on our criteria, a total of 153 articles were included in our MBI library: 9 articles published during 1980–1989, 29 articles published during 1990–1999, and 115 articles published during the 2000–2013 period. Among the 153 articles, there were 104 empirical articles. Among these empirical articles, only 30 articles reported MOA. These 30 articles were analyzed in this study. In the following, we first report some basic characteristics of the articles, then report and draw on the results to address our research questions.

Study Characteristics

Table 1 lists some attributes and codes of the 30 articles. The publication years of these articles range from 1997 to 2013, indicating that the earlier studies did not include MOA. The content topics include biology ($n = 7$), chemistry ($n = 8$), physics ($n = 10$), and earth science ($n = 5$). The grade levels span from upper elementary to high school, with the majority of them ($n = 19$) focusing on or including high school students, and only 4 focusing on the elementary level. The majority of the MOA articles ($n = 25$) employed technology-enhanced MBI. There is not a discernable pattern, however, when comparing the MOA approaches used in MBI studies with a technology component and those without.

Although all of the 30 articles came from peer-reviewed research journals and met the minimum quality criteria, they do show varying degrees of research quality. On the positive side, approximately two-thirds ($n = 19$) of the articles reported extensive theoretical description of MBI (the highest ranking in theoretical description in our coding). On the negative side, in terms of experimental design, only three articles employed randomized design approaches; in terms of reporting statistical results, only four articles included effect sizes.

The MOA measures were employed to serve different types of research design in the articles (Figure 1), including (a) pre-/post-MBI intervention comparison, (b) comparison between an MBI intervention and a control group, (c) comparison between different versions of MBI treatments (e.g. dyads versus individuals, Manlove et al., 2009), (d) case studies, and (e) others types such as using MOA to report modeling outcome without any comparison. Moreover, 21 articles reported learning outcome in terms of MOA as a result of the MBI intervention, and out of these articles 13 reported gains in terms of MOA, 1 reported no gains, and 7 reported mixed results.

MOA Dimensions and Themes

To answer our first research question 'How has MOA been approached by researchers in K-12 settings in the last three decades?' we report the themes of MOA here. Figure 2 shows the distribution of the 30 MOA articles in terms of the 3 dimensions of modeling assessment. The majority of the articles (21 out of 30) reported a single MOA dimension, and only 1 article (Papaevripidou et al., 2007) included all 3

Table 1. MOA articles and coding results

Authors (year)	Participants *N* (grade/age)	Topic	Technology platform	Unit of analysis[a]	Assessment format[b]	Assessment themes[c]		
						PROD	PRAC	META
Barak and Hussein-Farraj (2013)	175 (12th)	Bio-chemistry	3D Models	I	P	1,3		
Barnea and Dori (1999)	169 (10th)	Chemistry	Desktop molecular modeler	I	S/C			1,2,4
Chang et al. (2010)	271 (7th)	Chemistry	Chemation	I	C/P	1,2	3	
Dori and Kaberman (2012)	614 (12th)	Chemistry	CCL & CMM	I	P	1,2		
Ergazaki et al. (2007)	6 (high school)	Biology	Models-creator	C/I	P		1,2,3	
Gobert and Pallant (2004)	360 (middle school)	Earth science	WISE	I	C			1,2,4
Gobert (2000)	47 (5th)	Earth science	–	I	P	1,2,3		
Hogan and Thomas (2001)	10 (11th & 12th)	Biology	STELLA	C	P		2,3	
Kaberman and Dori (2009)	414 (12th)	Chemistry	CCL & CMM	I	P	1,2		
Komis et al. (2007)	2 (high school)	Biology	Models-creator	I	P		1,2,3	
Kurtz dos Santos et al. (1997)	(11–18 years old)	Biology	VISQ	I	P	1,3		
Levy and Wilensky (2009b)	904 (high school)	Chemistry	NetLogo & Pedagogica	I	S/P	2		1,2,4
Liu (2006)	33 (high school)	Chemistry	Model simulations from websites	I	S/P		2,3	1,4
Louca et al. (2011)	38 (11- and 12-year olds)	Physics	Stagecast creator	I	P	1,2	3	
Louca et al. (2011)	40 (6th)	Physics	Stagecast Creator & Microworlds Logo	C	P	1,2,3		

(*Continued*)

Table 1. Continued

Authors (year)	Participants *N* (grade/age)	Topic	Technology platform	Unit of analysis[a]	Assessment format[b]	Assessment themes[c]		
						PROD	PRAC	META
Manlove et al. (2009)	42 (high school)	Physics	Co-Lab	C/I	P	1		
Mulder et al. (2010)	31 (junior and senior high)	Physics	Co-Lab	I	P	1	2,3,4	
Nuhoglu (2008)	10 (7th)	Physics	STELLA	I	P	1,2,3		
Papaevripidou et al. (2007)	33 (5th)	Biology	Stagecast Creator	I	P	1	2,3,4	2,3,4
Pluta et al. (2011)	324 (7th grade)	Earth science	–	I	C			1,4
Prins et al. (2009)	18 (11th)	Chemistry	–	C	I		1	
Saari (2003)	122 (7th and 9th)	Physics	–	I	C			1,2,4
Schwarz and White (2005)	72 (7th)	Physics	Thinker tools	I	I/C			1,2,3,4
Schwarz et al. (2009)	NA (5th and 6th)	Physics	–	I	I/P			1,2,3,4
Sins et al. (2009)	26 (11th)	Physics	Powersim®	I	C/I		2,3,4	1,2,3,4
Stratford et al. (1998)	16 (9th)	Biology	Model-It	C	I		1,2,3	
Sun and Looi (2013)	46 (high school)	Physics	WiMVT	C	P	1,2	2,3	
Treagust et al. (2004)	36 (11th)	Chemistry	The Chemistry Set	I	S			1,2,3,4
Van Borkulo et al. (2012)	74 (16–19-year olds)	Earth science	CoLab	I	C	1	3,4	
Wu (2010)	29 (10th)	Earth science	APoME	I	I		1,2,3,4	

[a]Codes: C, collective; I, individual.

[b]Codes: S, selective test; C, constructive test; I, interview or other oral format; P, performance.

[c]Codes: PROD (1, PROD-construct; 2, PROD-representation; 3, PROD-coherence), PRAC (1, PRAC-plan; 2, PRAC-generation; 3, PRAC-interpretation; 4, PRAC-evaluation), META (1, META-model; 2, META-modeling; 3, META-evaluation; 4, META-purpose).

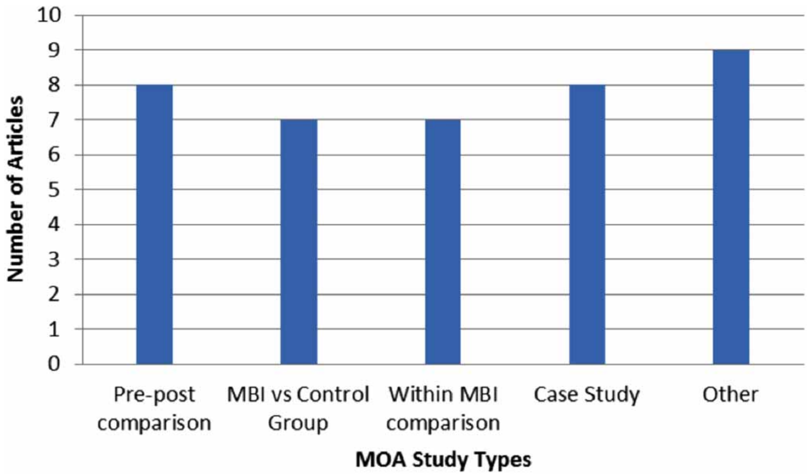

Figure 1. Frequency of study types (some studies are categorized in more than one study type).

dimensions. In the following, we explain in detail the types of assessment used in each dimension. Examples are included to help readers understand these assessment approaches.

Assess modeling products. We found that 15 articles (50%) assessed students' modeling products. Among these articles, we identified the following three themes (Table 2): quality of model construct (PROD-construct), quality of model representation (PROD-representation), and coherence of a model as a whole product (PROD-coherence).

The first theme *PROD-construct* refers to the procedures which assess whether a model contains relevant components that are considered necessary by experts to

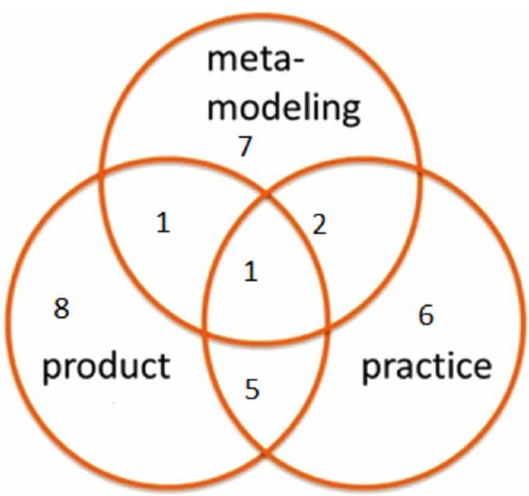

Figure 2. Distribution of articles in the three MOA dimensions.

Table 2. Types of assessment of modeling products from the literature

Theme	Explanation
1. Quality of model construct (PROD-construct)	Assess the quality, quantity, characteristics, and connections of the components of a model such as correct names for model elements, types of the variables involved, number of variables and their relationship
2. Quality of model representation (PROD-representation)	Assess the representational aspect of a model including the correctness and completeness of symbolic representations such as the accuracy of mathematical equations and chemical representations, and the proficiency of explanatory texts used in a model
3. Coherence of a model as a whole (PROD-coherence)	Assess the overall quality of a model in terms of how well the whole model coherently reflects the real-world phenomenon under investigation

represent structural connections in a system to be modeled (Barak & Hussein-Farraj, 2013; Dori & Kaberman, 2012; Kaberman & Dori, 2009; Louca, Zacharia, Michael, & Constantinou, 2011; Mulder, Lazonder, & de Jong, 2010; Sun & Looi, 2013) and the accuracy of the connections among those components (Kurtz dos Santos, Thielo, & Kleer, 1997; Gobert, 2000; Manlove et al., 2009; Mulder et al., 2010; Nuhoglu, 2008; Papaevripidou et al., 2007; van Borkulo et al., 2012). Different scoring mechanisms have been used to assess the accuracy of model constructs such as scoring the levels of correctly specified model components (Chang et al., 2010), awarding points to correctly named variables (Manlove et al., 2009), assessing the cause and effect relationship in causal models (Nuhoglu, 2008), and correctly locating the model components (Gobert, 2000).

The second theme *PROD-representation* refers to the procedures which assess the quality of specific representations carrying the communicative features of models, including textual representations (Chang et al., 2010; Gobert, 2000), and symbolic representations such as mathematical equations (e.g. Nuhoglu, 2008) and chemical formulas (e.g. Barak & Hussein-Farraj, 2013; Dori & Kaberman, 2012). Mathematical equations, emphasizing the abstract representation of a model, can be viewed as a special case of assessing simultaneously the model components (the variables) and their relationships (the formulas). Levy and Wilensky (2009a, 2009b), for instance, described a unit on gas laws and kinetic molecular theory in the Connected Chemistry Curriculum. This curriculum aimed to support students in making transitions between the submicroscopic and macroscopic levels, connecting conceptual and symbolic descriptions, and understanding models as a representation of physical world. They investigated student learning by analyzing how students constructed mathematical equation of the ideal gas law ($PV = kNT$). In the assessment, the equation was deemed correct if at least three variables and all the dependencies were correct.

The third theme *PROD-coherence* refers to the procedures which assess whether a model is coherent, interpreted in any one or the combination of the following two aspects: (a) if the model is internally coherent as a whole with respect to some established rules (Barak & Hussein-Farraj, 2013; Nuhoglu, 2008) or (b) if the model can be coherently mapped onto the external real world (Gobert, 2000; Louca, Zacharia, & Constantinou, 2011). For the former, Nuhoglu (2008) assessed whether student-generated models could run correctly in STELLA, a system dynamic-modeling tool. This means that all of the linkages among model components and mathematical equations needed to be correct, and the system model as a whole could generate meaningful graphic output in STELLA. For the latter, Kurtz dos Santos et al. (1997) assessed the coherence of the final models created by students. Using a semi-quantitative system modeling tool VISQ, the students assigned qualitative variables and connected them to model a natural system behavior (e.g. rat population explosion). Their final models were classified as either fully or partially coherent as a whole. A fully coherent model referred to one that contained correct qualitative values for the variables as expected in science and life.

The most popular theme in the category of assessing modeling products is PROD-construct (14 out of 15 articles) and the least popular one is PROD-representation (5 out of 15 articles). About two-thirds of the articles covered more than one theme. For instance, Gobert (2000) investigated fifth graders' learning on plate tectonics. Students were asked to draw different earth layers and their movements under different conditions. Gobert (2000) assessed whether students correctly identify the components of volcanic eruption (PROD-construct), whether students correctly labeled the attributions on the model components (PROD-representation), whether their drawn models of interior of the earth are spatially correct (PROD-construct), and whether these drawings reflect the real-world phenomenon of characteristics of interior of the earth (PROD-coherence).

Assess modeling practices. We found a total of 14 articles (47%) reporting assessments of modeling practices. These were organized in four themes: quality of model planning (PRAC-plan), quality of model generation (PRAC-generation) that includes model construction, testing, and revision, quality of model interpretation (PRAC-interpretation), and quality of model evaluation (see Table 3).

The *PRAC-plan* assessment theme focuses on the decision-making processes about the constituents, function, conditions, and scope of a model. This assessment theme includes a variety of indicators, as model planning may cover an array of practices such as,

- defining the problem and the purpose of the model (Prins et al., 2009),
- proposing justified plans for models (Stratford et al., 1998),
- indicating possible limitations and errors of models (Wu, 2010),
- creating model hypothesis based on the domain specification (Mulder et al., 2010),
- using appropriate scaffolds in model planning (Manlove et al., 2009), and
- finding proper objects and content areas for modeling (Ergazaki et al., 2007; Komis, Ergazaki, & Zogza, 2007; Prins et al., 2009).

These model planning practices ensure that the modelers understand the problem, the context, and the scope of a model, locate the relevant resources

Table 3. Types of assessment of modeling practices from the literature

Theme	Explanation
1. Quality of model planning (PRAC-plan)	Assess model planning practices, including proposing a justified and efficient plan to create a model, defining the modeling problem and context, and deciding the scope, purpose, and limitations of a model
2. Quality of model generation, including construction, testing, and revision (PRAC-generation)	Assess model generation practices including model construction such as identifying, selecting, creating, adding, deleting, and changing model elements as well as their properties and connections; model testing such as running a model in a modeling environment; and model revision such as adding/deleting/changing model parts, relationships, and quantities, and modifying and improving a model based on testing
3. Quality of model interpretation (PRAC-interpretation)	Assess model interpretation practices, including discussing and commenting on model properties, exploring the relationship between model structure and its output, and describing, explaining, and critiquing a model
4. Quality of model evaluation (PRAC-evaluation)	Assess model evaluation practices, including evaluating a model or comparing multiple models based on one or more criteria

and knowledge, and have a sound, efficient strategic procedure. Wu (2010), for instance, reported the design and implementation of the Air Pollution Modeling Environment (APoME), a system modeling tool designed to help students to model air quality. In order to design the system, in her first study, she interviewed 18 participants with different levels of expertise in atmospheric science and asked them to design an investigation about the air quality change in Taipei during a daily cycle and a seasonal cycle. They were then asked to define air quality, describe their research plan, identify major variables for the investigation and relationships between those variables, and define how they would manipulate those variables. In terms of assessing their research plans, two criteria were identified: (a) providing a research plan that is justified by scientific theories, and (b) indicating possible errors and limitations in their plans. The interview results indicated that the majority of the participants did not propose a plan that met the criteria. Though most of them brought up experimental and human errors, none of them indicated limitations. In her second study, APoME was implemented with 29 tenth-grade students to investigate their modeling practices and conceptual understanding on air pollution dispersion. It was found that the students could not provide a plan that met the two criteria specified above during the activity. However, follow-up interviews with 14 focus students showed improved understanding of model planning.

The *PRAC-generation* assessment theme focuses on the actual model generation process, which includes the sub-processes of construction, testing, and revision of models. Although we coded these sub-processes individually, it is often hard to distinguish them in actuality.

In terms of model construction, the practices assessed may include adding and deleting model components and manipulating their properties to observe model behavior during construction (Ergazaki et al., 2007; Komis et al., 2007), selecting, identifying, and connecting relevant variables (Liu, 2006; Sun & Looi, 2013), and defining model components and quantifying them (Hogan & Thomas, 2001; Sins, Savelsbergh, van Joolingen, & van Hout-Wolters, 2009). Most assessments of model construction focused on students' cognitive processes. For instance, in their study with 16 ninth graders, Stratford et al. (1998) investigated students' modeling practices with Model-It, a dynamic-modeling tool that allows students to create qualitative models. The researchers identified five cognitive strategies for students' modeling practices (analyzing, relational reasoning, synthesizing, testing and debugging, and explaining) and listed specific behaviors as evidence for each of these strategies. We conceived several of the listed behaviors as particularly pertinent in model construction such as identifying factors, creating objects, and selecting relationships. To compare the modeling performances of small groups, they identified low, moderate, and high-quality cognitive strategies for modeling.

Model testing is conceptualized here as a mechanism of model evaluation during the model generation process, whereas the PRAC-evaluation theme (see below) is conceptualized as a process independent of model generation. During testing, students run their models according to some established rules in a modeling environment (Mulder et al., 2010). The assessment of student model testing practices included checking if the finished models can run properly in modeling programs (Komis et al., 2007) and looking at the quality and quantity of conducted experiments when running the model (Mulder et al., 2010). Ergazaki et al. (2007), for instance, reported a case study of high school students collaborating on a plant growth modeling task in the computer-based modeling environment ModelsCreator. Based on the framework proposed by Stratford et al. (1998), Ergazaki et al. (2007) identified four cognitive strategies in this study (analysis, synthesis, testing and interpretation, and cognitive and technical support). In terms of assessing the model testing practice specifically, this study focused on the quantity of testing when the students ran their completed models in the software. Researchers counted the instances when students actually tested their models and also provided verbal explanations.

Model revision can be interpreted at different levels. At a micro-level, model revision co-occurs with model construction when students add and delete model components and change their properties (Hogan & Thomas, 2001). These micro-level revisions are often achieved through trial-and-error. At a macro-level, model revision occurs after an explicit testing and evaluation phase (Stratford et al., 1998), or after comparing multiple models (Papaevripidou et al., 2007). It is often carried out deliberately (i.e. knowing what to change and why).

The PRAC-interpretation theme focuses on model interpretation, a sense-making process based on models provided or models generated by students themselves. It may include assessment practices that focus on interpretation of different sorts, such as,

- interpreting model components, for example, in terms of accuracy of content, thoroughness of content, and coherence between the model and interpretation (Mulder et al., 2010; Papaevripidou et al., 2007; Sins et al., 2009),
- interpreting the relationships between model components (Ergazaki et al., 2007; Komis et al., 2007; Liu, 2006),
- interpreting model representation, for example, discussing and commenting on model representation (Chang et al., 2010; Stratford et al., 1998),
- interpreting model output, for example, using model output to explore the relationship between model output and model structure (Hogan & Thomas, 2001), and
- interpreting data, for example, collecting data to support or reject hypotheses (Wu, 2010).

Chang et al. (2010) investigated 271 seventh graders' use of Chemation, a modeling tool for building 2D molecular models and flipbook style animations on Palm devices. Students were randomly assigned to three conditions: (a) design-interpret-evaluate animations ($n = 101$), (b) design-interpret animations ($n = 96$), or (c) view-interpret teacher-made animations ($n = 74$). In the interpreting activity, students wrote down their interpretation of the animation and its relevance to the macroscopic chemical phenomena. Three criteria were used to assess students' interpretation: (a) accuracy of content, which refers to whether or not the interpretation included accurate science content, (b) thoroughness, which refers to the detail of the discussion of atom rearrangement, and (c) coherence, which refers to the coherent explanation between student interpretation and the animation. All three criteria were scored at three levels (0-unsatisfactory, 1-satisfactory, and 2-proficient). The study found that interpretation of models was significantly better for students in the design-interpret-evaluate condition than those in the design-interpret condition. However, there was no significant difference between interpretation of models in the design-interpret condition and view-interpret condition.

The *PRAC-evaluation* theme assessed in the literature includes critiquing a model or comparing multiple models based on a certain set of criteria (Stratford et al., 1998), challenging, rejecting, or accepting the justifications for model creation decisions (Komis et al., 2007), and using scientific knowledge to evaluate the model (Sins et al., 2009). For instance, when examining how students evaluate their models, Papaevripidou et al. (2007) developed an assessment task where students were asked to choose and explain the best model out of four givens ones on ant colony. Students' model evaluation proficiency was evaluated based on the number of good criteria they used, including the following: the model represents parts of the phenomenon, the model represents the way the phenomenon evolves over time, and the model includes an underlying structural mechanism.

Overall, of the 14 articles assessing modeling practices, PRAC-interpretation is the most popular theme (13 out of 14). This is understandable as model interpretation occurs in all phases of modeling. On the contrary, model planning, model evaluation, and model revision were much less assessed ($n \leq 5$). These processes require additional time in a modeling curriculum, and therefore, posing challenges in implementation.

Assess meta-modeling knowledge. A total of 11 (37%) articles included some forms of the assessment of students' meta-modeling knowledge (Table 4). This assessment dimension refers to the procedures that explicitly examine students' understanding of the nature of models and modeling in science. All the assessed meta-modeling knowledge types fell under the categories (and subcategories) proposed by Schwarz and White (2005). Therefore, we adopted the assessment themes as nature of models (META-model), nature of modeling (META-modeling), evaluation of models (META-evaluation), and purpose or utility of models (META-purpose).

The *META-model* theme includes assessment items or tasks focusing on students' meta-modeling knowledge on the following aspects: (a) *types of models and model attributes* such as what a model is and how students would describe it (Gobert & Pallant, 2004; Saari, 2003) or how a model predicts events that happen in reality (Sins et al., 2009); (b) *model content* such as what models and their components represent (Schwarz & White, 2005); (c) *multiple models* such as whether there could be different models of the same phenomena (Saari, 2003; Schwarz et al., 2009; Treagust, Chittleborough, & Mamiala, 2004); and (d) *constructed nature of models* such as whether a model is an accurate duplicates of reality (Treagust et al., 2004), whether a model is a mental or representative structure (Barnea & Dori, 1999), whether a model represents the reality (Schwarz & White, 2005), and the proximity of a model to the real things (Gobert & Pallant, 2004).

Table 4. Types of assessment of meta-modeling knowledge (META) from the literature

Theme	Explanation
1. Nature of models (META-model)	Assess students' epistemologies about the following: types of models and model attributes, model content, multiplicity of models, constructed nature of models
2. Nature of process of modeling (META-modeling)	Assess students' epistemologies about the following: modeling processes, designing and creating models, changing nature of models
3. Evaluation of models (META-evaluation)	Assess students' epistemologies about the following: whether there is a way to decide one model is better than the other, the criteria used to evaluate models, and the steps needed to improve models
4. Purpose and utility of models (META-purpose)	Assess students' epistemologies about the following: purposes of models, utility of models, and utility of multiple models

The *META-modeling* theme includes assessment items or tasks focusing on students' meta-modeling knowledge on the following aspects: (a) *modeling processes*, including the specific individual modeling processes (Schwarz et al., 2009), the definition of modeling (Sins et al., 2009), and the importance of modeling in understanding science (Levy & Wilensky, 2009b); (b) *designing and creating models*, including constructing models based on experiments and/or observations (Barnea & Dori, 1999), choosing the components of models (Gobert & Pallant, 2004), planning about models (Saari, 2003), and improving models through including new variables and relationships (Sins et al., 2009); and (c) *changing models*, including whether scientists change a model (Treagust et al., 2004), existence of instances (Levy & Wilensky, 2009b) and reasons (Saari, 2003) for changing models, conditions for scientists' acceptance of a new model (Treagust et al., 2004), and evaluating models and making revision to understand models are changeable entities (Schwarz et al., 2009).

The *META-evaluation* theme includes assessment items or tasks focusing on students' meta-modeling knowledge on the following aspects: (a) determining the existence of a way to evaluate and compare models (Sins et al., 2009; Schwarz et al., 2009); and (b) specific *criteria* for model evaluation (Schwarz & White, 2005).

The *META-purpose* theme includes assessment items or tasks focusing on students' meta-modeling knowledge on the following aspects: (a) *purposes of* (*multiple*) *models* such as helping students to picture the scientific phenomenon and understand science (Levy & Wilensky, 2009b), representing something (Liu, 2006; Treagust et al., 2004), determining relations and drawing conclusion based on a model (Sins et al., 2009), and representing the same phenomenon from multiple perspectives (Treagust et al., 2004); and (b) *utility of models in science and science classes* such as how models can be helpful for scientists and students (Schwarz & White, 2005), how models are used to describe and explain phenomena (Barnea & Dori, 1999), how models are used to help make and test predictions about scientific events (Levy & Wilensky, 2009b; Sins et al., 2009), how models help understand and present the phenomenon better (Papaevripidou et al., 2007).

Whereas all of the 11 articles reported META-purpose theme and most of them also reported either META-model ($n = 10$) or META-modeling ($n = 9$), only 5 of them reported META-evaluation. In terms of the format of the meta-modeling knowledge assessment, some studies used questionnaire (e.g. Barnea & Dori, 1999; Levy & Wilensky, 2009b; Liu, 2006; Sins et al., 2009), some used surveys (e.g. Gobert & Pallant, 2004; Treagust et al., 2004), some used interviews (e.g. Saari, 2003; Schwarz & White, 2005; Schwarz et al., 2009), and only one article used observations of student activities (e.g. Papaevripidou et al., 2007).

Patterns and Gaps in the MOA Literature

This section reports relevant findings to address our second research question 'What are the patterns and gaps regarding MOA in the literature?' Our results indicate that MOA has not been adequately utilized in MBI studies (i.e. only 30 out of 104 empirical MBI articles reported MOA). The majority of the MBI studies focused on other learning outcomes such as students' conceptual learning and affective domains. The MOA literature does show an increasing trend: 5 articles were published in 1997–2002, 10 in 2003–2008, and 15 in 2009–2013. We are optimistic that more MOA studies will be carried out in the coming years.

Although most articles reported multiple MOA themes, the majority of articles ($n = 21$) only reported MOA within a single dimension. Within the nine articles that reported multiple dimensions, only two articles made connections between them. For instance, Sins et al. (2009) found a positive relationship between students' meta-modeling knowledge (i.e. epistemological understanding) and students' modeling practices. Similarly, within the 15 articles that also reported other assessment types, only 5 reported the connection between MOA and those assessments. For instance, Schwarz and White (2005) found that as a result of participating in the modeling curriculum, a positive correlation existed between students' meta-modeling knowledge and inquiry skills, as well as between meta-modeling knowledge and physics understanding.

Due to its nature, the approaches used for the assessment of modeling practices were much more diverse compared to the other two dimensions. As a result, the coding of this dimension took much more time than the other two. In terms of sequence, some researchers assessed students' modeling practices performed during the modeling activities (e.g. Ergazaki et al., 2007; Komis et al., 2007; Liu, 2006; Stratford et al., 1998), whereas others designed specific tasks to assess students' modeling practices after the modeling activities (e.g. Papaevripidou et al., 2007). In terms of format, some researchers assessed students' modeling practices by looking into students' actions (e.g. Hogan & Thomas, 2001), whereas others assessed the contents of students' written interpretations (e.g. Chang et al., 2010) or verbal discourses (e.g. Louca et al., 2011). Considering the intertwined nature of modeling practices, our collection of articles also suggest some promising strategies. One good way is to provide an overview of student modeling practices by identifying the patterns (e.g. sequencing, Ergazaki et al. 2007; or 'frames', Louca et al., 2011) of the practices in students' modeling activities. Another good way is to triangulate multiple sources of data such as cross-checking students' modeling practices observed through video recording and their verbal reflections through interviews (e.g. Wu, 2010).

Most of the MOA used individuals as their unit of analysis. In fact, all the assessments of meta-modeling knowledge were individual based. In terms of modeling practices, seven articles reported students' collaborative actions during modeling but only three of them reported student performance at a collective level (e.g. Hogan & Thomas, 2001; Prins et al., 2009; Stratford et al., 1998). When assessing student-generated models, although students' collaborative modeling actions were described in most articles, only two of them assessed models collectively produced (Louca et al., 2011; Sun & Looi, 2013).

Our analysis showed that the majority of the articles used computer-based modeling tools in their MBI interventions ($n = 25$). These tools included computer animation, virtual experiments, agent-based simulation, system dynamics modeling tools, and technology-enhanced curriculum platform. More than half of these articles reported the use of technology in their MOA (n=15), assessing either modeling practices or modeling products. Of these 15 articles, the majority of them used MOA in the formative sense.

Discussion

Our review of the 30 MOA articles results in 11 themes across 3 dimensions and reveals some major gaps in the literature. To address our third research question, we weave our practical suggestions into the following three guiding principles for developing high-quality MOA: (a) framing MOA in an ecology of assessment, (b) providing authentic modeling contexts for assessment, and (c) spelling out the connections between MOA items and the essential aspects of modeling to be assessed. The latter two principles can be conceptualized as further elaboration on two important aspects of the first one.

First, we encourage researchers to consider framing distinctive MOA approaches within an ecology of MOA. As a starting point to consider, the core of the ecology comprised the multiple dimensions within MOA as suggested in this study (Figure 3). Our review revealed that little research has been devoted to understand the connections among the specific measures within and across the MOA dimensions. Considering the theme of model evaluation, for instance, students' meta-level understanding of model evaluation may correspond to, or predict to a certain degree, their model evaluation practices, and the evaluative criteria they can generate for the produced model.

The ecology also includes the multi-layered links between MOA and other assessments foci such as other scientific practices (e.g. constructing explanations), conceptual knowledge, and affective domains. Consider conceptual understanding here. Intuitively, the quality of students' modeling practices on a particular topic depends on their conceptual understanding of the knowledge domain. But the question is to what extent and under what conditions. This is particularly important when we consider the recent science education reform initiatives. For example, the U.S. Next Generation of Science Standards (NGSS) is written as a core set of performance expectations that couple practices and content, or more precisely, integrate the three dimensions of science learning: practices, crosscutting concepts, and disciplinary core ideas (National Research Council, 2014). Developing valid assessments that target these NGSS performance expectations entails a clear understanding of the connections among practices and content assessments, which is not readily available in the field of MOA.

To paint a more complete picture, appropriate MOA should be developed and evaluated at the elementary level, especially at the lower grades. This is woefully inadequate in the field at the moment (only four articles in our collection focused on the elementary level and all these four focused on either fifth or sixth graders). Developing MBI curriculum and corresponding MOA approaches for younger children will help illuminate the starting point or the lower limit of the modeling progression in K-12 (Louca et al., 2011; Schwarz et al., 2009). It is also emphasized in the NGSS that modeling should be dealt with throughout the entire K-12 experience (NGSS Leads State, 2013). The fact that very few MOA have been developed at the

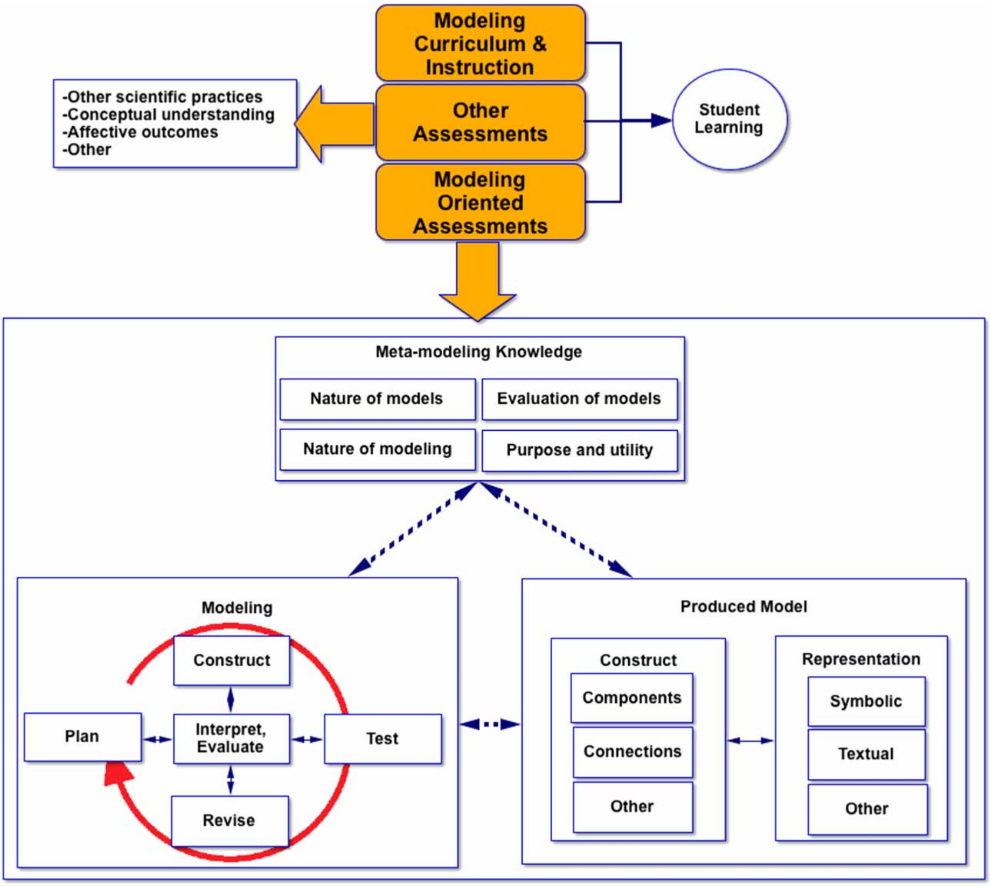

Figure 3. Interlinked MOA dimensions as part of the MOA ecology.

elementary level is simply because very few MBI curricula have been developed at that level. This brings up another important aspect of the ecology—understanding that MOA is conditioned upon MBI, which will be explained further when we describe the next two principles.

A better understanding of an MOA ecology can inspire researchers to develop new MOA (or other) instruments that are based on a balanced mixture of specific measures. A systematic way of framing may help researchers to compare and contrast their own approach with others and reevaluate the particular MOA approach they are adopting. This is why we coined the term MOA instead of treating the individual MOA dimensions separately. However, taking a systematic view does not mean that an individual MOA study should cover more breadth over depth. Neither do we envision that the MOA ecology can be established in any single study or a few attempts. Rather, we see the establishment of a viable and evolving MOA ecology as a long-term commitment from a community of devoted researchers and developers.

Second, we urge that future work on MOA should seriously consider authenticity. By authenticity, we mean an MOA approach that is consistent with the corresponding

MBI curriculum and learning theory (National Research Council, 2001). Here we elaborate on three issues that emerged from the literature review that speak to the authenticity of MOA: unit of analysis, assessment medium, and complexity.

In the review, we found a discrepancy between the unit of analysis of MOA and how students conducted their modeling practices. More specifically, although collaboration is an inherent element of modern science practice and often promoted in MBI (Gilbert & Boutler, 2000; Gobert & Pallant, 2004; Ioannidou et al., 2010; Lehrer & Schauble, 2006), the unit of analysis of MOA in the list of papers we reviewed was predominately individual based. Therefore, we suggest future research should explore the assessment of collaborative modeling and investigate questions such as 'Is it practical to assess collaborative modeling? If so, what challenges does it have?', 'What indicators can be used to assess modeling at the collective level?', and 'Do these indicators apply in all of the three MOA dimensions?'

Another discrepancy we found was the mismatch between the media students use to carry out their modeling activities and the media through which they were assessed afterwards. Paper-and-pencil is still the primary delivering medium for MOA; but most modeling activities rely heavily on the use of physical or computer-based materials (the majority of the articles used some form of technology in our literature collection). The performance expectations laid out in NGSS emphasize that students need to *demonstrate* what they understand. In modeling, students can better express their understanding and competency if equipped with an appropriate medium. The use of physical or digital modeling materials in assessment may pose major logistic and economic challenges, especially for large-scale assessment. Some innovative ways of handling modeling materials in MOA have been brought up. In terms of physical materials, Rivet and Kastens (2012) provided a middle-ground solution for MOA at the whole class level. Instead of using paper-and-pencil or a student-manipulated model, they managed to demonstrate a model dynamically in front of the whole class. In terms of computer-based MOA, there have been emerging studies that incorporate computer simulations in MOA with a particular emphasis on large-scale assessments (Buckley & Quellmalz, 2013; Liu, Waight, Gregorius, Smith, & Park, 2012). Using SimScientists, Quellmalz, Timms, Silberglitt, and Buckley (2012), for instance, incorporated simulation-based assessments in Force and Motion and Ecosystem units. In these units, teachers insert two to three embedded formative and one benchmark assessment in the unit. While manipulating the simulations students make predictions, interpret data, evaluate their predictions, and explain results. One of the unique aspects of this approach is that the system can give automated feedback to students' responses, for instance, to the correctness of student-generated diagrams and experimental designs.

The issue of complexity naturally arises when MOA is framed in an ecology that contains all the layers previously discussed. Another big source of complexity comes from assessing a 'messy' learning process. Although in this paper, the themes of modeling practices are reported separately and sequentially, good assessments of modeling practices need to recognize that these practices are very dynamic and intertwined in nature. For instance, model revision requires making some alterations to the models (constructed or given). To do that, students need to test their model, interpret how the model behaves, and evaluate the current state of the model based on some criteria. Then, students need to plan what kind of changes they need to make and then make revisions accordingly, which may lead to the construction of a new model. The messy processes pose challenges in both data collection (faithfully capture the complexity) and data analysis (make sense of the complex data). The former has been remediated by data capturing technologies such as video recording, screen recording, and computer logging (e.g. Buckley et al., 2004). Unfortunately, these data collection methods produce an even larger amount of raw data to be evaluated. The latter has been typically approached from data reduction—researchers come up with labels, codes, or categories that reduce the complexity of the modeling process data. Future MOA should consider incorporating emerging technologies such as educational data mining and learning analytics (e.g. Gobert, Sao Pedro, Raziuddin, & Baker, 2013) to better capture and analyze the fine-grained, dynamic-modeling processes, and be able to provide automated feedback (Linn, 2013). Since the assessment of modeling practices is often contextualized and conditioned upon many factors, we believe that assessors should choose the most suitable approach to balance the complexity and feasibility of MOA.

The complexity of MOA points directly to the importance of the third guiding principle—spelling out the connections between MOA items and the essential aspects of modeling to be assessed. This principle is a validity question and may sound trivial, but requires much more work and caution. It is a reality, and perfectly acceptable, that scholars hold different views on what essence of modeling is to be assessed. But the essence of modeling to be assessed has to be made crystal clear: Is it the meta-level knowledge on specific aspects of modeling, the cognitive strategies relevant to modeling, or the scientific practices in building models? What are the possible levels of the assessed aspect? Some articles in our collection did not clearly explain this. Furthermore, the connections between the essential aspects of modeling to be assessed and the corresponding MOA items have to be spelled out clearly: How are the proposed MOA items connected with the essential aspects of modeling to be assessed? Does the empirical evidence support the connections? To what degrees can we interpret the results?

The construct-centered approach (Shin, Stevens, & Krajcik, 2010; Wilson, 2005) is a good assessment model that can address many of these questions. In this approach, the assessment items and interpretations of results are built upon clearly specified, hierarchical constructs to be assessed. A good example using this approach in MOA is the work done by Rivet and Kastens (2012). They developed and tested a construct-based assessment to examine eighth and ninth graders' analogical reasoning around physical models on Earth science topics. They spelled out the modeling construct based on analogical reasoning and considered two important aspects of it: correspondences and non-correspondences (identifying matches and mismatches when comparing the behavior of the expressed model against that of the referent system) and direction of mapping (reasoning from model to Earth or from Earth to model). Their construct included the following four levels from the least to the most sophisticated analogical reasoning: no mapping, entries/attributes (i.e. students' ability to do object-to-object mapping between the model and the Earth system), configuration/motion (i.e.

students' reasoning about the analogical correspondences between the relative positioning or distribution in space of entities in the model and those in the Earth System), and mechanism/causation (i.e. students' reasoning about what kind of mechanisms or causes included in the model maps onto the same cause or mechanism that crates the phenomena in the real earth system). The item design targeted to elicit students' responses that support the knowledge included in the construct map. Design of the items included initial interview of the students to gather information, iterative design of the items and the assessment materials, and implementation of the items (for detailed explanation of the item design, see Rivet & Kastens, 2012, pp. 724–726).

Conclusion and Future Research Directions

In this study, we reviewed 30 MOA empirical research articles from 1980 to 2013 in the K-12 context. We summarized the findings in an array of themes in three dimensions (modeling products, modeling practices, and meta-modeling knowledge) and identified the major gaps in the field: (1) In general, MOA has not been widely used in the MBI literature, even though there is an increasing interest and acknowledgement about its significance. The review also reveals that within the limited body of the MOA literature, very little work has been conducted at the elementary level. (2) There is no comprehensive framework depicting the three MOA dimensions, their internal connections, and their connections to other assessments. Such a framework is critical for establishing a solid foundation for MOA. (3) Even though collaboration is typically conceived and implemented as an important element in modeling, very few studies take this into consideration during assessment. (4) Even though many MBI approaches have used computer technologies, these technologies have not been widely utilized in large-scale MOA.

Based on the review, we propose three guiding principles for future work in MOA. Table 5 summarizes the practical recommendations tied to these guiding principles. We invite further discussion and debate on these principles, and expect more empirical studies from the research community to validate them. On top of these guiding principles, we urge that all future quantitative studies on MOA (and other topics) should include adequate information (e.g. effect sizes) in their statistical reports so that the reader can understand better the impact of these findings and future researchers can conduct meta-analysis (Lipsey & Wilson, 2001). We hope this synthesis work marks a first step towards a more fruitful pursuit of MOA research in the future.

Limitations

There are several limitations concerning our review study. First of all, despite our effort to make the review comprehensive, the resulting sample is relatively small. This is partially due to the fact that MOA has not been extensively pursued in K-12 settings, and partially due to the criteria we used to include or exclude the articles. For instance, in this study, we included papers only describing an explicit MBI intervention and studies assessing students' externalized models. This, however, does not necessarily entail that the authors

Table 5. MOA guiding principles and future research directions.

MOA guiding principles	Future MOA research and development directions
Framing MOA in an ecology of assessment	• Build the connections within MOA themes and dimensions • Build the connections between MOA and other assessments • Establish a full MOA spectrum, including early grades at the elementary level
Providing authentic modeling contexts for assessment	• Explore how to assess collaborative modeling • Develop creative ways to incorporate physical materials or computer technologies in (large-scale) MOA • Applying advanced technology such as learning analytics to analyze complex modeling practices
Spelling out the connections between MOA items and the essential aspects of modeling	•Further explore the essential nature of modeling to be assessed • Employ approaches that make explicit the connections between the modeling construct and the MOA items

excluded their theoretical emphasis on mental models (e.g. Gobert, 2000). In our review, this criterion was a practical decision as the papers on mental modeling literature intersect greatly with conceptual change literature. For instance, in many conceptual change studies, the authors may focus on students' development of mental models as a way to define conceptual understanding or mental structure. They could use modeling-based assessment instruments such as FCI (Hestenes et al., 1992) to assess students' conceptual learning, which was not the focus of this paper. There are also articles that only used terms such as 'computer simulation' without referring to modeling. These articles will not be found through our search methods. Also excluded were articles that did not include an extensive report of the MBI component when they conducted their MOA study (e.g. Rivet & Kastens, 2012). To compensate for this, we made connections to these references in our discussion section. Another major limitation is that this study applied one way to code the modeling themes, especially for the modeling practices, and the coding scheme was based on our own theoretical perspectives on modeling. Obviously, there are many other alternatives. Using a different scheme, the coding will most likely result in new findings and different insights.

Acknowledgements

This material is based upon work supported by the US National Science Foundation (NSF) under grant number DRL1019866. Any opinions, findings, and conclusions expressed in this poster are those of the authors and do not necessarily reflect the views of the NSF. We thank our collaborators who contributed significantly in the larger project from which a part of the data was drawn: Jing Lei, Heng Luo, Ye Chen, and Sung Hye Lee. We also thank the anonymous reviewers who provided valuable feedback to an earlier version of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Note

1. The original project reviewed literature from 1980 to 2010. We expanded the search to 2013 in this paper.

References

Angell, C., Kind, P. M., Henriksen, E. K., & Guttersrud, Ø. (2008). An empirical-mathematical modelling approach to upper secondary physics. *Physics Education*, *43*(3), 256–264. doi:10.1088/0031-9120/43/3/001

Barak, M., & Hussein-Farraj, R. (2013). Integrating model-based learning and animations for enhancing students' understanding of proteins structure and function. *Research in Science Education*, *43*(2), 619–636. doi:10.1007/s11165-012-9280-7

Barnea, N., & Dori, Y. J. (1999). High-school chemistry students' performance and gender differences in a computerized molecular modeling learning environment. *Journal of Science Education and Technology*, *8*(4), 257–271. doi:10.1023/A:1009436509753

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rubmle, M. (2012). Defining twenty-first century skills. In P.Griffin, B.McGaw, & E.Care (Eds.), *Assessment and teaching of twenty-first century skills* (pp. 17–66). London: Springer.

Black, P., & William, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90.

Branch, R. M., & Kopcha, T. J. (2013). Instructional design models. In J. M. Spector, M. D. Merill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed.). (pp. 77–87). New York: Springer.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. doi:10.1191/1478088706qp063oa

Bravo, C., van Joolingen, W. R., & de Jong, T. (2009). Using Co-Lab to build system dynamics models: Students' actions and on-line tutorial advice. *Computers & Education*, *53*(2), 243–251. doi:10.1016/j.compedu.2009.02.005

Buckley, B. C. (2000). Interactive multimedia and model-based learning in biology. *International Journal of Science Education*, *22*(9), 895–935. doi:10.1080/095006900416848

Buckley, B. C., Gobert, J. D., Kindfield, A. C. H., Horwitz, P., Tinker, R. F., Gerlits, B., … Willett, J. (2004). Model-based teaching and learning with BioLogica™: What do they learn? How do they Learn? How do we know? *Journal of Science Education and Technology*, *13*(1), 23–41. doi:10.1023/B:JOST.0000019636.06814.e3

Buckley, B. C., & Quellmalz, E. S. (2013). Supporting and assessing complex biology learning with computer-based simulations and representations. In D. F. Treagust & C.-Y. Tsui (Eds.), *Multiple representations in biological education* (pp. 247–267). Dordrecht: Springer.

Chang, H.-Y., Quintana, C., & Krajcik, J. S. (2010). The impact of designing and evaluating molecular animations on how well middle school students understand the particulate nature of matter. *Science Education*, *94*(1), 73–94. doi:10.1002/sce.20352

Cheng, M. F., & Brown, D. E. (2010). Conceptual resources in self-developed explanatory models: The importance of integrating conscious and intuitive knowledge. *International Journal of Science Education*, *32*(17), 2367–2392. doi:10.1080/09500690903575755

Crawford, B. A., & Cullin, M. J. (2004). Supporting prospective teachers' conceptions of modeling in science. *International Journal of Science Education*, *26*(11), 1379–1401. doi:10.1080/09500690410001673775

Dori, Y. J., & Kaberman, Z. (2012). Assessing high school chemistry students' modeling sub-skills in a computerized molecular modeling learning environment. *Instructional Science*, *40*(1), 69–91. doi:10.1007/s11251-011-9172-7

Ergazaki, M., Zogza, V., & Komis, V. (2007). Analysing students' shared activity while modeling a biological process in a computer-supported educational environment. *Journal of Computer Assisted Learning*, *23*(2), 158–168. doi:10.1111/j.1365-2729.2006.00214.x

Fazio, C., Guastella, I., Sperandeo-Mineo, R. M., & Tarantino, G. (2008). Modelling mechanical wave propagation: Guidelines and experimentation of a teaching–learning sequence. *International Journal of Science Education*, *30*(11), 1491–1530. doi:10.1080/09500690802234017

Fretz, E. B., Wu, H., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Research in Science Education*, *32*, 567–589. doi:10.1023/A:1022400817926

Gilbert, J. K., & Boutler, C. J. (2000). *Developing models in science education*. Dordrecht: Kluwer Academic.

Gilbert, J. K., Boutler, C. J., & Elmer, R. (2000). Positioning models in science education and in design and technology education. In J. K. Gilbert & C. J. Boutler (Eds.), *Developing models in science education* (pp. 3–17). Dordrecht: Kluwer Academic.

Gobert, J. D. (2000). A typology of causal models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education*, *22*(9), 937–977. doi:10.1080/095006900416857

Gobert, J. D., & Buckley, B. C. (2000). Introduction to model-based teaching and learning in science education. *International Journal of Science Education*, *22*(9), 891–894. doi:10.1080/095006900416839

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:10.1023/B:JOST.0000019635.70068.6f

Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences*, *22*(4), 521–563.

Grosslight, L., Unger, C., Jay, E., &Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, *28*(9), 799–822. doi:10.1002/tea.3660280907

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, *22*(9), 1011–1026. doi:10.1080/095006900416884

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, *55*(5), 440–454. doi:10.1119/1.15129

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158. doi:10.1119/1.2343497

Hogan, K., & Thomas, D. (2001). Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology*, *10*(4), 319–345. doi:10.1023/A:1012243102249

Hsu, Y.-S., Lin, L.-F., Wu, H.-K., Lee, D.-Y., & Hwang, F.-K. (2012). A novice-expert study of modeling skills and knowledge structures about air quality. *Journal of Science Education and Technology*, *21*(5), 588–606. doi:10.1007/s10956-011-9349-5

Ioannidou, A., Repenning, A., Webb, D., Keyser, D., Luhn, L., & Daetwyler, C. (2010). Mr. Vetro: A collective simulation for teaching health science. *International Journal of Computer-Supported Collaborative Learning*, *5*(2), 141–166. doi:10.1007/s11412-010-9082-8

Johnstone, A. H., & Mahmoud, N. A. (1981). Pupils' response to a model for water transport. *Journal of Biological Education*, *15*(3), 203–208. doi:10.1080/00219266.1981.9654379

Kaberman, Z., & Dori, Y. J. (2009). Question posing, inquiry, and modeling skills of chemistry students in the case-based computerized laboratory environment. *International Journal of Science and Mathematics Education*, *7*(3), 597–625. doi:10.1007/s10763-007-9118-3

Klopfer, E., & Um, T. (2005). Teaching complex dynamic systems to young students with StarLogo. *Journal of Computer in Mathematics and Science Teaching*, *24*(2), 157–178.

Komis, V., Ergazaki, M., & Zogza, V. (2007). Comparing computer-supported dynamic modeling and "paper & pencil" concept mapping technique in students' collaborative activity. *Computers & Education*, *49*(4), 991–1017. doi:10.1016/j.compedu.2005.12.007

Kurtz dos Santos, A. C., Thielo, M. R., & Kleer, A. A. (1997). Students modelling environmental issues. *Journal of Computer Assisted Learning*, *13*(1), 35–47. doi:10.1046/j.1365-2729.1997.00005.x

Lehrer, R., & Schauble, L. (2000). Developing model-based reasoning in mathematics and science. *Journal of Applied Developmental Psychology*, *21*(1), 39–48. doi:10.1016/S0193-3973(99)00049-0

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning. In K. R.Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371–387). New York: Cambridge University Press.

Levy, S. T., & Wilensky, U. (2009a). Crossing levels and representations: The connected chemistry (CC1) curriculum. *Journal of Science Education and Technology*, *18*(3), 224–242. doi:10.1007/s10956-009-9152-8

Levy, S. T., & Wilensky, U. (2009b). Students' learning with the connected chemistry (CC1) curriculum: Navigating the complexities of the particulate world. *Journal of Science Education and Technology*, *18*(3), 243–254. doi:10.1007/s10956-009-9145-7

Linn, M. C. (2013). *Automated scoring and adaptive guidance*. Poster session presented at the annual conference of the American educational research association. San Francisco, CA.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

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:10.1007/s10956-006-0359-7

Liu, X., Waight, N., Gregorius, R., Smith, E., & Park, M. (2012). Developing computer model-based assessment of chemical reasoning: A feasibility study. *Journal of Computers in Mathematics and Science Teaching*, *31*(3), 259–281.

Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, *64*(4), 471–492. doi:10.1080/00131911.2011.628748

Louca, L. T., Zacharia, Z. C., & Constantinou, C. P. (2011). In quest of productive modeling-based learning discourse in elementary school science. *Journal of Research in Science Teaching*, *48*(8), 919–951. doi:10.1002/tea.20435

Louca, L. T., Zacharia, Z. C., Michael, M., & Constantinou, C. P. (2011). Objects, entities, behaviors, and interactions: A typology of student-constructed computer-based models of physical phenomena. *Journal of Educational Computing Research*, *44*(2), 173–201. doi:10.2190/EC.44.2.c

Manlove, S., Lazonder, A. W., & de Jong, T. (2009). Collaborative versus individual use of regulative software scaffolds during scientific inquiry learning. *Interactive Learning Environments*, *17*(2), 105–117. doi:10.1080/10494820701706437

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, *47*, 474–496.

Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2010). Finding out how they find it out: An empirical analysis of inquiry learners' need for support. *International Journal of Science Education*, *32*(15), 2033–2053. doi:10.1080/09500690903289993

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on conceptual framework for the New K-12 science education standards. Committee on a conceptual framework for New K-12 science Ed. Washington, DC: The National Academies Press.

National Research Council. (2014). *Developing assessments for the next generation science standards*. Committee on a conceptual framework for new K-12 science education standards. Board on science education, Division of behavioral and social sciences and education. (J. W. Pellegrino, M. R. Wilson, J. A. Koenig, & A. S. Beatty, Eds.). Washington, DC: The National Academies Press.

Nersessian, N. J. (2008). *Creating scientific concepts*. Cambridge: MIT Press.

Nersessian, N. J. (2009). How do engineering scientists think? Model-based simulation in biomedical engineering research laboratories. *Topics in Cognitive Science*, *1*, 730–757.

NGSS Leads States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Norman, D. A. (1983). Some onservations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nuhoglu, H. (2008). Modeling spring mass system with system dynamics approach in middle school education. *The Turkish Journal of Educational Technology*, *7*(3), 22–34.

Pallant, A., & Tinker, R. F. (2004). Reasoning with atomic-scale molecular dynamic models. *Journal of Science Education and Technology*, *13*(1), 51–66. doi:10.1023/B:JOST.0000019638.01800.d0

Papaevripidou, M., Constantinou, C. P., & Zacharia, Z. C. (2007). Modeling complex marine ecosystems: An investigation of two teaching approaches with fifth graders. *Journal of Computer Assisted Learning*, *23*(2), 145–157. doi:10.1111/j.1365-2729.2006.00217.x

Penner, D. E., Lehrer, R., & Schauble, L. (1998). From physical models to biomechanics: A design-based modeling approach. *The Journal of the Learning Sciences*, *7*(3&4), 429–449. doi:10.2307/1466793

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:10.1002/tea.20415

Popham, W. J. (2010). *Classroom assessment: What teachers need to know*. Boston, MA: Pearson/Allyn and Bacon.

Prins, G. T., Bulte, A. M. W., Driel, J. H., & Pilot, A. (2009). Students' involvement in authentic modelling practices as contexts in chemistry education. *Research in Science Education*, *39*(5), 681–700. doi:10.1007/s11165-008-9099-4

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, *49*(3), 363–393. doi:10.1002/tea.21005

Rivet, A. E., & Kastens, K. A. (2012). Developing a construct-based assessment to examine students' analogical reasoning around physical models in earth science. *Journal of Research in Science Teaching*, *49*(6), 713–743. doi:10.1002/tea.21029

Roth, W., Woszczyna, C., Smith, G., Universiq, S. F., & Va, B. C. (1996). Affordances and constraints of computers in science education. *Journal of Research in Science Teaching*, *33*(9), 995–1017. doi:10.1002/(SICI)1098-2736(199611)33:9<995::AID-TEA3>3.0.CO;2-Q

Saari, H. (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. doi:10.1080/0950069032000052081

Scalise, K., Timms, M., Moorjani, A., Clark, L., & Holtermann, K. (2011). Student learning in science simulations. Design futures that promote learning gains. *Journal of Research in Science Teaching*, *48*(9), 1050–1078. doi:10.1002/tea.20437

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., … Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. doi:10.1002/tea.20311

Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, *23*(2), 165–205. doi:10.1207/s1532690xci2302_1

Shen, J. (2006). *Teaching strategies and conceptual change in a professional development program for science teachers of K-8* (Unpublished doctoral dissertation). Washington University, St. Louis.

Shen, J., & Confrey, J. (2007). From conceptual change to transformative modeling: A case study of an elementary teacher in learning astronomy. *Science Education*. *91*(6), 948–966. doi:10.1002/sce.20224

Shen, J., & Confrey, J. (2010). Justifying alternative models in learning the solar system: A case study on K-8 science teachers' understanding of frames of reference. *International Journal of Science Education*, *32*(1), 1–29.

Shen, J., Lei, J., Chang, H., & Namdar, B. (2014). Technology-enhanced, modeling-based instruction (TMBI) in science education. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communication and technology* (4th ed., pp. 529–540). New York: Springer. doi:10.1007/978-1-4614-3185-5_41

Shen, J., Lei, J., Namdar, B. & Chang, H., (2013, April). *Synthesizing modeling-based instruction in science education from 1980 to 2010*. Poster presented at National Association for Research in Science Teaching (NARST) Conference 2013, Puerto Rico.

Shin, N., Stevens, S. Y., & Krajcik, J. S. (2010). Tracking student learning over time using construct centered design. In S. Rodrigues (Ed.), *Using analytical frameworks for classroom research: Collecting data and analyzing narrative*. (pp. 39–58) New York: Routledge.

Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R., & van Hout-Wolters, B. H. A. M. (2009). The relation between students' epistemological understanding of computer models and their cognitive processing on a modelling task. *International Journal of Science Education*, *31*(9), 1205–1229. doi:10.1080/09500690802192181

Smetana, L. K., & Bell, R. L. (2012). Computer supported simulations to support science instructions and learning. A critical review of the literature. *Journal of Science Education*, *34*(9), 1337–1370. doi:10.1080/09500693.2011.605182

Stern, L., Barnea, N., & Shauli, S. (2008). The effect of a computerized simulation on middle school students' understanding of the kinetic molecular theory. *Journal of Science Education and Technology*, *17*(4), 305–315. doi:10.1023/A:1009436509753

Stratford, S. J., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology*, *7*(3), 215–234. doi:10.1023/A:1021840407112

Sun, D., & Looi, C.-K. (2013). Designing a web-based science learning environment for model-based collaborative inquiry. *Journal of Science Education and Technology*, *22*(1), 73–89. doi:10.1007/s10956-012-9377-9

Treagust, D. F., Chittleborough, G. D., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24*(4), 357–368. doi:10.1080/09500690110066485

Treagust, D. F., Chittleborough, G. D., & Mamiala, T. L. (2004). Students' understanding of the descriptive and predictive nature of teaching models in organic chemistry. *Research in Science Education*, *34*(1), 1–20. doi:10.1023/B:RISE.0000020885.41497.ed

van Borkulo, S. P., van Joolingen, W. R., Savelsbergh, E. R., &de Jong, T. (2012). What can be learned from computer modeling? Comparing expository and modeling approaches to teaching dynamic systems behavior. *Journal of Science Education and Technology*, *21*(2), 267–275. doi:10.1007/s10956-011-9314-3

Vosniadou, S. (1994). Capturing and modelling the process of conceptual change. *Learning and Instruction*, *4*, 45–69. doi:10.1016/0959-4752(94)90018-3

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118. doi:10.1207/s1532690xci1601_2

Wilson, M. (2005). *Constructing measures: An item response approach*. Mahwah, NJ: Erlbaum.

Wu, H.-K. (2010). Modelling a complex system: Using novice-expert analysis for developing an effective technology-enhanced learning environment. *International Journal of Science Education*, *32*(2), 195–219. doi:10.1080/09500690802478077