The potential of learning from erroneous models: comparing three types of model instruction

Frances M. Wijnen,[a]* Yvonne G. Mulder,[b] Stephen M. Alessi[c] and Lars Bollen[b]

Abstract

Learning from computer models is a promising approach to learning. This study investigated how three types of learning from computer models can be applied to teach high-school students (aged 14–17) about the process of glucose–insulin regulation. Two traditional forms of learning from models (i.e. simulating a predefined model and constructing a model) were compared to learning from an erroneous model. In this innovative form of learning from computer models, students are provided with a model that contained errors to be corrected. As such, students do not have to engage in the difficult task of constructing a model. Rather, they are challenged to work with and correct the model in order for the simulation to generate correct output. As predicted, learning from erroneous models enhances learning of domain-specific knowledge better than running a simulation or constructing a model.

Syst. Dyn. Rev. **31**, 250–270 (2015)

Introduction

Models are often used in education as they can represent dynamic processes, features or relationships that are notoriously difficult to understand (Groesser and Schaffernicht, 2012). Models (i.e. simplified representations of complex systems) are easier to work with than with actual systems (Ford, 1999). Coll and Lajium (2011) describe three main goals for using models: (i) to represent simpler forms of objects or concepts; (ii) to stimulate the learning of a concept and to support the visualization of a phenomenon; and (iii) to offer explanations of scientific phenomena. In this study a model refers to a graphical system dynamics computer model that describes and simulates time-dependent changes in a system (such as a human body, economy or market).

In the system dynamics literature there are many examples of studies that aim to improve the teaching of modeling. Similarly, there are examples in the field of education that use modeling to enhance learning about scientific phenomena (Mulder *et al*., 2015b). First we provide an overview of studies regarding learning and

[a] Institute for Teacher Education, Science Communication and School Practices (ELAN), University of Twente, PO Box 2177500 AE Enschede, Netherlands.
[b] Department of Instructional Technology, University of Twente, PO Box 2177500 AE Enschede, Netherlands.
[c] University of Iowa, Iowa City, IA 52242, U.S.A.

* Correspondence to: Frances M. Wijnen, Institute for Teacher Education, Science Communication and School Practices (ELAN), University of Twente, PO Box 2177500 AE Enschede, Netherlands. E-mail: f.m.wijnen@utwente.nl

Accepted by Andreas Größler, *Received 19 January 2015*; *Revised 9 December 2015*; *Accepted 10 December 2015*

System Dynamics Review
System Dynamics Review vol 31, No 4 (October-December 2015): 250–270
Published online 11 February 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/sdr.1546

teaching modeling from the literature in education. Then we discuss examples that aim to improve the teaching of modeling from the system dynamics perspective.

In education, two methods of learning with models are typically used: learning by simulating (i.e. by running) existing models (de Jong, 1991); and learning by creating one's own models (Van Lehn, 2013). The first, simulating existing models, requires students to run a model which has been built by someone else or the class. Students can use such models to investigate time-dependent changes in the process that the models represent. Compared to more traditional forms of learning, like studying a book or listening to a lecture, learning by running a simulation engages students in inquiry-learning activities. The learner can test hypotheses by performing experiments and observing outcomes, which is generally associated with higher learning outcomes (Alessi, 2000; Alfieri *et al.*, 2011; Scalise *et al.*, 2011). Furthermore, compared to constructing a model, running a simulation is very efficient, as it is easier to understand the underlying structure of a model than it is to program one (Alessi, 2000). The disadvantage of running simulations is that it may not result in the same quality, or depth of knowledge, as learning by constructing a model because it does not require the student to understand all the component variables and their relationships.

In contrast, learning by creating models requires students to first construct a model before it can be simulated. Students engage in iterative cycles of model construction, testing and revising (Stratford *et al.*, 1998; Hogan and Thomas, 2001). This is related to inquiry learning, which involves generating hypotheses, exploring hypotheses through experimentation and evidence evaluation (Hagemans *et al.*, 2013). When students simulate a model they do not engage in the processes of inquiry learning per se but, like model construction, model simulation can be very convenient for inquiry-learning activities. Supported by principles of constructivism, which suggest that students draw their own conclusions through creative experimentation, model construction should engage students in thoroughly thinking about a scientific phenomenon and that should facilitate a deeper understanding of the phenomenon (Alessi, 2000; Jonassen *et al.*, 2005). One reason for this deeper understanding is that learners become aware of knowledge gaps they had not noticed before (Kolloffel *et al.*, 2010), or that would not surface when simply running a simulation.

However, creating a model is not an easy task. The literature in system dynamics is replete with researchers' attempts to improve not only the teaching of *modeling*, but also the even more basic *understanding* of the concepts underlying modeling. Concerning the latter, the problem of *stock–flow failure* or understanding of accumulations is a persistent theme at the annual system dynamics conference and in articles published in *System Dynamics Review* and other journals (e.g. Cronin and Gonzalez, 2007; Cronin *et al.*, 2009; Brunstein *et al.*, 2010; Sterman, 2010; Groesser and Schaffernicht, 2012). Concerning the teaching of modeling, the many articles in which system dynamics researchers strive to *improve* the teaching of modeling (as we are

DOI: 10.1002/sdr

doing) is testament that it is by no means perfected and requires improvement. Following are the most well-known examples of such work.

Beginning almost two decades ago, Richardson (1996) discussed the challenges that lay ahead for the field of system dynamics, saying that examples of good instructional practice in the field are hard to find and that many models are too complex for any but expert modelers to understand.

Andersen *et al*. (1997), suggesting ways to improve *group* model building, states that it is still more art than science, and goes on to suggest improvements that would make it more a science. Although that was 1997, 15 years later Richardson (2012) was still discussing the difficulty of group model building, especially by beginners, and suggested the method of *concept modeling* as a way to facilitate beginners' modeling.

Haslett (2001), suggesting ways to improve not only students' modeling in the classroom but also the transfer of what they learn to real-world contexts, states that "modeling is time consuming for students and that their modeling processes are flawed" (p. 162).

Booth Sweeney and Sterman (2000), suggesting that improvements are needed in the teaching of system dynamics, reporting that even sophisticated MIT students exhibit poor understanding of stock and flow principles, and in their conclusions said, "System dynamics educators can learn much from attempts to overcome these misconceptions in science and mathematics education." Furthermore, Sterman (2002) in his seminal article "All models are wrong: reflections on becoming a systems scientist" lamented that even after 30 years of teaching system dynamics, "I'm sure I don't know the best way."

The last ten years have seen considerable work to improve the teaching of system dynamics modeling. Wolstenholme (2004), developed "generic system archetypes" as a way to improve the teaching of modeling. Hines *et al*. (2010) states that modeling is costly, time consuming and difficult to do, and suggests another technique, *construction by replacement*, with the goal of improving it. Kunc (2012) presents a method he calls *aided modeling* because "modeling takes specific training and a lot of time, even using user-friendly software" (p. 31).

Fisher (2011) has been the main system dynamics researcher focusing on teaching it to younger learners. Discussing her years of teaching system dynamics in a variety of K-12 classrooms, she maintains, and we agree, that while system dynamics is not too difficult for young learners and is very worthwhile for their education, "the process takes practice and requires discipline" (p. 404). A significant portion of her work in books and articles has focused on improving the teaching of modeling to younger learners. Skaza *et al*. (2013), like Fisher, focuses on K-12 instruction about system dynamics modeling. They discuss the barriers that K-12 teachers face introducing system dynamics in their classrooms. In addition to administrative and logistic barriers, teachers identify their own understanding of system dynamics principles and confidence in their own modeling ability as barriers to teaching system dynamics to their students.

DOI: 10.1002/sdr

Most recently, Richardson (2014a) states that "Conducting a model-based study of a personally chosen problem is the last, most difficult stage in a modeling course" (p. 86) and proposes that having students correct flawed models is a good way to ease into modeling one's own problems. In a subsequent article, Richardson (2014b) elaborates on his suggestion on how students should be given flawed models for correction and improvement. Not only does Richardson repeatedly point out how difficult it is to become a modeler, but also he specifically suggests learning from erroneous models, a method that our study addresses and lends support to.

To construct a model, learners must use a variety of resources to investigate the subject matter. They must select the information that is important and determine their relationships (Louca and Zacharia, 2012). They must also learn to use modeling software to create a model (Alessi, 2000). Therefore, students often find it difficult to create a model without additional help (Mulder *et al.*, 2010). Researchers cited in the previous paragraphs demonstrate their recognition that teaching modeling is hard and needs improvement.

Recent research suggests that completing a partial model is a more effective model construction activity that leads to higher learning gains (Mulder *et al.*, 2015a) compared to constructing a model 'from scratch'. However, it is still unclear how learning by completing a partial model relates to other forms of model-based learning (i.e. simulation, correction).

A third and new approach to model-based learning is providing students with erroneous models. Richardson (2014a) describes a "canonical sequence" to teach students modeling. This seven-step canonical sequence proposes an easy to complex modeling approach in which students start with exploring existing models and end with modeling personally chosen problems. Richardson explicitly suggests working with erroneous models as an intermediate activity, between simulating a model and creating one. Since students are provided with an already constructed model, the learning activity does not require much model construction. However, due to the errors in the model, students must engage in model testing and revising activities, which requires attention to the actual structure of the underlying model, before they can correctly simulate the behavior of the system. Since students engage in these additional activities, it is predicted that correcting a model leads to a deeper understanding of the phenomenon than just running (simulating) a model, but is less difficult than creating a model.

Because using erroneous models as a means of learning subject matter content is rather new, little is known about its effect on learning subject matter, compared to what is known about model simulation and construction. However, the literature does report on studies assessing the instructional value of erroneous examples in general. Erroneous examples are examples that contain one or more errors for learners to detect and correct (Tsovaltzi *et al.*, 2010). The literature reports mixed results on learning subject matter content from erroneous examples. On the one hand there are positive learning effects. For

DOI: 10.1002/sdr

example, compared to learning from correct examples, erroneous examples demonstrated an advantage for learning fractions (Tsovaltzi *et al.*, 2010) and when learning to make diagnoses in medical education (Stark *et al.*, 2011). On the other hand, some studies were not able to replicate these findings. For instance, Isotani *et al.* (2011) found no significant learning effects in the domain of decimals for erroneous examples compared to correctly worked-out examples and partially supported problem solving where students were able to see the correct answer if they did not succeed in solving the presented problem. Furthermore, Hilbert *et al.* (2008) found that participants in their incorrect worked-out map condition produced many false conclusions on a comprehension test. This casts some doubt on the effectiveness of using erroneous examples in instruction. As a result, teachers are suspicious of presenting errors to students because they fear that presenting errors to students will make them more inclined to make these errors (Tsamir and Tirosh, 2005; McLaren *et al.*, 2012).

In reaction to the mixed results, McLaren *et al.* (2012) describe three basic conditions that should make erroneous examples helpful for learning subject matter content. First, the errors should not be attributed to anybody, so students are not embarrassed by being confronted with their own errors. Second, the erroneous examples should be interactive and engaging: students should be asked to explain the errors (i.e. analyze the errors). It is important that students give possible solutions to correct the error, and students should receive feedback on their explanations and possible solutions. Finally, the errors should focus on the deepest misconceptions and misunderstandings students have, which should make erroneous examples more helpful to students with low prior knowledge. If students are confronted with their misconceptions and are asked to explain them they are forced to reconsider their first incorrect assumption (i.e. misconception), which helps them to learn how the complex process of interest functions (Mayer, 2006). Thus, when students correct errors they should engage in three cognitive processes: detecting errors, analyzing the errors (e.g. what makes this an error?) and correcting the errors.

An attempt to incorporate erroneous examples in model-based learning is described by Mulder *et al.* (2014). They conducted a study of learning from erroneous models approach to determine the effects of finding and correcting the errors in models. Although they found that the erroneous models enhanced students' learning, no significant improvement was attributable to the processes of detecting and/or correcting errors. Although informative, the Mulder *et al.* study did not incorporate all three basic conditions as formulated by McLaren *et al.* (2012). They did not include the processes of asking students to explain the errors and therefore the cognitive process of analyzing the error did not take place. Additionally, the question remains how this learning from erroneous models approach compares to the more traditional model-based learning approaches where students either run an existing model or construct their own models.

DOI: 10.1002/sdr

Design and hypotheses

The purpose of the present study was to assess the instructional effectiveness of the erroneous model-based learning approach when it meets all conditions described by McLaren *et al.* (2012), as compared to the two more traditional model-based learning approaches. The study employed a between-group design with three conditions for a lesson in which students learned about glucose regulation of the blood. In the model simulation condition, students worked with a correct pre-constructed model that could be simulated (run). In the model correction condition, students worked with an erroneous model. Students in the model construction condition worked with a partially developed model that had to be completed. To assess the relative instructional effectiveness of the erroneous model approach, students' modeling behavior and their increases in domain knowledge in the model correction condition were examined and contrasted to students in both the model simulation and model construction conditions. It was predicted that the more interactive and engaging learning methods (i.e. model correction and model construction) would result in more model testing and revising activities, as evidenced by the number of times students run their model. Moreover, it was predicted that students in the model correction condition would acquire more domain-specific knowledge than participants in the model simulation condition or the model construction condition, as the erroneous model is an interactive and engaging learning method that is less difficult than model construction.

Method

Participants

Participants were 70 Dutch high school students (58.57 % female) following the science track, with an average age of 15.74 (SD = 0.72). Class ranked pre-test scores were used to assign students to the model simulation condition ($n = 22$), the model correction condition ($n = 23$), or the model construction condition ($n = 25$), so that prior knowledge was comparable in every condition. A review of school curricula showed that the students had not yet learned about the processes of glucose regulation in the blood. Therefore, it was expected that the students had little or no prior knowledge.

Materials

Learning environment

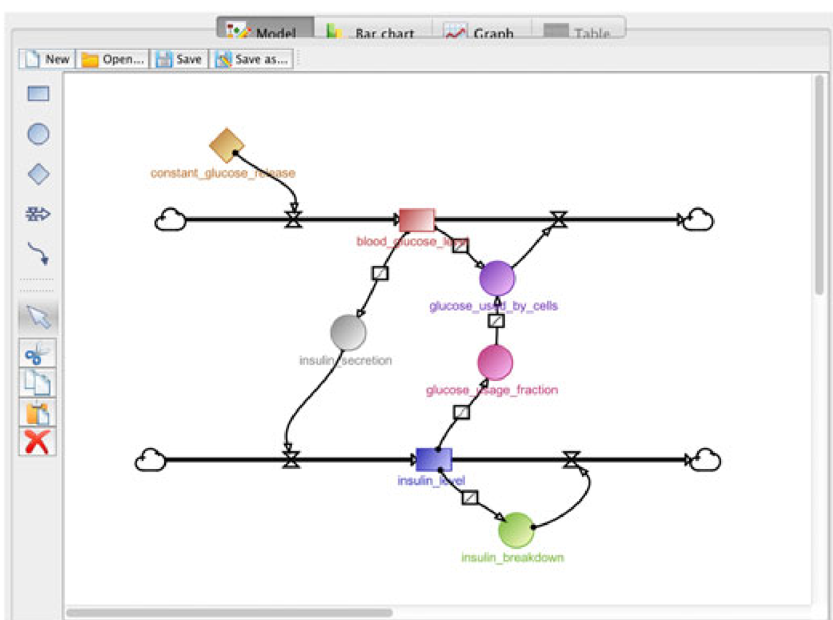
All participants worked with modeling software called SCYDynamics (Mulder *et al.*, 2014). It contains a model editor tool, as well as a bar chart

DOI: 10.1002/sdr

and a graph tool. There are three reasons for using SCYDynamics in this study: (i) it is appropriate for educational purposes because it provides a compact overview of the glucose-insulin process, making it easier for students to understand how this process works; (ii) it is possible to log all the actions students perform in the model, which provides insight into how students approach learning with this model; and (iii) it is possible to prepare variations of the model in order to create the three different conditions. Figure 1 shows the *model editor* tool and the model that was used in the model simulation condition.

The model itself and its variations have been created based on earlier, existing models from Niles *et al.* and Hamid (Halbower *et al.*, 1992; Hamid, 2009), on (theoretical) biomedical research results (Cobelli *et al.*, 1982; Tolic *et al.*, 2000; Makroglou *et al.*, 2006; Man *et al.*, 2007; Liu and Tang, 2008) and on biology text books used in Dutch secondary education. For pedagogical reasons, the complexity of the model has been lowered to match the target group (aged 14–17) and the time constraints (see "Procedure" section) imposed by the study design. Metabolic features like active/inactive insulin or the influence of glucagon on the liver have been omitted. The relations between glucose level and insulin secretion, and between insulin level and insulin usage, have been implemented as generalized logistic growth functions ("Richards' curve"; Richards, 1959). An exemplary result of simulating the model can be seen in Figure 3, where a glucose peak (e.g. from eating a candy bar) is being regulated back to a state of homeostasis.

Fig. 1. Model editor tool, displaying the model used in the simulation condition



DOI: 10.1002/sdr

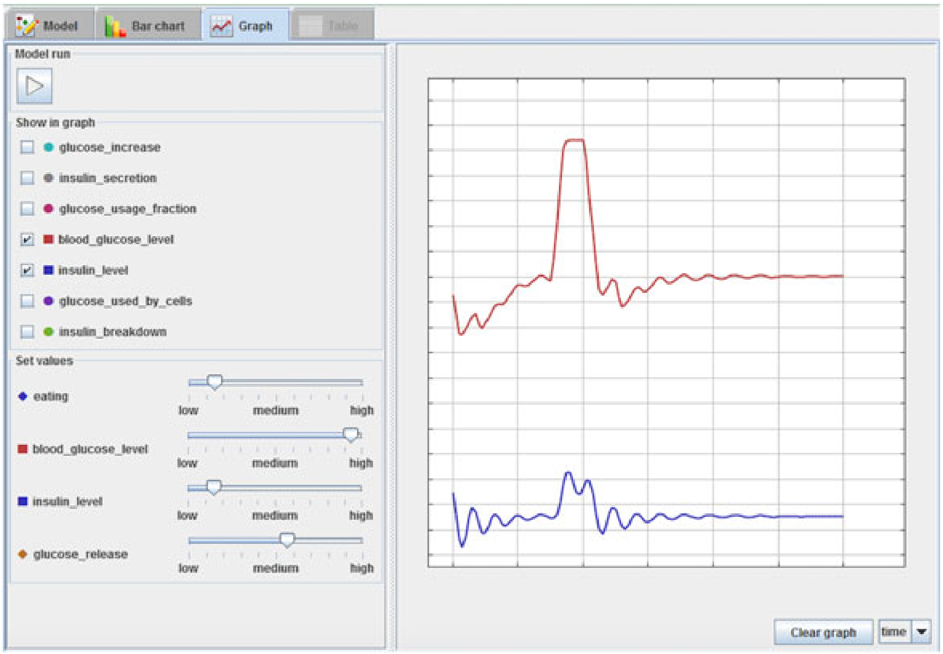
The students were spared from the complex, mathematical details of the model. Instead, they were able to select predefined qualitative relations between variables and define starting values as being “low”, “medium” or “high”. For example, they could select linear rising: if X is increasing, then Y is increasing. The selection of such a specification resulted in the application of mathematical formulas and variable values by SCYDynamics so as to create meaningful output in the form of graphical diagrams. The mathematical expressions and variable values were fetched from a so-called “reference model”, which also contained lists of alternative variable names (e.g. “glucose level” was treated to be semantically indifferent to “blood sugar level”) and alternative specifications of relations (e.g. the relation between blood glucose level and insulin secretion can be “constant low”, instead of a logistic growth relation, to indicate a diabetic disease). The reference model was hidden from the learners, and was only editable by teachers and domain experts.

The model editor tool enabled participants to run the model and analyze its output in the bar chart or the graph tool. Figure 2 shows the *bar chart*. It provided feedback on the model structure by displaying the number of correct and incorrect variables and relations. For each relation, it also showed the number of correct and incorrect causal directions. The bar chart indicated correct and incorrect model features based on the reference model mentioned above. This means the bar chart could also accept potentially alternative variable names and alternative specifications of relations as being correct, if available and described in the reference model. Figure 3 shows the *graph* tool, which gave information on the model’s behavior. All variables that students

Fig. 2. Bar chart

DOI: 10.1002/sdr

Fig. 3. Graph tool



included in their model could be selected to inspect their behavior in the system over time. If one or more relations were defined incorrectly, the graph would provide incorrect output or an error message.

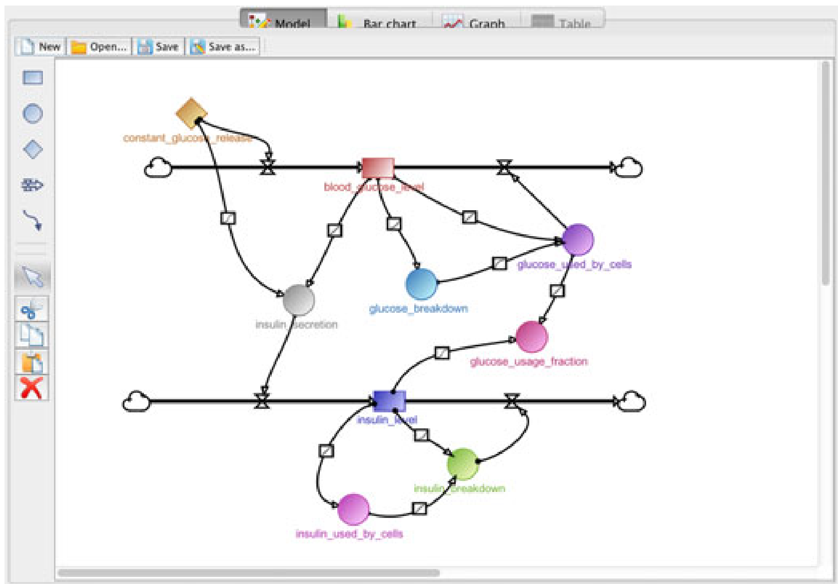
Erroneous model

The erroneous model (see Figure 4) was presented as if created by another student. Participants were told that the model contained errors and were asked to detect, analyze and correct the errors. Three types of errors can be introduced (in a correct model) to create an erroneous model: (i) a link or variable is redundant; (ii) a link or variable is presented incorrectly; and (iii) a link or variable is missing. The third type of error was not used in this study because of its similarity to the model construction activity. Consistent with studies on erroneous concept maps by Chang *et al*. (2002) and Hilbert *et al*. (2008), 30 percent of the model was incorrect, and errors were introduced regarding the concepts, the relations and the relation specifications. This resulted in a total of six errors. To determine what errors to include, the questions in Mulder *et al*. (2015a) that students most frequently answered incorrectly about glucose insulin regulation were used.

This erroneous model was designed to engage students in the three relevant cognitive processes as described in the introductory session of this

DOI: 10.1002/sdr

Fig. 4. Erroneous model



article (i.e. error detection, error analysis and error correction). When an incorrect model is given and students are instructed to correct this model, error detection and correction are activated. To ensure error analysis, the system provided a prompt screen when participants deleted or changed something in the model. The prompt asked: "What do you think is incorrect?" Only after answering this question could students continue with correction of the error. Because learners answered this prompt we were able to identify whether they engaged in the process of error analysis. This is important in order to interpret the data more appropriately.

Partial model

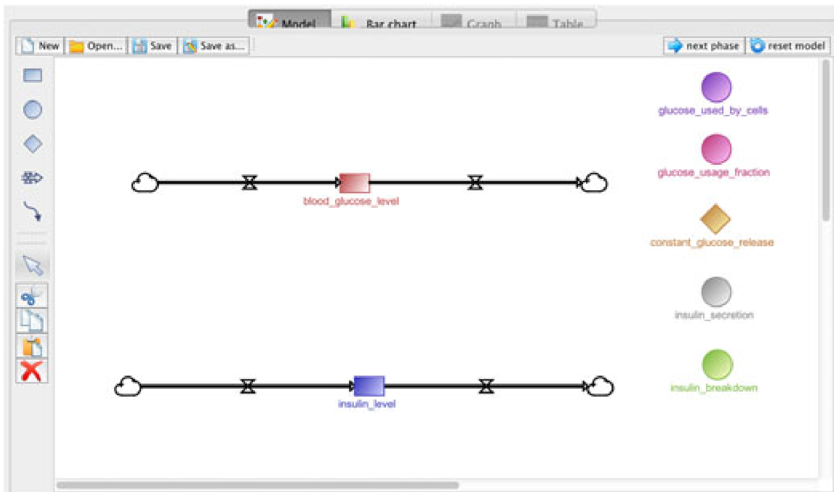
Students in the model construction condition received a partial model to complete (see Figure 5). As in Mulder *et al.* (2015a), the partial model provided an overview, displayed the relevant stocks (i.e. glucose and insulin) and listed the relevant constants and auxiliaries. To complete the model, participants had to correctly position the remaining variables and specify (i.e. draw) their relationships.

Domain knowledge test

A domain knowledge test was used to assess participants' knowledge of the glucose–insulin domain. The knowledge test consisted of four parts addressing key domain concepts (i.e. insulin and glucose; two questions),

DOI: 10.1002/sdr

Fig. 5. Partial model



themodel structure (four questions), the model system's behavior (three questions), and the recognition of errors (six errors). Model structure questions addressed relationships described in one of the instructional text's sections. Students could answer these questions by drawing the shape of the relationship in a graph. The model system's behavior questions addressed students' knowledge of glucose–insulin regulation in the three scenarios (i.e. in homeostasis, when eating high-calorie food, and the effects of diabetes). In recognition of error questions, students received a picture of the erroneous model and had to indicate what was presented incorrectly and how it should be presented.

Instructional text

All participants received an instructional text (five pages) that provided all relevant information about the glucose regulation of the blood. This text described the different processes that take place in the human body to regulate glucose and insulin levels. The text explained what homeostasis is, how the body regulates the levels of glucose and insulin and how this regulation is affected by diabetes.

Assignment

For each condition, students were given an assignment consisting of three parts. The first part concerned the glucose–insulin regulation in the steady state of homeostasis. The second and third parts were about disruptions of the steady state (i.e. eating a pizza and the effects of diabetes).

DOI: 10.1002/sdr

In the *model simulation condition,* the first part of the students' assignment was to explain the relations between variables. Additionally, students had to display processes in homeostasis and explain what happens with the glucose or insulin levels over time. For the *model correction condition,* the first part of the students' assignment required adjustment of the model in homeostasis so all errors were corrected. That included detecting, analyzing and correcting all errors in the model. The first part of the students' assignment for the *model construction condition* required them to complete the model in homeostasis.

The second part of the assignment required students to think about what happens with the glucose and insulin levels if you eat a pizza. Students had to add two new variables so the model correctly represented this situation. The third part of the assignment required students to think about what happens if a person has diabetes. To correctly answer that question, one relation had to be changed.

Procedure

The complete experiment consisted of two parts scheduled over the course of a week and a half. We did not prevent communication among the participants between the two parts of the experiment. We thought this was unnecessary, since participants followed the sessions as a whole class and therefore received the same information as fellow classmates. The introductory part (50 minutes) started with instructions explaining the goal of the experiment and participants' activities. Subsequently, participants were given 10–15 minutes to complete the pre-test. Participants then completed a tutorial about the model editor and basic concepts of modeling.

The second part of the experiment (90–100 minutes) started with reading the instructional text, which was also available for the rest of the experiment. After 10 minutes the assignment was provided so students could work with the model to complete the assignment. After 60 minutes the participants terminated the program and started the post-test, for which they had 20 minutes. After finishing the post-test the experiment ended.

Coding and scoring

Data were assessed using the knowledge tests and log files. Variables under investigation were students' knowledge scores, model testing activities and prompt answers. Students' model testing activities denote the number of times students ran their model with the bar chart tool (to obtain feedback on their model) or the graph tool (to gain insight into the process of glucose–insulin regulation). The numbers of model runs were assessed from the log files and indicated how many times participants clicked "play" in the bar chart and the graph tool.

DOI: 10.1002/sdr

Knowledge scores were based on students' knowledge test performance and indicated students' comprehension of the glucose–insulin domain. For every correct answer (and every corrected error in the final part) one point was given, allowing a maximum score of 14 points. (Question 8 was omitted from the analysis because a technical failure prevented many students from finding the answer.) The knowledge test was used as a pre- and post-test. Part four was not used in the pre-test, because the detection of errors is not useful for measuring prior knowledge. Inter-rater reliability was calculated based on data from previous studies and was high for all parts of the knowledge test (i.e. greater than 0.92 Cohen's k) (Mulder *et al.*, 2014, 2015a).

The prompt answers were also extracted from the log files and indicated participants' answers to the prompt question: "What do you think is incorrect?" in the model correction condition. A scoring scheme was developed for scoring the prompt answers. Each prompt answer was classified according to an increasing level of reasoning, using a hierarchical rubric consisting of high-level, low-level or no reasoning. High-level reasoning answers reflected knowledge gained either from the model or from the instructional text (e.g. "when insulin levels are rising more glucose shall be decomposed"). Two points were given for this reasoning. Low-level reasoning answers related to the model's syntax or the assignment's wording (e.g. "the arrow should be the other way around"). One point was given for this reasoning. Statements of ignorance (e.g. "blabla") were scored as no reasoning and no points were given. To determine inter-rater reliability, two raters scored 10 percent of the prompt answers. The inter-rater reliability was 0.91 (Cohen's k).

Results

The pre-test scores were used to distribute participants across the three conditions, so that prior knowledge was similar. Table 1 reports descriptive statistics of students' performance. Using Pillai's trace, multivariate analysis of variance (MANOVA) on the three aspects of the pre-test indicated no significant differences between the conditions, $V = 0.85$, $F(6, 132) = 0.77$, $p = .598$.

To determine the learning effect on participants' domain-specific knowledge, the knowledge pre-test and post-test was compared using a paired-sample t-test. Because the mean score of the pre-test (2.46, SD = 1.19) was low, the students can be considered novices. Students' mean post-test score for the first three parts was 3.40 (SD = 1.43), which differed significantly from students' knowledge pre-test scores, $t(69) = -5.93$, $p < 0.001$(two-tailed), $r = 0.58$.

A MANOVA, with all four aspects of the knowledge post-test as dependent variables, was used to analyze differences in post-test scores between conditions. Pillai's trace showed significant differences between the three conditions, $V = 0.40$, $F(8, 112) = 3.48$, $p = 0.001$. Subsequent univariate analysis of variance (ANOVA) showed significant differences between the conditions

DOI: 10.1002/sdr

Table 1. Descriptive statistics of students' performance

	Model simulation		Model correction		Model construction	
	M	SD	M	SD	M	SD
Pre-test						
1. Key domain concepts	1.41	0.59	1.17	0.58	1.40	0.65
2. Model structure	0.82	0.73	0.96	0.83	0.72	0.84
3. Model system's behavior	0.27	0.46	0.35	0.49	0.28	0.46
4. Error recognition	—	—	—	—	—	—
Total (parts 1,2 and 3)	2.50	1.06	2.48	1.20	2.40	1.32
Post-test						
1. Key domain concepts	1.68	0.58	1.70	0.47	1.55	0.67
2. Model structure	1.00	0.88	1.35	0.81	0.73	0.70
3. Model system's behavior	0.68	0.67	0.85	0.75	0.77	0.69
4. Error recognition	1.42	1.17	3.30	1.13	2.14	1.28
Total (parts 1,2 and 3)	3.77	1.88	4.30	1.61	3.44	1.61
Model testing behavior						
Model runs	2.63	2.31	12.70	7.97	10.33	12.62
Graph runs	22.89	16.34	23.86	19.83	18.71	23.75

regarding students' knowledge of the model structure, $F(2, 58) = 3.20$, $p = 0.048$, partial $\eta^2 = 0.10$ and regarding students' error recognition, $F(2, 58) = 12.26$, $p < 0.001$, partial $\eta^2 = 0.30$, but not on students' knowledge of key domain concepts $F(2, 58) = 0.45$, $p = 0.641$, nor regarding students' knowledge of the model system's behavior, $F(2, 58) = 0.27$, $p = 0.762$. Repeated contrasts analysis determined which conditions differed significantly. Concerning knowledge of the model structure, the model correction condition significantly outperformed the model construction condition, $t(69) = 0.62$, $p = 0.014$, $r = 0.08$, but the difference between the model simulation and model correction condition was not significant, $t(69) = 0.35$, $p = 0.176$, $r = 0.04$. Concerning error recognition, the model correction condition significantly outperformed both the model simulation condition, $t(69) = 1.88$, $p < 0.001$, $r = 0.22$, and the model construction condition, $t(69) = 1.16$, $p = 0.003$, $r = 0.14$.

Model testing behavior was analyzed using the Kruskal–Wallis test because the distribution of the data on students' bar chart runs and graph runs differed significantly from a normal distribution (bar chart runs: Shapiro–Wilk, $p < 0.001$; graph runs: Shapiro–Wilk, $p < 0.001$). That analysis showed significant differences between the three conditions, $H(2) = 16.74$, $p < 0.001$ regarding students' bar chart runs, but not regarding the number of times students ran their model with the graph tool, $H(2) = 2.55$, $p = 0.279$. Mann–Whitney U-tests indicated that students from both the model correction and model construction conditions ran their model more often with the bar chart than students in the model simulation condition (model simulation vs. model correction: $U(40) = 385.50$, $z = 4.24$, $p < 0.001$, $r = 0.65$; model simulation vs. model construction:, $U(41) = 320.50$, $z = 2.28$, $p = 0.023$, $r = 0.35$). There was

DOI: 10.1002/sdr

no significant difference in bar chart runs between the model correction and model construction condition, $U(45) = 199.50$, $z = -1.63$, $p = 0.103$.

Of final interest was the quality of the students' reasoning concerning corrections, as shown by their answers to the error analysis prompts. Students were prompted to analyze the error when they deleted or changed something in the model. In total, the 23 students answered 483 prompts; 47.62 percent of those answers were not guided by any reasoning, and 36.43% showed low-level reasoning based either on the syntax of the model or based on the assignment. Only a small percentage of the prompt answers (15.93%) showed high-level reasoning based on knowledge students gained from the model or the instructional text.

Correlations between students' domain knowledge gains and students' answers to the prompts indicated a significant positive correlation between the number of prompts a student had answered and his/her learning gain, $r = 0.44$, $p = 0.024$ (one-tailed). For each student, a total prompt answer score was calculated, indicating students' reasoning quality. The moderately positive correlation between this score and students' gain in knowledge was nearly significant, $r = 0.34$, $p = 0.064$ (one-tailed). However, a partial correlation, controlling for the number of prompts, showed a non-significant correlation between students' prompt score and knowledge gains, $r = 0.035$, $p = 0.442$ (one-tailed). This suggests that the positive relation between students' answers to the prompts and their knowledge gains can be attributed to the number of prompts they received, and thus the number of changes they had made to the model.

Discussion and conclusion

This study investigated the effects of an erroneous model learning approach on students' domain knowledge acquisition. To assess these effects, the learning from an erroneous model approach was contrasted to two more traditional model-based learning approaches where students either simulate a predefined model or construct a model themselves. It was predicted that learning from erroneous models would benefit learning as students need not engage in the difficult task of constructing a model, whereas they are challenged to work with the model more thoroughly (than when simply running the model) to generate correct output.

Results on the knowledge tests indicate that students in all three conditions improved their understanding of glucose–insulin regulation. In line with expectations, students working with an erroneous model learned more about the model structure than those who constructed the model themselves. Additionally, participants in the model correction condition were better at recognizing and correcting errors than in the simulation and model construction conditions. This indicates that students who had to correct an erroneous model learned more about its structure, both in that they were better able to answer questions about this structure and that they were able to recognize and correct errors,

DOI: 10.1002/sdr

which they had previously seen. These findings are consistent with some of the literature on learning from erroneous examples (Tsovaltzi *et al.*, 2010; Stark *et al.*, 2011; Durkin and Rittle-Johnson, 2012). However, these findings are by no means self-evident. For example, it could be that students in the model simulation condition were able to recognize the differences between the model they used and the erroneous model that was used in the model correction condition and presented to them afterwards. That might have been an easier task than recognizing the errors after finding and correcting the errors first.

Regarding learning about the behavior of the system, we see that there is an improvement in knowledge between the pre-test and the post-test. However, this difference is statistically non-significant. Furthermore, the results regarding learning about the behavior of the system indicated no differences between the conditions. There might be two explanations for these results. The first is that the activity of constructing and correcting a model in this study is mainly focused on the model structure and not so much on looking into model behavior. Because of this, learners might not have carefully observed the model behavior. The second explanation is that the questions used in the assignment that would require learners to focus on model behavior were the same in the three conditions. To help learners improve their understanding of the structure–behavior relationship it is important that learners focus on the model structure *and* observe model behavior. For the correction condition this would be the easiest to accomplish, by observing, recording and reflecting upon the model behavior after correction. This could be done by letting learners answer the questions: What is changing? How is it changing? Why is it changing? What else changes this way? (Quaden *et al.*, 2007). In the construction condition this could be accomplished by observing, recording and reflecting upon the model behavior after (every) modification. In the simulation condition this is more difficult to accomplish because the structure is not changed. It might help learners to focus on the model structure by asking specific questions about it. For example, how does the blood glucose level relate to insulin secretion? This could also be done for model behavior. For example, how does insulin secretion change if the blood glucose level is high versus low?

A possible explanation for the benefits of learning subject matter content from erroneous models is that this task, like simulating a pre-existing model, is easier than learning by constructing a (partial) model (Alessi, 2000). Like constructing models, students must engage in model testing and debugging activities that are essential for learning by modeling as they help students become aware of knowledge gaps (Stratford *et al.*, 1998; Hogan and Thomas, 2001; Kolloffel *et al.*, 2010). This explanation is partially supported by the results of students' model testing behavior. As expected, both participants in the correction and construction conditions tested their model more often with the bar chart compared to those in the simulation condition. This confirms that students in the model correction and model construction conditions engaged in more model testing and debugging activities regarding the model's

DOI: 10.1002/sdr

structure. However, contrary to expectations, there were no significant differences between conditions regarding graph use, which was high in all conditions. This could be explained from the feedback function of the graph tool, which provides information about the system's behavior. Most likely, this tool helped students in all conditions, as students in all conditions could use this tool to learn about the system behavior of the glucose–insulin regulation. This would also explain why there were no differences between conditions in learning affordances regarding the questions about system behavior. All model-based learning activities enhanced students' learning of the system. Identifying and correcting errors had no additional learning effect regarding students' knowledge about the behavior of the system.

The results of this study suggest conditions for successful model-based learning and help explain the mixed results of studies on learning from erroneous examples. In the latter, some studies reported positive results (Tsovaltzi *et al.*, 2010; Stark *et al.*, 2011; Durkin and Rittle-Johnson, 2012) and others did not (Hilbert *et al.*, 2008; Isotani *et al.*, 2011). Responding to these mixed results, McLaren *et al.* (2012) described three basic conditions, which should improve learning from erroneous examples (errors should be fictitious, erroneous examples should be interactive and engaging so students participate in the process of analyzing the errors, and errors should aim at students' deepest misconceptions). In this study all three conditions are met and positive results for learning from an erroneous model were found, lending support to the value of the conditions described by McLaren *et al.* (2012). Additionally, this shows that the benefits of learning from erroneous examples apply to complex learning domains such as used in system dynamics modeling.

The second of McLaren's conditions, that erroneous examples should be interactive and engaging and that students should be asked to explain the errors, is most often lacking. Therefore, this study provided prompts the moment students wanted to change something in the model, triggering them to explain the errors. To assess the value of these explanations, this study examined the relationship between students' reasoning to explain the errors and students' knowledge improvement. As predicted, a significant correlation was found between the number of prompts and students' knowledge gains. However the correlation between the quality of students' explanations and students' gain in knowledge (controlling for the number of prompts) was not significant. This indicates that the number of changes students made in the model is related to students' knowledge gains, but that the quality of the reasoning has little or no added effect. This suggests that students should be encouraged to make as many changes as they think are necessary.

Another way to address the second of McLaren's conditions is the use of the bar chart, which provides the learner with feedback about the correctness of the structure of the model. As is described above, the participants in the correction and construction condition use the bar chart more often than participants in the simulation condition, which indicates that participants in those

DOI: 10.1002/sdr

conditions engaged in more model testing and debugging activities. This outcome also supports the benefit of McLaren's second condition for learning from correcting and constructing models.

These results have practical implications for education. Learning subject matter content from erroneous models can be a valuable way to teach students about complex processes. Learning subject matter content from models is consistent with the canonical sequence described by Richardson (2014a). His framework proposes a sequence from easy to complex modeling activities, starting with exploring existing models (simulating), then incorporating learning from erroneous models, and ending with modeling of personally chosen problems (construction). The present study emphasizes the value of learning from erroneous models as part of that sequence. More research is necessary to address: (i) the effects of individual tasks in this canonical sequence; and (ii) the effects of these tasks on students' modeling skills. Additionally, future research could focus on whether additional support (to make learning from erroneous models easier) can produce the same learning effects as learning from erroneous models without support. That will make learning by modeling more accessible to students at other educational levels.

Acknowledgements

This study was conducted in the context of the project "Learning through modeling and self explanations", which is part of the National Initiative Brain and Cognition (NIHC) funded by the Dutch Organization for Scientific Research (NWO), grant 056- 31-011.

Biographies

Frances M. Wijnen is a junior researcher at the Institute for Teacher Education, Science Communication and School Practices (ELAN) at the University of Twente. Currently, she focuses on the use of robots in primary education to support inquiry-based learning.

Yvonne G. Mulder is a postdoctoral researcher at the Department of Instructional Technology of the University of Twente. She specializes in learning by modeling. In her current project she investigates whether and how principles from explanation-based learning can be applied to learning by modeling.

Stephen Alessi is an Associate Professor of Educational Psychology at the University of Iowa. His teaching and research emphasize the application of cognitive learning theory to the design of educational software and online instruction, especially the design of instructional simulations. He is co-author (with Stanley Trollip) of *Multimedia for Learning: Methods and Development*.

DOI: 10.1002/sdr

Lars Bollen is a postdoctoral researcher at the Department of Instructional Technology of the University of Twente. In his research he focuses on learning by modeling, learning with simulations, modeling and sketching, visual modeling languages and environments, mobile devices and pen-based devices in learning scenarios, and (inter)action analysis.

References

Alessi S. 2000. Building versus using simulations. In *Integrated and Holistic Perspectives on Learning, Instruction and Technology*, Spector JM, Anderson T (eds). Springer: Berlin; 175–196.

Alfieri L, Brooks PJ, Aldrich NJ, Tenenbaum HR. 2011. Does discovery-based instruction enhance learning? *Journal of Educational Psychology* **103**(1): 1–18.

Andersen DF, Richardson GP, Vennix JAM. 1997. Group model building: adding more science to the craft. *System Dynamics Review* **13**(2): 187–201.

Booth Sweeney L, Sterman JD. 2000. Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review* **16**(4): 249–286.

Brunstein A, Gonzalez C, Kanter S. 2010. Effects of domain experience in the stock–flow failure. *System Dynamics Review* **26**(4): 347–354.

Chang KE, Sung YT, Chen ID. 2002. The effect of concept mapping to enhance text comprehension and summarization. *Journal of Experimental Education* **71**(1): 5–23.

Cobelli C, Federspil G, Pacini G, Salvan A, Scandellari C. 1982. An integrated mathematical-model of the dynamics of blood-glucose and its hormonal control. *Mathematical Biosciences* **58**(1): 27–60.

Coll R, Lajium D. 2011. Modeling and the future of science learning. In *Models and Modeling*, Khine MS, Saleh IM (eds). Springer: Berlin; 3–21.

Cronin MA, Gonzalez C. 2007. Understanding the building blocks of dynamic systems. *System Dynamics Review* **23**(1): 1–17.

Cronin MA, Gonzalez C, Sterman JD. 2009. Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes* **108**: 116–130.

de Jong T. 1991. Learning and instruction with computer simulations. *Education and Computing* **6**(3–4): 217–229.

Durkin K, Rittle-Johnson B. 2012. The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction* **22**(3): 206–214.

Fisher DM. 2011. "Everybody thinking differently": K-12 is a leverage point. *System Dynamics Review* **27**(4): 394–411.

Ford A. 1999. *Modeling the Environment: An Introduction to System Dynamics Modeling of Environmental Systems*. Island Press: Washington, DC.

Groesser SN, Schaffernicht M. 2012. Mental models of dynamic systems: taking stock and looking ahead. *System Dynamics Review* **28**(1): 46–68.

Hagemans MG, van der Meij H, de Jong T. 2013. The effects of concept map-based support tool on simulation-based inquiry learning. *Journal of Educational Psychology* **105**(1): 1–24.

Halbower M, Niles R, Sudnick T. 1992. Glucose regulation: a pre-college instruction guide employing dynamic computer simulation. Department memorandum D-4261. System Dynamics Group, MIT.

DOI: 10.1002/sdr

Hamid TKA. 2009. *Thinking in Circles about Obesity*. Springer: New York.
Haslett T. 2001. Experiences and reflections on the transition from classroom to practice. *System Dynamics Review* **17**(2): 161–169.
Hilbert T, Nückles M, Matzel S. 2008. Concept mapping for learning from text: evidence for a worked-out-map-effect. In *Proceedings of the 8th International Conference for the Learning Sciences*, Vol. **1**. International Society of the Learning Sciences: Utrecht: Netherlands; 358365.
Hines J, Malone T, Goncalves P, Herman G, Quimby J, Murphy-Hoye M, Rice J, Patten J, Ishii H. 2010. Construction by replacement: a new approach to simulation modeling. *System Dynamics Review* **27**(1): 64–90.
Hogan K, Thomas D. 2001. Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology* **10**: 319–345.
Isotani S, Adams D, Mayer R, Durkin K, Rittle-Johnson B, McLaren B. 2011. Can erroneous examples help middle-school students learn decimals?. In *Towards Ubiquitous Learning*, Kloos CD, Gillet D, Crespo Garcia RM, Wild F, Wolpers M (eds). Springer: Berlin; 181–195.
Jonassen D, Johannes S, Gottdenker J. 2005. Model building for conceptual change. *Interactive learning environments* **13**(1–2): 15–37.
Kolloffel B, Eysink TH, de Jong T. 2010. The influence of learner-generated domain representations on learning combinatorics and probability theory. *Computers in Human Behavior* **26**(1): 1–11.
Kunc M. 2012. Teaching strategic thinking using system dynamics: lessons from a strategic development course. *System Dynamics Review* **28**(1): 28–45.
Liu W, Tang F. 2008. Modeling a simplified regulatory system of blood glucose at molecular levels. *Journal of Theoretical Biology* **4**(21): 608–620.
Louca LT, Zacharia C. 2012. Model-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. *Educational Review* **64**(4): 471–492.
Makroglou A, Li J, Kuang Y. 2006. Mathematical models and software tools for the glucose–insulin regulatory system and diabetes: an overview. *Applied Numerical Mathematics* **56**(3–4): 559–573.
Man CD, Rizza RA, Cobelli C. 2007. Meal simulation model of the glucose–insulin system. *IEEE Transactions on Biomedical Engineering* **54**(10): 1740–1749.
Mayer RE. 2006. *Learning and instruction*, Pearson, New Jersey.
McLaren B, Adams D, Durkin K, Goguadze G, Mayer R, Rittle-Johnson B, Sosnovsky S, Isotani S, *et al*. 2012. To err is human, to explain and correct is divine: a study of interactive erroneous examples with middle school math students. In *21st Century Learning for 21st Century Skills*, Ravenscroft A, Lindstaedt S, Kloos CD, Hernández-Leo D (eds). Springer: Berlin; 222–235.
Mulder YG, Lazonder AW, 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): 20332053.
Mulder YG, Bollen L, de Jong T. 2014. Learning from erroneous models using SCYDynamics. In *Proceedings of the International Conference of the System Dynamics Society*. Delft, Netherlands.
Mulder YG, Bollen L, de Jong T, Lazonder A. 2015a. Scaffolding learning by modelling: the effects of partially worked-out models. *Journal of Research in Science Teaching* 2015. doi: 10.1002/tea.21260.

DOI: 10.1002/sdr

Mulder YG, Lazonder AW, de Jong T. 2015b. Key characteristics of successful science learning: the promise of learning by modelling. *Journal of Science Education and Technology* **24**(2–3): 168–177.
Quaden R, Ticotsky A, Lyneis D. 2007. *The Shape of Change*. Creative Learning Exchange: Acton, MA.
Richards FJ. 1959. A flexible growth function for empirical use. *Journal of Experimental Botany* **10**(2): 290–300.
Richardson GP. 1996. Problems for the future of system dynamics. *System Dynamics Review* **12**(2): 141–157.
Richardson GP. 2012. Concept models in group model building. *System Dynamics Review* **29**(1): 42–55.
Richardson GP. 2014a. “Model” teaching. *System Dynamics Review* **30**(1–2): 81–88.
Richardson GP. 2014b. “Model” teaching III: examples for the later stages. *System Dynamics Review* **30**(4): 291–299.
Scalise K, Timms M, Moorjani A, Clark L, Holtermann K, Irvin PS. 2011. Student learning in science simulations: design features that promote learning gains. *Journal of Research in Science Teaching* **48**: 1050–1078.
Skaza H, Crippen KJ, Carroll KR. 2013. Teachers’ barriers to introducing system dynamics in K-12 STEM curriculum. *System Dynamics Review* **29**(3): 157–169.
Stark R, Kopp V, Fischer MR. 2011. Case-based learning with worked examples in complex domains: two experimental studies in undergraduate medical education. *Learning and Instruction* **21**(1): 22–33.
Sterman JD. 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review* **18**(4): 501–531.
Sterman JD. 2010. Does formal system dynamics training improve people’s understanding of accumulation? *System Dynamics Review* **26**(4): 316–334.
Stratford SJ, 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.
Tolic IM, Mosekilde E, Sturis J. 2000. Modeling the insulin–glucose feedback system: the significance of pulsatile insulin secretion. *Journal of Theoretical Biology* **207**(3): 361–375.
Tsamir P, Tirosh D. 2005. In-service elementary mathematics teachers’ views of errors in the classroom. *Focus on Learning Problems in Mathematics* **27**(3): 30–42.
Tsovaltzi D, Melis E, McLaren B, Meyer A, Dietrich M, Goguadze G. 2010. Learning from erroneous examples: when and how do students benefit from them?. In *Sustaining TEL: From Innovation to Learning and Practice*, Wolpers M, Kirschner PA, Scheffel M, Lindstaedt S, Dimitrova V (eds). Springer: Berlin; 357–373.
Van Lehn K. 2013. Model construction as a learning activity: a design space and review. *Interactive Learning Environments* **21**(4): 371–413.
Wolstenholme E. 2004. Using generic system archetypes to support thinking and modeling. *System Dynamics Review* **20**(4): 341–356.

DOI: 10.1002/sdr