

What Can Be Learned from Computer Modeling? Comparing Expository and Modeling Approaches to Teaching Dynamic Systems Behavior

Sylvia P. van Borkulo · Wouter R. van Joolingen ·
Elwin R. Savelsbergh · Ton de Jong

Published online: 18 May 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Computer modeling has been widely promoted as a means to attain higher order learning outcomes. Substantiating these benefits, however, has been problematic due to a lack of proper assessment tools. In this study, we compared computer modeling with expository instruction, using a tailored assessment designed to reveal the benefits of either mode of instruction. The assessment addresses proficiency in declarative knowledge, application, construction, and evaluation. The subscales differentiate between simple and complex structure. The learning task concerns the dynamics of global warming. We found that, for complex tasks, the modeling group outperformed the expository group on declarative knowledge and on evaluating complex models and data. No differences were found with regard to the application of knowledge or the creation of models. These results confirmed that modeling and direct instruction lead to qualitatively different learning outcomes, and that these two modes of instruction cannot be compared on a single “effectiveness measure”.

Keywords Assessment · Computer modeling · Dynamic systems · Instructional technology · Simulation-based learning environments

Introduction

Computer modeling involves the construction or modification of models of (dynamic) systems that can be simulated (Penner 2001). Constructing models and experimenting with the resulting simulations helps learners to build their understanding about complex dynamic systems. Although modeling of dynamic systems appears to be difficult for secondary education students (Cronin and Gonzalez 2007; Fretz et al. 2002; Hmelo et al. 2000; Sins et al. 2005; Serman 2002; Wilensky and Resnick 1999), its potential benefits make it a worthwhile activity to include in the science curriculum (Magnani et al. 1998; Mandinach 1989; Qudrat-Ullah 2010; Stratford et al. 1998).

An example of a computer modeling environment is shown in Fig. 1. This modeling environment, called Co-Lab (van Joolingen et al. 2005), provides a modeling language as well as tables and graphs for displaying the results of executing the model.

Recently, the debate about the effectiveness of constructivist and, in particular, inquiry approaches to learning has gained new momentum (Kirschner et al. 2006; Klahr and Nigam 2004; Rittle-Johnson and Star 2007). Opponents of constructivist approaches to learning argue that the proposed benefits of inquiry approaches do not find support in experimental research. Indeed, no unequivocal evidence for the benefits of inquiry learning can be found in the literature. Some studies find no improvement from inquiry learning (e.g., Lederman et al. 2007), whereas others do find gains on inquiry-specific learning outcomes such as process

S. P. van Borkulo · W. R. van Joolingen (✉) · T. de Jong
Department of Instructional Technology, Faculty of Behavioral
Sciences, University of Twente, PO BOX 217, 7500 AE
Enschede, The Netherlands
e-mail: w.r.vanjoelingen@utwente.nl

S. P. van Borkulo
e-mail: s.vanborkulo@uu.nl

T. de Jong
e-mail: a.j.m.dejong@utwente.nl

S. P. van Borkulo · E. R. Savelsbergh
Freudenthal Institute for Science and Mathematics Education,
Utrecht University, Utrecht, The Netherlands
e-mail: e.r.savelsbergh@uu.nl

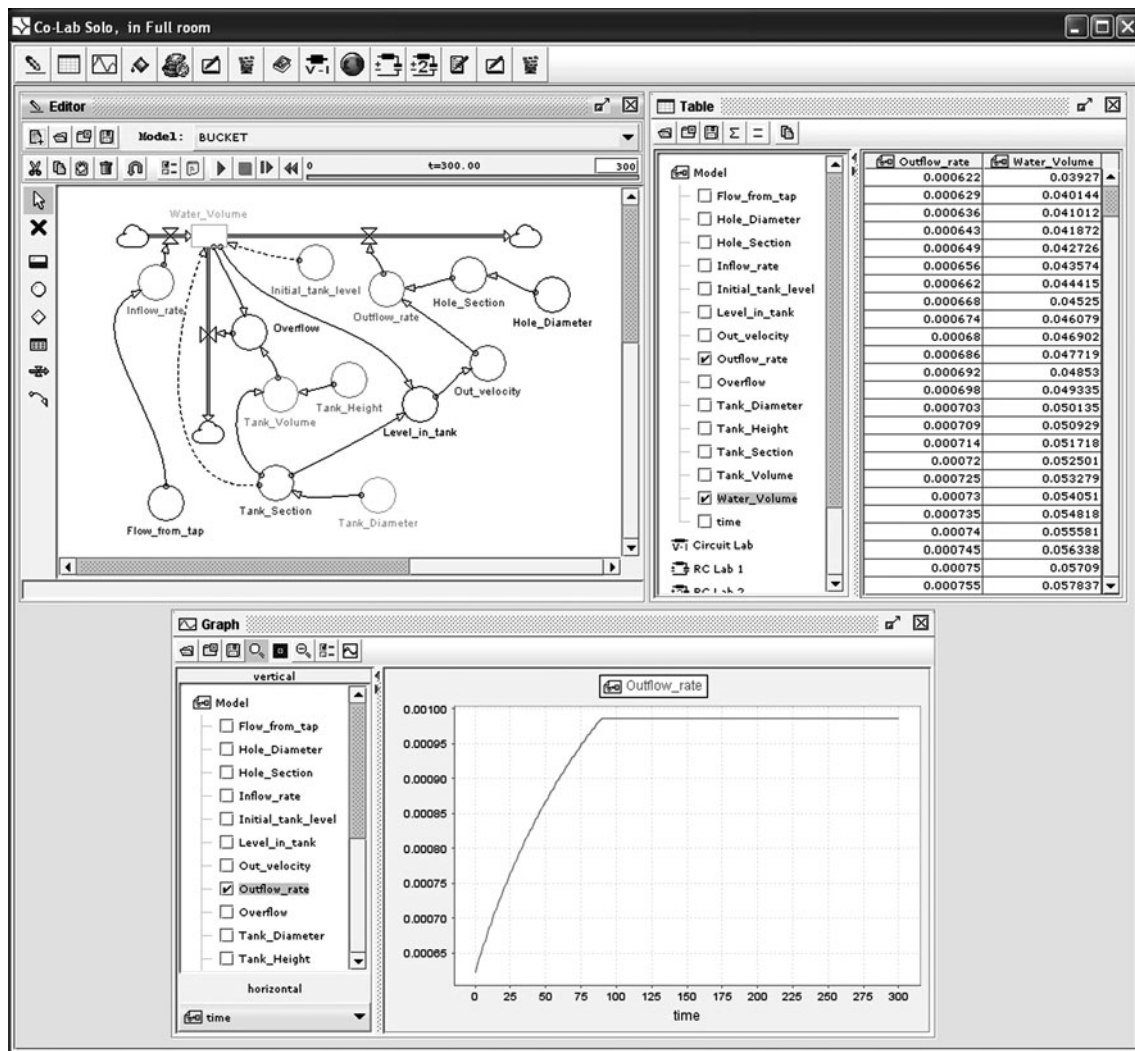


Fig. 1 The learning environment Co-Lab with its modeling tool (top left window)

skills (Geier et al. 2008), ‘more sophisticated reasoning abilities’ involved in solving complex, realistic problems (Hickey et al. 1999), and scientific thinking skills in guided inquiry (Lynch et al. 2005). A recent study by Sao Pedro and colleagues (Sao Pedro et al. 2010) shows that there are indications that inquiry learning can result in better long-term retention of inquiry skills, such as the *control of variables* strategy. A recent meta-analysis confirms that overall, inquiry learning is more effective than expository teaching, as long as the inquiry is scaffolded (Alfieri et al. in press). In the current article we contribute to this general discussion by comparing a specific form of inquiry learning, namely learning by modeling, with expository teaching. We also take the discussion to the next level by raising the issue of what it means to be “better”; in other words, what to measure when comparing different modes of instruction? Learning approaches are designed with the intention of improving specific learning processes and learning

outcomes. This means that when comparing one approach with another, one should expect changes in the specific learning outcomes for which each approach was designed. In other words, if a specific mode of instruction claims to improve reasoning skills, its effects are not properly measured by a memory test. This means that we need to use measures that are appropriate for the learning outcomes we expect from computer modeling as well as for those expected from expository teaching. In order to do this we need to describe in more detail what the expected learning outcomes of learning by modeling are.

Various benefits of computer modeling have been claimed in the literature. First, modeling is a method for understanding the behavior and characteristics of complex dynamic systems (Booth Sweeney and Sterman 2007; Sterman 1994). Second, modeling is assumed to enhance the acquisition of conceptual knowledge of the domain involved (Clement 2000). Modeling has the potential to help learners

develop high-level cognitive skills and thereby to facilitate conceptual change (Doerr 1997). Third, modeling is assumed to be especially helpful for the learning of scientific reasoning skills (Buckley et al. 2004; Mandinach and Cline 1996). Key model-based scientific reasoning processes are *creating*, *evaluating*, and *applying* models in concrete situations (Wells et al. 1995).

In comparing the outcomes from the two contrasting modes of instruction, expository instruction and computer modeling, we expect specific differences on the model-based reasoning processes of applying, creating, and evaluating models. The expository mode of instruction in this study directly presents the information to the learners, primarily in a textual format. Guidance is provided in the form of assignments, but without any dynamic tools such as simulations or concept maps and without explicit model building. The modeling mode of instruction comprises a guided inquiry approach supported by modeling and simulation tools. The two modes of instruction were compared using a test intended to detect the specific forms of knowledge gained by both modeling activities and expository instruction.

The test distinguishes two dimensions of knowledge: type of reasoning and complexity. The first dimension comprises *declarative* knowledge, the ability to remember facts from the information provided. It also includes the core reasoning activities of a modeling activity: *applying* knowledge of relations in a model by making predictions and giving explanations, *creating* a model from variables and relations between variables, and *evaluating* models and experimental data produced by a model (Wells et al. 1995).

The second dimension concerns the aspect of complexity. Modeling is typically used to understand complex dynamic systems and understanding complex systems is fundamental for understanding science (Assaraf and Orion 2005; Hagemayer and Waldmann 2000; Hmelo-Silver et al. 2007; Hogan and Thomas 2001; Jacobson and Wilensky 2006). We distinguish *simple* and *complex* model units based on the number of variables and relations involved. A *simple* unit is the smallest meaningful unit of a model, with only one dependent variable and only direct relations to that variable. A *complex* unit is a larger chunk that contains indirect relations and possibly (multiple) loops and complex behavior (see Fig. 2). Because the derivation of indirect relations in a causal network is often complex and computationally more demanding (Glymour and Cooper 1999), a test item about indirect relations will invoke more complex reasoning.

Two versions of the test were developed to cover the difference between domain-dependent and domain-independent modeling skills. In principle, model-based reasoning can be largely domain-independent. For instance, if a model contains a relation stating that when the water

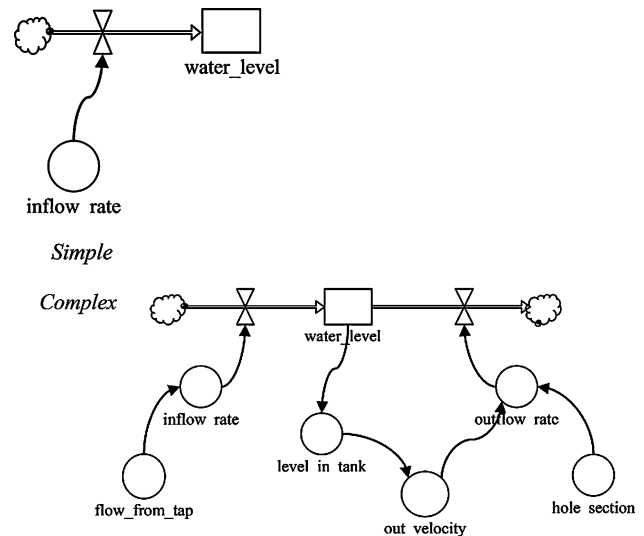


Fig. 2 Example of a simple and a complex model part

level in a tank increases, the water flow will increase, predicting what will happen to the water flow when the water level changes can be done independently of the meaning of the variables involved. However, reasoning with a model can be influenced by the availability of relevant domain knowledge (Fiddick et al. 2000) and thus may be different in a familiar versus an unfamiliar domain. In an unfamiliar domain, the only information learners have is the model itself. The learner must reason by following the relations in the model in a step-by-step way, building a chain of reasoning. In a familiar domain, reasoning steps may be bypassed because the outcome of the reasoning chain as a whole can be retrieved from memory. For instance, in a model that includes a capacitor, a person with knowledge of electronics will be able to reason that the voltage over the capacitor will increase as a consequence of a charging current, stepping over the charge as an intermediate variable. In an unfamiliar domain such a reasoning shortcuts will not be possible.

Research Question

The main research question for the current study was whether the two contrasted instructional approaches of modeling and expository teaching will result in specific differences in knowledge acquisition as measured by subscales of our test. Because learners in our study worked on a modeling problem in a specific domain, for a relatively short period of time, we expected effects mainly in their knowledge related to that domain, rather than in their more general modeling skills. Therefore, we focused on the domain-specific test to assess outcomes, and used domain-independent modeling skills as a pretest. We expected differences in learning outcomes on several subscales. Being able to run their own models and

having a simulation tool available enabled the learners in the modeling condition to perform experiments and to evaluate experimental data. Moreover, a large part of evaluation based on experiments is making predictions, and thereby applying the rules of system dynamics by reasoning with the relations. Therefore, we expected the modelers to perform better on the subscales that measure the reasoning processes of *evaluation* and *application*. Furthermore, a substantial amount of time should be spent on constructive activities such as translating concepts into variables and creating relations between variables. Thus, we also expected differences in favor of the modelers on the *create* scale. The expository learners are more directly and explicitly exposed to the concepts in the domain. Therefore, we expected the expository teaching to cause learners to be more efficient in remembering *declarative* simple and complex domain knowledge. Because the modelers had tools that support the creation and exploration of conceptual structures with a concrete artifact that provides a structural overview of the model, we expected the predicted advantages of the modelers to be more prominent with the *complex* models than the *simple* models.

Method

Participants

Seventy-four (51 males and 23 females) eleventh grade students from two upper track secondary schools participated in this study. The participants were between 16 and 19 years old ($M = 17.20$, $SD = .55$) and all were in a science major.

Materials

Co-Lab Learning Environment

The Co-Lab software (van Joolingen et al. 2005) provides a learning environment for each of the two conditions. The domain chosen was global warming. One version of the environment was configured for modeling-based instruction and consisted of a simulation of the basic energy model of the Earth, a modeling editor to create and simulate models, graphs and tables to evaluate the data produced by the model, and textual information about the domain. A second version of the environment was set up for expository instruction and consisted of the textual and pictorial information needed for writing a summary report on the topic of global warming.

Worksheets with assignments about factors in global warming were given to all participants as scaffolds. Their work was subdivided into three parts. The first part was

about *climate models* in general and included questions about the quality and accuracy of making global warming predictions using models. The second part concerned the factors *albedo* and *heat capacity*, and included questions about the influence of these factors on the temperature on Earth. This was implemented in different ways for the modelers (who created a model to support their reasoning) and the expository learners (who used the information provided to solve the problems in the text). For example, an assignment about the influence of the albedo on the equilibrium temperature asked both groups to predict what would happen with the equilibrium temperature if the albedo was high or low respectively. Subsequently, the modelers were asked to investigate their hypotheses with their model whereas the expository learners answered the question based on the information given. The third part was about evaluating one's understanding of the domain structure. The modelers were asked to compare their own model's behavior with the given simulation of the Earth's basic energy model. The expository learners were asked to compare their findings about the influential factors with given global warming scenarios. These scenarios specified a number of plausible future climates under the assumption of different values of future emissions of greenhouse gases. The expository learners wrote a report about the factors influencing the temperature on Earth as a final product, while the final product for the modelers was represented by the model they created.

The Modeling Knowledge Tests

Two paper-and-pencil tests for modeling knowledge were constructed according to the considerations introduced above (van Borkulo et al. 2008). This means that both tests had 4 (Remember declarative knowledge, Apply, Create, Evaluate) \times 2 (Simple, Complex) subscales. One test was domain-independent and the other test was specific for the domain of energy of the Earth. The domain-independent test was used as a pretest and the domain-specific test as posttest. The scores on the domain-generic pretest were used to match participants in the experimental groups for prior modeling ability. The results on the domain-specific posttest were analyzed using the domain-general pretest as a covariate to control for individual differences in prior modeling skills.

The domain-independent test introduced the fictitious phenomenon of the "harmony of the spheres". Because this test was about a fictitious phenomenon, it was impossible that students would have any relevant domain knowledge or experiential knowledge to rely on. The domain-specific test was about the domain of global warming, where students would have relevant domain knowledge after the intervention. Both tests introduced a model of the domain about

Table 1 Distribution of the number of items in pre—and posttest among the framework dimensions

Number of items	Pretest		Posttest	
	Harmony		Black sphere	
	Simple	Complex	Simple	Complex
Declarative	3	3	3	3
Application	3	4	3	3
Creation	2	4	3	3
Evaluation	3	3	3	3
Total	11	14	12	12
	25		24	

which different kinds of questions were asked. The model structures for both tests were isomorphic, meaning that the models presented were identical, except for the names of the variables.

The “harmony of the spheres” test consists of 25 items distributed over the eight subscales (see Table 1). The declarative knowledge items measured students’ prior knowledge about modeling formalism, and the application, creation, and evaluation categories contained problems about the harmony model that was introduced to the students. Figure 3 shows the model that was given in the pretest. Figure 4 shows examples of a simple application item and a complex evaluation item.

The domain-specific “black sphere” posttest concerned the modeling of global warming and hence involved the domain of energy of the Sun and the Earth. The black sphere test consists of 24 items again covering the eight subscales introduced above (see Table 1). Figure 5 shows the introductory model that was given in the posttest. Figure 6 shows examples of a simple declarative item and a complex create item.

The tests were scored by giving participants 0–1 point for each item. Partial credit was given for partly correct answers. The maximum score on the harmony test was 25. The maximum score on the black sphere test was 24.

In order to ensure equivalence in test circumstances between conditions, the models in the test were not represented in the system dynamics notation used in the modeling tool, so that students in the modeling condition would not experience an advantage. Instead, a causal concept map notation was used. Variables were represented by circles labeled with a variable name, causal relations were represented by arrows, and the quality of the relation was expressed by a plus or minus sign (see Figs. 3, 5).

Scoring Method

We developed a scoring scheme based on an analysis of the item responses of students at different levels of modeling

proficiency. An answer model was derived for each item, with elements defining the correct answer and elements representing common errors.

The expected answer for many items was the specification of a relation. In these cases we used a detailed scoring algorithm, giving points for the specification of the *existence* of a relation, the *direction* of a relation (causality), and the *quality* of a relation (positive or negative influence). A relation could be expressed not only textually in a written explanation, but also schematically in the drawing of a model. The threefold scoring of a relation provided a detailed view of the elaborateness of students’ reasoning.

Procedure

The experiment consisted of two sessions of 200 min each, at an interval of 2 or 4 weeks depending on the school program. The lessons were led by the experimenter, were additional to the regular curriculum and were compulsory for all students. Participants from one school were awarded course credit for their participation.

All participants attended an initial session of 150 min in which modeling was introduced, using examples on the spreading of diseases and a leaking water bucket. Following this session, participants had 50 min to complete the harmony pretest. For the second session, the students were divided into two groups, based on equal distribution of the harmony pretest modeling knowledge scores. We included all combinations of school, teacher, class, and gender for both conditions. In the second session, both conditions were given information and assignments about the factors influencing the temperature on Earth. In addition to the assignments, the students in the modeling condition ($N = 38$) performed an modeling task. The students in the expository condition ($N = 36$) wrote a report on the factors in global warming. After 150 min all participants completed the black sphere posttest, which took 50 min.

Results

We computed analyses of variance with the pretest subscore as a covariate. In the analysis of black sphere test *subscores*, the corresponding pretest *subscores* were used as a covariate. For the declarative knowledge scale, the pretest scale was not comparable, and no covariate was used (see Table 2).

No significant main effect of condition on total score on the black sphere test was found, although there was a trend in favor of the modeling condition ($F(1, 72) = 2.972$, $p = .089$).

Fig. 3 The fictitious model of the harmony of the spheres that was given in the pretest

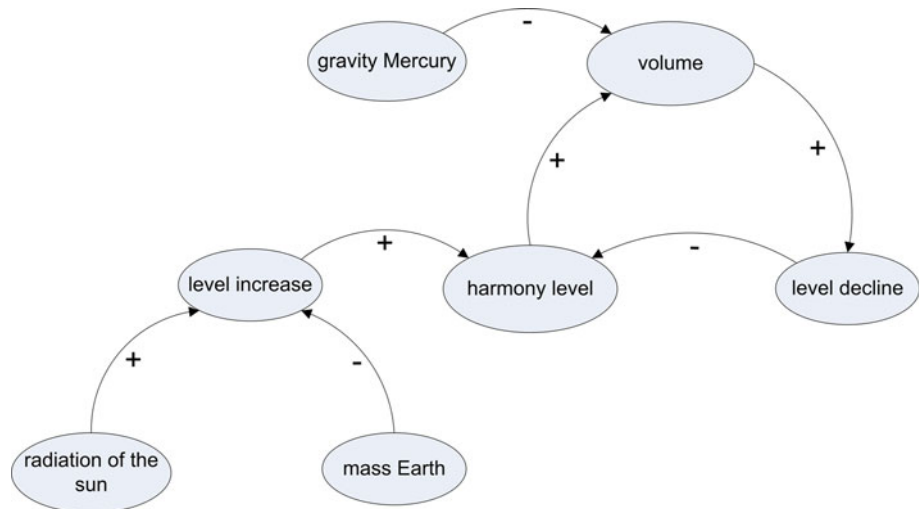


Fig. 4 Two examples of fictitious pretest items

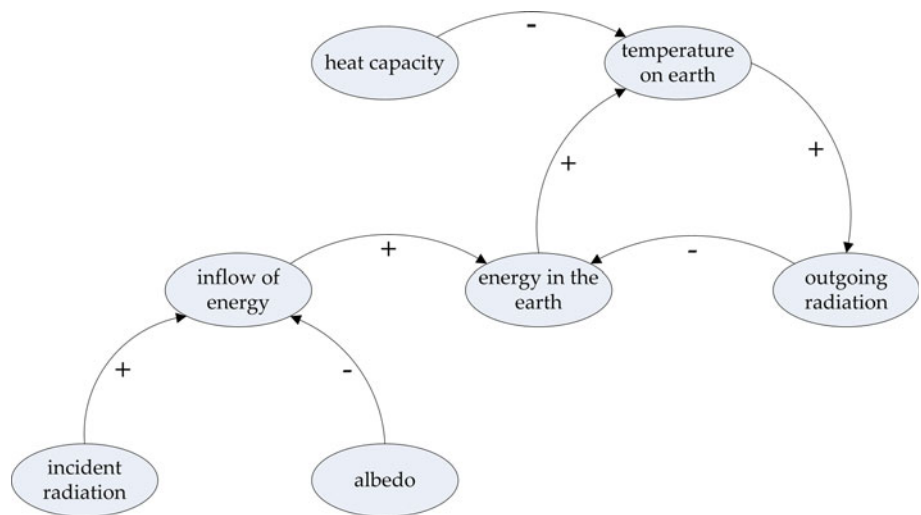
Application - simple
 Choose the right statement.

- A. The lower Mercury's gravity, the smaller the music volume.
- B. The lower Mercury's gravity, the larger the music volume.
- C. Mercury's gravity does not influence the music volume.

Explain your answer.

Evaluation - complex
 Is it correct that increasing Mercury's gravity will cause a faster decrease of the harmony of the spheres? Explain your answer.

Fig. 5 The black sphere model that was given in the global warming posttest



We expected differences on the subscales. We first took the subscales for the different skills (Remember declarative knowledge, Apply, Create, Evaluate). When looking at each scale overall (taking the simple and complex items together), we found no differences. When looking at the complex items across all subscales, we found a significant difference in favor of the modeling condition

($F(1, 72) = 8.780, p = .004, \text{partial } \eta^2 = .110$). More specifically, students in the modeling condition performed significantly better on both the complex declarative items ($F(1, 72) = 7.065, p = .010, \text{partial } \eta^2 = .089$) and the complex evaluation items ($F(1, 72) = 3.966, p = .050, \text{partial } \eta^2 = .053$). For the other subscales in the framework no significant differences were found (see Table 3).

Fig. 6 Two examples of black sphere posttest item

Declarative - simple

What is albedo (or reflectivity) of a substance?

Creation - complex

Suppose you have a watch running on solar cells. Just like a solar collector, the solar cells pick up the visible light of the sun, convert it to electricity, and store the electricity in the watch's battery. If you use the watch's stopwatch function often, the battery will empty earlier. Also, the intensity of the display's lighting is adjusted to the amount of energy in the battery. When the lighting is weak, you know the battery is almost empty.

Draw a model of this situation in order to be able to predict when the battery will be empty. Explain your drawing.

Table 2 Means and standard deviations of the harmony pretest (sub)scores for the two conditions

	Harmony pretest		
	Expository (<i>n</i> = 36)	Modeling (<i>n</i> = 38)	Max
Overall			
Simple	5.11 (1.39)	5.01 (1.45)	11
Complex	5.89 (2.27)	6.08 (2.46)	14
Total	11.00 (3.29)	11.09 (3.67)	25
Declarative			
Simple	0.38 (0.61)	0.31 (0.52)	3
Complex	0.91 (0.60)	0.75 (0.47)	3
Total	1.29 (0.91)	1.06 (0.81)	6
Application			
Simple	1.30 (0.50)	1.18 (0.65)	3
Complex	2.21 (1.22)	2.36 (1.14)	4
Total	3.51 (1.56)	3.54 (1.58)	7
Creation			
Simple	1.68 (0.57)	1.70 (0.48)	2
Complex	1.54 (0.79)	1.51 (0.99)	4
Total	3.23 (1.18)	3.21 (1.26)	6
Evaluation			
Simple	1.75 (0.71)	1.82 (0.59)	3
Complex	1.24 (0.92)	1.46 (0.95)	3
Total	2.99 (1.37)	3.28 (1.24)	6

Table 3 Means and standard deviations of the black sphere posttest (sub)scores for the two conditions

	Black sphere posttest		
	Expository (<i>n</i> = 36)	Modeling (<i>n</i> = 38)	Max
Overall			
Simple	6.59 (1.38)	6.58 (1.68)	12
Complex	3.67* (1.50)	4.72* (1.72)	12
Total	10.26 (2.48)	11.30 (3.10)	24
Declarative			
Simple	2.00 (0.80)	1.74 (0.84)	3
Complex	1.06* (0.71)	1.50* (0.69)	3
Total	3.06 (1.09)	3.23 (1.28)	6
Application			
Simple	1.04 (0.69)	1.21 (0.70)	3
Complex	0.90 (0.69)	1.12 (0.71)	3
Total	1.95 (1.16)	2.33 (1.18)	6
Creation			
Simple	2.09 (0.69)	2.07 (0.85)	3
Complex	1.16 (0.74)	1.26 (0.76)	3
Total	3.24 (1.34)	3.33 (1.48)	6
Evaluation			
Simple	1.46 (0.59)	1.56 (0.66)	3
Complex	0.54* (0.53)	0.85* (0.63)	3
Total	2.00 (0.79)	2.41 (0.99)	6

* Means differ at *p* < .05 in the analysis of variance

Discussion

The aim of this study was to investigate the specific learning outcomes of computer modeling compared to expository instruction. Although no significant overall differences in posttest scores between the two conditions occurred, clear differences were found with respect to the complex items. In line with our expectations, the modeling condition performed significantly better on the overall complex items. More specifically, the difference in performance concerned the complex evaluation items and the complex declarative items.

An explanation for modelers' better performance on the complex items is that the model created by learners in the modeling condition provides an overview of the complete model structure, allowing for a better integration of the various facts and relations that are present in the domain. This can also explain the unexpected advantage modelers had on the complex declarative items. Apparently, complex facts are not simply reproduced, but are reconstructed during the test. So, possibly because of a better developed ability for reasoning with the domain structure, the

modelers were better able to remember or reconstruct relevant facts in the domain.

Against our expectations, we found no differences related to the application and creation of models. The creation items in the posttest required the modeling of phenomena that were similar to the phenomena modelers had practiced with. We expected the modelers to be able to perform well on these items with similar model structures. Explanations for this unexpected lack of difference include the amount of time available for the modeling activity, which could have been too brief for a difference to emerge, and the possibility that the actual behavior by students engaged in the modeling could have been ineffective. This would be the case when learners merely copied their models from given examples rather than creating models from scratch. For instance, a common error for the modelers during the second session was to omit the temperature variable from the models they created. Apparently, the modelers copied the familiar model structures superficially instead of reasoning and experimenting with the model and discovering mistakes with respect to the new context. Ideally, the modelers had the opportunity to learn from their mistakes by receiving feedback from the simulation of their model, as opposed to the expository learners who did not receive feedback.

Relational reasoning seems to be an important factor in creating and evaluating a model. Applying knowledge of a model is not obviously involved in creating a relation. In this study, the participants were creating relations, but seemed not to learn how to reason with them. It is worthwhile to further investigate how the acquisition of creation skills can be supported and how the support for the different parts of creation skills can be implemented in the instruction.

In conclusion, computer modeling appears to result in qualitatively different learning outcomes between modeling and expository instruction. Differences arose in reasoning with complex knowledge structures, with respect to remembering complex conceptual knowledge and evaluating models. Proper tests with relevant subscales can reveal the differences in knowledge that can be acquired using a particular teaching method. The test introduced here serves as an example. As a consequence, the discussion on the benefits and drawbacks of constructivist teaching methods such as inquiry learning and modeling, as triggered by Kirschner and others (Kirschner et al. 2006; Klahr and Nigam 2004; Mayer 2004), can gain depth by devising such tests to address specific effects on specific types of knowledge.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alfieri L, Brooks PJ, Aldrich NJ, Tenenbaum HR (2010) Does discovery-based instruction enhance learning? *J Educ Psychol* 103(1):1–18
- Assaraf OBZ, Orion N (2005) Development of system thinking skills in the context of earth system education. *J Res Sci Teach* 42:518–560
- Booth Sweeney L, Sterman JD (2007) Thinking about systems: student and teacher conceptions of natural and social systems. *Syst Dyn Rev* 23:285–311
- Buckley BC, Gobert JD, Kindfield ACH, Horwitz P, Tinker RF, Gerlits B et al (2004) Model-based teaching and learning with *biologicaTM*: what do they learn? How do they learn? How do we know? *J Sci Educ Technol* 13:23–41
- Clement J (2000) Model based learning as a key research area for science education. *Int J Sci Educ* 22:1041–1053
- Cronin MA, Gonzalez C (2007) Understanding the building blocks of dynamic systems. *Syst Dyn Rev* 23:1–17
- Doerr HM (1997) Experiment, simulation and analysis: an integrated instructional approach to the concept of force. *Int J Sci Educ* 19:265–282
- Fiddick L, Cosmides L, Tooby J (2000) No interpretation without representation: the role of domain-specific representations and inferences in the wason selection task. *Cognition* 77:1–79
- Fretz EB, Wu HK, Zhang BH, Davis EA, Krajcik JS, Soloway E (2002) An investigation of software scaffolds supporting modeling practices. *Res Sci Educ* 32:567–589
- Geier R, Blumenfeld PC, Marx RW, Krajcik JS, Fishman B, Soloway E et al (2008) Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *J Res Sci Teach* 45:922–939
- Glymour CN, Cooper GF (eds) (1999) *Computation, causation, and discovery*. American Association for Artificial Intelligence Press, Menlo Park
- Hagmayer Y, Waldmann MR (2000) Simulating causal models: the way to structural sensitivity. In: *Proceedings of the twenty-second annual conference of the cognitive science society*. pp 214–219
- Hickey DT, Kindfield ACH, Horwitz P, Christie MA (1999) Advancing educational theory by enhancing practice in a technology-supported genetics learning environment. *J Educ* 181:25–55
- Hmelo CE, Holton DL, Kolodner JL (2000) Designing to learn about complex systems. *J Learn Sci* 9:247–298
- Hmelo-Silver CE, Marathe S, Liu L (2007) Fish swim, rocks sit, and lungs breathe: expert-novice understanding of complex systems. *J Learn Sci* 16:307–331
- Hogan K, Thomas D (2001) Cognitive comparisons of students' systems modeling in ecology. *J Sci Educ Technol* 10:319–344
- Jacobson MJ, Wilensky U (2006) Complex systems in education: scientific and educational importance and implications for the learning sciences. *J Learn Sci* 15:11–34
- Kirschner PA, Sweller J, Clark RE (2006) Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ Psychol* 41:75–86
- Klahr D, Nigam M (2004) The equivalence of learning paths in early science instruction. *Psychol Sci* 15:661–667
- Lederman J, Lederman N, Wickman P-O, Lager-Nyqvist L (2007) An international, systematic investigation of the relative effects of inquiry and direct instruction. Paper presented at the ESERA
- Lynch S, Kuipers J, Pyke C, Szesze M (2005) Examining the effects of a highly rated science curriculum unit on diverse students: results from a planning grant. *J Res Sci Teach* 42:912–946

- Magnani L, Nersessian NJ, Thagard P (eds) (1998) Model-based reasoning in scientific discovery. Kluwer Academic/Plenum Publishers, New York
- Mandinach EB (1989) Model-building and the use of computer-simulation of dynamic-systems. *J Educ Comput Res* 5:221–243
- Mandinach EB, Cline HF (1996) Classroom dynamics: the impact of a technology-based curriculum innovation on teaching and learning. *J Educ Comput Res* 14:83–102
- Mayer RE (2004) Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *Am Psychol* 59:14–19
- Penner DE (2001) Cognition, computers, and synthetic science: building knowledge and meaning through modelling. *Rev Res Educ* 25:1–37
- Quadrat-Ullah H (2010) Perceptions of the effectiveness of system dynamics-based interactive learning environments: an empirical study. *Comput Educ* 55:1277–1286
- Rittle-Johnson B, Star JR (2007) Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *J Educ Psychol* 99: 561–574
- Sao Pedro M, Gobert JD, Raziuddin JJ (2010) Comparing pedagogical approaches for the acquisition and long-term robustness of the control of variables strategy. In: Proceedings of the international conference on the learning sciences. Chicago, IL, pp 1024–1031
- Sins PHM, Savelsbergh ER, van Joolingen WR (2005) The difficult process of scientific modelling: an analysis of novices' reasoning during computer-based modelling. *Int J Sci Educ* 27:1695–1721
- Sterman JD (1994) Learning in and about complex systems. *Syst Dyn Rev* 10:291–330
- Sterman JD (2002) All models are wrong: reflections on becoming a systems scientist. *Syst Dyn Rev* 18:501–531
- Stratford SJ, Krajcik J, Soloway E (1998) Secondary students' dynamic modeling processes: analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *J Sci Educ Technol* 7:215–234
- van Borkulo S, van Joolingen WR, Savelsbergh ER, de Jong T (2008) A framework for the assessment of learning by modeling. In: Blumschein P, Stroebel J, Hung W, Jonassen D (eds) Model-based approaches to learning. Sense Publishers, Rotterdam, pp 179–195
- van Joolingen WR, de Jong T, Lazonder AW, Savelsbergh ER, Manlove S (2005) Co-lab: research and development of an online learning environment for collaborative scientific discovery learning. *Comput Hum Behav* 21:671–688
- Wells M, Hestenes D, Swackhamer G (1995) A modeling method for high-school physics instruction. *Am J Phys* 63:606–619
- Wilensky U, Resnick M (1999) Thinking in levels: a dynamic systems approach to making sense of the world. *J Sci Educ Technol* 8:3–19