


RESEARCH

Open Access



Comparison between behavioral and structural explanation in learning by model-building

Tomoya Horiguchi^{1*} , Tetsuhiro Masuda², Takahito Tomoto³ and Tsukasa Hirashima⁴

* Correspondence: horiguti@maritime.kobe-u.ac.jp

¹Graduate School of Maritime Sciences, Kobe University, 5-1-1, Fukaeminami, Higashinada, Kobe, Hyogo 658-0022, Japan
Full list of author information is available at the end of the article

Abstract

In science education, building models of dynamical systems is a promising method for understanding various natural phenomena and artifacts with scientific concepts. It is, however, difficult to learn skills and concepts necessary for modeling. Though several model-building learning environments (MBEs) have been developed with potentially useful methods for assisting students, the verification of them has been limited so far. Most studies evaluated their effectiveness by measuring the degree of model completion by students, or total learning effect that consists of several types of assistance. In this study, we investigated how students learn modeling skills and concepts of system dynamics through modeling dynamical systems, focusing on how students' behavior and understanding are influenced by the type of assistance and students' prior knowledge. We implemented the function that detects the difference of a model by students from the correct model and gives one of the two types of feedback: *structural explanation* indicates structurally erroneous parts of a model by students to promote students' model completion, while *behavioral explanation* suggests erroneous behavior of a model by students to promote students' understanding about the cause of error. Our experiment revealed the following: (1) Students assigned to structural explanation showed high model completion, but their understanding depended on whether they used the feedback appropriately or not. (2) Students assigned to behavioral explanation showed less model completion, but once they completed models, they acquired a deeper understanding.

Keywords: Learning by modeling, Model-building learning environment, System dynamics, Adaptive feedback, Behavioral explanation and structural explanation

Introduction

The purpose of this study is to investigate how students learn modeling skills and concepts of system dynamics through modeling dynamical systems, focusing on how students' behavior and understanding are influenced by the type of assistance and students' prior knowledge.

In science education, building models of dynamical systems is a promising method for understanding various natural phenomena and artifacts with scientific concepts. Learning to formulate, test, and revise models is crucial for understanding science. Supporting students in articulating models and refining them through experience and reflection leads them to a deeper, systematic understanding of science (Collins, 1996).

Therefore, several model-building learning environments (MBEs) have been developed in which students are given a set of model components and build models of dynamical systems by combining them (Biswas, Leelawong, Schwartz, & Vye, 2005; Bravo, van Joolingen, & de Jong, 2006; Bredeweg, Linnebank, Bouwer, & Liem, 2009; Forbus, Carney, Sherin, & Ureel II, 2005; isee systems, 1985; Vanlehn, Wetzel, Grover, & Van De Sande, 2016). They can also simulate their model to see whether it behaves as they expected. If it does not, they modify the model and try simulation again.

It is, however, a difficult task for most students to build correct models in MBEs. They often have difficulty in building the basic structure of models, combining components appropriately, and connecting the representation of models to their behavior (Bravo et al., 2006; Bredeweg et al., 2009; Forbus et al., 2005; Gracia et al., 2010). Therefore, several methods for assisting students have been implemented in MBEs. For example, some MBEs enable qualitative modeling and simulation based on qualitative reasoning technique (Bredeweg et al., 2009; Forbus et al., 2005). Others have a help system that explains mathematical/physical concept model components stand for (Forbus et al., 2005; Gracia et al., 2010), a syntax checker of models (Forbus et al., 2005), a function that detects the difference between the model by students and the correct model (Bravo et al., 2006; Gracia et al., 2010), and a function that gives causal explanation for models' unexpected behavior (Beek & Bredeweg, 2012a; Beek & Bredeweg, 2012b).

However, the verification of these methods' usefulness has been limited so far (Bravo et al., 2006; Forbus et al., 2005; Gracia et al., 2010; Vanlehn et al., 2016). Though some types of students' interesting behavior were reported (e.g., students were passive about using assistance, or modified models in an ad hoc way) (Bravo et al., 2006; Gracia et al., 2010), most studies evaluated their effectiveness by measuring the degree of model completion by students (Bravo et al., 2006, Gracia et al., 2010) or the total learning effect that consists of several types of assistance (Vanlehn et al., 2016). Few studies investigated the relation among the type of assistance, students' understanding, their behavior, and prior knowledge in detail. Especially, it is unclear what kind of knowledge was acquired through learning by modeling. That is, was the knowledge students acquired sufficiently generalized to be applied to other tasks, or was it based on memorization and task-dependent?

In this study, we examined such an issue by comparing the effect of two types of assistance: one aims at promoting students' model completion and the other aims at promoting students' reflection on their model. Through an experiment, we investigated how students used each type of assistance to complete models, what kind of knowledge they acquire through modeling, and how their prior knowledge influenced learning. Because most of the assistance in the current MBEs can be classified into these two types and they are often in the relation of trade-off (we discuss this issue in the "[Assistance in the current systems](#)" and "[Function for assistance](#)" sections), our findings would contribute to designing functions for assistance in MBEs that adaptively use these types of assistance according to learning contexts, students' characters, and prior knowledge.

Related work

Purpose of learning by modeling

VanLehn et al. classified the purpose of educational use of MBEs as follows (Vanlehn et al., 2016): (a) to deepen domain knowledge (i.e., fundamental principles and

concepts) by building models that involve them (e.g., learning the principle of Newtonian dynamics by building the models of falling blocks), (b) to understand a particular system (e.g., learning global warming by building its submodels), (c) to understand the role of models in science, and (d) to learn modeling skills and concepts necessary for building models (rather than domain knowledge). In the field of intelligent educational systems and learning sciences, the purpose (c) has been often focused on (Bredeweg et al., 2013; Bredeweg & Forbus, 2003; Marx et al., 2004), while the purpose (d) is a prerequisite for using model building for the other purposes (Bravo et al., 2006; van Borkulo, van Joolingen, Savelsbergh, & de Jong, 2012; Vanlehn et al., 2016). However, since STELLA (which is the first modeling and simulation software for system dynamics developed by Barry Richmond and distributed by isee systems inc.) pioneered the educational use of model building (isee systems, 1985), many studies have reported it was difficult for students to build models except very basic or simple models (Beek & Bredeweg, 2012a, Beek & Bredeweg, 2012b, van Borkulo et al., 2012, Bravo et al., 2006, Bredeweg et al., 2013, Forbus et al., 2005, Gracia et al., 2010, Marx et al., 2004, Vanlehn et al., 2016).

In this study, therefore, aiming at the purpose (d), we focus on the learning of modeling skills and concepts of system dynamics through modeling dynamical systems. That is, based on Hopper & Stave (2008), the target ability is defined as follows: (1) to differentiate stocks, flows, and other parameter types and recognize their local connections and emergent phenomena (i.e., local behavior); (2) to identify feedback and recognize emergent phenomena (i.e., global structure and behavior); (3) to understand and explain dynamic behavior of the whole system, especially important behaviors (such as equilibrium); and (4) to use conceptual models to explain the effect of parameter manipulation (i.e., change of condition) on the behavior of systems.

Assistance in the current systems

Several methods for assisting students in MBEs have been implemented (Beek & Bredeweg, 2012a; Beek & Bredeweg, 2012b; Bravo et al., 2006; Bredeweg et al., 2009; Forbus et al., 2005; Gracia et al., 2010; Vanlehn et al., 2016). Though their purposes and target students are wide-ranged, here, we focus on those that aim the learning of modeling skills and concepts of system dynamics through the modeling of dynamical systems.

Bravo et al. developed a function that enumerates the differences between a model by students and the correct model by a teacher (i.e., the differences are the erroneous parts in the model by students) and gives advice on them (Bravo et al., 2006). The content of advice is adaptively controlled depending on the progress of models, in which errors in parameter types, errors in parameters dependency, and other errors are indicated. Experiments revealed most advices were valid and contributed to the completion of models. However, the effect on students' understanding of dynamical systems was not measured.

Bredeweg et al. developed several functions that give intelligent feedback about the errors of models in their Garp3/DynaLearn project (Beek & Bredeweg, 2012a; Beek & Bredeweg, 2012b; Gracia et al., 2010), for example, a function that indicates the components lacking/unnecessary in the models by students and a function that gives causal explanation about unexpected behavior of models. Experiments were made to evaluate some of these functions, which clarified some types of students' interesting behavior in

building models: students were passive about using the functions so that their thinking would not be interrupted, or would not correct errors in spite of appropriate suggestion. However, the effect on students' understanding of dynamical systems was not clear.

VanLehn et al. implemented several types of assistance in their MBE called Dragoon (Vanlehn et al., 2016), for example, a function that shows the difference between the behavior of models by students and the correct behavior and a function that guides students' model building (immediate feedback is provided about erroneous input) and gives advice on what to do next. An experiment in classroom revealed students who learned modeling with these functions acquired better understanding of dynamical systems. It also gave an interesting suggestion about the progress of students' modeling ability. However, this experiment evaluated the total learning effect that consists of several types of assistance by the functions in their MBE (i.e., students could use different types of assistance together with). It is, therefore, not clear how each function influenced students' behavior and understanding.

Preceding studies have revealed students' several inappropriate behaviors in building models. In addition to the above behaviors, students often modified models in an ad hoc way or overused assistances to complete models without understanding why their models were erroneous. It is, therefore, necessary to clarify what types of assistance cause what types of students' behavior, how they influence students' understanding, and what kind of influence students' prior knowledge has. These factors should be related to each other to investigate the effect of learning by modeling.

Evans: a model-building learning environment

Outline

In this study, we used a model-building learning environment called Evans we have been developing (Horiguchi, Hirashima, & Forbus, 2012; Horiguchi & Masuda, 2017). In Evans, students can build qualitative models of dynamical systems and observe its behavior by qualitative simulation (Weld & de Kleer, 1990). A set of model component classes are provided that stand for basic concepts of qualitative reasoning, such as object, quantity (constant or variable), proportional relation, integral relation, qualitative operator, corresponding values, and so on. Students instantiate these classes to make model components and combine them into a model. In qualitative modeling and simulation, it is possible to assign qualitative values to quantities and to deal with systems with incomplete quantitative information. The framework of Evans is based on QSIM (Kuipers, 1986) that is one of the most popular methods for qualitative modeling and simulation. We elaborate some important components/concepts used in Evans/QSIM:

1. Quantity (variable or constant): it stands for a quantitative attribute of an object (e.g., the amount of water in a bathtub). The value of a quantity consists of its amount and derivative both of which are represented qualitatively. For example, the amount of water in a bathtub can be *zero*, $[zero, amt_{ini}]$, amt_{ini} , $[amt_{ini}, amt_{max}]$, or amt_{max} ($[a, b]$ means the interval between a and b ; amt_{ini} and amt_{max} are the initial and maximum amount of water respectively that are qualitatively important values and are called "landmarks"). The derivative can be "+," "0," or "-" (the sign of derivative of the quantity).

2. Qualitative proportional relation (P+/P-): if the amount of a quantity x increases (decreases) whenever that of another quantity y increases (decreases), they are connected with qualitative proportional relation “P+ (P-).” The degree of the relation (e.g., linear or quadratic) is not considered.
3. Qualitative integral relation (I+/I-): if the amount of a quantity x increases/decreases/is steady (decreases/increases/is steady) whenever that of another quantity y is positive/negative/zero, they are connected with qualitative integral relation “I+ (I-).” The rate of integration is not considered.
4. Qualitative operator: if a quantity z is the sum of the other quantities x and y , their relation is represented as “ADD(x, y, z)” (other arithmetic relations are similarly defined). Note that since the values of quantities are qualitative, the result of calculation sometimes becomes ambiguous. For example, when x is positive and y is negative, the sum of them can be any of positive, negative, and zero.

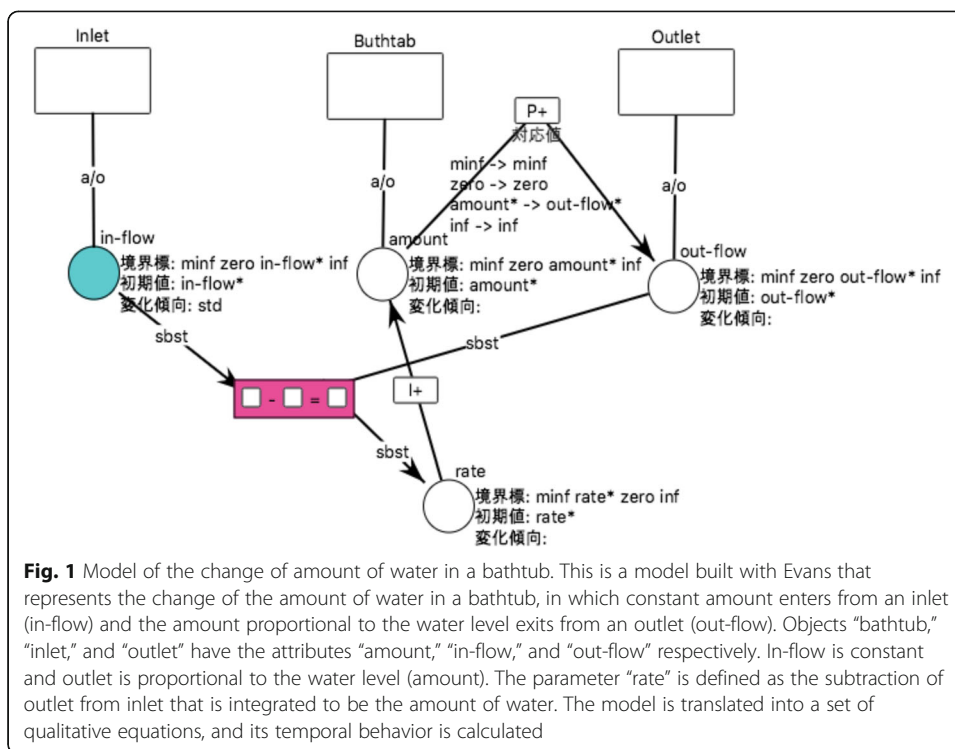
A qualitative state of a system is the set of qualitative values of all quantities. If any qualitative value does not change, the qualitative state of the system does not change (even if physical time progresses). When any qualitative value changes, the qualitative state of the system changes to another (called state transition). In Evans/QSIM, by using a set of qualitative constraints between quantities (i.e., the model of a system) and a set of state transition rules (Weld & de Kleer, 1990), the changes of the qualitative state of the system are inferred along time (i.e., qualitative simulation of the behavior).

Figure 1 shows a model built with Evans that represents the change of the amount of water in a bathtub (in which constant amount enters from an inlet and the amount proportional to the water level exits from an outlet). The model is translated into a set of qualitative equations, and its temporal behavior is calculated. Figure 2 shows one of the possible behaviors in which the amount of water gradually decreases to become steady (i.e., input and output are in equilibrium). The behavior is represented as a sequence of qualitative states (each of which is a set of qualitative values of amounts of the system). When sufficient information is not provided for determining the unique behavior, all of the possible behaviors are enumerated because of ambiguity. In this example, depending on the initial condition, the amount of water could gradually increase to become steady or become almost zero.

Function for assistance

In order for learning in MBE to be beneficial, students need to understand the mathematical/physical roles and usage of model components and build at least syntactically correct (i.e., calculable) “initial models.” Most of the current MBEs provide functions for assisting this. Evans also has such functions as a help system that explains the mathematical/physical roles and usage of model components on students’ demand and a syntax checker that manually/automatically detects syntactic errors in models (a preliminary test revealed these functions helped students build syntactically correct models (Horiguchi & Masuda, 2017)).

In addition, in order for simulation (i.e., calculated behavior of models) to provide useful information for testing a model, students need to build a model with a certain

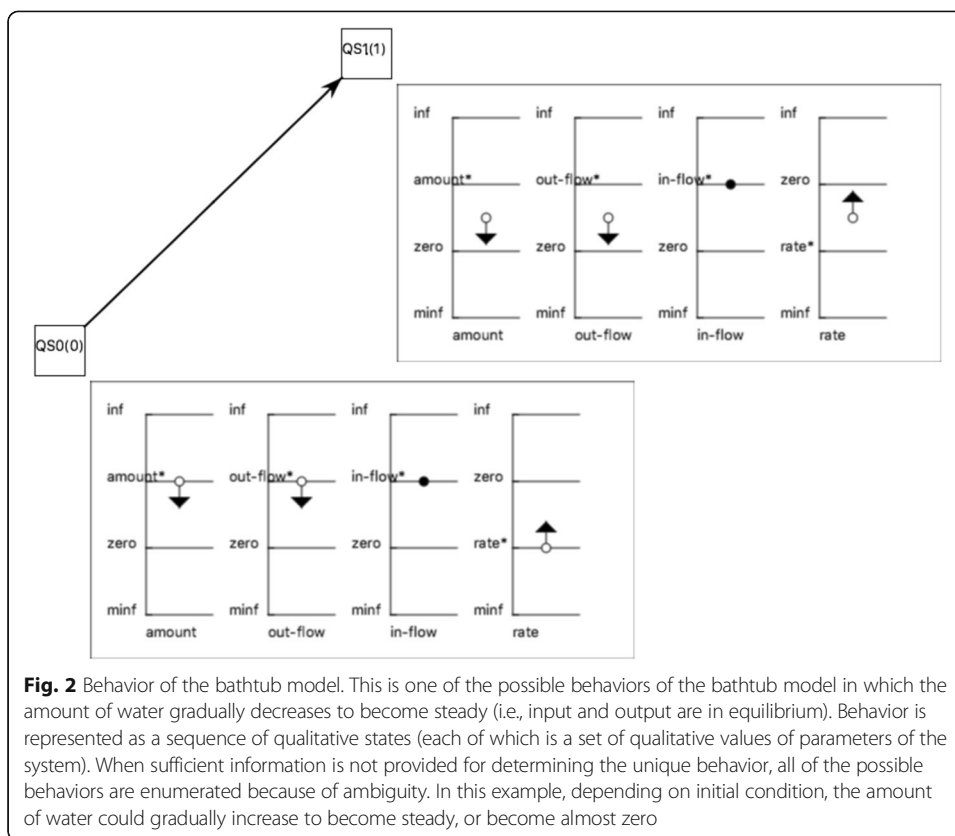


degree of completion that includes a certain number of constraints on models' behavior. It is, however, reported students often build a syntactically incorrect or very "sparse" initial model that includes few constraints (Gracia et al., 2010). We think this is one of the reasons why many MBEs provide functions that directly guide students in model building (and are evaluated by the degree of model completion). Even if students initially build a model without understanding, it is expected they improve their understanding through repeated observation of simulation and modification. There is, however, still a possibility that such functions are overused by students whose concern is the completion of models.

On the other hand, some MBEs provide functions that less directly assist students (e.g., by suggesting the cause of unexpected behavior of models). Such functions promote students' reflection that would deepen their understanding. However, it does not necessarily lead to the correction of errors, that is, there is a possibility students cannot receive useful feedback because of the low degree of model completion.

There is a tradeoff between these two types of assistance, and as to each, several types of students' appropriate and inappropriate behavior have been reported. However, few studies have compared these two types of assistance considering the relation with students' behavior, model completion, and the effect of learning.

In this study, therefore, through the comparison of these two types of assistance, we investigate the relation among the types of assistance, students' behavior, the degree of their model completion, and their understanding. In Evans, we implemented a function called "difference-list" that compares a model by students and the correct model by teacher and enumerates their differences (i.e., erroneous parts of the former). Difference-list provides students one of the two types of explanation about the differences: (1) *Structural explanation* aims at the increase of model completion that simply



indicates structural differences from the correct model (e.g., lacking/unnecessary amounts, reverse direction of a relation between amounts). For example, as to the bathtub model in Fig. 1, suppose a student erroneously connected the variables “out-flow” and “amount” with “I+ (integral relation)” link instead of correctly connecting them with “P+ (proportional relation)” link. In structural explanation, the following explanation is shown to the student: “Though ‘out-flow’ and ‘amount’ are connected with ‘I+’ link in your model, that is incorrect. Consider removing the link or replacing it with another link.” According to such indication, students can easily (i.e., without understanding the cause of errors) increase the degree of their model completion. (2) *Behavioral explanation* aims at the correction and reflection on semantic errors in students’ models that indicates the unnaturalness of models’ behavior caused by the errors (e.g., when students’ model lacks a “promotional” relation between two amounts, it indicates one of them does not necessarily increase even if the other increases). For example, as to the bathtub model in Fig. 1, suppose a student erroneously connected the variables “out-flow” and “amount” with “I+ (integral relation)” link instead of correctly connecting them with “P+ (proportional relation)” link. In behavioral explanation, the following explanation is shown to the student: “In your model, ‘amount’ increases whenever ‘out-flow’ is positive. Is it true?” In order to correct models according to such indication, students need to understand the relation between the structure and the behavior of models, which would promote their reflection on the cause of semantic errors. If the effects of these types of assistance are clarified, we can extend difference-list to a function

that provides adaptive feedback according to students' progress, understanding, character, and prior knowledge.

Experiment

Design

Purpose

In learning by modeling, through the comparison between students who received feedback that promotes the completion of models (structural explanation group) and students who received feedback that promotes reflection on the cause of errors (behavioral explanation group), we investigate how their behaviors, the degree of their model completion, and their understanding of dynamical systems differ (or the same).

Hypotheses

Based on the discussion in the “[Function for assistance](#)” section, we built up the following two opposite working hypotheses. As we discussed, structural explanation directly teaches how to correct the models (called direct instruction) while behavioral explanation encourages to reflect on the errors (called indirect instruction). In learning sciences literature, indirect instruction has been traditionally believed more effective for deeper understanding (Chi, Bassok, Lewis, Reimann, & Glaser, 1989, for example). Therefore, the following hypothesis is built.

Hypothesis 1: Students in the behavioral explanation group improve their understanding better than the students in structural explanation group.

On the other hand, recent literature has pointed out the effect of indirect instruction is learning-context dependent, and direct instruction is sometimes more effective (Klahr, 2009, for example). Additionally, as we discussed, the completion of models by students is important in improving their understanding. In MBE, students cannot receive useful feedback for learning until they build models with a certain degree of completion. Such memorized experience would work well especially immediately after the learning (i.e., in post-test). Because structural explanation's promotion of model completion is expected to be stronger than behavioral explanation's, the following hypothesis is built.

Hypothesis 2: Students in the structural explanation group complete the models with more degree than students in the behavioral explanation group. The more degree students complete the models with, the better they improve their understanding (especially immediately after the learning).

Subjects

We recruited 17 graduate and undergraduate students whose major was engineering for the experiment. We paid them the fee. Their age distributed from 20 to 24. Fifteen of them were male and two of them were female. Though they had finished the courses of basic physics and mathematics at university, they had no experiences of modeling dynamical systems in MBE.

Instruments

We used a set of teaching/testing materials in addition to Evans.

Evans It is a MBE introduced in the “[Outline](#)” section in which students can build qualitative models of dynamical systems with GUI and observe their behavior by qualitative simulation. As described in the “[Function for assistance](#)” section, a help system, a syntax checker, and the difference-list (either structural or behavioral explanation is provided) were implemented. Students could use these functions freely during modeling. Operations by students were recorded in log files.

Booklet for tutorial After brief introduction to dynamical systems, basic usage of Evans to make and test models was explained with some examples. Students practiced building models by using two simple tasks (i.e., simplified “bathtub systems” with only either inlet or outlet) in this booklet.

Modeling task We made students build the model of “bathtub system” in the “[Outline](#)” section by using Evans. As indicated in the “[Function for assistance](#)” section, students often build very “sparse” models with few constraints. In such a case, they cannot receive useful information from simulation nor functions for assistance. In this experiment, therefore, we gave students all the components necessary for building the model. They were made by disassembling the correct model by the experimenter (who was one of the authors), and necessary parameters (e.g., initial values) were input. Therefore, what students had to do was to assemble them into the complete model. Thus, we removed students’ load in making necessary components by analyzing the system structure (which is often difficult for students) in order to focus on the role of functions for assistance in completing models.

Test on system dynamics This is used for measuring students’ understanding of system dynamics described in the “Purpose of learning by modeling” section. It consists of problem-1 and problem-2 that deal with the “bathtub system” in the “[Outline](#)” section and a simple (RC) electric circuit respectively. Problem-1 is a “learning task” that deals with the same system as the modeling task students work on in Evans, while problem-2 is a “transfer task” that deals with an isomorphic system to that of modeling task. Problem-1 and problem-2 include eight and seven questions respectively that test the understanding explained in the “[Purpose of learning by modeling](#)” section. Sample questions are shown in Fig. 3. The questions ask either the local characteristics of model components (e.g., Q1 of both problems), the global behavior of the system (e.g., Q2 of both problems), or the change in system behavior due to the change of conditions (e.g., Q4 of problem-1 and Q5 of problem-2). This test was used as pre-, post-, and delayed post-test. All tests were written tests. Full marks are 37.

Procedure

The experiment was made in our laboratory that was sufficiently quiet. First, students worked on the written pre-test on system dynamics (about 20 min). Then, after a briefing on the outline of the experiment (5 min), they practiced building models by using the booklet (about 30 min). After that, they worked on the modeling task with Evans (modeling session). Eight of them were assigned to the structural explanation group (who received structural explanation from the difference-list), while nine of them were

Problem-1
Consider the amount of water in a bathtub. Constant amount is entering from the inlet per a unit of time and the amount proportional to the water level is exiting from the outlet per a unit of time. Suppose the capacity of the bathtub is not limited.

Q1. How does the amount of water exiting from the outlet per a unit of time change when the water level increases?

Q2. After a certain period of time passed, the change of the water level stopped. Why?

Q5. Suppose the water level first increased and then became constant. If the outlet was replaced by a bigger one and restart from the same water level,
(a) Is it possible the water level keeps constant from the beginning? Why?
(b) Is it possible the water level first increases and then becomes constant? Why?

Problem-2
Consider an electric circuit in which a battery, a resistor, a capacitor and a switch are wired in series. The battery supplies the constant voltage V_{in} . The voltage across the capacitor is proportional to the amount of electric charge in it. Suppose the capacitor was initially empty and the switch has been turned on.

Q1. How does the voltage across the capacitor change when the amount of electric charge in it increases?

Q2. After a certain period of time passed, the current flowing into the capacitor became zero. What value is the voltage across the capacitor at this time?

Q4. Suppose the capacitor was initially charged to a certain degree. When the current flowing into the capacitor became zero (after a certain period of time passed),
(a) Is it possible the amount of electric charge in the capacitor keeps constant from the beginning? Why?
(b) Is it possible the amount of electric charge in the capacitor first decreases and then becomes constant? Why?

Fig. 3 Test on system dynamics (extract). This is the sample of the test that was used for measuring students' understanding of system dynamics. It consists of problem-1 and problem-2 that deal with the "bathtub system" (learning task) and a simple electric circuit (transfer task) respectively. Problem-1 and problem-2 include eight and seven questions respectively. The questions ask either the local characteristics of model components, the global behavior of the system, or the change in system behavior due to the change of conditions. This test was used as pre-, post-, and delayed post-test. Full marks are 37

assigned to the behavioral explanation group (who received behavioral explanation from the difference-list). During the session, students could use the difference-list any-time to receive explanations on demand. Most students could complete the model around 20 min after they started modeling. Five students (one in the structural explanation group and four in the behavioral explanation group) who could not complete the model within 30 min were instructed to stop the modeling. Finally, they worked on the written post-test that was the same as pre-test (about 20 min). About a month later, they worked on the written delayed post-test that was the same as the preceding two tests (about 20 min).

Measure

The improvement of students' understanding of system dynamics was measured by the increase of scores between tests. The immediate effect of learning was measured by the increase between pre- and post-tests, while the effect of learning on knowledge generalization was measured by the increase between post- and delayed post-tests. The degree of students' model completion was calculated based on the degree of correspondence between the "final model" by students (i.e., the model at the end of modeling session) and the correct model (full marks are 3). In addition, the frequency of assistance (i.e., the number of times students used the difference-list per 1 min) was calculated by using log files. Because there is a possibility that the factors "explanation (structural/behavioral)" and "test (pre-/post-/delayed post-test)" influence each other, we used two-way mixed ANOVA instead of t test to analyze students' scores of tests (the scores of students in the same condition were averaged). For considering the influences of other factors (i.e., the degree of model completion and the frequency of using assistance), we also used correlation analysis.

Result and discussion

The answers of tests were marked as follows: For each question, 1 point was given to the correct answer. If a question asked the reason of the answer (i.e., “why?”), 2 points were added to the correct reason. In this way, the full marks were 37 (the full marks of problem-1 (learning task) and problem-2 (transfer task) were 20 and 17 respectively). Two experimenters (the first and second authors) first marked the answers of tests independently and then negotiated the final scores (Since the criteria for marking were clearly defined beforehand, most of the scores by them corresponded. In fact, Kappa statistics of pre-, post-, and delayed post-test were .9982, .9992, and .9992 respectively). They adopted the same procedure in marking the degree of model completion (The marking method is described in the previous subsection. Kappa statistic was .9992).

The average scores of tests and the result of statistical analysis are shown in Fig. 4 and Table 1. First, we conducted a *t* test between the pre-test score of the structural group and that of the behavioral group and found there was no significant difference between them ($t = -.8529$, $p = .4071$). That is, it was confirmed that the baselines (i.e., understandings before the learning) of the two groups were not different. Then, we conducted a two-way mixed ANOVA of 2 (explanation: structural/behavioral) \times 3 (test: pre-/post-/delayed post-test). Because the interaction of the factors was significant ($F = 3.315$; $p < .05$), we tested the simple main effect of each factor. As a result, the factor “explanation” was not significant while the factor “test” was significant (test(structural): $F = 23.783$; $p < .01$, test(behavioral): $F = 7.039$; $p < .01$). Multiple comparison revealed the following facts: In the structural explanation group, there were significant differences between pre- and post-test ($p < .01$) and between pre- and delayed post-test ($p < .01$). In the behavioral explanation group, there were significant differences between post- and delayed post-test ($p < .05$) and between pre- and delayed post-test ($p < .01$).

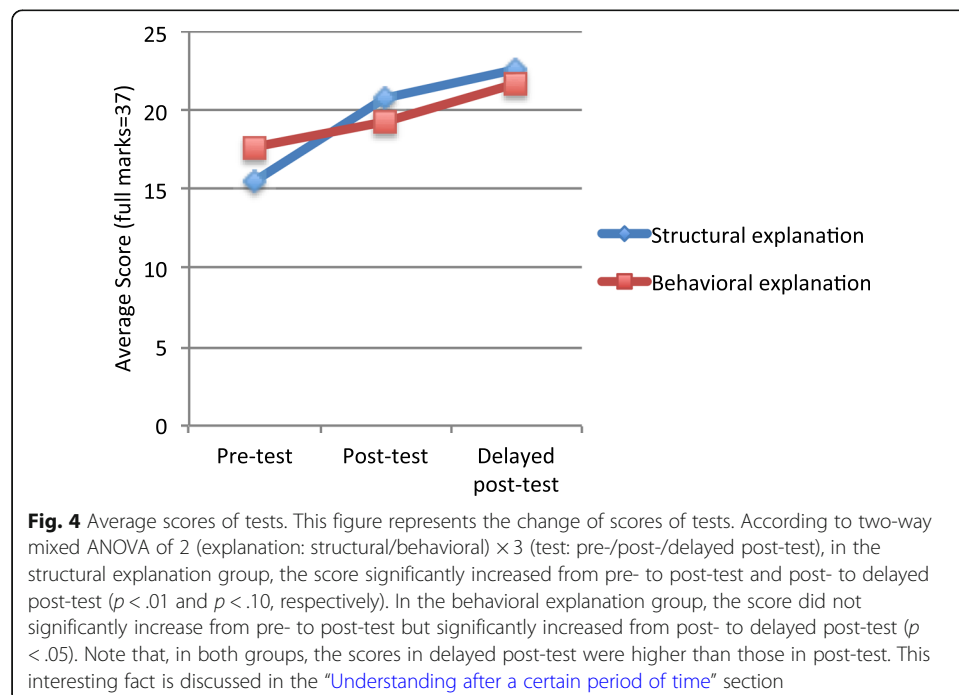


Table 1 Result of tests

	Pre-test	Post-test	Delayed post-test	Simple main effect of test	Increase between pre and post-test	Increase between post and delayed post-test	Increase between pre and delayed post-test
Structural explanation ($n = 8$)	15.50 (4.42)	20.75 (4.97)	22.63 (3.77)	$F = 23.783$ $p < .01^{**}$	$t = 4.764$ $p < .01^{**}$	$t = 1.701$ $p < .10^{\dagger}$	$t = 6.465$ $p < .01^{**}$
Behavioral explanation ($n = 9$)	17.67 (5.31)	19.33 (4.74)	21.67 (6.73)	$F = 7.039$ $p < .01^{**}$	$t = 1.604$ $p > .10$	$t = 2.246$ $p < .05^{*}$	$t = 3.850$ $p < .01^{**}$

$\dagger p < .10$, $*p < .05$ and $**p < .01$

This is the average scores of tests and the result of statistical analysis. In two-way mixed ANOVA of 2 (explanation: structural/behavioral) \times 3 (test: pre-/post-/ delayed post-test), because the interaction of the factors was significant, we tested the simple main effect of each factor. As a result, the factor "explanation" was not significant while the factor "test" was significant (test(structural): $F = 23.783$; $p < .01$, test(behavioral): $F = 7.039$; $p < .01$). Multiple comparison revealed the following facts: In the structural explanation group, there were significant differences between pre- and post-test ($p < .01$) and between pre- and delayed post-test ($p < .01$). In the behavioral explanation group, there were significant differences between post- and delayed post-test ($p < .05$) and between pre- and delayed post-test ($p < .01$).

Note that, in this experiment, there are some factors that could influence the scores of post- and delayed post-test such as the degree of model completion, the frequency of using assistance, and the items of tests. Though we here addressed the influence of these factors by using correlation analysis (see the next section), considering a two-factor ANCOVA is our important future work.

Understanding immediately after learning by modeling

According to Table 1, *Hypothesis 1* was not supported because the simple main effect of "explanation" was not significant. In addition, the fact that the increase of students' score between pre- and post-test was significant only in structural explanation group rather suggests the contrary to *Hypothesis 1*. (Note that it was confirmed the baselines of the two groups were not different.) Therefore, we need an explanation about why the understanding of students in the structural explanation group significantly improved immediately after learning while that in the behavioral explanation group did not. Here, *Hypothesis 2* gives a suggestion. According to *Hypothesis 2*, if the degree of model completion of students in structural explanation group is significantly greater, we can explain why only their score significantly increased between pre- and post-test.

For this purpose, we made correlation analysis between several factors (see Table 2). Though there was a marginally significant medium positive correlation ($p < .10$) between the increase of all students' score from pre- to post-test and the degree of their model completion ($R = 0.476$, see Table 2a), there was no significant difference in the degree of model completion between the two groups (U test, $p > .10$). However, the result of correlation analysis gave some interesting suggestions: In the behavioral explanation group, there was a medium positive correlation between the degree of model completion and the increase of score from pre- to post-test ($R = 0.412$, see Table 2c). Though this correlation is not significant ($p > .10$), it is at least considerably greater than that of the structural explanation group ($R = 0.183$, see Table 2b). That is, in the behavioral explanation group, model completion contributed to the improvement of students' understanding, while it did not in structural explanation group. In addition, in the behavioral explanation group, there was no correlation between the frequency of assistance (by difference-list) and the increase of score from pre- to post-test ($R =$

Table 2 Result of correlation analysis

	Pre-test	Increase between pre- and post-test	Degree of model completion	Frequency of assistance
(a) All students				
Pre-test	1.000	–	–	–
Increase between pre- and post-test	– 0.383	– 1.000	–	–
Degree of model completion	– 0.313	0.476 [†]	1.000	–
Frequency of assistance	0.152	– 0.101	0.442 [†]	1.000
(b) Structural explanation group				
Pre-test	1.000	–	–	–
Increase between pre- and post-test	– 0.066	1.000	–	–
Degree of model completion	– 0.728**	0.183	1.000	–
Frequency of assistance	0.145	– 0.550	0.022	1.000
(c) Behavioral explanation group				
Pre-test	1.000	–	–	–
Increase between pre- and post-test	– 0.475	1.000	–	–
Degree of model completion	– 0.128	0.412	1.000	–
Frequency of assistance	0.113	0.190	0.673**	1.000

These are the results of correlation analysis between several factors. Table 2a is of all students, Table 2b is of students in the structural explanation group, and Table 2c is of students in the behavioral explanation group. There was a medium positive correlation between the increase of all students' score from pre- to post-test and the degree of their model completion ($R = 0.476$, Table 2a). In the behavioral explanation group, there was a medium positive correlation between the degree of model completion and the increase of score from pre- to post-test ($R = 0.412$, Table 2c), while there was not in the structural explanation group ($R = 0.183$, Table 2b). In addition, in the behavioral explanation group, there was no correlation between the frequency of assistance and the increase of score from pre- to post-test ($R = 0.190$, Table 2c), while there was a medium negative correlation in the structural explanation group ($R = -0.550$, Table 2b)

[†] $p < .10$

** $p < .05$

0.190, see Table 2c), while there was a medium negative correlation in the structural explanation group ($R = -0.550$, see Table 2b). Though these correlations are not significant ($p > .10$), the former is at least considerably greater than the latter.

On the other hand, the following analysis suggested structural explanation's promotion of model completion was considerably stronger than behavioral explanation's. That is, as to all students, there was a weak negative correlation between the degree of model completion and the score of pre-test ($R = -0.313$, see Table 2a), and especially in the structural explanation group, the correlation was strong ($R = -0.728$, see Table 2b) and significant ($p < .05$). That is, the higher students' prior knowledge was, the lower their degree of model completion was. The reason was suggested by a log file analysis that revealed several students whose score of pre-test was high were trying to build an alternative correct model that was different from the correct model by the experimenter. Two models were equivalent, but the usage of "integral relation" components was different. In the structural explanation group, such students tended to modify the "integral relation" part they once built to complete the correct model by the experimenter guided by the assistance, while in the behavioral explanation group, they did not. That is, the ratio of such modification to all the modification during modeling was significantly greater in the structural explanation group (the average of the ratio was

.488) than in the behavioral explanation group (the average of the ratio was .170; t test, $t = 3.193$, $p < .05$). This fact suggests model completion of students with high prior knowledge was negatively influenced when the model they tried to build was different from the model by the experimenter (because the assistance was given based on the latter). Even in such a case, structural explanation still promoted model completion while behavioral explanation did not.

These analyses suggest a possibility that a certain number of students in the structural explanation group overused assistance to complete the models without understanding why their models were erroneous. Based on the above discussion, therefore, we can integrate *Hypothesis 1* and *Hypothesis 2* into the follow findings.

Findings:

1. As for students who were assisted by behavioral explanation, completing models contributed to improving their understanding. However, because behavioral explanation's promotion of model completion was relatively weak (there were a few students whose models were in low degree of completion), their scores between pre- and post-test did not significantly increase.
2. As for students who were assisted by structural explanation, the improvement of their understanding through model completion depends on how they utilized the assistance. Because structural explanation's promotion of model completion was strong (almost all students' models were in high degree of completion), some students appropriately utilize assistance to complete models with understanding, while there were a certain number of students who overused assistance to complete the models without understanding. (In this experiment, at least two students were identified as the former while two students were identified as the latter. The former students indicated a large increase of score from pre-test to post-test (almost double of the average), and their frequency of using assistance was below average. The latter students indicated only a small increase of score from pre-test to post-test (below average), and their frequency of using assistance was quite large (above double of the average as to one of them).)

Understanding after a certain period of time

In this experiment, the score of 82% of (all) students increased between post- and delayed post-test, and the increase was significant as to students in the behavioral explanation group (see Table 1).

In general, when learning effect is measured with tests, the score of delayed post-test usually decreases compared to that of post-test. This is easily understood as the attenuation of memory of learning. However, in studies on second-language learning, the increase of delayed post-test score is occasionally reported (Fukuda, 2016; Miles, 2014). Though the reason is not clearly verified, a possibility is indicated that explicit learning in which students are conscious of grammatical rules tends to be memorization-centered and its effect does not last long, while unconscious learning in which they are unconscious of grammar gradually forms generalized/conceptualized knowledge with time (Fukuda, 2016).

Though the reproducibility of this result should be carefully verified, we can suggest the following possibility at this point:

- (a) Learning by modeling in Evans contributed to the acquisition of not only declarative/explicit knowledge of a specific model and components of dynamical system, but also procedural/implicit knowledge of relation among components and relation between structure and behavior of models.
- (b) The acquired knowledge is not temporarily memorized but generalized one that is not easily lost with time.

Based on the above suggestion, the result of delayed post-test is explained as follows: Students in both groups showed good knowledge retention in delayed post-test because they acquired it in generalized form. Such knowledge, however, could not be acquired if students overused assistance. Therefore, the increase of students' score between post- and delayed post-test was not significant in the structural explanation group in which overuse of assistance tended to occur. (Note that this is just a suggestion because, according to ANOVA, the increase was not significantly different between groups.)

Contribution and implications

The results of this experiment not only have commonality with those of previous studies but also add some new findings. First, Bravo et al. showed that the advice about the differences between a model by students and the correct model by a teacher improved students' model completion (Bravo et al., 2006), which is supported by our result especially on the structural explanation group. Our contribution is to have clarified the fact that the model completion does not necessarily improve students' understanding about system dynamics and the reason of the fact. Second, inappropriate use of the assistance (i.e., overuse of structural explanation) observed in our experiment provides another example of students' inappropriate behavior in building models, some of which were reported by Bredeweg et al. (Beek & Bredeweg, 2012a; Beek & Bredeweg, 2012b; Gracia et al., 2010). Third, VanLehn et al. reported students' understanding about system dynamics improved through modeling by the combination of several types of assistance. They also reported the improvement was quite different depending on the person. We clarified the effect of each type of assistance (i.e., structural and behavioral explanation) on students' understanding and gave a suggestion about why the improvement was individually different considering the usage of assistance. Finally, the result of our experiment suggested that the understanding of students who received direct instruction (i.e., structural explanation) improved immediately after learning and that the understanding of students who received indirect instruction (i.e., behavioral explanation) improved after a certain period of time. These facts add a case to the discussion about the effect of direct and indirect instruction (Klahr, 2009) to be further analyzed.

We think these findings are useful for both educators in teaching modeling and researchers in designing the functions for assistance in MBE. By knowing the features of direct and indirect instruction in assisting modeling, they could choose appropriate feedback according to students' behavior (including the usage of assistance), progress of modeling (including the impasse), and prior knowledge. It could be possible to implement a function that provides adaptive feedback according to learning contexts.

Though the scale of our experimental data is currently small, the findings here could work at least as a case study or the “anchor” for comparison with other experimental data.

Concluding remark

In this study, we investigated how students’ behavior and understanding were influenced by the type of assistance and students’ prior knowledge in learning by modeling. As a result, students who received feedback that promotes the model completion (structural explanation) improved their understanding immediately after learning, while a certain number of them overused the assistance (i.e., their frequency of using assistance was quite large) to complete models without understanding. On the other hand, students who received feedback that promotes the reflection on the cause of errors (behavioral explanation) improved their understanding gradually with time.

The generality of the findings here is currently limited because the sample size was small. However, we think these findings are useful to a certain degree at least as a case study. We chose students who had basic knowledge of physics and mathematics as the subjects in this study. That is, they were typical target students to whom modeling skill for engineering is taught. Therefore, other students in modeling would be likely to show the similar behavior to those in this study. Even if other students showed different behavior, our findings could be used as the “anchor” for the comparison to analyze their behavior. Based on these accumulated findings, we could build a function that provides adaptive feedback according to students’ progress, understanding, characters, and prior knowledge. Our important future work is, therefore, to scale up the experimental data and to verify the reproducibility of this result for clarifying the process of acquisition of modeling skills and concepts of system dynamics.

Abbreviations

MBE: Model-building learning environment

Acknowledgements

Not applicable

Funding

This work is supported in part by Grant-in-Aid for Scientific Research (B) (No.16K12558) from the Ministry of Education, Science, and Culture of Japan.

Availability of data and materials

Not applicable

Authors’ contributions

This research was designed by THo and TM based on the discussion with the TT and THi. The system was primarily implemented by TM. The experiment was designed and conducted by THo and TM. The analyses were primarily conducted by the TM but were verified by all the co-authors. The first manuscript draft was written by THo but was jointly reviewed multiple times and complemented by all the co-authors. All the authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Graduate School of Maritime Sciences, Kobe University, 5-1-1, Fukaeminami, Higashinada, Kobe, Hyogo 658-0022, Japan. ²KEYENCE Corporation, 1-3-14, Higashinakajima, Higashiyodogawa, Osaka, Osaka 533-8555, Japan. ³Faculty of

Engineering, Tokyo Polytechnic University, 1583, Iiyama, Atsugi, Kanagawa 243-0297, Japan. ⁴Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan.

Received: 29 August 2018 Accepted: 7 March 2019

Published online: 18 March 2019

References

- Beek, W. and Bredeweg, B. (2012a) Context-dependent help for novices acquiring conceptual systems knowledge in DynaLearn. Proc of ITS12.
- Beek, W. and Bredeweg, B. (2012b) Providing feedback for common problems in learning by conceptual modeling using expectation-driven consistency maintenance. Proc of QR12.
- Biswas, G., Leelawong, K., Schwartz, D., & Vye, N. (2005). Learning by teaching: a new agent paradigm for educational software. *Applied Artificial Intelligence*, 19, 363–392.
- Bravo, C., van Joolingen, W. R., & de Jong, T. (2006). *Modeling and simulation in inquiry learning: checking solutions and giving intelligent advice simulation*, Vol82, Issue11 (pp. 769–784).
- Bredeweg, B., & Forbus, K. D. (2003). Qualitative modeling in education. *AI Magazine*, Winter, 2003, 35–46.
- Bredeweg, B., Liem, J., Beek, W., Linnebank, F., Gracia, J., Lozano, E., Wißner, M., Bühling, R., Salles, P., Noble, R., Zitek, A., Borisova, P., & Mioduser, D. (2013). DynaLearn - an intelligent learning environment for learning conceptual knowledge. *AI Magazine*, Winter, 2013, 46–65.
- Bredeweg, B., Linnebank, F., Bouwer, A., & Liem, J. (2009). Garp3 - workbench for qualitative modelling and simulation. *Ecological Informatics*, 4(5–6), 263–281.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, 13(2), 145–182.
- Collins, A. (1996). Design issues for learning environments. In S. Vosniadou, E. D. Corte, R. Glaser, & H. Mandl (Eds.), *International perspectives on the design of technology-supported learning environments* (pp. 347–361). Lawrence Erlbaum, Mahwah.
- Forbus, K. D., Carney, K., Sherin, B. L., & Ureel II, L. C. (2005). VModel: a visual qualitative modeling environment for middle-school students. *AI Magazine*, 26, 63–72.
- Fukuda, J. (2016) Impacts of awareness in incidental learning on acquisition of conscious/unconscious knowledge: a perspective from salience of form-meaning links. Doctoral dissertation of Nagoya University (in Japanese).
- Gracia, J., Liem, J., Lozano, E., Corcho, O., Trna, M., Gómez- Pérez, A., & Bredeweg, B. (2010). Semantic techniques for enabling knowledge reuse in conceptual modelling. In *Proceedings of ISWC2010* (pp. 82–97).
- Hopper, M., & Stave, K. A. (2008) Assessing the effectiveness of systems thinking interventions in the classroom. Session presented at the 26th international conference of the system dynamics society.
- Horiguchi, T., Hirashima, T. and Forbus, K.D. (2012) A model-building learning environment with explanatory feedback to erroneous models, Proc. of ITS 2012, pp.620–621.
- Horiguchi, T., & Masuda, T. (2017). *Evaluation of the function that detects the difference of learner's model from the correct model in a model-building learning environment*. Proc. of HCI2017 (LNCS, volume 10274) (pp. 40–49).
- isee systems (1985) STELLA, <http://www.iseesystems.com/>.
- Klahr, D. (2009). "To everything there is as season, and a time to every purpose under the heavens" what about direct instruction. In S. Tobias & T. H. Duffy (Eds.), *Constructivist theory applied to instruction: success or failure?* (pp. 291–310). Routledge, New York.
- Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence*, 29, 289–338.
- Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in the middle grades: assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, 41(10), 1063–1080.
- Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, 42, 412–428.
- van Borkulo, S. P., van Joolingen, W. R., Savelsbergh, E. R., & de Jong, T. (2012). What can be learned from computer modeling? Comparing expository and modeling approaches to teaching dynamic systems behavior. *Journal of Science Education and Technology*, 21, 267–275.
- Vanlehn, K., Wetzel, J., Grover, S., & Van De Sande, B. (2016). Learning how to construct models of dynamic systems: an initial evaluation of the dragoon intelligent tutoring system. *IEEE Transactions on Learning Technologies*, 10(2), 154–167.
- Weld, D. S., & de Kleer, J. (1990). *Readings in qualitative reasoning about physical systems*. Morgan Kaufmann, San Mateo.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com