J Sci Educ Technol (2015) 24:168–177
DOI 10.1007/s10956-014-9537-1

# Key Characteristics of Successful Science Learning: The Promise of Learning by Modelling

**Yvonne G. Mulder · Ard W. Lazonder · Ton de Jong**

Published online: 2 December 2014


**Abstract** The basic premise underlying this research is that scientific phenomena are best learned by creating an external representation that complies with the complex and dynamic nature of such phenomena. Effective representations are assumed to incorporate three key characteristics: they are graphical, dynamic, and provide a pre-specified outline of the domain. This study examined the impact of these characteristics on performance and learning. High school students first read an instructional text about glucose–insulin regulation and then created a representation of its content. Representations differed regarding the key characteristics such that the summary ($n = 15$), concept map ($n = 16$), model ($n = 23$), and outlined model ($n = 21$) all incorporated one additional characteristic compared to their precursor. Main results indicated learning effects in each of these four conditions. Furthermore, creating a model was found to enhance students' learning more than creating a concept map, and students who completed an outlined model were found to learn more than those who created a model from scratch. In conclusion, this study does not univocally verify the necessity of all key characteristics individually, but the results do show that a representational format that combines all key characteristics enhances learning more than other formats.

**Keywords** Learning by modelling · Representations · Biology · System dynamics

Y. G. Mulder (✉) · A. W. Lazonder · T. de Jong
Department of Instructional Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: y.g.mulder@utwente.nl

## Introduction

One plausible reason why most students consider science a difficult subject is that scientific phenomena generally consist of multiple components that are interrelated in intricate ways (Graesser et al. 1994). Complexity increases even further when the phenomena change dynamically over time. These so-called dynamic phenomena, or dynamic systems, are multidimensional which means that students need to learn the various individual components and their relations as well as the dynamic behaviour of the system as a whole (Wilensky and Resnick 1999). An example of such a complex dynamic system is the human glucose regulation system, a topic commonly found in many high school biology curricula. The human body constantly needs glucose, as it is the basic source of energy. However, the glucose level should stay within a narrow range as either too high or too low blood glucose levels can cause damage to nerves, blood vessels, and organs. Students need to learn about this complex phenomenon, on the level of the individual components and the complex relations of the system and on the level of the (dynamic) behaviour of the system as a whole. To comprehend all aspects of this multidimensional system, students have to identify the individual mechanisms that change the state of the glucose and insulin levels (e.g., glucose production, insulin secretion, and the hormone regulation processes) and also understand the dynamic behaviour of the system as a whole in stable situations (i.e., in homoeostasis) and in unstable situations like when eating a pizza or in case of disruptions like diabetes.

In line with constructivists' notions of learning, deep understanding of dynamic systems is likely to improve when students are actively engaged in their knowledge construction (cf. Hmelo-Silver et al. 2007; Kafai and

Resnick 1996; Mayer 2004; Palincsar 1998). Meaningful understanding presumably takes the form of internal mental replicas of the relations among objects of the learning material (Johnson-Laird 1983). Creating an external representation based on this mental picture requires students to actively engage with the learning material more than working with pre-defined representations (e.g., Cox 1999). There are different known formats for creating these external representations. Among these are creating a summary, a concept map, or a runnable computer model. There are also different features or key characteristics we can distinguish when we examine and compare these different formats. The effect of these key characteristics on task performance and learning is the main focus of the current paper.

A self-constructed summary is a commonly used form of external representation. Writing summaries has long since been found to improve comprehension (Armbruster et al. 1987; Wittrock and Alesandrini 1990) as it requires students to focus their attention on the important information in a text. Writing a summary also helps students to reflect on their own knowledge of the domain and, as a result, improves metacomprehension accuracy (Thiede and Anderson 2003). However, these affordances are not limited to creating a summary, but most likely apply to other representational formats as well (e.g., a concept map or a runnable computer model). Since these representational formats differ with regard to their affordances on learning (Suthers and Hundhausen 2003), this study examines whether and how the presence of three key characteristics in different representational formats promote meaningful learning of dynamic science phenomena.

First, as understanding science texts often requires students to draw causal inferences (Graesser et al. 1994), the representational format needs to focus on such relationships. Compared to text, a *graphical*[1] representation of a complex dynamic system can provide an overview of all the relevant components and their intricate relations within the system (e.g., Schnotz 2002). For instance, a concept map[2] is a node-link diagram that represents concepts and the relationships between these concepts (e.g., Novak 2005). Constructing concept maps therefore directs attention to concepts and their mutual relationships (Nesbit and Adesope 2006). When learning about complex dynamic phenomena from text, concept maps thus have the advantage that their graphical structure emphasizes the relationship between concepts (e.g., Hilbert and Renkl 2009). A variety of studies have demonstrated the effectiveness of creating concept maps as a learning method to enhance students' comprehension of the domain, as they encourage learner engagement and require students to explicitly identify concepts and the relations among them (e.g., Horton et al. 1993; Nesbit and Adesope 2006). The graphical nature is considered the first key characteristic of a representational format.

Second, the created representational format should do justice to the *dynamic* nature of the system. This dynamic nature is considered the second key characteristic. Simulating a model of a complex dynamic system provides insight in the behaviour of the system over time. For instance, it can show equilibrium states or growth. The increased use of computers in science and science education has enabled students to engage in modelling activities to construct their own dynamic representations (e.g., models; for an overview of learning by modelling see Fretz et al. 2002; Louca and Zacharia 2011; VanLehn 2013). Although modelling instruction can include conceptual models (both graphical and physical), in this paper, we reserve the term 'model' for a system dynamics computer model that is runnable and adheres to the system dynamics modelling formalism as proposed by Forrester (1961). The construction of models, much like the construction of a concept map, requires students to identify key components and create links that indicate how these components are related. However, a unique additional quality of models is that they offer students the opportunity to simulate their representation, whereas concept mapping tools do not offer this simulation facility. This provides students both feedback on the correctness of their created representation and insight into the dynamic behaviour of the entire system over time. A systematic comparison by Löhner (2005) showed that the ability to simulate models enhances students' reasoning of the behaviour of system as a whole.

Several learning environments offer modelling platforms. Some of the more well-known examples include STELLA (Steed 1992), Model-It (Jackson et al. 1994), Co-Lab (van Joolingen et al. 2005), more recently, Betty's Brain (Leelawong and Biswas 2008), and SCY-Lab (de Jong et al. 2010). Research on the representational format of models showed that novice learners benefit more from graphical modelling structures compared to textual modelling structures (Löhner et al. 2003). Learning environments that offer modelling tools that adhere to the system dynamics modelling formalism as proposed by Forrester (1961) implement both the previously identified *graphical* and the *dynamic* key characteristics that allegedly promote meaningful learning.

However, students often have difficulty constructing system dynamics models (hereafter: models). Students who have to create such a model from scratch often create

[1] By graphical representations, we do not mean 'graphical only', and as such graphical representations can incorporate textual information.

[2] In this paper, no distinction is made between a concept map or a mind map as both representations have a graphical structure consisting of nodes and links.

models which only reflect the relevant components, but not the complex relations between these components (Mulder et al. 2010). To compensate for these difficulties, some modelling learning environments (e.g., Löhner et al. 2003; Wu 2010) offer students a list of variables to help identify the major model variables and thus provide students with an *outlined* model (i.e., the third characteristic). Such a model completion task enables students to concentrate on the relations between components and the dynamic system as a whole (Sins et al. 2008). Studies on concept mapping show that providing an outline of the concept map generally enhances learning of the domain (Chang et al. 2001, 2002; Gouli et al. 2004).

The present study examined the affordances of the three identified key characteristics (i.e., graphical, dynamic, and outlined). Four representational formats were compared which increasingly incorporate these characteristics: a summary, a concept map, a model, and an outlined model. The summary representation functioned as a baseline; subsequent representational formats incorporate one additional characteristic compared to its precursor. Comparison of the concept map condition with the summary condition aimed to determine the influence of the 'graphical' characteristic. The comparison of the model condition with the concept map condition served to assess the influence of the 'dynamic' characteristic, and finally a comparison among both model conditions sought to reveal the influence of the 'outline' characteristic. It was expected that each characteristic enhances both the quality of the representations students create and their learning of the phenomenon. Thus, the highest quality of created representation and knowledge gains is expected in the outlined model condition.

## Method

### Participants

Seventy-five Dutch high school students participated in the experiment as part of their regular biology lessons. The sample comprised 29 boys and 46 girls with a mean age of 15.71 (SD = 0.74). Students were randomly assigned to either the summary condition ($n = 15$), the concept map condition ($n = 16$), the model condition ($n = 23$), or the outlined model condition ($n = 21$). The difference in number of students per condition can be explained by the allocation to conditions within classrooms: in some of the classrooms, students were given a short introduction to modelling. After this, they were assigned to either one of the two modelling conditions. Students in the other classrooms were assigned to either the summary or the concept map condition.

### Materials

#### *Instructional Text*

Students in all four conditions received a six-page instructional text (1,997 words) about the 'supply and demand' mechanisms, which ensures that the cells in the human body receive blood containing the right amount of sugar. This topic was chosen because it is a complex and dynamic process and because it is incorporated in the Dutch high school biology curriculum. The instructional text was organized into four sections that addressed (1) how glucose is produced and how much should be in our blood, (2) how the organs in our body balance the glucose level by secreting insulin, (3) how the brain controls this regulation process, and (4) how exactly this regulation process proceeds through time.

#### *Assignment*

All students received an assignment on glucose–insulin regulation that was based on three scenarios: (1) the case of homoeostasis, where the blood glucose level reaches an equilibrium, (2) eating high-calorie food, which creates a spike of glucose in the bloodstream, and (3) the case for people with diabetes Type 1, where the body cannot control the blood glucose level. The first scenario, homoeostasis, served as a starting point for students to construct the external representation associated with their experimental condition. Toward this end, students had to identify important concepts and processes of the glucose–insulin regulation during homoeostasis and include these in their external representation. The subsequent scenarios provided students with real-life cases that affect the glucose–insulin regulation process. These cases aimed to challenge students to create a representation of the glucose–insulin regulation process that accounts for the systems' behaviour in different, more complex situations. As such, these scenarios required the students to evaluate their representations and to make changes when necessary.

These assignments were identical for all students, but the created representational formats differed across conditions. Students in the *summary* condition had to write a summary: a written sequential representation that contained a reduction of the instructional text to its main points. The summary thus incorporated none of the identified key characteristics that were expected to be essential for learning of complex dynamic systems. To check the instructional materials, a small pilot study was conducted with a participant that was familiar with the domain and two pilot participants that were unfamiliar with the domain (comparable to students in the study). Based on this pilot study, students were advised to write a summary of at least one page.

Students in the *concept map* condition created a concept map. Being a *graphical* representational format, concept maps incorporate one of the identified key characteristics. Concept maps were defined as a graphical structure consisting of nodes and links that show the relationships between concepts. There was no distinction made between mind maps and concept maps, meaning that students could choose the structure pattern (e.g., hierarchical or radial) as they liked and were free to use text labels on their connecting links.

Students in the *model* condition were asked to create a model. A model is a *graphical dynamic* representational format and thus incorporates two of the identified key characteristics that were hypothesized to enhance the learning of complex dynamic systems. Students created their models with SCYDynamics (de Jong et al. 2010), a modelling tool that makes use of the graphical structure of a stock and flow diagram that consists of variables and relations students could add via the menu buttons on left side of the screen (see Fig. 1). Variables are the constituent components of a model and can be of three different types: variables that do not change over time (i.e., constants), variables that specify the integration of other variables (i.e., auxiliaries), and variables that accumulate over time (i.e., stocks). Relations define how two or more variables interact. Each relation is visualized by an arrow connector to indicate the causal link between model components and can be further specified by selecting a pre-defined, qualitative relation from a drop-down menu.

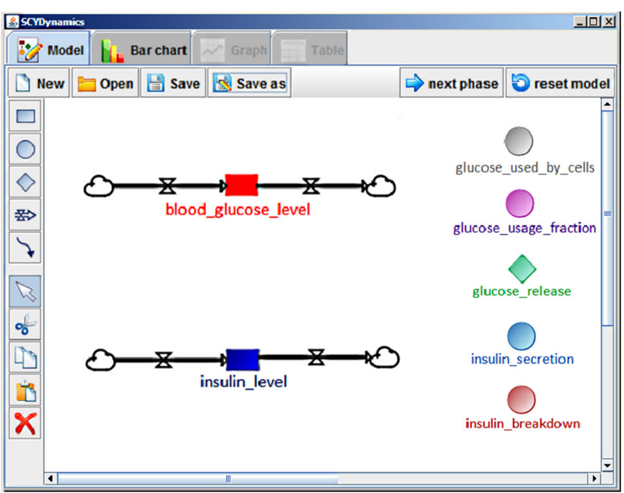


**Fig. 1** Learning environment with the outlined model. Students in the model condition started with a blank screen, whereas students in the outlined model condition received the depicted partial model (that show the major components plus a list of variables)

SCYDynamics enabled students to simulate their model and analyse the output in a bar chart and graph tool. The bar chart provided feedback on the model structure by showing the number of correct components and relations. The graph tool showed the behaviour over time of the model that the students built. Students could compare the graph output to what they thought the behaviour of the system would be and therefore judge the quality of their model. Students could use this output to revise/refine their model. Students in the model condition started with a blank screen in the model editor.

Students in the *outlined model* condition also created a model with SCYDynamics but, unlike students in the model condition, received the incomplete model shown in Fig. 1, which contained all relevant variables of a model of the glucose–insulin regulation in homoeostasis. Such an incomplete model provides a referent knowledge structure and can thus serve as a framework for model construction. This enables students to concentrate on creating the links between variables during their modelling processes. The learning environment in the outlined model condition was identical to that in the model condition except for the availability of the incomplete model.

### *Knowledge Test*

A nine-item paper-and-pencil test addressed students' knowledge of the domain. Three types of questions addressed the key domain *concepts*, the *local* interactions of the components in the system, and the function of the entire *system*. Students' knowledge of the domain concepts (i.e., glucose and insulin) was assessed by two open-ended questions about the function of these concepts in the human body. Students' local reasoning was assessed by four items, each addressing glucose and insulin regulated in a homoeostatic situation. This required students to think about influences that cause increase or decrease glucose and insulin during homoeostasis. These relationships were described in one of the instructional text's sections. Students could answer these questions by drawing the shape of the relationship as a graph. Students' system reasoning was assessed by three items that addressed the glucose–insulin regulation in the three scenarios (i.e., in homoeostasis, when eating high-calorie food, and when having diabetes Type 1). These questions required students to regard the behaviour of the glucose–insulin regulation over time, corresponding to the three scenarios of the assignment. Examples of the three types of items are given in Fig. 2. The knowledge test was administered on two occasions: as a pretest before students engaged in their learning activities and as a posttest to indicate knowledge gains that resulted from these activities.

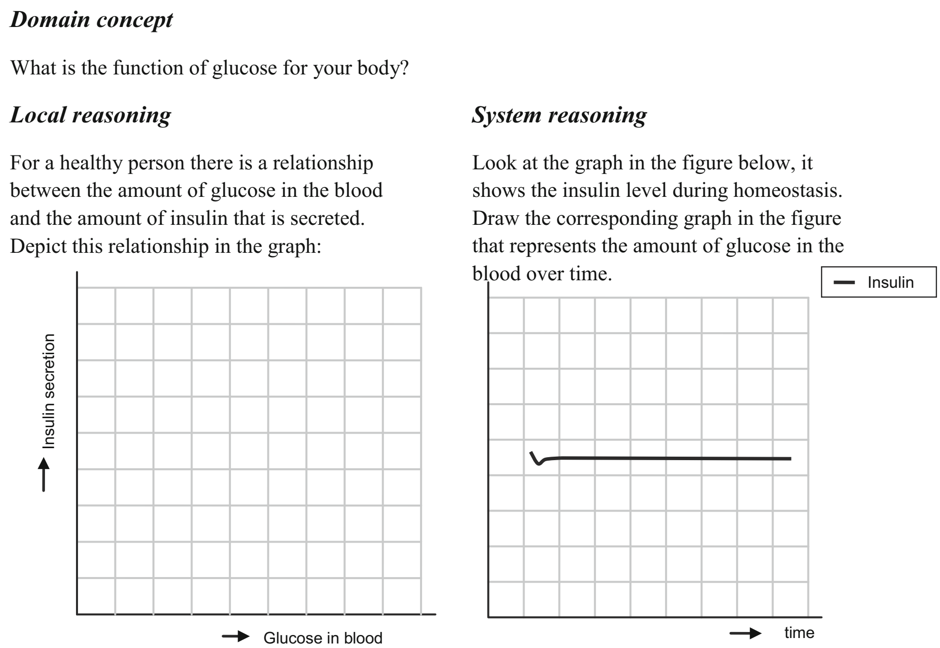


**Fig. 2** Examples of the three types of items in the knowledge test

## Procedure

The study was carried out during the students' regular biology lessons. Data were collected in two parts that was scheduled over the course of 1 week. During the introductory part, participants completed the prior knowledge test. Participants from the two modelling conditions also received a short introduction to modelling. They were familiarized with the basic concept of system dynamics modelling and completed a brief tutorial to learn this modelling language and the operation of the modelling tool.

During the second part, all students first read the instructional text and then worked on their assignment (i.e., the creation/completion of the summary, concept map, or model). They were instructed to describe the processes related to the glucose–insulin regulation in the representation they had to create and to focus on the most important concepts and how they are related. The instructional text remained available while students created their representations. Participants could stop ahead of time if they had completed their assignment. After this learning phase, the posttest was administered.

## Data Analyses

Dependent variables of the study were product quality and knowledge scores. Product quality was considered an indication of how well students expressed their comprehension of the topic during the task. It was assessed from the representations that had been created by the students and computed using a rubric based on Manlove et al. (2009)'s model coding rubric which has excellent reliability (i.e., Cohen's $\kappa > .90$). This rubric assessed whether the representation included relevant concepts related to the glucose–insulin regulation, whether these concepts were linked, and whether the scenarios of the assignment were addressed. One point was awarded for each correct concept, one additional point for each correct link between two concepts, and a third point for the correct implementation of a scenario. The maximum product quality score was 23. All scores were converted into percentages to compensate for the number of correct variables provided in the outlined model condition.

Inter-rater reliability of the summary and concept map quality coding rubric was established by having a second rater score seven randomly selected summaries and concept maps. The Cohen's $\kappa$ reliability estimates for the summaries reached 0.94 (variables), 0.84 (relations), and 1.00 (scenarios). Inter-rater reliability for the concept maps was 0.75 (variables), 0.91 (relations), and 1.00 (scenarios). The models created by the students in both modelling conditions were judged by a software agent that was designed to recognize specified variables, the relations between variables, and the correct implementation of scenarios. The agent was sensitive to alternative terms for variables and variant forms of spelling and used the Levenshtein distance (Levenshtein 1966) to correct for orthographic mistakes. After the experiment, the first author coded all terms in the students' models and reached high inter-rater agreement with the software agent (Cohen's $\kappa = 0.94$).

**Table 1** Summary of participants' performance

| | Summary | | Concept map | | Model | | Outlined model | |
|---|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD | M | SD |
| Pretest score | | | | | | | | |
| Concepts | 1.27 | 0.70 | 1.25 | 0.58 | 1.26 | 0.75 | 1.48 | 0.51 |
| Local reasoning | 1.13 | 0.99 | 1.19 | 0.75 | 0.83 | 0.98 | 1.19 | 0.87 |
| System reasoning | 0.73 | 0.80 | 0.75 | 0.86 | 0.30 | 0.56 | 0.62 | 0.67 |
| Total | 3.13 | 1.25 | 3.19 | 1.56 | 2.39 | 1.56 | 3.29 | 1.42 |
| Posttest score | | | | | | | | |
| Concepts | 1.73 | 0.94 | 2.00 | 0.00 | 1.74 | 0.45 | 1.86 | 0.36 |
| Local reasoning | 1.20 | 1.01 | 1.44 | 0.73 | 1.48 | 0.73 | 1.57 | 0.87 |
| System reasoning | 1.33 | 0.62 | 1.31 | 0.95 | 0.91 | 0.60 | 1.86 | 0.73 |
| Total | 4.33 | 1.59 | 4.94 | 1.39 | 4.52 | 1.31 | 5.43 | 1.47 |
| Gain scores | | | | | | | | |
| Concepts | 0.47 | 0.74 | 0.75 | 0.58 | 0.48 | 0.79 | 0.38 | 0.50 |
| Local reasoning | 0.07 | 0.46 | 0.25 | 0.58 | 0.65 | 0.65 | 0.38 | 0.59 |
| System reasoning | 0.60 | 0.99 | 0.56 | 0.63 | 0.61 | 0.89 | 1.24 | 0.89 |
| Total | 1.20 | 1.78 | 1.75 | 2.15 | 2.13 | 2.42 | 2.14 | 1.80 |
| Product quality (%) | 58.55 | 14.22 | 47.01 | 19.97 | 26.47 | 15.62 | 77.38 | 20.20 |

Knowledge scores were considered a reflection of students' comprehension of the glucose–insulin domain and were indicated by students' performance on the knowledge test prior to the task (i.e., pretest) and following the assignment (i.e., posttest). A rubric was developed to score participants' answers to the nine questions, and one point was allocated to each correct response. This led to a maximum score of 2 points for the domain *concept* questions, 4 points for *local* reasoning questions, and 3 points for *system* reasoning questions. The Cohen's κ inter-rater reliability of this rubric was 0.89. Between-group analyses of students' knowledge scores were addressed by Multivariate analysis of variance (MANOVA), to take into account all types of knowledge measured by the knowledge test.

## Results

Table 1 reports the scores on the knowledge tests. The overall mean score on the pretest was 2.96 (SD = 1.48), which indicates that the students in our sample had little prior knowledge of glucose–insulin regulation. Multivariate analysis of variance (MANOVA) confirmed that the marginal cross-condition differences in prior knowledge were not statistically significant, $F(9,\ 213) = 0.88$, $p = .548$.

The representations that were created by the students were analysed on product quality. This measure reflects the extent to which a representation includes the relevant concepts and relations. Raw scores were converted into percentages to compensate for the number of correct variables provided in the outlined model condition. Table 1 shows that the lowest mean product quality score was observed in the model condition and the highest score was obtained by students in the outlined model condition, whereas student in the summary condition and concept map condition held the middle ranks. Univariate analysis of variance (ANOVA) found a significant between-group effect on these scores, $F(3,71) = 31.31$, $p < .001$. Planned contrasts revealed no significant difference between the summary and concept maps condition, $t(71) = 0.25$, $p = .806$. However, the concept maps included significantly more relevant concepts and relations than the models made by students in the model condition, $t(71) = 2.03$, $p = .046$, $r = .23$, who in turn were outperformed by students from the outlined model condition, $t(71) = -3.06$, $p = .003$, $r = .34$.[3]

The knowledge posttest scores shown in Table 1 reflect students' understanding of glucose–insulin regulation after they had created their representation. Gain scores were calculated by subtracting students' pretest scores from their posttest scores. To examine whether students in all conditions had learned during the experiment, four one-sample *t*-tests were conducted using Bonferroni-adjusted one-tailed alpha levels of .0125 per test (.05/4). Results showed

[3] To compensate for possible ceiling or floor effects, ANOVA was performed to analyse between-group effects on a arcsine transformation of the product quality scores. $F(3,71) = 26.77$, $p < .001$. As these analyses produced similar results, the original data were reported to ease interpretation.

that the gain scores differed significantly from zero in all conditions (summary: $t(14) = 2.61$, $p = .010$, $r = .57$; concept map: $t(15) = 3.26$, $p = .003$, $r = .64$; model: $t(22) = 4.23$, $p < .001$, $r = .67$; outlined model: $t(20) = 5.47$, $p < .001$, $r = .77$).

Having established that all students had improved their understanding of glucose–insulin regulation, additional analyses sought to reveal whether these knowledge gains were comparable among conditions. MANOVA produced a significant effect for experimental condition, $F(9,213) = 2.75$, $p = .005$, indicating that the magnitude of students' overall knowledge gains depended on the type of representation they had to create. Subsequent univariate analysis of variance (ANOVA) showed that condition significantly affected students' knowledge gains in local reasoning, $F(3,71) = 3.38$, $p = .023$, and system reasoning, $F(3,71) = 2.83$, $p = .044$, but not their increase in concept knowledge, $F(3,71) = 1.00$, $p = .400$. This means that the type of representation influenced how much students learned of both the local effects within the system and the behaviour of the system as a whole, but it did not influence their increase in concept knowledge. Next, planned contrasts were performed to pinpoint the between-group differences in local and system reasoning. Significant differences in local reasoning scores were found between the concept map condition and the model condition, $t(71) = 2.12$, $p = .037$, $r = .24$. The other differences in local reasoning were not statistically significant (summary vs. concept map: $t(71) = 0.87$, $p = .384$; model vs. outlined model: $t(71) = -1.54$, $p = .127$). Regarding system reasoning, the only significant difference occurred between the model condition and the outlined model condition, $t(71) = 2.42$, $p = .018$, $r = .28$. The other comparisons did not reach statistical significance (summary vs. concept map: $t(71) = -0.12$, $p = .904$, concept map vs. model: $t(71) = 0.17$, $p = .870$). Together these results indicate that students learn more of the local effects of the system when they create a model, compared to when they create a concept map. Likewise, students learn more of the system as a whole when they complete a model outline, compared to when they build a model from scratch.

Bivariate correlations were computed to examine whether students' knowledge gains were associated with the quality of the representation they created. As can be seen in Table 2, the product quality scores were positively related to knowledge gains in system reasoning but unrelated to gains in conceptual knowledge and local reasoning. Additionally, students' knowledge gain in local reasoning was positively associated with their knowledge gains in system reasoning. No significant correlations were found between students' knowledge gains on conceptual knowledge and either students' local reasoning improvement or systems reasoning improvement.

**Table 2** Correlations between performance measures

| | 1. | 2. | 3. | 4. |
|---|---|---|---|---|
| 1. Product quality score | – | | | |
| 2. Concepts gain score | .035 | – | | |
| 3. Local reasoning gain score | .046 | .094 | – | |
| 4. Systems reasoning gain score | .334* | .105 | .355** | – |

* $p < .05$; ** $p < .01$

## Discussion

The aim of this study was to assess the affordances of three identified key characteristics of representations (i.e., graphical, dynamic, and outlined) on students' performance and learning. Four external representational formats were compared that increasingly incorporated these characteristics: a summary, a concept map, a model, and an outlined model. The results confirm the positive effects of the characteristics as representational formats differed with regard to their affordances on students' performance and learning.

First it was expected that graphical representations such as concept maps and models, due to their graphical structure, direct students' attention to concepts and their mutual relationships (cf. Nesbit and Adesope 2006). Therefore, it is slightly surprising that the learning effect failed to show when comparing summaries to concept maps, but did show when comparing concept maps to models. A closer inspection of local reasoning scores shows that students in the summary condition did not increase on this measure and thus reached an average score of 1.20 on this measure. In contrast, students in the concept map, model, and outlined model condition did show knowledge gains and reached an average score of 1.50. Although not significant, this pattern is in line with the expectation that graphical representational formats aid students in learning about concepts and their mutual relationships. The statistical difference between summaries and concept maps in gain scores on this measure might incidentally have been an artefact of marginally (but not significantly) lower scores on the prior knowledge measure.

Second, dynamic representations were expected to increase students' knowledge of the behaviour of the system as a whole as they offer the possibility simulate complex dynamic systems over time. However, results in this study show that novice students that create a model from scratch do not reap these specific benefits of the dynamic representation, compared to an alternative graphical representation. The low quality of the created products in the model condition is a plausible explanation for the absence of an effect on students' knowledge gains.

Given that students have to interact with a high-quality model in order to learn about the correct behaviour of a dynamic system (e.g., Alessi 2000).

Overall, looking at product quality, the observed quality was lowest for students who created a model from scratch. Significant between-group differences were found that partially confirmed the predicted influence of representational format on product quality. As expected, the outlined model representation evidenced most of the relevant concepts and their relations. However, contrary to the expectations, the quality of the concept maps was higher than the quality of the models created by students that modelled from scratch. Likewise, the quality of the summaries was higher than the quality of the concept maps.

Since there were no differences in prior knowledge, a likely explanation for these performance findings is that the difference in quality of the created representation is the result of both the ease of constructing the representation and students' learning gains. The first part of the explanation suggests that some representational formats are easier to construct than others. Looking at syntactical constrains, it might be easier to construct a summary than a concept map or a model. In this comparison, a model offers the most syntactical constraints; students have to identify concepts and relations between these concepts. By definition, these relations are causal in nature as they specify how one variable influences the other. Concept maps are less constrained, because—among others—students can choose the type of relations themselves, which can be of various sorts (e.g., belonging to a class, indicating a predicate or prepreposition; cf. Kozma 1992). Finally, summaries are the least syntactically constrained. Consequently, it might have been easier for students to include the relevant concepts and their relations in less constrained representational formats, as for instance you can copy large parts of the text into a summary. This would explain why the concept map condition outperformed the model condition.

Evidence for the second part of the explanation comes from correlations between students' product quality and students' knowledge gains. Students whose products incorporated more relevant concepts and relations learned more from the system behaviour as a whole. Most likely, this explains why the students in the outlined model condition, even though they created the most constrained representational format, performed best and learned most.

Finally, it was confirmed that outlined representations enhance students' knowledge of the behaviour of a system even more than modelling from scratch. The dynamic nature of correctly constructed models offers students a unique learning experience to observe a systems' behaviour changing over time, which likely enhances students' knowledge of the behaviour of systems as a whole. The fact that the learning effect only showed when comparing model students to outlined model students confirms the necessity of support (i.e., the model outline) for modelling to be effective in learning.

This is an interesting finding as it is by no means evident that this knowledge of the systems' behaviour over time is in fact enhanced when students have the opportunity to see the behaviour of the system over time as they work with a model that can be run: first, because all the students were asked to think about the behaviour of the system over time in the assignment; second, because students need to build a correct model and plot the correct graph (i.e., select the right variable of interest), in order to be able to inspect the systems behaviour correctly; and finally, because students have to correctly evaluate their model run and have to recognize that their model output shows the behaviour of the system over time.

In conclusion, this study does not univocally verify the necessity and effect of all key characteristics individually, but the results do show that a representational format which includes all key characteristics (i.e., an outlined model) resulted in the highest learning gains compared to the other representational formats. This study demonstrated how creating representations of different kinds can support learning of complex systems and that it is important to use the representation best suited for the task at hand.

## Implications

With the increasing use of computers in science and science education, it has become possible to actively engage students in modelling activities where they construct dynamic representations that change over time. Nowadays, there is an important role for models and model construction in science education (see for instance the Next Generation Science Standards).

The present research suggests that the promise of a learning by modelling approach extends beyond the known benefits of this approach (e.g., that it offers students a means to learn like scientists do), as it has great potential to promote meaningful learning of science domains that have a dynamic nature. However, this study also suggests that in order for students to reap the benefits from learning by modelling, support is necessary. Such support can take the form of a list of relevant variables that students have to include in their models. But future research should focus on the effects of alternative support mechanisms as well. Additionally, more research is called upon to identify the individual effects of the key characteristics of learning by modelling, especially on the specific effects of support mechanisms on students' learning by modelling. Furthermore, as learning by modelling becomes a more integrated and common part of students' science curriculum, it might be that students' support needs change as well. Moreover, if students gain more modelling skills, a learning by modelling approach will likely even reach a higher potential.

**Acknowledgments** This study was conducted in the context of the project 'Learning through modelling and self explanations' which is part of the National Initiative Brain and Cognition (NIHC) funded by the Dutch Organization for Scientific Research (NWO), Grant No. 056-31-011. The authors gratefully acknowledge Theresa Kersjes, Jannis Kraiβ, Franziska Von Petersenn, and Christina Zottmann for their help in collecting the data.

## References

Alessi SM (2000) Building versus using simulations. In: Spector JM, Anderson TM (eds) Integrated & holistic perspectives on learning, instruction & technology: improving understanding in complex domains. Kluwer, Dordrecht, pp 175–196

Armbruster BB, Anderson TH, Ostertag J (1987) Does text structure/summarization instruction facilitate learning from expository text? Read Res Q 22:331–346. doi:10.2307/747972

Chang KE, Sung YT, Chen SF (2001) Learning through computer-based concept mapping with scaffolding aid. J Comput Assist Learn 17:21–33. doi:10.1111/j.1365-2729.2001.00156.x

Chang KE, Sung YT, Chen ID (2002) The effect of concept mapping to enhance text comprehension and summarization. J Exp Educ 71:5–23. doi:10.1080/00220970209602054

Cox R (1999) Representation construction, externalised cognition and individual differences. Learn Instr 9:343–363. doi:10.1016/s0959-4752(98)00051-6

de Jong T, van Joolingen WR, Giemza A, Girault I, Hoppe U, Kindermann J, van der Zanden M (2010) Learning by creating and exchanging objects: the SCY experience. Br J Educ Technol 41:909–921. doi:10.1111/j.1467-8535.2010.01121.x

Forrester J (1961) Industrial dynamics. Pegasus Communications, Waltham

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. doi:10.1023/a:1022400817926

Gouli E, Gogoulou A, Papanikolaou K, Grigoriadou M (2004) COMPASS: an adaptive web-based concept map assessment tool. Paper presented at the first international conference on concept mapping, Pamplona, Spain

Graesser AC, Singer M, Trabasso T (1994) Constructing inferences during narrative text comprehension. Psychol Rev 101:371–395. doi:10.1037/0033-295x.101.3.371

Hilbert TS, Renkl A (2009) Learning how to use a computer-based concept-mapping tool: self-explaining examples helps. Comput Hum Behav 25:267–274. doi:10.1016/j.chb.2008.12.006

Hmelo-Silver CE, Duncan RG, Chinn CA (2007) Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). Educ Psychol 42:99–107. doi:10.1080/00461520701263368

Horton PB, McConney AA, Gallo M, Woods AL, Senn GJ, Hamelin D (1993) An investigation of the effectiveness of concept mapping as an instructional tool. Sci Educ 77:95–111. doi:10.1002/sce.3730770107

Jackson SL, Stratford SJ, Krajcik J, Soloway E (1994) Making dynamic modeling accessible to precollege science students. Interact Learn Environ 4:233–257

Johnson-Laird PN (1983) Mental models: toward a cognitive science of language, inference, and consciousness. Harvard University Press, Cambridge

Kafai YB, Resnick M (eds) (1996) Constructionism in practice: designing, thinking, and learning in a digital world. Lawrence Erlbaum Associates, Mawhaw

Kozma RB (1992) Constructing knowledge with learning tool. In: Kommers PM, Jonassen D, Mayes JT, Ferreira A (eds) Cognitive tools for learning, vol 81. Springer, Berlin, pp 23–32

Leelawong K, Biswas G (2008) Designing learning by teaching agents: the Betty's Brain system. Int J Artif Intell Educ 18:181–208

Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions, and reversals. Sov Phys Dokl 10:707–710

Löhner S (2005) Computer based modeling tasks, the role of external representation. PhD, University of Amsterdam, Amsterdam

Löhner S, van Joolingen WR, Savelsbergh ER (2003) The effect of external representation on constructing computer models of complex phenomena. Instr Sci 31:395–418. doi:10.1023/A:1025746813683

Louca LT, Zacharia ZC (2011) Modeling-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. Educ Rev 64:471–492. doi:10.1080/00131911.2011.628748

Manlove S, Lazonder AW, de Jong T (2009) Trends and issues of regulative support use during inquiry learning: patterns from three studies. Comput Hum Behav 25:795–803. doi:10.1016/j.chb.2008.07.010

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. doi:10.1037/0003-066X.59.1.14

Mulder YG, Lazonder AW, de Jong T (2010) Finding out how they find it out: an empirical analysis of inquiry learners' need for support. Int J Sci Educ 32:2033–2053. doi:10.1080/09500690903289993

Nesbit JC, Adesope OO (2006) Learning with concept and knowledge maps: a meta-analysis. Rev Educ Res 76:413–448

Novak J (2005) Results and implications of a 12-year longitudinal study of science concept learning. Res Sci Educ 35:23–40. doi:10.1007/s11165-004-3431-4

Palincsar AS (1998) Social constructivist perspectives on teaching and learning. Annu Rev Psychol 49:345

Schnotz W (2002) Commentary: towards an integrated view of learning from text and visual displays. Educ Psychol Rev 14:101–120. doi:10.1023/a:1013136727916

Sins PHM, van Joolingen WR, Savelsbergh ER, van Hout-Wolters B (2008) Motivation and performance within a collaborative computer-based modeling task: relations between students' achievement goal orientation, self-efficacy, cognitive processing, and achievement. Contemp Educ Psychol 33:58–77. doi:10.1016/j.cedpsych.2006.12.004

Steed M (1992) Stella, a simulation construction kit: cognitive process and educational implications. J Comput Math Sci Teach 11:39–52

Suthers DD, Hundhausen CD (2003) An experimental study of the effects of representational guidance on collaborative learning processes. J Learn Sci 12:183–218. doi:10.1207/s15327809jls1202_2

Thiede KW, Anderson MCM (2003) Summarizing can improve metacomprehension accuracy. Contemp Educ Psychol 28:129–160. doi:10.1016/S0361-476X(02)00011-5

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. doi:10.1016/j.chb.2004.10.039

VanLehn K (2013) Model construction as a learning activity: a design space and review. Interact Learn Environ 21:371–413. doi:10.1080/10494820.2013.803125

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. doi:10.1023/a:1009421303064

Wittrock MC, Alesandrini K (1990) Generation of summaries and analogies and analytic and holistic abilities. Am Educ Res J 27:489–502

Wu H-K (2010) Designing a technology-enhanced learning environment to support scientific modeling. Turk Online J Educ Technol 9:58–65