

Students' reasoning during modeling in an inquiry learning environment

Simone Löhner^{a,*}, Wouter R. van Joolingen^{a,c},
Elwin R. Savelsbergh^{a,b}, Bernadette van Hout-Wolters^a

^a *Graduate School of Teaching and Learning, University of Amsterdam, Wibautstraat 2-4, 1091
GM Amsterdam, The Netherlands*

^b *Utrecht University, The Netherlands*

^c *University of Twente, The Netherlands*

Available online 23 November 2004

Abstract

In an inquiry-learning task, computer modeling can be a powerful tool to enhance students' reasoning and help them improve their understanding. Optimal learning effects in such an environment require a high quality of students reasoning activities. It is not straightforward, however, which type of modeling tool will be most helpful to students' reasoning. In order to identify the effects of different tools, students working with two common types of tools were compared to a normative description of the modeling process. Also the influence of reasoning activities on the achieved modeling result was examined. Different modeling tools did induce significantly different reasoning activities. Students working with a graphical representation designed more experiments with their own model, formulated more qualitative hypotheses, and spent more time evaluating their own model than students working with a textual representation. Results also indicate that many students have serious difficulties performing this task in a systematic manner. The paper concludes with suggestions for support students might need. © 2004 Elsevier Ltd. All rights reserved.

* Corresponding author.

E-mail address: S.Lohner@uva.nl (S. Löhner).

1. Introduction

An important goal in science education is to help students make sense of the world in a scientific way (e.g., Penner, 2001), and to offer them an authentic view of doing science, which goes beyond the much-held misconception of science as a collection of facts (Hestenes, 1992).

Inquiry learning (Bruner, 1961; Dewey, 1938) has been a long advocated approach to offer such an authentic experience by engaging learners in a knowledge construction process. Inquiry became much more feasible as a learning approach with the advent of user-friendly computer simulations, which can provide a supportive environment for inquiry learning (Njoo & de Jong, 1993). Although many educators take a positive stance towards inquiry learning, it is also clear that successful inquiry learning poses high demands on cognitive skills that at least part of the students may be lacking (Kuhn, Black, Keselman, & Kaplan, 2000). Effective inquiry learning requires students to construct and evaluate their own hypotheses, and derive their own conclusions. By contrast, students were found to design inconclusive experiments, display insufficient experimentation behavior and confirmation bias, take an engineering approach instead of a scientific one, and display unstructured behavior driven by local decisions rather than an overall plan (e.g., de Jong & van Joolingen, 1998).

1.1. Combining modeling and inquiry

A method that can possibly support and enhance inquiry learning is computer modeling. Computer models are executable external representations of a domain or subject. Students can express their ideas and observe the resulting model output. This enables students to simulate their view of a subject matter (Jackson, Stratford, Krajcik, & Soloway, 1996).

Most of the studies on computer modeling can be classified into two types of modeling, defined by Bliss (1994) as *explorative modeling* and *expressive modeling*. In explorative modeling, learners explore a given model representing someone else's ideas by trying it out and perhaps modifying it. In some cases this means that the students actually explore a model in a computer-modeling environment, in other cases it is another name for inquiry learning with computer simulations. In expressive modeling, on the other hand, learners construct their own computer model of a domain. This can vary from just expressing their understanding in a model without worrying about 'the correct' model, to making a computer model on the basis of formulas found in books or experimental data.

Another possibility for the use of computer models is a combination of these two types of computer modeling into what we will label *inquiry modeling*. Inquiry modeling refers to integrating a modeling environment into an inquiry-learning environment. In an inquiry-modeling task, learners can learn about the domain by doing experiments. They have a modeling tool where they can express their acquired ideas in a model. Fig. 1 presents an example of such an inquiry-modeling environment that

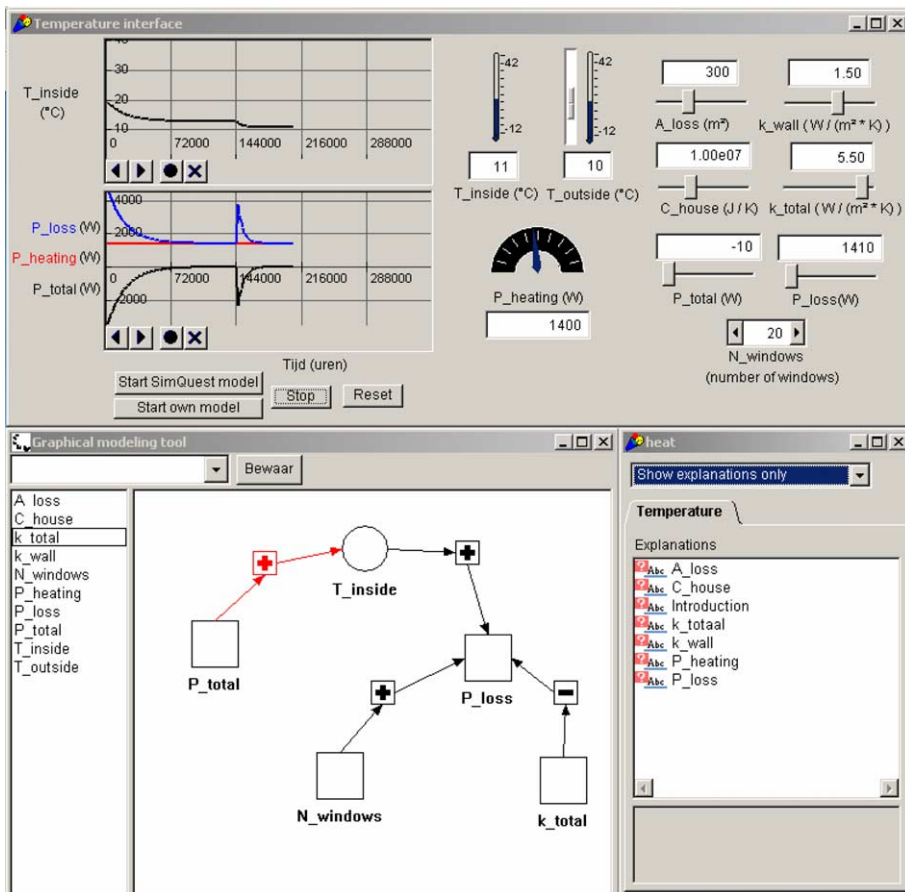


Fig. 1. Screenshot of an inquiry learning environment with the simulation of the temperature of a house. On top the simulation window in which the students can conduct experiments and on the bottom the modeling window.

has been implemented in the Simquest learning environment with a simulation and a modeling window, so that the model output can be easily compared to the 'real data' of the simulation.

1.2. Model representations

A computer modeling tool that is being used for inquiry modeling serves several cognitive functions (Löhner, van Joolingen, & Savelsbergh, 2003), and it can be expected that the way in which these functions are fulfilled, depends on properties of the tool. There is a wide range of available modeling tools and there are substantial differences between them. Some allow for a maximum of precision and full control

over the model whereas others provide more intuitive modeling taking over many decisions from the user, and as a consequence allow less precision and control. These two types of modeling tools: the ‘high precision’ and the ‘intuitive’ tools each form a category with a set of properties that tends to cluster together. The ‘high-precision’ tools tend to be text-based, to have a specified relation as the basic unit of the model, to ask for quantitative specification, and to have the user responsible for filling out all details. The more intuitive tools on the other hand, tend to use a diagram representation, to have variables as basic units of the model, to allow qualitative specification, and to fill out a lot of detail for the user.

Because computer models are external knowledge representations, and because external representations tend to have a strong influence on reasoning (e.g., Lewalter, 2003; Stern, Aprea, & Ebner, 2003; Suthers, 1999; Zhang, 1997), we expect that the different representations used in modeling tools may have differential effects on students’ reasoning processes. For example, a diagram representation of a model may help to make the model structure more salient. Because students can point out exactly which relation they want to discuss they will more easily discuss these relations (Suthers, 1999). However, the effect of different model representations is not always straightforward, for instance, in a high precision tool, adding a new relation is quite laborious because the relation has to be entered in formula form, which also includes its full specification. In an intuitive tool this is less costly because relations can be added without further specification. However, it may happen that less costly interface operations lead to more, but less efficient activity (O’Hara & Payne, 1998; Schär, Schierz, Stoll, & Krueger, 1997).

In our study, we compared the effects two different modeling tools have on the modeling process in a collaborative inquiry-modeling task. The first, which we will refer to as ‘textual representation’, is a text-based high precision tool that requires the students to fill in all relations in the form of mathematical equations. The other, the ‘graphical representation’, is a diagrammatic intuitive tool.

In a previous paper, we reported the effects these model representations have on student activity in the environment (Löhner et al., 2003). In this paper we will focus on differences in reasoning activities and, more specifically, differences in the way the collaborating students communicate about the models they are constructing.

1.3. The inquiry and modeling process

Several researchers have proposed descriptions of the processes that take place in inquiry learning. Njoo and de Jong (1993) distinguish communicative, transformative, and regulative processes during inquiry learning. They deem the transformative processes to be the actual activities of scientific inquiry as they relate specifically to conducting experiments or constructing relations between variables. Examples of these processes include analyzing, generating hypotheses, testing hypotheses, and drawing conclusions.

Although many researchers make their own classification of reasoning activities in inquiry learning, depending on the specific research questions, most of the identified

patterns resemble the so-called inquiry cycle (e.g., White & Shimoda, 1999) containing the following steps: question, hypothesize, investigate, analyze, model, and evaluate. This is a normative description of the inquiry process, built on assumptions about how the process should look like ideally.

Several attempts have been made to describe the modeling process, and most descriptions have remained rather generic (Hogan & Thomas, 2001; Schecker, 1998; Stratford, Krajcik, & Soloway, 1997). A description that links modeling activities to reasoning during inquiry, is proposed by de Jong et al. (2002). In this approach, modeling processes and activities are seen as realizations of processes of inquiry; hence there is a strong link between them. Sketching the model is for instance seen as a way to express the inquiry activity of orientation, specifying the model is seen as hypothesizing, and evaluating the model is seen as drawing conclusions (see Fig. 2).

Because inquiry modeling as we describe it closely resembles the way scientists work, the process of building a model should be similar to the inquiry cycle. The expected reasoning activities during inquiry modeling are orientation, hypothesizing, experimenting, and evaluating the model. In the descriptions of the modeling and inquiry process that were cited above we indeed see that these activities are present, although the terminology may differ. In Fig. 3 we link these terminologies to ours, to show the recurrence of the inquiry cycle in the various contributions.

As can be seen most descriptions indeed contain reasoning activities in four categories: orientation, hypothesis generation, experimentation, and evaluation. Two of the descriptions also contain documentation activities. The differences in terminology include different definitions of the term ‘modeling’. For instance, while some descriptions contain modeling activities in the early stages of the process, in the description of White and Shimoda (1999) modeling is only mentioned at a later stage. The reason for this is that their modeling activity is mainly model

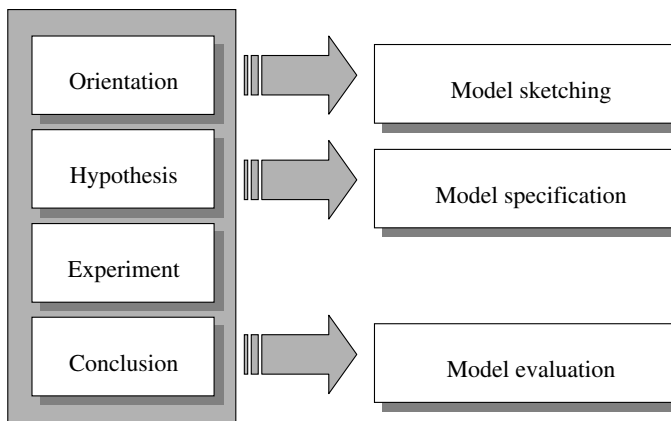


Fig. 2. The links between inquiry and modeling activities as proposed by de Jong et al. (2002, p. 40, adapted figure).

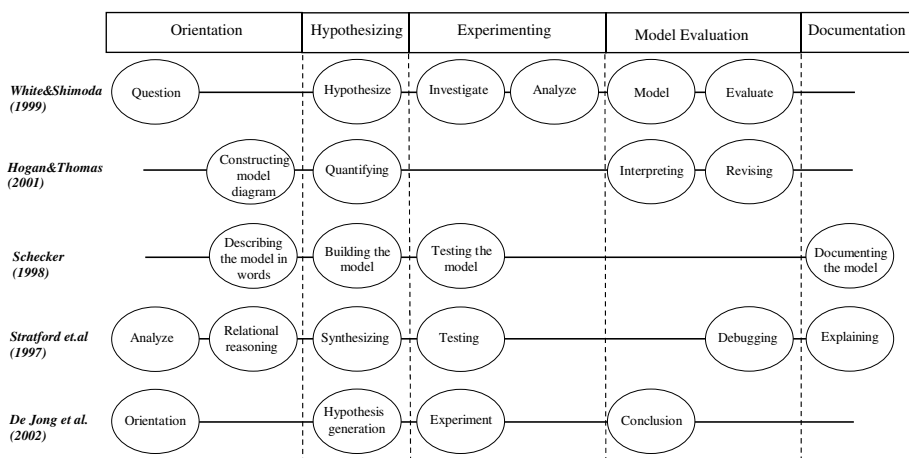


Fig. 3. Overview of reasoning activities for modeling and inquiry as proposed by different authors.

implementation, while in other descriptions the term modeling also refers to the generation of hypotheses. All descriptions see the stages and activities in modeling as a cyclic process.

Several other researchers do not classify the whole inquiry cycle, but zoom in on specific parts of it. [Kuhn et al. \(2000\)](#) for example focus on the strategies for hypothesis generation. They divide these into the stages of inquiry, analysis, inference, and argument. [Hmelo-Silver, Nagarajan, and Day \(2002\)](#) describe experimental strategies for working with a simulation. They distinguish use of prior knowledge or analogies, metacognitive control (monitoring, evaluation and reflection), and scientific reasoning. Within the scientific reasoning category they further distinguish between the planning of the experiments and the interpretation of data at several levels. They find that novice experimenters often have sub optimal experimentation strategies and often do not go further than 'low level' interpretation of their results.

What seems to be missing in empirical research on computer modeling thus far is insight into the validity of the normative description of the modeling process. Also it is still not clear whether this description captures all the reasoning activities of students during inquiry modeling.

In this study we address the following questions:

- How does the normative description of the modeling process compare to the reasoning activities students engage in during inquiry modeling?
- In which way do different representations influence the reasoning activities?
- What is the influence of these reasoning activities on the modeling product?

We will closely examine the reasoning activities of students while they build computer models of the built-in system simulation, and compare these activities to our normative model of the modeling process. For the different external representations

we expect that the students working with the graphical representation are able to generate model relations and to evaluate their model more easily, because the graphical representation makes it easier to express model relations and is easier to talk about. Also the fact that students have to formulate exact equations in the textual representation might interfere with the model building process.

2. Method

2.1. Participants

Forty-two secondary school students (grade 11, average age 16.8 years; 17 girls and 25 boys) from three schools in the Amsterdam area, who had chosen to follow science courses, participated in the experiment as part of their regular coursework. The students received 14 for participating. The experiment took 3 h in total.

The study was conducted in a random group pre-test design with two experimental conditions: the ‘textual’ condition using the textual representation and the ‘graphical’ condition using the graphical representation (see the following section) to build their models. Students worked on their task in small groups of two or three.

2.2. Materials

The *textual modeling representation*, which is shown on the left-hand side of Fig. 4, uses text as the primary representation. Its modeling language is quantitative and the primary model entities are variables and relations. The information is externalized in the form of a list of formulas. The textual representation is a high-precision representation. The right-hand side of Fig. 4 shows the *graphical representation*, a more intuitive model representation. Here the primary representation is graphical and the modeling language is qualitative (or semi-quantitative). The primary model entities are the variables and complex relations are handled by the system. The

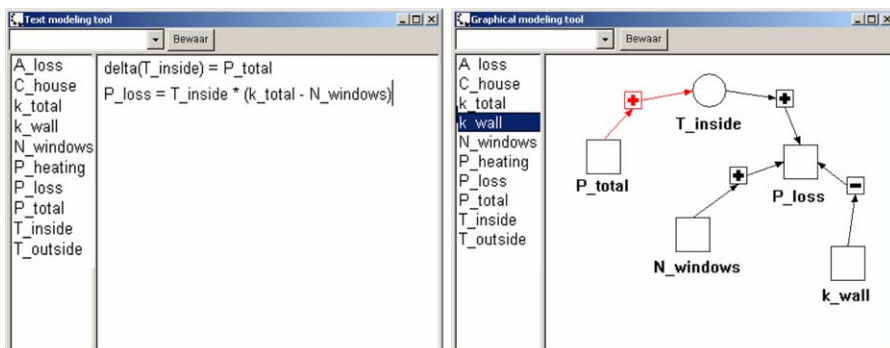


Fig. 4. Screenshots from the modeling tools using a textual representation (left) and a graphical representation (right).

learner indicates whether the relation is positive or negative, the system decides on the precise form of the relation. The model has a visible graphical structure.

The modeling tools for both representations were implemented in SimQuest, an authoring system for discovery learning simulations (van Joolingen & de Jong, 2003; van Joolingen, King, & de Jong, 1997).

The students' task was to build a model that fits the empirical data they could obtain from a simulation of the domain (see Fig. 1) that was also present in the SimQuest environment. The students could use this simulation to conduct experiments. The domain was 'the temperature regulation inside a house'. The students could, for instance, vary the number of windows of the house, the outside temperature, or the heat emitted by the radiators. The outcomes of the simulation, such as the temperature inside the room, were displayed in a graph.

When working in an inquiry-modeling environment, the output of the students' model should eventually resemble the output of the built-in system model. In the textual representation this is the case when the students enter the correct equation. In the graphical representation this is more difficult, because the students only indicate if the relationship is positive or negative. Therefore we chose to provide the students, in both groups, with a list of variables that were part of the model. This makes it possible for the simulation engine to compare the graphical qualitative student model with the system model and fill in the right mathematical equations. The students could only choose the variables for their model from this list. They also received a short explanation of the meaning of these variables.

2.3. Measures and tests

The students were pre-tested for their scientific reasoning skills, for which we selected and translated items from the scientific reasoning part of the ACT (2001). The test results were used post-hoc to verify equality of the two experimental groups. Besides that, to verify group equality we also compared the student's grades for math and physics.

During the modeling task all actions of the learners in the learning environment were logged and the students' conversations were recorded. From the log files we extracted all models the students made in the course of the task, and all actions they performed to run either the system simulation or their own models.

To determine the quality of the models constructed by the students, they were automatically analyzed with a program that weighed the correct and incorrect relations in each model in such a way that a perfect model would yield a score of 1; an empty model would yield a score of 0, and a model consisting of a number of randomly created relations, would also result in a score close to 0. We used this model score to express the quality of the models rather than just the number of correct relations, because this enabled us to correct the score for guessing and made it possible to give credit for partially correct relations. All the models the students simulated during the modeling session were analyzed in this way. To determine the quality of the modeling process, both the highest model score and

the average score over all models could be used. As the correlations between the average model score and the highest model score were very high in both the text and the graphical representation condition (Spearman rank correlation is 0.822 for the textual condition and 0.881 for the graphical condition), we decided to use only the highest model score as a measure for the quality of the modeling product of each group.

The verbal protocols of the students were transcribed and analyzed with the coding scheme described in the following section. The protocols were segmented into episodes in which the students were working on the same activity. The length of these episodes differed from just one or two utterances to several minutes. After coding, the percentage of time students spent on each activity was computed. As a unit of analysis, we used the small groups, because in collaborative reasoning students have great influence on each other's thinking. Also the students were working on a joint product – the computer model. The protocol of one of the groups in the graphical condition could not be transcribed due to bad audio quality. The protocol analysis was carried out using the computer program MEPA for Multiple Episode Protocol Analysis (Erkens, 1998).

Because of the relatively small sample size in this experiment ($n = 10$ groups per condition), all between-condition differences were analyzed by non-parametric Mann–Whitney U tests. Correlations were analyzed using the non-parametric Spearman rank method.

2.4. Instrument development

To interpret the activities of the students working with the inquiry-modeling environment, we developed a coding scheme in an iterative procedure starting from the students' protocols and using the normative description of reasoning processes during inquiry modeling from Fig. 3 – orientation, hypothesizing, experimentation, and model evaluation – as the main categories.

First, we skimmed the protocols to identify the kind of activities the students were engaged in and developed a first tentative analysis scheme using the global normative description of the modeling process as a starting point. In order to refine the analysis, we randomly chose parts of a protocol of a graphical group and parts of a protocol of a textual group and classified the activities we saw the students performing. These activities were grouped to give a preliminary description of the modeling process. This description was compared to the normative model of the inquiry modeling process (see Fig. 3) and adaptations were made at several points. For example a separate category for model implementation was added, to see how much time the students spend on the mechanics of implementing the model. In the process of studying the protocols, we also noticed several other salient features of the protocols, which we incorporated in the analysis scheme in order to be able to determine whether these were regularly occurring events.

The definitive coding scheme contains two main groups of activities, the scientific reasoning categories and “other categories”. In all categories we only coded the students' talk. The students' actions, registered in the computer log-files, were used to

understand the students' reasoning, but they were not coded. The *scientific reasoning categories* we distinguished after taking the adaptations into account are:

Orientation. Episodes in which students talk about the meaning of the different variables are coded as orientation activities. Also coded as orientation activities, are episodes in which students talk about own experiences related to the subject, prior knowledge they have from schoolbooks, and instruction they received about how to build a model. This is a slightly narrow interpretation of orientation activities. Often, part of experimenting is also counted as orientation (e.g., Njoo & de Jong, 1993). But we found we were unable to make a distinction between orienting experiments and other experiments on the basis of the protocols as a lot of the experiments the students performed had no real hypothesis. Therefore all experiments were coded as experimenting.

Hypothesizing. This category comprises of the generation of hypotheses and predicting output. A distinction is made between hypotheses of different precision (qualitative and quantitative). Every indication the students gave of a relation between two or more variables was counted as a hypothesis.

Because students rarely used causal mechanisms to account for the hypotheses they proposed and most often just proposed a hypothesis out of the blue without even trying to give any substantiation for it, it was decided to introduce a distinction between the formulation of hypotheses: (a) with no substantiation; (b) substantiated with experimental data; (c) substantiated with mechanism (see Table 1).

Experimenting. Episodes about the design of experiments for the system simulation as well as for the students' model were scored in this category. The description of the output of either simulation was also coded as experimenting. Both simulations (i.e., the SimQuest simulation and the own model) were scored in different subcategories. As students often did not explicitly use hypotheses or concrete experiments, but instead just pressed the start button of the system simulation without stating an objective, the experiments the students designed were divided into: (a) experiments with no hypothesis; (b) experiments with a qualitative hypothesis; (c) experiments with a (semi) quantitative hypothesis (see Table 2). A qualitative hypothesis as basis for experiment was also scored if one variable was indicated as a focus variable.

Table 1
Examples of the three types of hypothesis formulation

Hypothesis with no substantiation

A: Err... Lets try number of windows goes up, then temperature goes up

Hypothesis substantiated with experimental data

Students run a simulation and change the number of windows

A: Look, the temperature is going up, so it depends on the number of windows'

Hypothesis substantiated with a causal mechanism

C: The power of the radiator had something to do with the temperature inside. When the heating is turned up high...

A: ...it gets warmer inside

B: Yes. Then it works harder

All types are possible for qualitative as well as quantitative hypotheses.

Table 2

Examples of the three kinds of experiment design for either the simplest simulation or the simulation of the own model

Experiment designed with no hypothesis

A: Lets see what it does

Experiment designed with a qualitative hypothesis

A: You have to look what happens when you change the number of windows

Experiment designed with a semi-quantitative hypothesis

C: OK, if P_heating goes up, the temperature goes up. Lets see if that's right

Although this is a rather broad definition it is legitimate in this particular domain because the output variables of the simulation and thus the dependent variables of the hypothesis were fixed.

Model implementation. Episodes in which the students talk about their actions for building the model are scored in this category. Only episodes in which they actually talked about these activities were coded. Sometimes it could be seen in the log-files that students were adding or deleting relations, but it was not coded if they did not explicitly mention this in their discussion.

Model evaluation. In this category all episodes are scored in which students make evaluative remarks about their own model.

The other categories comprise of:

Actions. All episodes in which the students talk about how the modeling tool or the simulation works as well as episodes in which the students are reading or calculating were coded as actions.

Regulation. This category comprises of episodes in which the students plan what they will do, choose the activity they will engage in next, evaluate their progress, or talk about the task.

Off task. In this category all off task statements, such as talk about until what time the McDonalds was open or about grades the students got on their report, are scored.

Experimenter. In this category all experimenter interventions are scored. Whenever the students were really stuck, the experimenter helped them to get back on the right track.

A complete list of categories and subcategories can be found in [Appendix A](#).

A second coder coded part of the (segmented) protocols to determine the interrater reliability. The interrater reliability was found to be acceptable (77% agreement; Cohen's kappa = 0.75 for the main categories). The sub codes for the different types of experiment design and hypothesis generation described in [Tables 1 and 2](#) were subjected to a separate reliability analysis. The second rater received the protocols coded with the main categories and then coded the sub codes. The interrater reliability for these codes was poor (61% agreement, Cohen's kappa = 0.34 for the types of experimental design; 71% agreement, Cohen's kappa = 0.49 for the types of hypothesis generation).

2.5. Procedure

The students were first tested for scientific reasoning skills. This took about 20 min. The students were randomly assigned to one of the two conditions. To get acquainted with the modeling environment each student individually worked through an instruction manual on an example model of personal finance, ‘the contents of your wallet’, for ≈ 45 min. After a short break, within each condition students were randomly assigned to small groups (18 dyads and 2 triads, half of which in each condition) for the experimental modeling task. As the experiment was conducted in three separate sessions for the three schools, all groups consisted of classmates. They spent about 1 h on a task that required them to model the temperature regulation of a house, in a face-to-face setting. For this task they were given only a minimal instruction, in order not to influence their working process. The collaborative setting of the experiment was chosen to trigger verbalization of understanding and thinking and giving of explanations and justifications for actions in the environment. When the students were stuck and expressed that they did not know how to proceed, the experimenter intervened to help them continue.

3. Results

The pre-test results for scientific reasoning skills, revealed no significant difference between the students in the textual condition ($M = 5.83$, $SD = 1.65$) and the graphical condition ($M = 6.64$, $SD = 3.00$; $U = 169$, $p = 0.43$). Also the average grades in math ($M = 5.80$, $SD = 1.58$ for the textual group and $M = 5.89$, $SD = 1.54$ for the graphical group; $U = 218$, $p = 0.96$) and physics ($M = 6.65$, $SD = 0.88$ for the textual group and $M = 6.75$, $SD = 1.02$ for the graphical group; $U = 200.5$, $p = 0.60$) did not significantly differ between groups.

3.1. Students reasoning activities

Table 3 shows the percentages of time the groups spent on each activity. The standard deviations of these percentages are high, indicating large differences between the individual small groups. The amount of off task talk and experimenter interventions is high; in some cases even 50% of the total time. However, most students do spend the majority of their time on the scientific reasoning activities.

When looking at the differences between conditions in percentage of time spent on the modeling activities, the only significant difference occurs in the category model evaluation. The groups working with the graphical representation spend significantly more time (13.4%, $SD = 5.9$) evaluating their models than the group working with the textual representation (7.1%, $SD = 3.4$; $U = 13$, $p = 0.008$). The difference in percentage of time spent on regulative activities is marginally significant: the graphical group spent somewhat less time on regulation (4.9%, $SD = 2.1$) than the textual group (6.9%, $SD = 2.9$; $U = 23$, $p < 0.079$).

Table 3

Percentage of time spent on reasoning activities for the textual and graphical condition (standard deviation in parentheses)

	Graphical	Textual	<i>U</i>	<i>p</i>
Orientation	5.7 (2.2)	5.9 (4.9)	38	0.60
Hypothesizing	19.4 (9.8)	14.9 (7.6)	31	0.28
Experimenting	21.1 (6.7)	15.9 (11.3)	29	0.21
Model implementation	5.9 (2.5)	4.1 (1.6)	33	0.36
Model evaluation	13.4 (5.9)	7.1 (3.4)	13*	0.008
Actions	8.8 (3.7)	9.1 (6.3)	40	0.72
Regulation	4.9 (2.1)	6.9 (2.9)	23**	0.079
Off task	8.9 (10.2)	16.4 (15.1)	28	0.19
Experimenter	10.1 (8.3)	16.6 (12.0)	29	0.21

U and *p* scores were obtained with a non-parametric Mann–Whitney *U*-test.

* $p < 0.05$.

** $p < 0.1$.

On a lower level of aggregation (see the [Appendix A](#) for an overview of the sub-categories), more significant differences between the two conditions can be observed. Within the category hypothesizing, the groups in the graphical condition formulate more qualitative hypotheses ($M = 17.72$, $SD = 8.14$) than the textual group ($M = 4.48$, $SD = 2.50$; $U = 0$, $p = 0.001$), and formulate less quantitative hypotheses ($M = 1.47$; $SD = 1.90$) than the textual group ($M = 10.36$, $SD = 6.93$; $U = 10$, $p = 0.003$). [Table 4](#) shows the percentages of hypotheses according to the way they are substantiated. An average of 50.2% of the formulated hypotheses ($SD = 16.8$) is not substantiated at all. Closer inspection of the protocols revealed that in the textual group even quantitative hypotheses are often just formulated out of the blue. A student would for instance just come up with: “Well, I think that P_{loss} is the number of windows, plus the difference in temperature between inside and outside plus k_{wall} times the area of the house” (student protocol 1714T). The groups in the graphical condition substantiate significantly more hypotheses (30.1%, $SD = 6.3$) with a mechanism than those in the textual condition (13.7%, $SD = 7.9$; $U = 18$, $p = 0.028$). There are no significant differences for the other forms of substantiation. The results in [Table 4](#) must be interpreted with care, because the coding scheme was not sufficiently reliable to permit strong conclusions.

Table 4

Percentage of hypotheses with no substantiation, substantiation by experimental data, and substantiation by reasoning with a mechanism in the two representations (standard deviation in parentheses)

	Graphical	Textual
No substantiation	44.2 (11.1)	55.6 (19.6)
Substantiated with experimental data	25.6 (11.4)	30.8 (21.0)
Substantiated with mechanism*	30.1 (6.3)	13.7 (7.9)

* $p < 0.001$.

Table 5

Percentage of experiments designed with no hypothesis, a qualitative hypothesis, and a (semi) quantitative hypothesis in the two representations

	Graphical	Textual
No hypothesis	39.3 (15.4)	31.9 (20.1)
Qualitative hypothesis	48.7 (17.4)	48.9 (22.1)
(Semi) quantitative hypothesis	11 (8.6)	19.2 (19.4)

Within the category experimenting, the groups in the graphical condition design more experiments with their own model ($M = 2.53$, $SD = 1.70$) than the textual groups ($M = 0.64$, $SD = 0.61$; $U = 13$, $p = 0.008$). When designing their experiments with both the system simulation and the own model, in most cases they do that while expressing a qualitative hypothesis ($M = 48.7$, $SD = 17.4$ for the graphical condition; $M = 48.9$, $SD = 22.1$ for the textual condition; see Table 5). According to our definition this means that the students do not necessarily actually formulate a hypothesis, but that they perform the experiment while focusing on a particular variable (see also Table 2).

Within the category actions, there is no significant difference in the amount of calculating – but in the textual condition calculating is observed in 5 of the 10 groups, while in the graphical condition there is a calculating episode in only 1 of the 10 groups.

To study how students went through the inquiry process, the transitions between different activities were computed. A significant transition occurs if the number of successions between two cognitive activities is significantly higher than might be expected on the basis of a random distribution of coding. Fig. 5 shows the significant transitions occurring in the two experimental groups. The thickness of the arrow denotes how much the occurring transition frequency is higher than the frequency that can be expected on the basis of a random distribution of coding. To prevent that

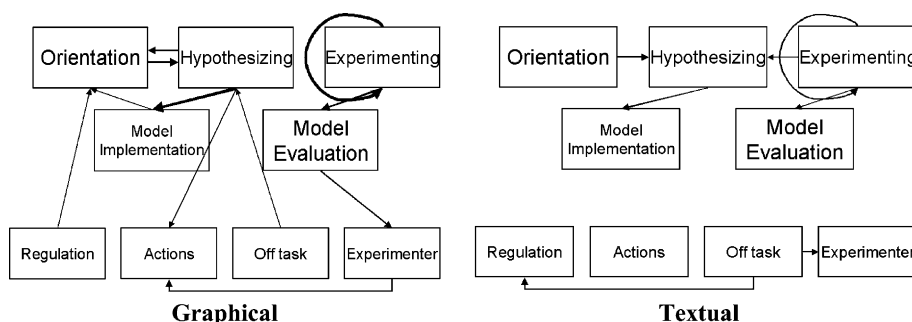


Fig. 5. Probability of transitions between different modeling activities in the graphical and the textual representation. The thickness of the arrow denotes how much the occurring transition frequency is higher than the frequency that can be expected on the basis of a random distribution of coding.

infrequently used codes nevertheless led to significant transitions, only those transitions of which the expected frequency was 5 or higher were included in the analysis. There are several similarities between the two groups. In both groups, a hypothesis is typically stated after orientation; a model is implemented after hypothesizing, and a model is evaluated after experimenting. Also after experimenting often another experimenting episode occurs. This means that they perform several experiments in a row, or that they interpret the experimental data after designing and performing an experiment.

In the textual representation, there is no cycle present involving more than one activity. For the groups working with the graphical representation, there is such a cycle (i.e., orientation – hypothesizing – model implementation – back to orientation) but experimenting is not included in it. This means that the students are not following our normative description of the process based on the inquiry cycle. The formulation of hypotheses is not systematically followed by performing experiments in either of the conditions and the transition between experimenting and the formulation of hypotheses can only be observed in the textual representation.

3.2. Relation with modeling product

The modeling result, indicated by the highest model score, differed significantly between the two conditions ($U = 0$, $p = 0.001$). The groups in the graphical condition produced better models ($M = 0.80$, $SD = 0.19$) than those in the textual condition ($M = 0.39$, $SD = 0.21$). To study whether there was a connection between

Table 6

Spearman rank correlations between percentage of time spent on particular reasoning activities and the highest achieved model score for all students and separately for students in the textual and the graphical condition

	Highest model score			
	Graphical	Textual	All students	
			Partial correlation	Full correlation
Orientation	−0.18	0.27	0.23	0.25
Hypothesizing	0.32	0.39	0.33	0.41**
Experimenting	0.56	0.06	0.11	0.29
Model implementation	0.03	−0.15	−0.14	0.07
Model evaluation	0.17	0.44	0.12	0.51*
Actions	−0.15	0.13	0.05	0.10
Regulation	0.08	−0.10	0.08	−0.26
Off task	−0.70*	−0.35	−0.49*	−0.55*
Experimenter	−0.24	0.29	0.12	−0.14

Correlations for all students are given as full correlation and as partial correlation with the effect of condition partitioned out.

* $p < 0.05$ (2-tailed).

** $p < 0.1$ (2-tailed).

the quality of the models the students designed and the reasoning activities they performed during modeling, correlations were computed between the reasoning activities and the highest model scores. Table 6 shows that, for all students, there is a significant positive correlation ($r_s = 0.51$) between the amount of time spent on model evaluation and the highest model score. However, when the effect of group assignment is partitioned out this effect disappears, meaning that this correlation must be attributed to the effect of the group, not to an independent effect of the amount of model evaluation itself. Also a significant negative correlation ($r_s = -0.55$) between the amount of time spent on off task communication and the model score is found. There also is a marginally significant positive correlation ($r_s = 0.41$) between the time spent on formulation of hypotheses and the model score. These latter correlations depend less on group assignment, as can be seen from the partial correlations. When looking at the separate conditions almost all significant correlations disappear, except the negative correlation between the amount of time spent on off task communication and the model score ($r_s = -0.70$), which is still present in the graphical condition.

4. Conclusion and discussion

With respect to our first research question, namely, how the students' reasoning activities compare to a normative model of the inquiry process, we found that students do spend most of their time during inquiry modeling on scientific reasoning activities, but not in a systematic temporal order. There is no observable inquiry cycle for the students working in either the textual or the graphical representation. Even though a cyclic pattern can be observed in the graphical group, this only includes orientation, hypothesis formulation, and model implementation. This is in accordance with the findings of several other researchers (e.g., Kuhn et al., 2000), indicating that students often have difficulties using sound strategies for inquiry.

We also find that students do most experiments without expressing a hypothesis that is specific enough to guide the inquiry process. Moreover, most hypotheses are formulated without substantiation, neither from experimental data nor from an assumed causal mechanism. In the textual condition, even quantitative hypotheses often appeared out of the blue. This tendency was probably strengthened by the predefined list of variables given to students in the current version of the modeling tool. This list may have stimulated students to regard the task as a puzzle rather than as an authentic inquiry task. Another factor in the current version of the inquiry environment that might stimulate puzzle-solving behavior is the presence of a system simulation. Because students have the "correct" data available, they spend their effort trying to match their model output to the system simulation output, rather than trying to explain the underlying mechanism.

The representations used, differed on more than one aspect, partly because aspects confound each other (e.g., a qualitative model is not easily represented

by formulas), and partly to implement the two kinds of modeling introduced in the beginning of this article: ‘high precision’ and ‘intuitive’. With respect to our second research question, in which ways external representations influence students’ reasoning activities, this leads us to expect differences in modeling activity, as was found in our previous analysis (Löhner et al., *in press*), and in reasoning processes, as discussed here.

We found that students in the graphical condition designed more experiments with their own model; formulated more qualitative hypotheses; spent more time evaluating their own model, and supported their hypotheses more often by reasoning with a mechanism. Students in the textual condition formulated more quantitative hypotheses than students in the graphical condition and formulated hardly any qualitative hypotheses. In our previous analysis of student activities we found that students in the graphical condition ran more models, tried more different relations and ran more system simulations. Moreover, in the graphical condition, the resulting students’ models were on average closer to the target model.

When compared, the outcomes of the current analysis are less salient than those in the previous one, in the sense that fewer significant differences were found between groups with respect to the occurrence of reasoning activities. A possible explanation may be that the actions needed to implement a model were different in both model representations. Implementing a textual model would take more time and effort than implementing an equivalent graphical model in our tool. Therefore, part of the differences in modeling activity and modeling result, as expressed in the model score, should be ascribed to these implementation differences, and not to the influence of the representation on students’ reasoning. However, we do see influences of the representation on the reasoning processes, in particular for the process of model evaluation and the kinds of hypotheses they formulate.

To answer our third research question, about the influence of the reasoning activities on the modeling product, we studied correlations between the students’ reasoning activities and their highest model score. We found a – non-significant – trend towards an effect of more hypothesizing leading to better models. This is in line with the previous finding by van Joolingen and de Jong (1993) that hypothesizing has a positive effect on the learning product in a simulation-based learning environment. In future studies, the role of hypothesizing in modeling needs to be explored further.

Students in the graphical condition attained a higher model score. Our results do not allow attributing this improvement to their increased model evaluation activity. Other research, however, has shown product evaluation and hypothesis generation to be important activities in reasoning processes. Veermans, van Joolingen, and De Jong (2000) found that stimulating the evaluation of hypotheses, contributed to the development of intuitive knowledge. Also others (e.g., Elshout-Mohr, van Hout-Wolters, & Broekkamp, 1998; Masui & de Corte, 1999) stress the importance of evaluation in knowledge construction. Further research is needed to confirm this effect in the current context.

As shown above, students' behaviors clearly deviated from the normative description of the inquiry process. Although all individual reasoning processes in this description did occur, the order in which they occurred, differed from the prescribed order. There are three issues that come into play here. One is that the normative model of inquiry can be seen as a post-hoc rationalization of a process and that even professional scientists do not strictly follow this order (Feyerabend, 1975). Moreover, the students in our study were novices with respect to inquiry modeling. Apart from the modeling introduction they received as part of the experiment, the students had no or little experience with computer modeling. It can be expected that after building more experience, they will also show behavior closer to the normative model. To provide this experience, modeling should be incorporated in the regular science lessons for a longer period of time, instead of being presented as a separate activity for a very limited amount of time as is the case in most methods used in schools. Finally, the modeling environment we used allowed the students to see the modeling task as a puzzle, especially in the graphical condition. It contained a list of variables that had to be used in the model. It was often obvious that – even with the explanations they received – students did not really know what these variables exactly stood for. Consequently, some of the students used the variables as pieces of a puzzle, which they rearranged until the model output resembled the system simulation. In future studies, it might be better to have the students invent the variables themselves.

In designing learning environments for inquiry modeling support should be built in for the students, directed at the process of modeling according to a systematic strategy, as well as to specific reasoning activities. An option is a process support tool that indicates a template of the steps students could take and (content) specific support for each of these steps, similar to the process worksheet as introduced by [van Merriënboer, Clark, & de Croock \(2002\)](#), or the process advisors in [White & Shimoda \(1999\)](#). Given the unclear status of the order of processes we would argue for the support to be non-prescriptive in this respect. In other words, a process support tool should offer a description of the processes that need to be performed as well as their interdependencies, but not fix the order of execution of these processes.

Reasoning activities that seem to be associated with more productive modeling are the generation of hypotheses and the evaluation of the students' model. The use of a suitable graphical representation seems to be supportive, however, the representation should not be 'too easy' because it provides variables and relations that can be effortlessly combined into a running model. Such representations invoke 'puzzle-like' behavior.

Acknowledgements

We thank the teachers Frank Schwertzel, Bruno Velthuis and Jan Rijswijk for their participation in the study reported in this article. This research project is funded by the Dutch Research Organization NWO under grant number 411.21.115.

Appendix A

Complete list of activities contained in the protocol coding scheme:

Scientific reasoning

Orientation

Hypothesizing

Hypothesis

Quantitative

- With no substantiation
- Substantiated with experimental data
- Substantiated with mechanism

Qualitative

- With no substantiation
- Substantiated with experimental data
- Substantiated with mechanism

Predicting

Experimenting

Experimental design

Of the SimQuest simulation

- With no hypothesis
- With qualitative hypothesis
- With quantitative hypothesis

Of the own model

- With no hypothesis
- With qualitative hypothesis
- With quantitative hypothesis

Data interpretation

Model implementation

Model evaluation

Other categories

Actions

Talking about the tool

Reading

Calculating

Regulation

Off Task

Experimenter

References

- ACT, Inc. (2001). ACT Assessment Homepage. Retrieved 03-01-2001. Available from <http://www.act.org/aap/index.html>.
- Bliss, J. (1994). From mental models to modeling. In H. Mellar, J. Bliss, R. Boohan, J. Ogborn, & C. Tompsett (Eds.), *Learning with artificial worlds: Computer based modeling in the curriculum*. London: The Falmer Press.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21–32.
- Dewey, J. (1938). *Experience and education*. New York: MacMillan.
- Elshout-Mohr, M., van Hout-Wolters, B., & Broekkamp, H. (1998). Mapping situations in classroom and research: eight types of instructional-learning episodes. *Learning and Instruction*, 9, 57–75.
- Erkens, G. (1998). *Multiple episode protocol analysis (MEPA 3.0) (Internal publication)*. Utrecht, The Netherlands: Department of Educational Sciences, Utrecht University.
- Feyerabend (1975). Against method: outline of an anarchistic theory of knowledge. London, NLB.
- Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics*, 60, 732–748.
- Hmelo-Silver, C. E., Nagarajan, A., & Day, R. S. (2002). It's harder than we thought it would be: a comparative case study of expert–novice experimentation strategies. *Science Education*, 86, 219–243.
- Hogan, K., & Thomas, D. (2001). Cognitive comparisons of students' systems modeling in ecology. *Journal of Science Education and Technology*, 10, 319–344.
- Jackson, S. L., Stratford, S. J., Krajcik, J. S., & Soloway, E. (1996). Making dynamic modeling accessible to pre-college science students. *Interactive Learning Environments*, 4(3), 233–257.
- de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68, 179–201.
- de Jong, T., van Joolingen, W. R., Lazonder, A., Ootes, S., Savelsbergh, E. R., & Wilhelm, P. (2002). Co-Lab specifications; Part 1 Theoretical background (Technical Report). Enschede, The Netherlands: University of Twente.
- van Joolingen, W. R., & de Jong, T. (1993). Exploring a domain through a computer simulation: Traversing variable and relation space with the help of a hypothesis scratchpad. In D. Towne, T. de Jong, & H. Spada (Eds.), *Simulation-based experiential learning* (pp. 191–206). Berlin: Springer.
- van Joolingen, W. R., & de Jong, T. (2003). SimQuest, authoring educational simulations. In T. Murray, S. Blessing, & S. Ainsworth (Eds.), *Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive, and intelligent educational software* (pp. 1–31). Dordrecht, The Netherlands: Kluwer.
- van Joolingen, W. R., King, S., & de Jong, T. (1997). The SimQuest authoring system for simulation-based discovery environments. In B. Du Boulay & R. Mizoguchi (Eds.), *Knowledge and media in learning systems* (pp. 79–87). Amsterdam, The Netherlands: IOS.
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, 18, 495–523.
- Lewalter, D. (2003). Cognitive strategies for learning from static and dynamic visuals. *Learning and Instruction*, 13, 177–189.
- Löhner, S., van Joolingen, W. R., & Savelsbergh, E. R. (2003). The effect of external representation on constructing computer models of complex phenomena. *Instructional Science*, 31, 395–418.
- Masui, C., & de Corte, E. (1999). Enhancing learning and problem solving skills: orienting and self-judging, two powerful and trainable learning tools. *Learning and Instruction*, 9, 517–542.
- Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: learning processes and instructional support. *Journal of Research in Science Teaching*, 30, 821–844.
- O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology*, 35, 34–70.
- Penner, D. E. (2001). Cognition, computers and synthetic science: building knowledge and meaning through modeling. *Review of Research in Education*, 25, 1–36.

- Schär, S. G., Schierz, C., Stoll, F., & Krueger, H. (1997). The effect of the interface on learning style in a simulation-based situation. *International Journal of Human–Computer Interaction*, 9(3), 235–253.
- Schecker, H. P. (1998). *Physik – Modellieren, grafikorientierte modelbildungssysteme im physikunterricht* [Physics – modeling, graphically oriented model building systems in physics instruction]. Stuttgart, Germany: Ernst Klett Verlag GmbH.
- Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction*, 13, 191–203.
- Stratford, S. J., Krajcik, J., & Soloway, E. (1997, March). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Paper presented at the annual meeting of the American educational research association*, Chicago, IL.
- Suthers, D. D. (1999, December). Effects of alternate representations of evidential relations on collaborative learning discourse. *Paper presented at the computer support for collaborative learning (CSCL) conference*, Palo Alto, CA: Stanford University.
- van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: the 4C/ID*-model. *Educational Technology, Research & Development*, 50(2), 39–64.
- Veermans, K., van Joolingen, W. R., & De Jong, T. (2000). Promoting self-directed learning in simulation based discovery learning environments through intelligent support. *Interactive Learning Environments*, 8, 257–277.
- White, B. Y., & Shimoda, T. A. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: computer support for metacognitive development. *International Journal of Artificial Intelligence in Education*, 10, 151–182.
- Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, 21, 179–217.