Routledge
Taylor & Francis Group

RESEARCH REPORT

The Difficult Process of Scientific Modelling: An analysis of novices' reasoning during computer-based modelling

Patrick H. M. Sins[a]*, Elwin R. Savelsbergh[a,b] and Wouter R. van Joolingen[a,c]
[a]*University of Amsterdam, The Netherlands;* [b]*University of Utrecht, The Netherlands;* [c]*University of Twente, The Netherlands*

Although computer modelling is widely advocated as a way to offer students a deeper understanding of complex phenomena, the process of modelling is rather complex itself and needs scaffolding. In order to offer adequate support, we need a thorough understanding of the reasoning processes students employ and of difficulties they encounter during a modelling task. Therefore, in this study 26 students, working in dyads, were observed while working on a modelling task in the domain of physics. A coding scheme was developed in order to capture the types of reasoning processes used by students. Results indicate that most students had a strong focus on adjusting model parameters to fit the empirical data with little reference to prior knowledge. The successful students differed from the less successful students in using more prior knowledge and in showing more inductive reasoning. These observations lead to suggestions for the design of appropriate scaffolds.

Introduction

Educational Value of Modelling

The educational value of computer-based dynamic modelling has been advocated by many authors. Some emphasize the importance of the model as an artefact that allows explicit visual representation of complex relations (for example, Bliss, 1994; Hogan & Thomas, 2001; Mandinach, 1988; Schecker, 1993, 1994; Steed, 1994). Others put more stress on the activity of constructing a model as a meaningful

*Corresponding author. Graduate School of Teaching and Learning, University of Amsterdam, Wibautstraat 2–4, Amsterdam 1091 GM, The Netherlands. Email: P.H.M.Sins@uva.nl

learning experience (for example, Milrad, Spector, & Davidsen, 2002; Penner, 2001; Spector, 2000; Stratford, 1997). This activity offers the opportunity for students to think scientifically about the behaviour of complex phenomena (for example, Bliss, 1994; Hestenes, 1997; Jackson, Stratford, Krajcik, & Soloway, 1996; Wild, 1996), to reflect upon their own understanding (for example, Gilbert, Boulter, & Rutherford, 1998; Jonassen, Strobel, & Gottdenker, 2005; Raghaven, Satoris, & Glaser, 1998; Schecker, 1993), and to test their mental models (Coon, 1988; Doyle & Ford, 1998; Penner, 2001). Crucial in this process is that modelling tools help students to externalize their ideas, so that they are open to criticism and discussion (Devi, Tiberghien, Baker, & Brna, 1996; Rouwette, Vennix, & Thijssen, 2000; Suthers, 1999). However, constructing a dynamic model is a complex task, and it may not be surprising that novice modellers encounter problems in performing this task. In the literature several types of difficulties are reported with regard to the task perception, the content addressed, and the tools used.

At the level of *task perception*, it has been found that students tend to view a modelling task as an engineering problem rather than a scientific one. Instead of focusing on the meaning of the model they focus on its output, and without proper reasoning the behaviour turns to model fitting; that is, tuning model parameters until the model output resembles the observed empirical data. Apart from a probable lack of success of this behaviour, no "deep" reasoning on the model elements or model structure will take place (Hogan & Thomas, 2001; Ogborn, 1999; Stratford, Krajcik, & Soloway, 1998). Such model fitting behaviour leads to a disconnection between the model and content knowledge: the model becomes an artefact that has to "work", not something that provides explanatory power in understanding a phenomenon (for example, Bliss, 1994; Hogan & Thomas, 2001). Moreover, some students even do not detect mismatches between the model output and the expected behaviour of the phenomenon being modelled (Doerr, 1996; Steed, 1994; Whitfield, 1988). Finally, using empirical data gathered to generate hypotheses is a difficult process for students (de Jong & van Joolingen, 1998).

At the *content* level, students have difficulties conceptualizing the complex phenomena that are typically addressed in computer-based modelling. Typical difficulties are the time dependence of variables and multiple processes that cancel out. Students tend to consider the influences of individual variables separately (for example, Doerr, 1995; Hogan & Thomas, 2001; Kainz & Ossimitz, 2002; Stratford et al., 1998). Also feedback mechanisms are profoundly difficult. If there is a feedback loop present in the system (e.g., the amount of interest on a bank account leads to an increase of the main sum, which in turn leads to a higher interest next year) students often fail to reason about interactive variables and display linear causal thinking instead, which means that students trace one cause to one effect (for example, Kurtz dos Santos & Ogborn, 1994; Löhner, Van Joolingen, & Savelsbergh, 2003; Riley, 1990; Zohar, 1995).

At the level of the *tool*, students find it difficult to express their ideas in a modelling formalism. A typical modelling formalism, also used in the study

presented in this article, is system dynamics (Forrester, 1961). System dynamics employs Stocks (represent the central quantity of the model that builds up over time), flows (determine how quickly the stock changes), auxiliaries (factors that influence the flows), constants (factors with fixed values), and connectors (linkage arrows to show direction of proposed relationships in a system). Expressing ideas in this (or other) formalisms requires learning and involves gaining some experience. It has been found, for example, that students find it particularly complex to represent flows as concrete elements in their model (Hogan & Thomas, 2001; Kurtz dos Santos & Ogborn, 1994). Also, students have difficulties in deciding the type of the variable they would like to implement into their model and, in addition, students frequently struggle in specifying the mathematical relationships between variables in the model (Cox & Webb, 1994; Ossimitz, 2002; Sweeney & Sterman, 2000; Tinker, 1993). At last, students find it difficult to translate their own knowledge of the phenomenon into a computer model using the formalism of an icon-based modelling tool (Ainsworth, 1999; Kurtz dos Santos & Ogborn, 1994).

These studies point to several structural difficulties students encounter during the modelling process. However, to frame these problems and to provide appropriate support, we need a thorough exploration of the reasoning processes students employ during modelling. As such, process-oriented studies can shed more light onto how levels of reasoning processes interact and contribute to successful or less successful modelling.

Reasoning Processes During Modelling Activities

Several studies have aimed at describing novices' reasoning processes. Most studies have taken inductive viewpoints, either in the form of individual case studies (Hogan & Thomas, 2001; Schecker, 1998; Stratford et al., 1998) or quantitative analyses (for example, Fretz et al., 2003; Löhner, Van Joolingen, Savelsbergh, & Van Hout-Wolters, 2005; Zhang et al., 2002). Other researchers have taken a normative stance, based on expert views, leading to descriptions of the ideal modelling process (de Jong et al., 2002; Hestenes, 1987; Schecker, 1998; White & Frederiksen, 1998). An overview of the outcomes from the aforementioned studies was presented by Löhner et al. (2005). This framework provides the basis for our explorative analysis. We reconsidered some labels and definitions to encompass representative global *reasoning processes* (see Figure 1).

This led to the following categories: Analyse, Inductive Reasoning, Quantify, Evaluate, and Collect Data. Collecting data involves the gathering of information through experimental inquiry. Because our focus is on the modelling process and because data collection and interpretation through inquiry involves many difficulties of its own (for example, Kuhn, 1989; Chinn & Brewer, 1993; de Jong & van Joolingen, 1998), collecting data will not be further addressed in this article. The category "Explain" was added to this framework to involve processes in which students articulate explanations to others. The way in which students justify model actions
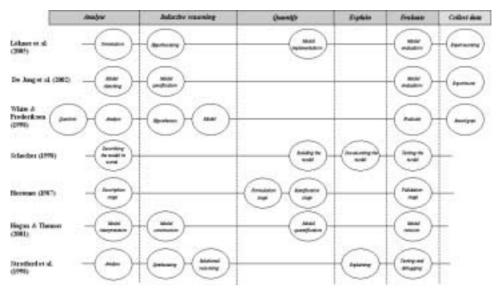
Figure 1.   Overview of reasoning processes found in studies on computer-based modelling
(adapted from Löhner et al., 2005)

demonstrate how elaborate they reason about the material (for example, Schecker, 1998; Stratford et al., 1998).

The reasoning processes specifically associated with modelling are as follows:

- *Analyse*: When students are analysing, they decompose the phenomenon they are studying into parts and identify important model elements (i.e., quantities, or relations between quantities) to be implemented in their model. Also, students interpret model output or empirical data that are presented in tables or graphs. Most of the modelling activities associated with this reasoning process are performed during the orientation phase of a modelling task. During Hestenes' (1987) "Description stage", for example, students decide on the type of model that will be constructed. In addition, they identify the variables and relations that have to be implemented in their model. Hogan and Thomas' (2001) define "Model interpretation" as a process involving exploring output, which is produced as tables or graphs. This involves students analysing graphs and tables identifying how variables increase or decrease over time. Both definitions were taken in our synthesis as indications of the process of analysing.
- *Inductive Reasoning*: Inductive reasoning occurs when students conjecture hypotheses on how model elements interact and on how the model should behave. This process implies a great deal of elaboration on the relationships between the model structure and the behaviour of the phenomenon being modelled, which makes it a complex process for students to perform (for example, Schecker, 1993; Tinker, 1993). Stratford et al. (1998) use the term "Synthesizing", which involves students making statements about the content,

behaviour, or structure of the model as a whole (e.g., considering how the model should behave, discussing the representation of their model, and discussing new relationships between quantities in the model), which can be categorized as a process of inductive reasoning.

- *Quantify*: When students construct a preliminary model, they can make their ideas about model elements and relations more precise by expressing them into an executable mathematical format. This implies that quantities in the model are specified with a starting-value and relations are worked out in equations. This process of quantifying a model comprises the processes employed during Hestenes' (1987) "Formulation stage" and "Ramification stage" to a great extent. The "formulation stage", according to Hestenes (1987), involves students using knowledge of physical laws to determine definite equations for the phenomenon that is modelled. During the "Ramification stage" the special mathematical properties and implications of the model are worked out. This stage implies that equations are solved out by experimenting with parameters and formulas.
- *Explain*: Involves the process in which students clarify to each other why model elements are related; that is, they document the reason(s) why one factor causes changes in another. The phase Schecker (1998) designates as "Documenting the model" clearly falls under this process.
- *Evaluate*: Finally, students have to connect between the output from their model and results obtained from experiments in order to evaluate and ultimately test their model. In evaluating their model, students determine whether their model is consistent with their own beliefs, with data obtained from experiments and/or with descriptions of behaviour about the phenomenon being modelled. Model evaluation leads to model revision activities, which involves modifying parts of the model so that it better describes or explains a given situation. The process of "Model revision" in Hogan and Thomas' (2001) study involves similar activities. They define "Model revision" as a process in which students assess the degree of fit between model output and expected or empirically confirmed patterns.

The purpose of the present study is to understand the specific reasoning processes that play a role during novice modellers' activities. In order to gain insight into these reasoning processes we need to investigate the occurrences as well as the quality of these processes. More in particular, as the paragraph on student difficulties made clear, students may focus on particular aspects of their model, while ignoring other aspects (such as individual variables versus interacting variables or global model structure) as well as use inappropriate arguments, or even no arguments at all, to justify their reasoning. Therefore, in addition to identifying types of reasoning, the focus of the conversation and the types of argumentations used need to be understood. For these features—focus and argumentation—we developed a scheme for analysis in a more inductive fashion on basis of the obtained protocol data. The main research question in this study therefore is:

- Which reasoning processes, employed by novice modellers during a computer-based modelling task, need support?

This question encompasses the following subquestions:

1. What features are relevant to describe novices' reasoning processes during computer-based modelling?
2. What distinguishes successful from less successful novice modellers?
3. Which reasoning processes are difficult for novice modellers to perform?

Method

Participants

This study involved 38 students from 11th-grade pre-university education, with a major in science. Students had no prior experience with system dynamics models. Students' age ranged between 16 and 18 years. During the task, participants worked in pairs, which were composed by having the students choose their own partners from within a group of familiar students.

Materials

Participants were presented with a task asking them to explore and revise a model that described the distance covered by an ice-skater.[1] Since participants had no prior experience with modelling, a completely open modelling task would be too complex for them to be successful within the time constraints of the experiment. Therefore, participants were given an incomplete model as a starting point. Such a model revision task enables the novice modeller to concentrate on trying to comprehend and improve a model without having to start from scratch. The modelling task was implemented in PowerSim, a modelling tool based on system dynamics (see Figure 2). PowerSim uses the five model building blocks characteristic for system dynamics modelling: stocks, rates, auxiliaries, constants, and connectors. To insert a modelling element, students can drag and drop the icons on the screen they think are relevant for the phenomenon being modelled, creating a qualitative diagram of the phenomenon. While creating this diagram, students can quantify these elements by entering values and formulas. Once the model is quantified it can be executed. When students run their model, PowerSim automatically generates the differential equations required to perform calculations. The results of simulations runs over time can be displayed as graphs or tables.

The modelling task was presented in a cover story in which a scientist attempted to construct a model of this phenomenon. Participants were provided with measurements that were obtained by the scientist, which they could use to test the model. The empirical data was presented in two graphs, one for the distance covered by the skater (see upper right-hand side of Figure 2) and one graph for the velocity of the skater (lower right-hand side of Figure 2). Participants' task was to revise their model in such a way in that it would provide a good match with the data. Successful
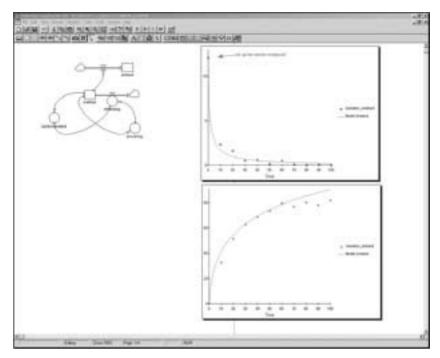
Figure 2.   Screenshot of a model in PowerSim

completion of the task would require the identification of two friction forces and of a feedback loop that runs from velocity to air resistance. This feedback implies that a skater at a higher velocity experiences more air friction, which consequently leads to a more rapid decrease in velocity.

The present study made use of data that were collected in order to compare the effects of different initially provided models and the effects of two sets of data that differed in quality. Therefore, dyads received slightly different versions of the modelling task. For the present purpose data could be pooled, since no significant differences were found between these settings on the dependent measures employed in this study.

Procedure

In order to become acquainted with system dynamics modelling in PowerSim, each student individually worked through an instruction manual. This manual was adapted from "Computer-supported modelling: Manual PowerSim",[2] which is developed and disseminated to Dutch schools by the Centre for Science and Mathematics Education at the University of Utrecht. In this instruction manual, students are presented with an example model of a water tank. The simplest system imaginable to illustrate fundamental aspects of modelling and the behaviour of dynamic systems is that of a water tank containing a faucet and a drain. The volume of water in the water tank is represented by the stock variable (i.e., reservoir variable), and

the flows represent respectively the inflow of water into the water tank via the faucet and the outflow of water from the water tank via a drain. The water tank model, its elements (i.e., variables and relationships between variables) and how it can be built in PowerSim were explained to students in the instruction. Also, students could execute and revise parts of the water tank model in the PowerSim environment. The instruction took about 1 hr. Subsequently, participants were grouped into dyads. Participants were informed that they were going to explore and subsequently revise models working in couples. Next, dyads read the modelling task and were presented with the initial model version and with the data. Dyads were asked to collaboratively revise the model for approximately 1.5 hr.

Data Collection

The primary source of data consists of the verbal protocols of the collaborating dyads. Modelling actions and verbal communication between students were obtained using the program Lotus ScreenCam™. This program recorded all onscreen actions and audio. Verbal protocols were obtained by transcribing these recordings. After an initial qualitative, exploratory analysis, the transcripts were subsequently segmented into episodes. Episodes were scored using an analysis scheme that was developed on the basis of the qualitative analyses of the protocol data using our framework of reasoning processes as reference. Scoring the protocols was performed employing the program MEPA (Erkens, 1998). Unfortunately, because of recording software failure the verbal protocols of only 13 dyads could be analysed.

The quality of revised models was also assessed. The scoring was based on both the degree of model fit and the conceptual structure of the model. The model fit score, ranging from one to five points, was based on face value. The score for structure consisted of a score for correct/incorrect quantities and a score for correct/incorrect relationships between quantities. For each correct quantity or relation in the model, one point was awarded. For each incorrect quantity or relation two points were subtracted. The model structure score was subsequently rescaled to a five-point scale. Addition of the two components led to a total model quality score on a 10-point scale.

Results

Exploratory Case Studies

The goal of the data analysis was to characterize the reasoning processes students employ during computer-based modelling activities. To that end, we started with an exploratory qualitative analysis of the performance of selected dyads. Selection was based on their model scores: we chose a high, a middle, and a low scoring dyad. As an initial scheme of analysis we used our synthesis of reasoning processes identified in the literature as a starting point. In examining these protocols, the focus of students' discussions and type of argumentation were also considered. Based on

these ideas and a detailed study of the protocol data we subsequently modified the analysis scheme in several rounds of analysis.

Case 1: High performing dyad

Dave and Roel created an excellent model (model score: 10), which was also reflected in the quality of the processes in which they engaged. Dave and Roel were very systematic in their approach, very elaborate in their reasoning, and critical in their discussions about their model. While implementing model revisions they carefully considered the semantics of the variables, relations, and the behaviour of their model. Also, they attempted to keep their model as simple as possible during the whole task.

   The critical stance of Dave and Roel towards their model is exemplified by the following episode. Directly before this episode they were trying to fit the output of their model to the given data by adjusting parameters. Since their attempts remained without success, they engaged in an elaborate model evaluation:

1	D	Because ice friction is a force, deceleration is acceleration, how much does the skater decelerate per time unit
2	R	Negative acceleration
3	D	Negative acceleration
4	R	So the ice friction
5	R	But look, deceleration in the model is equal to ice friction, while the ice friction also has to be negative, so that is not right
6	D	So that is a mistake in our model
7	R	Let's read the assignment again
8	D	Let's start with the most important, we have the ice friction
9	D	Because we have to go from force to uhh velocity
10	R	Force, is it not possible to do that with the Work law?
11	D	Yes, but it would come in handy if we had the formula's, but I do not know them by heart
12	R	It says here that ice friction is equal to deceleration, that is not right
13	D	That is absolutely not possible, there has to be a step in between

(Protocol: ac_16, start: 0′16′20, duration: ′52s, Process-code: evaluate; Focus-code: relation; Argumentation-code: physics knowledge)[3]

In this excerpt, Dave and Roel are *evaluating* a *relation* between variables. They conclude that the relation is incorrect and try to come up with a correct relation between these variables *using their knowledge of physics* as argumentation. They argue that the two variables in their model (i.e., "ice friction" and "deceleration") cannot be related in the way it was implemented (lines 1–5). Therefore, they try to find a physical law or formula that includes both variables in order to be able to compute one from the other (lines 8–13). Although they seem to know the relevant formulae, it appears that they had difficulty in translating these formulas into their computer model.

   Dave and Roel acknowledged the importance of focusing on how the different relationships in their model affected the model output, in order to figure out how their model works:

R   So we have to ask ourselves when the skater is at 0 m/s and how much he must skate
D   Yes, only he is not constant
D   If velocity decreases
R   The deceleration is constant
D   No, not any more
R   Why not?
D   Because air friction is now dependent on your velocity
R   Oh, let's delete air friction then
D   Let us first begin with uh
R   It is not realistic
D   Let's first investigate the influence of ice friction on the skater's velocity
(Protocol: ac_16, start: 0′37′40, duration: ′25s, Process-code: Inductive reasoning; Focus-code: model structure; Argumentation-code: experimental data)

In this episode, Dave and Roel *reason inductively* about relationships in their *model*, concluding that they have to proceed investigating individual relationships in order to comprehend the behaviour of their model. Therefore, Roel and Dave delete the variable "air friction" from their model in order to separately investigate how ice friction affects the velocity of the ice-skater.

An episode in which Dave and Roel *reasoned inductively* about a key *relationship* in their model, shows that they also base their arguments on *everyday experience*:

D   Yes do you see, you have to do something with air friction in order to relate it to velocity
D   Air friction depends on your velocity, if you go faster then your air friction is higher. If the wind goes faster, if you have a higher wind velocity, then your air friction is higher. When there is no wind, you do not notice air friction if you are riding your bike, when wind force is 12 then you will not be able to go forward
D   If you are riding your bike and you ride with 1 km/h than you would not feel a thing
R   The harder you ride, the better
D   But if you go with 20 km/h, for example from a bridge, then you will notice the air friction
D   So the air friction comes in handy, now you have to combine it with velocity
(Protocol: ac_16, start: 0′42′12, duration: ′42s, Process-code: Inductive reasoning; Focus-code: relation; Argumentation-code: experiential knowledge)

In this section, Dave and Roel reason about the relation between "air friction" and "velocity": air friction increases with higher velocity as we know from everyday experience. This leads Dave and Roel to recognize the need for the implementation of a feedback loop (i.e., between velocity and air friction) in their model.

Dave and Roel engaged in an elaborate approach to the modelling task, which was reflected in the use of high-quality reasoning processes. They often reasoned about relationships between quantities in their model and critically evaluated how their model worked. Mostly, their focus was on relationships between variables or on the structure of their model. In addition, they frequently referred to experiential knowledge as an argumentation for model revisions. Finally, they often referred to physics formulas they had learned in class.

Case 2: Medium performing dyad

Laurette and Hinde created a reasonable model (model score: 7). They mainly analysed and identified individual elements in their model without elaborating much on how their model could be improved. Instead of employing their own knowledge during modelling, they often referred to the degree of model fit as argumentation for model revisions. They also showed a concern for their model being a realistic representation of the phenomenon they were modelling.

In a significant part of their protocol Laurette and Hinde were engaged in unsuccessful model fit behaviour:

L	The green line {the model graph} has to run as the other
H	Yes! It has to go through the little dots {the data}
L	Oh
H	But uh
H	Yes, you know if you notice the graph, like the one we just had, than we had given ice-friction the value 1
L	Yes
H	And then the graph went through the first dots, this graph, look

(Protocol: an_23, start: 0′10′17, duration: ′24s, Process-code: quantify; Focus-code: quantity; Argumentation-code: correspondence model graph and data)

In this episode Laurette and Hinde *quantify* the *quantity* "ice friction" in their model to improve the *fit between the model graph and the data points*. Note that Laurette and Hinde implement this revision without further thinking about why their model yields this particular output.

Next to this superficial model fitting, they primarily *analysed* and discussed (relevant) modelling entities (i.e., *quantities*):

L	This one {the model graph} is more correct
H	Hmmhmm
L	Shall we include another variable, did we forget something?
H	Yes, you also have air resistance, is that right?
L	Eeeuh yes but
H	But that has nothing to do with the ice-skater in this problem

(Protocol: an_23, start: 0′17′58, duration: ′12s, Process-code: analyse; Focus-code: quantity; Argumentation-code: experiential knowledge)

Here, Laurette and Hinde identify a relevant variable that may be included in their model (i.e., "air resistance") but refute the idea on the basis of *experiential knowledge*.

In a few episodes in which Laurette and Hinde were *reasoning inductively* about their *model*, but they did not refer to relevant school physics knowledge:

H	And this arrow indicates that it, that the velocity
L	That it is a cycle, so if velocity is becoming less
H	The ice friction
L	Influences the distance
L	Yes, the distance becomes less but the velocity with which the distance decreases
L	That changes
H	Eeeeuhm
L	But

H   A little
L   Deceleration is equal to ice friction
H   Do we have something that also influences the deceleration?
(Protocol: an_23, start: 0′18′23, duration: ′31s, Process-code: Inductive reasoning; Focus-code: model structure; Argumentation-code: none)

In general, Laurette and Hinde were mainly engaged in analysing modelling elements and did not justify their model revisions. The focus of Laurette and Hinde was similar to that of Dave and Roel, but the quality of their reasoning processes was less, because of a lack of argumentation. Nonetheless, Laurette and Hinde showed a preference for applying experiential knowledge in their thinking about their model.

Case 3: Low performing dyad

Marije and Lola constructed a poor model (model score: 3). They spent much of their time attempting to adjust the model parameters to match the graph to the experimental data. They hardly came up with structural model revisions themselves. They relatively often requested guidance from the experimenter. These students did not seem to understand the purpose behind their model in specific and behind computer modelling in general. The main part of the protocol obtained from this dyad consists of episodes in which they were *quantifying* individual *quantities*:

M   Here, wait try this one
M   No, that one
L   No, that is not correct
M   No
L   There has to be something, ice friction
L   I set ice friction on 8
M   Let's see what it {graph} does
(Protocol: an_20, start: 0′08′02, duration: ′17s, Process-code: quantify; Focus-code: quantity; Argumentation-code: none)

Marije and Lola frequently quantified quantities in their model without consideration of why they chose particular values. It was quite difficult for Marije and Lola to see what specific kind of change was needed in order for them to get the model output fit the empirical data:

M   No
L   HHe {the graph} is becoming longer, much longer
M   Much longer indeed
L   This is not correct
M   Huh?
L   This is not okay
M   Eeeuhm
L   It must decrease much faster from here
(Protocol: an_20, start: 0′19′29, duration: ′13s, Process-code: evaluate; Focus-code: model fit/ model output; Argumentation-code: none)

This protocol fragment shows Marije and Lola *evaluating* the degree of *model fit*, but they do not know how to revise their model such that the fit improves.

In cases in which Marije and Lola *identified quantities* to include in their model, they did not provide any argument for this revision:

> M   Shall we make another square for the air-friction?
> L   huh?
> M   A square{i.e. constant}for air-friction?
> L   Yes
> (Protocol: an_20, start: 0′20′28, duration: ′06s, Process-code: analyse; Focus-code: quantity; Argumentation-code: none)

In this episode, Marije and Lola suggest to include the variable "air-friction" as a constant to their model, without clarifying why they want to add this quantity.

In general, Marije and Lola's approach to modelling can be represented as model fit behaviour. Their main focus was on individual quantities and they did not take into account relations between variables. Finally, they did not refer to their own experiential knowledge during modelling and did not provide evidence for their claims. Marije and Lola only shallowly processed the model task and they did not make the link between the behaviour and structure of their model.

These case descriptions indicate the range of reasoning processes that were to be found in the protocols of the other dyads participating in our study (see Table 1). In the example excerpts given earlier, it becomes clear how these episodes can be attributed to the reasoning categories listed in the introduction. However, for a great deal of episodes no clear reasoning process could be discerned, as can be seen in Table 1. Consequently, new categories had to be added to our initial scheme in order to capture these activities.

Table 1.   Percentage of time spent on global reasoning processes and model score for each dyad

| | Global reasoning process | | | | | | Model score (maximum score: 10) |
Dyad	Analyse	Inductive Reasoning	Quantify	Explain	Evaluate	Other	
Mark & Hugo	2.21	12.95	31.46	0.58	11.46	41.34	8
Jordy & Harry	7.04	9.96	22.68	1.93	8.97	49.42	7
Dave & Roel	**7.75**	**16.22**	**35.44**	**1.98**	**3.39**	**35.22**	**10**
Harma & Annemarie	6.99	7.41	32.82	0	8.51	44.27	5
Jaap & Robert	7.7	5.08	35.99	0	13.03	38.2	6
Rik & Corry	5.13	6.62	37.52	0	5.24	45.49	7
Gretha & Anique	3.31	8.19	39.07	0	4.88	44.55	5
Laurette & Hinde	**9.27**	**11.36**	**35.57**	**0**	**4.4**	**39.4**	**7**
Jos & Nadira	3.5	9.06	37.34	0	2.9	47.2	6
Marije & Lola	**6.38**	**3.86**	**49.34**	**0**	**11.05**	**29.37**	**3**
Paul & Benni	0.56	7.26	43.31	0	3.79	45.08	4
Denise & Astrid	4.23	4.62	38.53	1.1	6.46	45.06	4
Stephan & Sjoertje	2.38	0.74	31.03	0	1.94	63.91	8

The cases also show the need for additional scoring of students focus (i.e., the model elements students consider) during modelling and scoring of prior knowledge used, to grasp the differences in quality of the reasoning processes. For example, the less successful group (i.e., Marije and Lola) tended to focus on individual variables, which suggests that they did not see how model behaviour depends on the influence of interacting variables. This is in contrast to Dave and Roel who did reason about relations. From the aforementioned it becomes clear that, in order to describe the reasoning processes of novice modellers, we need a coding scheme that takes into account the *focus* of reasoning as well as the underlying *arguments* that are put forward by the modeller.

Coding the Data

Processes like analysing or explaining involve several turns by both partners in a dyad. Therefore, the unit of analysis was determined to be at the "episode" level, an episode being a period of coherent continuous talk on a single issue, rather than single utterances. As a practical operationalization, episodes were segmented on the basis of the following non-content criteria (cf. Chi, 1997):

- Following each run of the students' model.
- Following an interval of more than 15 s during which nothing is said.
- The interval during which the experimenter intervenes in the modelling process is a segment by itself.

In the majority of cases these criteria would lead to acceptable segments; however, some segments were of a much too long duration. Therefore, an additional criterion was applied:

- The maximum length for a segment is 1 min. If a segment lasts longer, the segment will be more closely analysed in order to see whether segmentation is possible on the basis of changes in reasoning process or changes in focus (often signalled by words as: "Okay …" or "Now ….").

The segments obtained using this procedure were classified according to the reasoning process that was employed by students. We employed the definitions of the five clusters of reasoning processes (i.e., Analyse, Reason, Quantify, Explain, and Evaluate) identified in our synthesis of the literature. Four categories were added to the coding scheme in order to characterize episodes in which no clear reasoning process was shown, namely:

- *Guiding by experimenter*: Guidance by the experimenter involves: (a) providing an explanation of the assignment, (b) explaining the tools and formalism used in PowerSim, and (c) encouraging the collaboration.
- *Read and paraphrase*: Students read, paraphrase, or discuss the assignment or modelling actions. When students are talking about the assignment or when they

mention the actions they are performing, that episode is coded with read and paraphrase.

- *Off task*: Students talk about subjects that are unrelated to the assignment at hand.
- *Other*: Other processes that are not included in the analysis scheme. For example, inaudible murmur is coded as other.

For coding the focus of reasoning we introduce the following categories:

- *Quantity*: Students discuss quantities (i.e., constant or stock).
- *Relation*: Students discuss relations or interaction(s) between quantities.
- *Model output*: Students discuss the output generated by running the model, including the degree of fit between the output generated by their model and the experimental dataset.
- *Data points*: Students explicitly discuss the experimental data in the graph.
- *Model structure*: Students discuss the global structure of the model, for example: How the quantities are linked. In order to score this category, the episode has to involve more than one relation in the model.
- *Modelling actions/the tool*: Students mention modelling actions, adding or deleting quantities, running the model, or adding or deleting relations. This category is also scored when students are trying to figure out how PowerSim (i.e., tools, buttons, and formalism) works.
- *The assignment*: Students talk about the modelling assignment (i.e., the ice-skater problem).

Finally, we categorized the types of argumentation employed in reasoning:

- *None*: Students make no reference to specific knowledge in their reasoning.
- *Physics knowledge*: Students use terminology, concepts, or formulas that are used in physics. For example, if units for variables in the model are mentioned.
- *Mathematics knowledge*: Students use terminology, concepts, or formulas that are used in mathematics. For example, if students talk about the mathematical function of the model graph.
- *Experiential knowledge*: Students refer to experiential knowledge. For example, if students mention their own experiences.
- *Correspondence between model graph and data*: Students refer to (mis)match between the model output and experimental data. For example, students quantify a quantity in their model and refer to the degree of match between the data and the model graph (i.e., model fit behaviour).
- *Experimental data*: Students refer to the data points without explicit reference to the model fit.

The complete coding scheme and the criteria for assigning the codes are presented in the Appendix. Inter-rater reliabilities for each of the three subcodes were determined by comparing the ratings of two independent judges ($n = 202$; Process-code, Cohen's kappa = 0.742; Focus-code, Cohen's kappa = 0.764; and Argumentation-code, Cohen's kappa = 0.533). The inter-rater reliability for the first two

categories can be regarded as satisfactory (Heuvelmans & Sanders, 1993). The Cohen's kappa for the argumentation code is low, and consequently findings with respect to this aspect should be taken with care.

Quantitative Results

The coding scheme was now applied to the entire set of protocols. Because protocols differ in length, and because protocol episodes are of different length as well, reporting frequencies give a skewed image. Therefore, frequencies were converted to proportion of total time for each dyad, and further analyses are based on these proportions. Data from these analyses are presented in Tables 2 and 3.

From Table 2 it becomes clear that the amount (re)reading and paraphrasing (of the modelling actions and/or modelling tool and of the assignment) and experimenter guidance is high for all dyads. Also, the time spent on quantifying is rather high. Many quantifying episodes have their focus on a single quantity, rather than interactions between quantities. Quite often in these episodes reference is made to the degree of correspondence between model graph and data, whereas only a few refer to prior knowledge (see Table 3). The conjunction of these features clearly indicates model fit behaviour. By contrast, most dyads did not spend much time on inductive reasoning; explaining seems to be almost a missing category; and use of prior knowledge is relatively rare.

In order to examine whether students follow a systematic approach, in that there is a preferred sequence of activities, a transition analysis was performed. The analysis was conducted on transitions between episodes, which means that a significant transition occurs when the number of observed transitions between two reasoning processes is significantly higher than may be expected on the basis of the distribution of coding (test of uni-directional transitions using z-scores, taking into account the conditional probabilities for every transition). The significant successions between the global reasoning processes found for the dyads in our study are shown in Figure 3. The line thickness indicates the magnitude of the difference between observed frequency and expected frequency.

What is apparent from Figure 3 is that students frequently switch from analysing, inductive reasoning, evaluating, or reading to guidance from the experimenter and vice versa. This means that after episodes in which students employed these processes, they inquired for support from the experimenter. In addition, Figure 3 reveals that quantifying, reading, and inductive reasoning are persistent activities, in that they tend to extend across multiple episodes.

Finally, to investigate whether particular types of reasoning processes are associated with the quality of the model students created, we computed Spearman rank correlations between types of reasoning processes and the quality of the students' model. There was a significant negative correlation between the model score and the amount of time spent on quantifying quantities without augmentation ($r = -51$, $p = 0.036$). In addition, a significant positive correlation was found between model score and inductive reasoning about relationships between quantities with reference to

Table 2. Percentage of time spent by type of reasoning process and type of focus averaged for all dyads

Reasoning process					Focus				
	Quantity	Relation	Model fit/model output	Data points	Model structure	Modelling actions/the tool	The assignment	n.a.	Total
Guidance	–	–	–	–	–	–	–	13.97	13.97
Evaluate	0.15	0.39	5.72	–	0.37	–	–	–	6.63
Explain	0.15	0.09	–	0.04	–	0.15	–	–	0.43
Quantify	32.80	3.41	–	–	–	–	–	–	36.21
Inductive reasoning	1.34	3.78	0.77	0.25	1.82	–	–	–	7.96
Analyse/identify	2.16	0.63	0.68	0.52	1.03	–	–	–	5.02
Read and paraphrase	–	–	–	–	–	10.70	9.08	–	19.78
Off task	–	–	–	–	–	–	–	3.95	3.95
Other	–	–	–	–	–	–	–	5.95	5.95
Total	36.60	8.30	7.17	0.81	3.22	10.85	9.08	23.87	100

Table 3. Percentage of time spent by type of reasoning process and type of argumentation averaged for all dyads

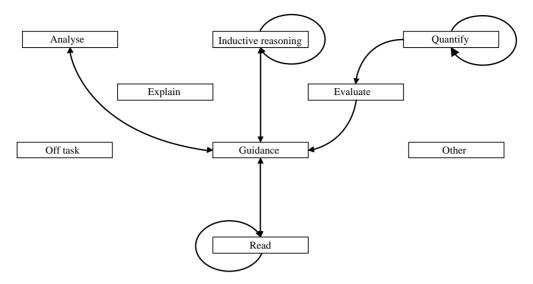| Reasoning process	Argumentation							
	None	Physics knowledge	Mathematics knowledge	Experiential knowledge	Correspondence between model graph and data	Experimental data	n.a.	Total
Guidance	–	–	–	–	–	–	13.97	13.97
Evaluate	4.42	0.28	0.60	0.70	–	0.63	–	6.63
Explain	0.22	0.08	-	0.13	–	–	–	0.43
Quantify	18.59	3.40	0.93	1.61	11.55	0.13	–	36.21
Inductive reasoning	2.50	1.32	0.47	2.52	0.76	0.39	–	7.96
Analyse/identify	3.80	0.48	0.07	0.27	0.21	0.19	–	5.02
Read and paraphrase	19.71	–	–	–	–	0.07	–	19.78
Off task	–	–	–	–	–	–	3.95	3.95
Other	–	–	–	–	–	–	5.95	5.95
Total	49.24	5.56	2.07	5.23	12.52	1.41	23.87	100

Figure 3. Transition diagram, displaying significant transitions between reasoning processes

experiential knowledge ($r = 0.46$, $p = 0.049$), and an almost signification positive correlation was found for inductive reasoning with reference to physics knowledge ($r = 0.45$, $p = 0.082$).

Conclusion and Discussion

The first part of our research concerned the features of reasoning processes that are needed to describe novices' computer-based modelling processes. In order to identify these features, we started from a framework based on available research, and then refined this framework on the basis of a qualitative analysis of protocol data. This resulted in a framework containing the following dimensions: (a) type of reasoning process, (b) topic focus, and (c) type of argumentation. Individual case studies made clear that all three dimensions are needed: apart from the *occurrence* of reasoning processes, it is important to also assess their *quality*. The first two dimensions could be reliably scored, whereas the type of argumentation remained difficult to decide on.

The second research question was how successful and less successful novice modellers differ in their reasoning. This question was first answered qualitatively, on the basis of three case studies, and next these answers were corroborated using a quantitative analysis. In the qualitative analysis we found that the more successful students, in contrast to the less successful ones, tended to justify their reasoning in terms of both experiential and physics prior knowledge. The less successful students were more narrowly focused on the model and the model output. Moreover, the more successful students regarded the model more as a whole, taking into account the model structure, whereas the less successful students mostly considered only one quantity or relation at a time. In summary, the weaker students spent a

relatively large part of their time manipulating parameters in order to let the model fit the given data, showing model fitting behaviour. The quantitative analysis of the protocols confirmed this picture. Correlation analysis revealed that students who spent much time on quantifying quantities without argumentation (indicative of model fitting behaviour), arrived at lower quality models. Students who spent their time on inductive reasoning with reference to prior knowledge arrived at better models.

The third research question was which reasoning processes are difficult for novice modellers to perform. It was found that, in general, students encountered a great deal of difficulties during the modelling task. This was indicated by the finding that the percentages of time spent on episodes in which students prompted for support or (re-)read the assignment were high. In addition transition analysis showed that episodes in which students were engaged in analysing, inductive reasoning, evaluating, or reading were often followed by episodes in which they asked for support from the experimenter.

The particular problems students had during the modelling task are discussed according to the three levels at which difficulties are known to occur. At the level of task perception, we found frequent evidence of model fitting behaviour. In addition, our case study revealed that the less successful students were found to be more engaged in model fitting compared with the more successful dyads. This model fitting behaviour was, furthermore, found to be negatively associated with the quality of the students' model. Similar patterns were found in a case study by Ogborn (1999). This indicates that most students were not able to go beyond employing the model as an artefact instead of using the model as a means to comprehend the behaviour of complex phenomena. At the content level, the quantitative analysis indicated that, on average, students did not often connect to prior knowledge, which was taken as an indication that students had difficulties with relating the computer model with their own conception of the phenomenon being modelled (cf. de Jong & van Joolingen, 1998). This is in line with findings in earlier case studies (Hogan & Thomas, 2001; Stratford et al., 1998; Zhang et al., 2002). In addition, it was found that less successful students focused more often on individual variables in their model, implying that they had difficulties with considering interactions between variables. Finally, at the level of the tool, the case study and the quantitative analysis revealed that, even though students had received an instruction in dynamic modelling with PowerSim, they still had difficulties in grasping the formalism used by PowerSim. This was reflected in the finding that students' reasoning focused on the modelling tool during a great deal episodes (cf. Cox & Webb, 1994; Tinker, 1993).

From these findings it can be concluded that modelling of dynamic phenomena is a complex undertaking for novice modellers, and that probably more experience is needed in order to obtain a learning benefit. Consequently, appropriate support should be provided, either in the modelling tool or in the classroom context to scaffold students' reasoning processes. Note that our sample consisted of students with no prior experience with computer-based models. Thus, results of the present

study and implications for scaffolding are applicable only to modellers who start to learn to use system dynamics models.

In the present study, it was found that when students employ their own knowledge during modelling activities, they constructed models of higher quality. Scaffolds should, thus, encourage students to activate their prior knowledge not only during modelling, but also before engaging in any modelling activities. When students are initially prompted to think about variables and relations that could play a role in explaining the behaviour of a dynamic phenomenon, activating whatever knowledge resources they have available, this may serve as an anchor for the further modelling process (for example, Clement, Brown, & Zeitsman, 1989; Hammer, 2000). Ideally, this knowledge activation takes place within a collaborative setting in which students can discuss their models with other groups (for example, Rouwette et al., 2000).

The identification and articulation of specific modelling (sub)goals also forms an essential aspect during the process of prior knowledge activation. When students have clear (sub)goals to attain, the modelling process will be more structured in the sense that students are guided in building an understanding of how the structure of their model influences the behaviour of the model. In the present study, students were found to be primarily engaged in superficial model fitting behaviour, instead of attempting to understand the association between the structure of their model and the phenomenon being modelled. By leading students to examine one model revision at a time, students should be able to discern what types of revisions produce what kinds of output, thereby building their understanding of the system they are modelling (for example, Hogan & Thomas, 2001).

In order to motivate students to reason more deeply about their model, scaffolds could be offered to enable students to test their model against multiple datasets. This means that they have to compare not only model output to data, but also different sets of data with each other. This may be a more fruitful learning experience for students, since multiple datasets trigger students to think of alternative variables or relations in the model revision process. As a result, students may not be primarily engaged in model fit behaviour since their model has to be tested against several datasets.

In addition, students should be asked to model phenomena of which they already have knowledge. This allows them to dedicate cognitive processing resources to translating their mental model into a system dynamics model instead of having to invest too much effort in identifying relevant variables and relationships between them. It is important to free up that capacity for mastering modelling techniques during early stages of learning to model (Hogan & Thomas, 2001). This stance is supported by the finding that students in the present study had difficulties with comprehending the system dynamics modelling formalism of PowerSim, even after they received an instruction.

Löhner et al. (2003) investigated the effect of two different external representations used in computer-based dynamic modelling tools on performance of novice modellers. They compared a text-based model representation, in which students

have to provide a list of equations before the model can be executed, with a graphical representation, in which the model is built by qualitatively linking variables. Results indicated that the different representations support different phases in the modelling process. Löhner et al. suggested that the graphical representation would be more suitable for the beginning of the modelling process, since this representation enables students to readily identify and implement variables and relationships into their model. The text-based representation may be more fit for more advanced modelling, since students have to provide complete mathematical expressions in order to implement model elements.

In the present study, students revised a graphical model within a quantitative system dynamics modelling environment. Given the finding that students, in the present study, had difficulties with using this tool, it is suggested that novice modellers should first be asked to construct a qualitative graphical model. When constructing this model, students can focus on important variables and relationships and on mastering the modelling formalism without having to be concerned about the mathematical form of the relationships. Subsequently, they can proceed to quantify the relationships between variables in a semi-quantitative form. This means that students can choose the qualitative form of each individual relation (such as "If A increases, then B also increases"). Finally, when students are more experienced with the modelling tool, relationships between variables can be filled in quantitatively. A computer-based learning environment in which this kind of model progression (i.e., from building qualitative, semi-quantitative, to quantitative models) is implemented is Co-Lab (Van Joolingen, De Jong, Lazonder, Savelsbergh, & Manlove, 2005; for more information see also: www.co-lab.nl). The modelling tool in Co-Lab resembles PowerSim in that the syntax is also based on system dynamics modelling. The modelling tool in Co-Lab enables students to specify variables by selecting pre-defined qualitative relations, drawing graphs or entering mathematical formulas.

Finally, in the present study it was found that less successful students primarily thought about individual variables and relationships while revising their model, which implies a bottom-up approach to modelling (Gobert & Discenna, 1997; Hogan, 1999; Hogan & Thomas, 2001). Students who are employing a bottom-up approach to modelling do not consider how local model revisions impact the behaviour of the model as a whole. More successful students employed a top-down approach to modelling, in contrast, which involves students considering interactions between variables in their model when revising their model. These students elaborate on the dynamics of their model when they change something in their model and, thus, take a more holistic view on their model. Students should be scaffolded to develop these more productive approaches to modelling, in which they learn to reflect on the impact of dependencies between variables on the dynamic behaviour of their model. When students are introduced to modelling, it is argued that the top-down approach could be scaffolded by offering an expert model in order for them to productively model a certain phenomenon. When students take a top-down approach to modelling, they enhance their understanding of the workings of their model and ultimately learn more about the phenomenon being modelled.

Important to consider here is that the aforementioned scaffolds are suggestions offered to tackle the difficulties novice modellers have during the initial phases of modelling. Further research is needed in order to test these proposals.

Notes

1. Adapted from "Computerondersteund modelleren natuurkunde: Een sportieve beweging" ["Computer-supported modelling physics: A sportive movement"] (courtesy of Koos Kortland, Kees Hooyman, and Development Group Dynamic Modelling, University of Utrecht).
2. Translated from the Dutch "Computerondersteund modelleren: Basishandleiding Powersim", which is available online (http://www.cdbeta.uu.nl/model/literatuur/basishandleiding.pdf).
3. Below each protocol segment the following information is provided: protocol number, start time, duration of the episode and the ultimate scoring of the episode employing our scoring scheme.

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33,* 131–152.

Bliss, J. (1994). From mental models to modelling. In H. Mellar, J. Bliss, R. Boohan, J. Ogborn, & C. Thompsett (Eds.), *Learning with artificial worlds: Computer based modelling in the curriculum* (pp. 27–33). London: The Falmer Press.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315.

Chinn, C. A,. & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49.

Clement, J., Brown, D., & Zeitsman, A (1989). Not all preconceptions are misconceptions: Finding "anchoring conceptions" for grounding instruction on students' intuitions. *International Journal of Science Education, 11,* 554–565.

Coon, T. (1988). Using STELLA simulation software in life science education. *Computers in Life Science Education, 5*(9), 56–71.

Cox, M., & Webb, M. (1994). In H. Mellar, J. Bliss, R. Boohan, J. Ogborn, & C. Thompsett (Eds.), *Learning with artificial worlds: Computer based modelling in the curriculum* (pp. 188–198). London: The Falmer Press.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Education Research, 68*(2), 179–291.

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.

Devi, R., Tiberghien, A., Baker, M., & Brna, P. (1996). Modelling students' construction of energy models in physics. *Instructional Science, 24,* 259–293.

Doerr, H. M. (1995). *An integrated approach to mathematical modelling: A classroom study.* Paper presented at the Annual Meeting of the American Educational Research Association, San Fransisco, CA, 18 April.

Doerr, H. M. (1996). STELLA ten years later: A review of the literature. *International Journal of Computers for Mathematical Learning, 1,* 201–224.

Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 14*(1), 3–29.

Erkens, G. (1998). *Multiple episode protocol analysis (MEPA 4.0)* (Internal Publication). Utrecht, The Netherlands: Department of Educational Sciences, Utrecht University.

Forrester, J. (1961). *Industrial dynamics.* Portland, OR: Productivity Press.

Fretz, E. B., Wu, H., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2003). An investigation of software supporting modelling practices. *Research in Science Education, 32,* 567–589.

Gilbert, J. K., Boulter, C. J., & Rutherford, M. (1998). Models in explanations, Part 1: Horses for courses ? *International Journal of Science Education, 20*(1), 83–97.

Gobert, J., & Discenna, J. (1997). *The relationship between students' epistemologies and model-based reasoning.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, Physics Education Research Supplement, 68,* 552–559.

Hestenes, D. (1987). Toward a modelling theory of physics instruction. *American Journal of Physics, 55*(5), 440–454.

Hestenes, D. (1997). Modelling methodology for physics teachers. In E. Redish, & J. Rigden (Eds.), *The changing role of the physics department in modern universities, American Institute of Physics Part II* (pp. 935–957).

Heuvelmans, A. P. J. M., & Sanders, P. F. (1993). Beoordelaarsoverstemming [Inter-judgement reliability measurement]. In T. J. H. M. Eggen, & P. F. Sanders (Eds.), *Psychometrie in de praktijk* [*Psychometrics in practice*] (pp. 443–469). Arnhem, The Netherlands: CITO.

Hogan, K. (1999). Relating students' personal frameworks for science learning to their cognition in collaborative contexts. *Science Education, 83*(1), 1–32.

Hogan, K., & Thomas, D. (2001). Cognitive comparisons of students' systems modelling in ecology. *Journal of Science Education and Technology, 10*(4), 319–345.

Jackson, S. L., Stratford, S. J., Krajcik, J., & Soloway, E. (1996). Making dynamic modelling accessible to pre-college science students. *Interactive Learning Environments, 4*(3), 233–257.

Jonassen, D. H., Strobel, J., & Gottdenker, J. (2005). Modelling for meaningful learning. Learning Sciences and Technologies Group (Ed.), *Engaged learning with emerging technologies.* (pp. 7–28). Dordrecht, The Netherlands: Springer Verlag.

Kainz, D., & Ossimitz, G. (2002). *Can students learn stock-flow-thinking? An empirical investigation.* Paper presented at the 2002 System Dynamics Conference, Palermo, Italy.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674–689.

Kurtz dos Santos, A. C., & Ogborn, J. (1994). Sixth form students' ability to engage in computational modelling. *Journal of Computer Assisted Learning, 10,* 182–200.

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*(6), 395–418.

Löhner S., van Joolingen, W. R., Savelsbergh, E. R., & van Hout-Wolters, B. H. A. M. (2005). Students' reasoning during modelling in an inquiry learning environment. *Computers in Human Behaviour, 21*(3), 441–461.

Milrad, M., Spector, J. M., & Davidsen, P. I. (2002). Model facilitated learning. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and practice* (pp. 13–27). London: Kogan Page.

Mandinach, E. B. (1988). *The cognitive effects of simulation-modelling software and systems thinking on learning and achievement.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Ogborn, J. (1999). Modelling clay for thinking and learning. In N. Roberts, W. Feurzeig, & B. Hunter (Eds.), *Computer modelling and simulation in science education* (pp. 5–37). Berlin: Springer Verlag.

Ossimitz, O. (2002). *Stock-flow-thinking and reading stock-flow-related graphs: An empirical investigation in dynamic thinking abilities.* Paper presented at the 2002 System Dynamics Conference, Palermo, Italy.

Penner, D. E. (2001). Cognition, computers, and synthetic science: Building knowledge and meaning through modelling. In W. G. Secada (Ed.), *Review of research in education* (pp. 1–35). Washington DC: American Educational Research Association.

Raghavan, K., Sartoris, M. L., & Glaser, R. (1998). Impact of the MARS curriculum: The mass unit. *Science Education, 82,* 53–91.

Rouwette, E. A. J. A., Vennix, J. A. M., & Thijssen, C. M. (2000). Group model building: A decision room approach. *Simulation & Gaming, 31*(3), 359–379.

Riley, D. (1990). Learning about systems by making models. *Computers and Education, 15*(1–3), 255–263.

Schecker, H. P. (1993). Learning physics by making models. *Physics Education, 28,* 102–106.

Schecker, H. P. (1994). *System dynamics in high school physics.* Proceedings of the 1994 International System Dynamics Conference, Stirling, Scotland (pp. 74–84).

Schecker, H. P. (1998). *Physik—Modellieren, grafikorientierte modelbildungssysteme im physikunterricht* [*Physics—Modelling, graphically oriented model building systems in physics instruction*]. Stuttgart, Germany: Ernst Klett Verlag GmbH.

Spector, J. M. (2000). System dynamics and interactive learning environments: Lessons learned and implications for the future. *Simulation & Gaming, 31*(4), 528–535.

Steed, M. B. (1994). *Effects of computer simulation construction on shifts in cognitive representation: A case study using STELLA.* Unpublished doctoral dissertation, University of Massachusetts, Amherst, MA.

Stratford, S. J. (1997). A review of computer-based model research in pre-college science classrooms. *Journal of Computers in Mathematics and Science Teaching, 16*(1), 3–23.

Stratford, S. J., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modelling processes: Analysing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology, 7*(3), 215–234.

Suthers, D. D. (1999). Effects of alternate representations of evidential relations on collaborative learning discourse. In C. M. Hoadley, & J. Roschelle (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference* (pp. 611–620). Palo Alto, CA: Stanford University.

Sweeney, L. B. & Sterman, J. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review, 16*(4), 249–286.

Tinker, R. (1993). Modelling and theory building: Technology is support of student theorizing. In D. L. Ferguson (Ed.), *Advanced educational technologies for mathematics and science* (pp. 91–113). Berlin: Springer-Verlag.

van Joolingen, W. R., De Jong, T., Lazonder, A. W., Savelsbergh, E. R., & Manlove, S. (2005). Co-Lab: Research and development of an online learning environment for collaborative scientific discovery learning. *Computers in Human Behaviour, 21,* 671–688.

White, B., & Frederiksen, J. (1998). Inquiry, modelling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–117.

Whitfield, A. (1988). STELLA and its impact on the teaching of mathematical modelling. In F. Lovis, & E. D. Tagg (Eds.), *Computers in education, Proceedings of the IFIP TC3 First European Conference on Computers in Education—ECCE 88, International Federation for Information Processing* (pp. 299–304). Amsterdam, The Netherlands: Elsevier Science Publishers.

Wild, M. (1996). Mental models and computer modelling. *Journal of Computer Assisted Reasoning, 12*(1), 10–21.

Zhang, B., Wu, H., Fretz, E. B., Krajcik, J. S., Marx, R., David, E. A., & Soloway, E. (2002). *Comparison of modelling practices of experts and novice students using a dynamic, student-centred modelling tool.* Paper presented at the annual meeting of the National Association of Research in Science Teaching, New Orleans, LA.

Zohar, A. (1995). Reasoning about interactions between variables. *Journal of Research in Science Teaching, 32*(10), 1039–1063.

Appendix: Coding scheme for modelling processes

Type of cognitive process (i.e. what are they doing?)

Guiding by experimenter	The experimenter provides guidance to the students. The scoring starts with a question of one of the students that is addressed directly to the experimenter or by a spontaneous utterance from the experimenter
Evaluate	Students positively/negatively evaluate an element(s) in relation to their model. Students make a (elaborate) value judgment on a modelling element.
Explain	Students explain to each other how elements within their model work or why they were included. An explanation must be preceded by a clear-cut question of one of the students
Quantify	Students talk about quantifying or specifying a quantity or relation within their model.
Inductive Reasoning	Students elaborate upon/about elements within or with respect to their model (involves mainly qualitative reasoning)
Analyse	Students talk about/interpret modelling elements *without further elaboration*. Or identify factors that may be relevant/included in their model *without further elaboration* (i.e. factors are uttered by the students without further discussion)
Read & paraphrase	Students read or paraphrase model elements
Off task (no further coding)	Students talk about subjects unrelated to the assignment at hand
Other (no further coding)	Other categories that are not included in the 'type of process' analysis scheme (only use this code when no other process-code can be applied!). Also inedible murmur is coded as other.

Focus (i.e. what are they talking about?):

Quantity	Quantity (i.e. constant or stock). In case students are specifying an auxiliary without talking about the relationship that is implied.
Relation	(Not yet implemented) relation/interaction between quantities. In case students are specifying an auxiliary and talk about the relationship that is implied.
Model fit/model output	Fit between the model output and experimental data. Students have to explicitly mention (the extent of) model fit or output the model generates (i.e. the model graphs or table).
Data points	Data points/data graph. Students have to explicitly talk about the data points
Model structure	Structure of the model at hand, how the quantities are (visually) linked (i.e. visual structure). How the quantities in the constructed model are causally linked to each other (i.e. causal structure). Or how their constructed model works over time (i.e. dynamics). *Students have to explicitly talk about their model at hand:* When students talk about more than one relation in their model

Modelling actions/the tool	Talk about/mentioning modelling actions: what the students are doing. Or the students are trying to find out how Powersim (i.e. tools, buttons, formalism etc.) works
The assignment	Students talk about the assignment (i.e. the ice-skater problem)

Argumentation (what type of knowledge/modelling element do they allocate in revising their model?)

None	No notice of modelling element or knowledge type
Physics knowledge	Use of terminology, concepts (i.e. units, quantities), formula's common in physics
Mathematics knowledge	Use of terminology, concepts, formula's common in mathematics
Experiential knowledge	Knowledge from everyday experience is used
Correspondence model graph and data	Students refer to (the extent of) correspondence between model output and experimental data.
Experimental data	Experimental data (i.e. data points/graph)