M. (TECHNOLOGY TRACK): ISSUES OF REPRESENTATION IN CSCL

The Effect of Representations on Communication and Product during Collaborative Modeling
Simone Löhner, Wouter van Joolingen
Graduate School of Teaching and Learning
University of Amsterdam
simone@ilo.uva.nl, wouter@ilo.uva.nl

ABSTRACT
In this paper we investigate the effect of different external representations on the process and product of collaborative computer modeling tasks. Shared representations can significantly influence the processes of modeling and communication. In order to find the specific benefits of two different representations, we compare pairs working on a collaborative modeling task using a text based model representation with others using a graphical representation. The learners, secondary school students, used the modeling representation for two hours working on a task in the domain of physics. Results indicate that the two representations support different phases of the modeling process.

Keywords
Computer modeling, representations, simulation, collaboration

INTRODUCTION
Computer modeling and CSCL are two applications of information technology in education that have recently become more and more important. In the study presented in this paper we investigate the combination of working collaboratively (in this case in a face-to-face setting) and constructing runnable dynamic models.

The creation and manipulation of models by learners is increasingly recognized as a potentially powerful technique within constructive learning environments (Mandinach, 1988). In modeling environments, learners create executable models of phenomena in, for instance, physics or biology. This requires coordination and integration of facts with scientific theory rather than a mere passive collection of facts and formulas (Hestenes, 1987). Because a model is a conceptual representation of a real system that behaves in accordance with physical laws, creating models will help learners focus on conceptual reconstruction of reality and thus help constructing a unified and coherent view of science (Doerr, 1995; Hestenes, 1987).

Model building has been associated with constructing accurate and appropriate mental models. Through model building learners are able to ‘run’ their own mental model of a phenomenon (Jackson, Stratford, Krajcik, & Soloway, 1996) and it provides a way of asking whether they can understand their own way of thinking about a problem (Doerr, 1995).

When learners construct their models collaboratively, there is an extra benefit, because they also have to make their assumptions about the model relation they are working on explicit before adding it to the model. Modeling environments then serve as a shared artifact with which and about which discussion and co-construction of knowledge can be shaped. In this paper we focus on the role of a modeling environment as a collaborative workspace. One important property of such a workspace is the shared representation that is used to build the models. We discuss the properties of these representations and present an experimental study in which we compare two representations for models.

Different ways of modeling
We distinguish two major categories of modeling in education, expressive modeling and explorative modeling. In expressive modeling learners try to externalize their thoughts about a domain by creating a model. Therefore expressive modeling makes ones mental models explicit, serving as a means for communication and negotiation of ideas. In expressive modeling there is no concept of a “correct” or “best” model. This can be the case where systems are considered for which no reference model is known or available or for which the model is too complicated to understand in detail, by the learners involved. Examples are models of populations, where the goal is to create and understand phenomena like the forming of clusters of population, with no claim that the model accurately describes the real world phenomena. The focus is on global understanding of phenomena and on the modeling process itself, and not so much on the rules of the domain.

In explorative modeling learners try to find a specific target model of a given domain. The target model is (more or less explicitly) present in the learning environment for example in the form of data or a simulation of the system to be modeled. The goal of explorative modeling is finding the rules governing the phenomenon under investigation using induction and...
thereby demonstrating an understanding of the domain that is being modeled. In our research we aim at explorative modeling. Learners collaboratively construct a model that explains given empirical data. Learners can retrieve the data that should be explained from a computer simulation that is also available in the environment. They can do experiments with a simulation and collect the data that can be compared with the output of the model they produce. Special to the situation is that in addition to the learner’s model also the system model that can generate the data, is present in the environment, although it is not presented to the learner in an explicit way.

Representations and collaboration
Model representations are a means to construct models, but representations also serve as a vehicle for thought. External representations are not simply inputs and stimuli to the internal mind; rather they are so intrinsic to many cognitive tasks that they guide, constrain and even determine cognitive behavior and the way the mind functions (Zhang, 1997). Zhang calls this phenomenon 'representational determinism'. Zhang did his research on the influence of representation in problem solving activities, but we believe his conclusions will also hold for modeling tasks.

As representations play a role in supporting, guiding and constraining the cognitive processes in model building, we can also assume that they will have a strong influence on the way learners will communicate and collaborate when constructing models together. Suthers (1999) states: „…the mere presence of representations in a shared context with collaborating agents may change each individual’s cognitive processes. One person can ignore discrepancies between thought and external representations, but an individual working in a group must constantly refer back to the shared external representation while coordinating activities with others…” (p.612). Tools in which learners can organize their knowledge, mediate collaborative learning discourse by providing the means to articulate emerging knowledge in a persistent medium, inspectable by all participants, where the knowledge then becomes part of the shared context.

As external representations can be tools for enabling and directing reasoning processes, the representation used for describing the model the learners are creating, is of paramount relevance to the way learners will engage in the modeling task. In (Löhner & Van Joolingen, 2001) a review is presented of several representations that are used in different modeling tools on the market, and an analysis is made of the different aspects of these representations. A distinction is made between the primary representation (text or graphics), qualitative or quantitative representations, primary model entities (variables or relations), the way complex relations are handled (by the modeler or by the system), the visibility of the simulation engine (need for programming by the learner), the amount of information that can be externalized and the amount of scaffolding a representation gives by preventing inconsistencies.

From the description of the characteristics it will be clear that representations can determine the modeling and collaboration processes to a rather large extent: representations determine the nature of the model that is constructed, e.g. qualitative or quantitative, and the process leading to it, e.g. by suggesting relations or offering sensible defaults.

Also it is clear that there is a trade-off between the various characteristics of the representations. For instance, it is impossible to let learners focus simultaneously on the structure of the model and the details of the relations constructed in a single representation. Choosing a graphical overview means emphasizing the qualitative model characteristics, choosing text implies a focus on the quantitative details of the relations. If the goal is to let the learner do both, the representation must offer different views of the model, like a zoom function on relations and/or variables. These are so called multiple external representations (MER’s) (Ainsworth, 1999). Ainsworth shows that different representations used simultaneously can constrain each other's interpretation, construct deeper understanding or complement each other. In modeling for example the interpretation of a qualitative graphical model can be constrained by a quantitative textual model. The problem with MER’s however is that, as Ainsworth shows, learners find it difficult to translate between the different representations.

There is also a trade-off between the ease of use of a representation and the expression power. An easy to use modeling representation may always yield a running model but the level of expression can probably not go deeper than semi-quantitative relations. A deeper specification could break down the internal simulation mechanism.

The two uses of modeling we identify, seem to put different requirements on the representations used for constructing the models. In the case of expressive modeling the optimal representation seems to emphasize qualitative views on the model and relations in the model also should be expressed qualitatively. Conversely, representations for explorative modeling should allow quantitative statements and should allow the system to generate quantitative data. However, the case is a bit more complicated. For some qualitative phenomena to occur in a model sometimes a more detailed specification of the model relation is necessary, for instance when phenomena depend on parameter values. In this case only qualitative input and output is not enough. On the other hand, qualitative representations used in models of a quantitative nature can help the learner in organizing the model and be an aid in finding the relations that should be specified.

THE MODELING ENVIRONMENT
In this paper we describe collaborative modeling by learners in a learning environment consisting of a simulation window and a modeling window (see Figure 13). The environment was built in SimQuest, an authoring system for discovery
learning simulations (Van Joolingen & De Jong, in press; Van Joolingen, King, & De Jong, 1997). For the purpose of this study, SimQuest has been extended with a modeling tool. In the simulation window, the learners can conduct experiments on the system simulation by changing the values of the variables and starting the simulation. The simulation is dynamical, so variable-values can also be changed during the simulation. In the modeling window the learners can construct their own model. They can also run a simulation of their own model (a so called learner simulation) and thus compare the outcomes of the simulations of the two different models. The learners task is to construct build a model that gives the same results as the system model.

Figure 13 The collaborative learning environment with at the top the simulation window and at the bottom on the left side the modeling window and on the right side the explanations. The domain of the simulation in this case is heat and energy. The language of the environment is Dutch.

In the modeling environment there were two different possible representations. These were chosen to be as far apart as possible on the characteristics of (Löhner & Joolingen, 2001) in order to obtain a maximal contrast. In the following paragraphs the two representations will be explained in more detail

Textual representation
In the textual representation (see Figure 14), the learners type in the relations using algebraic equations. There are two types of equations, direct equations and rate equations. In a direct equation the learner specifies how a variable can be computed from others, for instance: “force = mass*acceleration”. Rate equations take the form: “delta(velocity) = acceleration”, where the delta indicates that the equation computes the change over time of the variable, not the variable itself. In essence, a rate relation is a first-order differential equation. The equations are not statements in a computer program, like DMS (Robson & Wong, 1985). Instead a simulation engine uses them to generate data and, for instance, takes care of the order in which they are executed.

Figure 14 The textual modeling tool as present in the environment.
In the textual representation learners can, in principle, create variables by typing in their names. However only variables that are available in the underlying system simulation model in the learning environment can be made visible in the simulation interface. Here one of the consequences of the availability of a system simulation model becomes visible. The model defines a set of variables for modeling. This is different for modeling tools designed for expressive modeling, but inherent to the task at hand in which the model output needs to be compared with output from the simulation model.

**Graphical representation**

In the graphical representation (Figure 15), learners specify relations by drawing influence diagrams (inspired on Forbus (1984)), consisting of nodes and directed arcs. Each node represents a variable; arcs between two variables mean that the variable from which the arc is drawn influences the variable the arc points to. Influences are signed and exist in two flavors, similar to the rate and direct equations in the textual representation. Rate relations indicate that the influence specifies the change of the variable over time; direct relations indicate that value of the variable itself is affected. The sign indicates the direction of the influence. A positive sign means that if the source variable increases the (rate of change of the) target variable also increases. A negative sign means the opposite, i.e. a decrease of a variable causes an increase of (the rate of change of) another.

Figure 15 The graphical modeling tool with an example of a model. Rate relations are indicated in red, and point to a circle, indicating that the variable is a state variable.

To be able simulate the model in the graphical modeling tool and compare its output to the system simulation, the equations of the system simulation are used to determine the exact equations used for simulating the graphical model the learner creates.

As will be clear from the description, the graphical modeling language is qualitative. There is no precise specification of relations in the sense that the learner creates a single computational prescription that can compute the value of one variable from others. A feature of our graphical modeling tool, however, is that it can make non-local features of the model visible in the topology of the graph the learner is drawing. For instance, a feedback loop, an important modeling construct indicating that the change of a variable may be dependent on the size of the variable itself, is really visible as a loop in the graphical diagram as shown in Figure 16. The same model expressed as text does not emphasize the feedback loop character. Here the loop has to be constructed by substituting one relation in another.

![Graphical Modeling Tool](image)

Figure 16 Difference in representation of a feedback loop, on the left in the graphical representation, on the right in the textual representation. Both models represent the same model. The graphical representation emphasizes the loop character of the model, the textual description focuses on the computational precision.
EMPIRICAL STUDY

The experiment was designed to explore differences in communication and modeling processes, as well as differences in the product of modeling, under influence of different modeling representations for explorative modeling. As the goal of the modeling task is explorative modeling, pairs of learners are asked to recreate the model present in the system (system model) by comparing it to a model they build themselves (learners model). Through building this model they are expected to gain a better understanding of the domain being modeled (temperature inside a house). The learning environment requires the students to induce rules about the domain being modeled from the data in the system model (system simulation) and to come to an agreement about how to implement these rules in their learner model.

Method

41 secondary school students (grade 11) from three schools in the Amsterdam area participated in the experiment as part of their regular coursework. The students also received fl. 30,- (± 12 USD) for participating. The experiment took a total time of three hours.

First the students were tested for scientific reasoning skills with a test, adopted from the scientific reasoning part of the ACT (ACT, 2001). This took about 20 minutes. Then the students were tested for relevant domain knowledge on energy and heat (10 minutes). 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 approximately 45 minutes. The students were randomly assigned to the two different modeling environments. After a short break the students were then divided randomly into pairs for the final modeling task. They spent about an hour working on a task on the temperature of a house. For this task they were given only a minimal instruction, to give them as much freedom as possible. During this task all actions in the learning environment were logged and also the students conversation was recorded. Finally the students were again given a domain knowledge test (10 minutes).

As the goal of this study also was to gain understanding of the modeling process, the students collaborated in a face-to-face setting. From a pilot study we learned that the communication between students was much more explicit when they worked face-to-face, than when they worked in a CMC setting.

The quality of the models the pairs constructed during the final modeling task was determined using a method similar to the one (Vollmeyer, Burns, & Holyoak, 1996) use (structure score). The score was obtained by adding the proportion of correct relationships and the proportion of correct signs of the relationships. (There were 23 possible correct relationships.) The score was then corrected by subtracting a penalty for redundant relationships. The score of the models could be in a range from 0 to 2. In the text representation it would be possible to break down the correct specification of the relationships even further (correct mathematical operation, correct weight), but to be able to compare the two representations we did not do that.

Expectations

We expect effects of the representation on the communication of the pairs during the modeling, on the modeling process and also on the product of the collaboration.

Because the graphical representation emphasizes the structure of the model, we expect students working with that representation to talk more about structural aspects of the model. For students working in the textual representation the emphasis will be much more on the precise form of the relation. We also expect that there will be more discussion and disagreement in the textual representation because it is less easy to just add a relation. In the text representation the students will be much more inclined to reach an agreement about the relation they are about to add, whereas in the graphical representation they can easily draw an arc and later delete it.

Therefore, one of our expectations about the collaborative modeling process is that there will be more experimenting with the model (changing, adding and deleting relations) in the graphical representation. For the textual representation we expect more experimenting with the simulations (learner as well as system) because the pairs need more data to reach the higher precision of the relations that is necessary. We also expect the pairs working in the textual representation to take longer before they actually start their first learner model simulation.

Finally for the product of the collaboration we expect better models in the graphical representation due to the better ability to experiment with the model, but on the other hand we expect the pairs working in the textual representation to have a better understanding of the found relationships.

RESULTS

Comparison of the two groups (graphical and textual) yielded no significant difference between the groups on the scientific skills pretest. Also there were no significant difference on grades in math and physics. Therefore we can assume equivalence of the two groups. The results of the domain test turned out to be unreliable. Also no differences were found
between pre- and posttest on the domain for both groups, so the decision was made to discard the results of the domain tests.

Analysis of the data logged during the modeling session shows that the pairs working in the graphical representation run simulations of more different models (M=25.8, SD=11.1) than those working in the textual representation (M=16.4, SD=14.4). This difference is significant at an alpha level of 0.05.

The pairs working in the graphical representation also constructed more complex models. In their first models on average they used 6.1 (SD=3.7) relations, as compared to an average of 2.6 (SD=0.7) relations in the textual representation. The final models in the graphical representation also consisted of more relationships (M=10.7, SD=4 compared to M=7.5, SD=2.7 in the textual representation).

Not only were the models the pairs in the graphical representation constructed more complex, they also scored higher on our model structure score (score on the last model: graphical M=1.3, SD=0.6 and textual M=0.5, SD=0.4). All aforementioned differences between the representations are significant at an alpha level of 0.05. We found no correlation between the average score of the pairs on the scientific reasoning test and the model score of the last model.

Table 1 Overview of means and standard deviations (in parentheses) of some modeling process measures for the two representations (textual and graphical)

<table>
<thead>
<tr>
<th></th>
<th>Number of different learner models</th>
<th>Number of relations in the first model</th>
<th>Number of relations in the last model</th>
<th>Model structure score of the last model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textual</td>
<td>16.4 (14.4)</td>
<td>2.6 (0.7)</td>
<td>7.5 (2.7)</td>
<td>0.5 (0.4)</td>
</tr>
<tr>
<td>Graphical</td>
<td>25.8 (11.1)</td>
<td>6.1 (3.7)</td>
<td>10.7 (4)</td>
<td>1.3 (0.6)</td>
</tr>
</tbody>
</table>

Table 2 Overview of means and standard deviations (in parentheses) of measures for the use of the simulation during the modeling process for the two representations (textual and graphical)

<table>
<thead>
<tr>
<th></th>
<th>Number of system simulations</th>
<th>Total time running system simulations</th>
<th>Number of learner simulations</th>
<th>Total time running learner simulations</th>
<th>Number of simulations (system and learner) per compiled model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textual</td>
<td>17.9 (11.7)</td>
<td>10.8 min (8.0)</td>
<td>21.7 (14.7)</td>
<td>7.3 min (3.5)</td>
<td>3.2 (1.8)</td>
</tr>
<tr>
<td>Graphical</td>
<td>28.3 (12.6)</td>
<td>9.8 min (3.9)</td>
<td>39.0 (19.0)</td>
<td>10.8 min (4.1)</td>
<td>2.7 (0.6)</td>
</tr>
</tbody>
</table>

Influence of the representations on the communication

The two representations also give rise to different behavior of the collaborating pairs. In the graphical representations relationships can be drawn very easily, whereas in the textual representation the precise form had to be determined. Therefore in the graphical representation the focus of the conversation often jumps quickly from one relation to another.

A typical 'graphical' pair makes its relations on the basis of 'correlations' between the behaviors of variables. They look what happens when they change a variable, and then add those relation(s) to their model. In the textual representation often initially the same kind of reasoning is followed, but that is not enough for them to be able to add the relation to their model. They have to go deeper into specifying their relations. Sometimes the textual pairs express the need to just quickly sketch a relation. But also the graphical pairs are sometimes hindered by their representation, because it does not give them enough insight in what is actually happening in the model.

The following protocol fragment gives an example of the type of reasoning of a pair of learners working with a graphical modeling tool.

Students start a simulation of the system model (SimQuest model) and play around with the variables

A: OK, lets see. You see that P_total goes up, P_loss goes down.
B: Yes. P_total ... P_loss and T_inside.
A: Try what you can change here, in the SimQuest model. You can just ... with those sliders
A: No, not that one, not that one, not that ... That one. C_house. A_loss, A_loss?
B: Sure
A: K_wall? K_wall?
B: That too
A: K_wall has a relation to k_total. K_wall up, k_total up.
B: When k_total goes up, P_total goes... where? Up.
A: And temperature? I mean temperature outside? Temperature outside, OK, errr temperature outside that...
B: Number of windows
A: {writes} number of windows
B: That has an effect on k_total.
A: Wait a minute, what does it have?
B: It has an effect on k_total.
A: {writes} K_total, yes
B: K_total has an effect on P_total

In this protocol fragment, the students are playing with the simulation and are noting what happens when they change the value of an input variable on a purely phenomenological level. They do not actually think about the meaning of their findings, they just look for correlations in the behavior of variables. They also jump very quickly from one relation to the next. After this fragment, the students quickly add the relations they 'found' to their model. In the text representation this type of modeling is impossible. In the following fragment the learners would like to be able to sketch their model, but are hindered by their representation.

B: Let's start with P_heating, because … if that is higher then …
A: Huh? But that's not possible
B: {Asks researcher} So you write for instance … if the heating is turned up, the temperature inside increases … if it's low then the temperature decreases
R: Yes?
A: So how do I write that?
R: What exactly do you want to say? Because this is a textual modeling-tool. You have to be precise in what you want to say.
B: You can't say P_heating, the bigger …
R: No, you have to look at exactly what is happening

If they had been working in the graphical representation, this would not have been a problem. They could have just drawn the relation between the heating and the inside temperature, without having to think about an exact formulation. The student’s working with the textual representation thus often have to spend a lot of time thinking about the mathematics behind the model, as can be seen in the following fragment.

A: So P_loss is…
B: Less
A: But what kind of relation is it? Isn’t it a ehhh what’s it called, exponential? Isn’t it? That’s this kind of formula, we have to make an exponential formula right?
B: Why don’t we …
A: Well, that’s why I said exponential thing. He but that is right, isn’t it? That you’re closer to a value, so the formula… No it isn’t exponential it’s zero.
B: Then it has to stay
A: No, because then it’s a what’s it called
B: Asymmetrical
A: Then it’s a valley, valley yes then it’s an asymmetrical. It’s not x squared. Then you would get a downwards parabola right? That’s not what it is.

But the graphical representation does present other problems. What the model is ‘exactly’ doing is not always obvious from the representation, as can be seen in the following fragment.

B: We probably should do more with the outside temperature. Cause look, if the temperature outside is higher, A_loss is also higher
A: Yes
B: Or errr lower. Should I put a minus?
A: Yes

Students add a relation between T_outside and A_loss to the model
A: A _loss is arealoss right? Squared meters
B: (unintelligible)
A: But what you're doing here, T_outside, that's in degrees Celsius, and you're subtracting it from squared meters.
What the students cannot see is that the negatively signed arrow they have drawn in their model does not represent a subtraction, but a division. Some students even express their dissatisfaction with the 'easy' graphical representation.
B: (Looking at another group) Oh, it looks difficult with the text tool
A: Is it difficult in text? I don't think so. Then you can at least see what you're doing!

CONCLUSION & DISCUSSION
The aim of this study was to explore the influences of different modeling representations on the communication process between collaborating learners, the explorative modeling process and on the results of the process. In a learning environment consisting of a simulation window and a modeling window, learners could experiment with a system simulation and try to build their own model of the domain (heat and energy). The two modeling representations used in the study were chosen to make the expected differences as large as possible.

The data indicates that the representation has a strong influence on both the modeling process and its results. This is apparent from the differences in activity and quality of the models. Activity is higher for the graphical group both for modeling and experimenting with the simulation, as well as the scores indicating the quality of the resulting models. From these results the graphical representation seems easier to use. The pairs made more complex models, which also scored higher on the quality measure we defined. But this representation also allows for a less deep specification of the relationships. Therefore the question is whether the graphical representation really leads to better understanding of the domain or if it just gives the students a better ability to try out possibly correct relations. The graphical representation seems to invite more experimenting with the model (more changes), probably because the commitment to a learner model relation is not as high as in the textual representation. In the textual representation the pairs have to spend so much time on formulating a relation they deem correct, that they are probably much more reluctant to delete it. Nevertheless the relations the pairs use in the graphical environment mostly seem very reasonable. They do not seem to be making their relations randomly, just to see what will happen, but seem to base their relations on 'common sense' reasoning.

The results on the use of the simulation in the two representations were not in line with our expectations. We expected the pairs working in the textual representation to use the simulation more than those in the graphical representation, and also we expected a higher overall use of the simulation. A reason for this minimal use of the simulation could be that the students are not used to using experimental data in such a way. Also the difficulty of the textual representation might have been a problem.

In the analysis of the communication of the pairs we have also seen the great influence the representations have on the communication. The different representations seem to support to different phases in the modeling process. The graphical representation leads the students too switching quickly from one relation to the next, and trying out every idea that seems to come up. This seems like a viable strategy for the beginning of a modeling process, when the learners do not yet have a clear idea about the model they are making. In the text representation this kind of modeling is virtually impossible. Learners have to actually think a relationship through before they can implement it. This seems like a preferable strategy for the latter part of the modeling process, when the model is brought into its definite form. Both forms of representation investigated in this study have their own particular role in the modeling process.

Therefore in realistic settings learners should use a mixed representation that providing the benefits of both the expression power of the textual representation and the easy experimenting of the graphical representation. But that alone is not enough. The modeling process has to be regulated by either the system or the learner, to make sure the representations are used for the right purposes and to help overcome problems that can be introduced by using MER’s.

In our future research we plan on combining the representations, and manipulating the amount and form of the regulation of the modeling process.

ACKNOWLEDGMENTS
This research project is funded by the Dutch research organization NWO under grant number 411.21.115

REFERENCES


