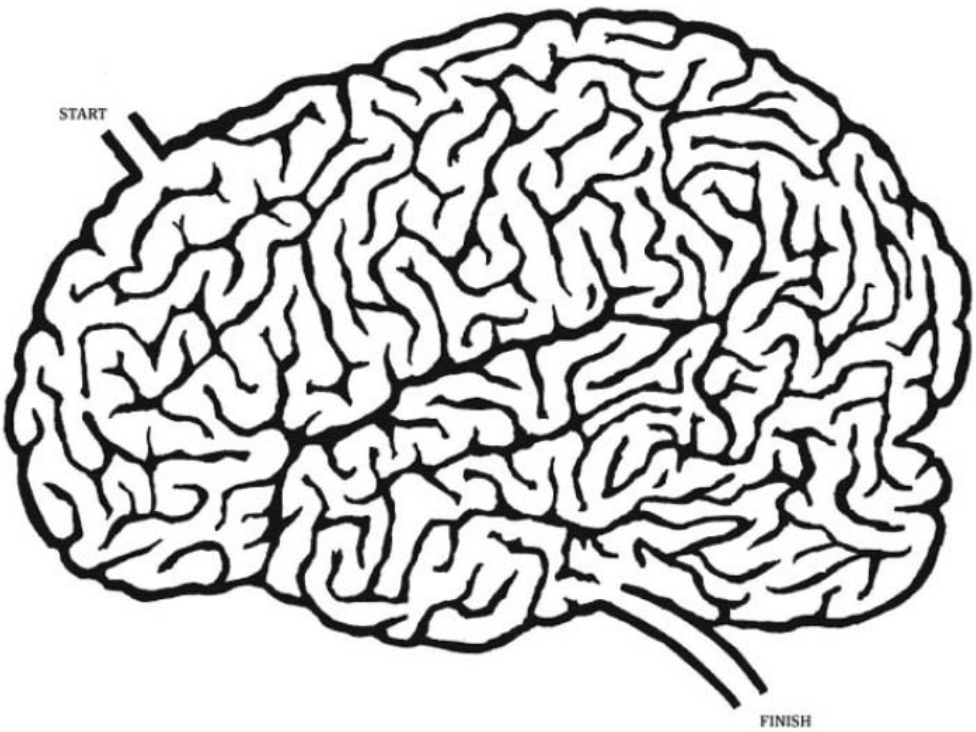
# How experts reason during modeling an ill-defined task

*An exploratory study*




Graduation thesis for the Master of Science degree in Psychology

Author: T.A van den Broek (0023523)
Supervisors: Dr. W.R. van Joolingen
Drs. S.P. van Borkulo

Date: 11 September 2007

Faculty of Behavioural Sciences, University of Twente

Universiteit Twente
de ondernemende universiteit

# How experts reason during modeling an ill-defined task

*Tijs van den Broek, University of Twente*

## Abstract

This study investigates the reasoning processes of experts while modeling a case in the domain of management science. Five experts in human resource management and five novices participated in this study. They produced an external representation of a HRM case, while verbalizing their thoughts. The resulting protocols were segmented and coded with a scheme based on descriptive theories of reasoning activities during inquiry modeling. The findings show that experts spent more time on orientation and used more experiential and conceptual domain knowledge during the entire modeling process. Furthermore, experts generated more hypotheses and separated the hypothesis generation from the modeling implementation. The results implicate that instruction on modeling an ill-defined task needs to enhance the learner to use prior knowledge by scaffolding. In addition, an environment for modeling ill-defined tasks needs a representational system, which can handle the interpretative nature of such tasks.

## 1 Introduction

The use of models is imperative to cope with the current myriad of knowledge-intensive problems. These problems include scientific problems, such as the development of the climate, or problems related to an economical or societal goal, like scheduling trains. Models cope with this complexity to extract an abstract representation of these problems, such that one can structure, predict or control complex phenomena.

In this thesis, modeling is defined as forming an explicit, conceptual representation of a phenomenon. In other words, a model is a simplified or idealized version of a part of the real world. Nowadays, computers are important tools for modeling dynamic phenomena. Ogborn (1994) defines computer modeling as “constructing and simulating external representations of dynamic phenomena by learners”. This definition stresses highly the educational purpose of models. The current study focuses on models as a way to investigate a certain problem or situation within the domain of Management Science.

Models have several cognitive functions. First, models are used to externalize thought in order to solve a problem (Bliss, 1994). A problem solver creates a mental model of the problem situation. When problems are complex, external representations of such models can serve as external memory and prevent an overload of the working memory (Zhang et al., 2006). Second, models help structuring the problem by visualizing the problem space. The construction of a model is essential to come to a solution of a problem (Simon, 1973). Therefore, models can be an important element in learning, which support learners in constructing a mental representation. Learners can form conceptual knowledge of qualitative or quantitative subjects by constructing a model in modeling environments like Model-It, PowerSim or Stella. Moreover, these learners learn about modeling and how to reason scientifically (Löhner et al., 2005).

In a social context external representations can function as a communication means. By constructing an shared external representation, learners can articulate their knowledge, discuss and form an agreed common knowledge base (Suthers, 2005). So, modeling can play a role in discussing a certain problem or situation. An example is a group-decision support system

(GDSS). Such a system facilitates the decision process during a face-to-face meeting by structuring discussions, visualizing facts with graphs, a shared workspace and whiteboard, etc. The shared representations within the GDSS support the communication and equilibrate the different opinions of the participators.

## 1.1 How experts reason during modeling

What exactly does someone who makes a model think and how should this modeler think to be most effective? The major focus in this study is on modeling as a way to investigate a certain problem, situation or issue. Scientific reasoning is the basis of this kind of inquiry and thus should be taken in account when studying modeling.

Löhner et al. (2005) compared five classifications of reasoning activities during scientific reasoning. In general the following reasoning processes take place while investigating a scientific problem:

1. Orientation: the scientist studies the situation and decomposes the situation into elements that have importance for the model.
2. Hypothesis generation: the scientist forms a hypothesis possibly using prior knowledge, or experimental evidence.
3. Experimentation: the scientist designs and executes an experiment to test the hypothesis.
4. Evaluation: the scientist compares the experimental data with the hypothesis of phase 2 and evaluates.

Although never observed and criticized by some authors, these processes form an inquiry cycle (Klahr & Dunbar, 1988; Löhner et al., 2005): the output of the evaluation process should function as input for the hypothesis generation process when the experimental data falsifies the hypothesis. This cycle continues until the right hypotheses are found.

These four processes of scientific reasoning are linked with three modeling activities (de Jong et al., 2002):

1. Model sketching
2. Model specification
3. Model evaluation

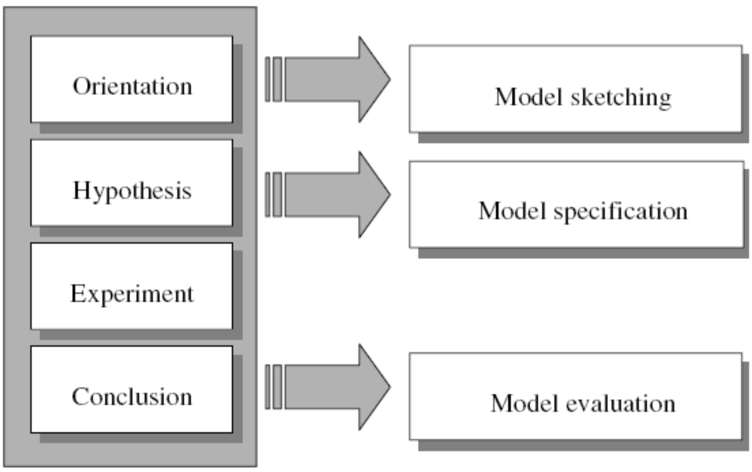



*Figure 1. The link between scientific reasoning and modeling activities*

These reasoning processes can function as a basis for a prescriptive model of how to model scientifically; a descriptive model. However, not everybody models in a scientific way. For instance, Löhner et al. (2005) conducted a study on the reasoning process of students who where modeling a physics subject using a computer modeling tool. The students followed a data-driven approach instead of an inquiry cycle and did not use any substance for their hypotheses.

A lot of research has been done on expertise in problem solving, especially in solving well-defined problems. Novice-expert research on problem solving is conducted in domains such as Physics, medical diagnosis, Advocacy, performing arts (Chi et al., 1985), but not specifically on the reasoning process during modeling. In the novice-expert research of the last decades, some discriminating principles of expertise in problem solving emerged:

- Experts make use of highly developed bodies of knowledge, called schemata. Experts have more schemata than novices and these schemata are also more specialized (Chi et al., 1985).
- Experts are highly effective in recognizing and encoding the underlying problem structure (de Groot, 1965). Therefore they should excel in the orientation phase of the scientific reasoning model.
- Experts select and apply appropriate problem solving strategies with minimal cognitive effort. They reason for example forwards instead of backwards (Gick, 1986).
- Experts have better meta cognitive ability to monitor their own progress when completing a task (Alexander, 2003; Glaser, 1996).

On the other hand, little research has been done on how experts solve ill-defined problems. Simon (1973) assigns the following properties to an ill-defined problem, which are opposed to the properties of a well-defined problem:

- Failing to present one or more problem elements.
- Having vaguely defined or unclear goals and unstated constraints.
- Possessing multiple solutions, solution paths, or sometimes no solutions at all.
- Possessing multiple criteria for evaluating solutions.
- Represent uncertainty about which concepts, rules, and principles are necessary.
- Having no explicit means for representing the problem and determining appropriate actions.
- Requiring learners to make judgments about the problem and problem solutions.

Reitman (1965) and Simon (1973) suggest that experts in solving ill-defined problems would excel in decomposing the problem and selecting the critical information, which supports de Groot (1965). In modeling activities, this would show up in decompositing the problem situation into variables and defining the relations between the variables. This suggests that experts should display a more sophisticated orientation process. Voss & Post (1985) provide some examples of research on expertise in solving ill-defined problems. They compared the problem structuring between the judgment processes of physicians and the judgment processes of magistrates. Moreover, they investigated the problem solving process of agricultural experts in the former Soviet-Union. The results pointed out that structuring ill-defined problems is highly domain-specific: the search process for appropriate variables was for example significant quicker for physicians than for magistrates. The researchers concluded that novices have difficulty processing divergent information in ill-structured problems, whereas experts have developed ways to structure the confusing, ill-defined problem for their domain of expertise. So, retrieving domain-specific conceptual knowledge of the problem domain plays an important role in the expertise of structuring the ill-defined problem (Simon, 1973).

The findings of Voss & Post (1985) are in line with research on factors that influence modeling positively. Van Joolingen & de Jong (1997) describe a few factors influencing the reasoning process of students that were working on a computer simulation. Their findings included the claims that domain specific prior knowledge and generic prior knowledge on the modeling process itself have both a positive effect on problem solving. Sins et al. (2005) compared a high performing dyad of pupils working on a modeling task in a modeling environment with a low performing dyad. The high performing dyad used significantly more own experience and theoretical knowledge during the reasoning and had a more holistic view on the. The low performing dyad used a highly data-driven approach instead of a theory-driven approach of the high performing dyad. These results should give some indications on how experts reason during modeling.

## 1.2 Modeling in the Management Science domain

This study focuses on modeling in the domain of Management Science. Pidd (2005) defines models in Management Science as representations of reality, used to understand, change, manage and control reality. This definition links directly to the task-oriented definition of management (Fayol, 1949), namely planning, organizing, commanding, coordinating, and controlling.

Ackoff (1979) used a distinction of problem types, which is similar to ill-defined and well-defined (Roberston, 2001). Problems in Management Science can vary on two dimensions:

- Problem formulation agreement: to what extent is there an agreement on the problem formulation, like the initial state and possible operators of the problem space?
- Problem solution agreement: to what extent is there an agreement on the problem solution space?

*Table 1. Problem types in management science*

| **Problem type** | **Characteristic Ackoff (1979)** | **Example in Management Science domain** |
|---|---|---|
| Puzzles | Agreed formulation and solution | Defining an optimization strategy for the distribution of goods |
| Problems | Only agreed formulation | The allocation of a new supermarket |
| Messes | No agreed formulation and solution | The evaluation of the failure of a new marketing campaign |

These different problem types and an example from the domain of Management Science are summarized in Table 1. The Management Science sub domains deal mostly with one particular problem type. One can state the general rule that the more quantitative the domain, the more puzzles one has to deal with. For example, Operations Research focuses more on puzzle problems and Strategic Management usually deals with messes.

Different problems like the ones described above require different models. There is a distinction in Management Science between interpretative models and quantitative models (Pidd, 2005), or soft system modeling and hard system modeling (Hicks, 1999). Quantitative models are appropriate for strongly well-defined problems or the so-called puzzles. These models have the purpose to strive for unarguable optimal solutions. However, quantitative models are less useful in solving messy problems (Hicks, 1999). In case of messy situations, interpretative models like cognitive mapping and Strategic Options Development and Analysis (Pidd, 2005) are more applicable than quantitative models, because these models are less strict and more open for discussion. Pidd (2005) calls these interpretative models “a way of generating debate and insight about the real world”. Such models offer the possibility to discuss and communicate group wise about a messy situation (Langfield-Smith, 1992), like a problem in Human Resource management, which has strong links with the constructivistic vision of Suthers (2005) on external representations.

A typical example of a quantitative model within the field of Management Science is linear programming, which is a subset of mathematical programming (Pidd, 2005). Linear programming is a mathematical method to model well-defined problems. The initial state of the problem is formed by reformulating the problem in a function, setting up constraints and choosing a goal (like maximize or minimize). An initial state looks as following:

Maximize $c_1x_1 + c_2x_2 + \ldots + c_nx_n$ ‘ goal and function
$x_1 + x_2 = 10$ ‘ constraint

This model is solved by making a graph or by linear algebra. The solution includes the values in which the variables, like $x_1$ $x_2$, are optimal. Linear programming is a common model in operations research and is often used for modeling production schedules, telecommunication networks or financial products.

An example of an interpretative model in the domain of Management Science is cognitive mapping, which is a subset of causal maps. A cognitive map is a model in which different concepts, a term for ideas or constructs, are connected with each other by arrows. These connections symbolize a causal relation between two concepts. The direction of the arrow means the direction of the causal relation. A minus sign means a negative relation and no minus sign or a plus sign means a positive relation.

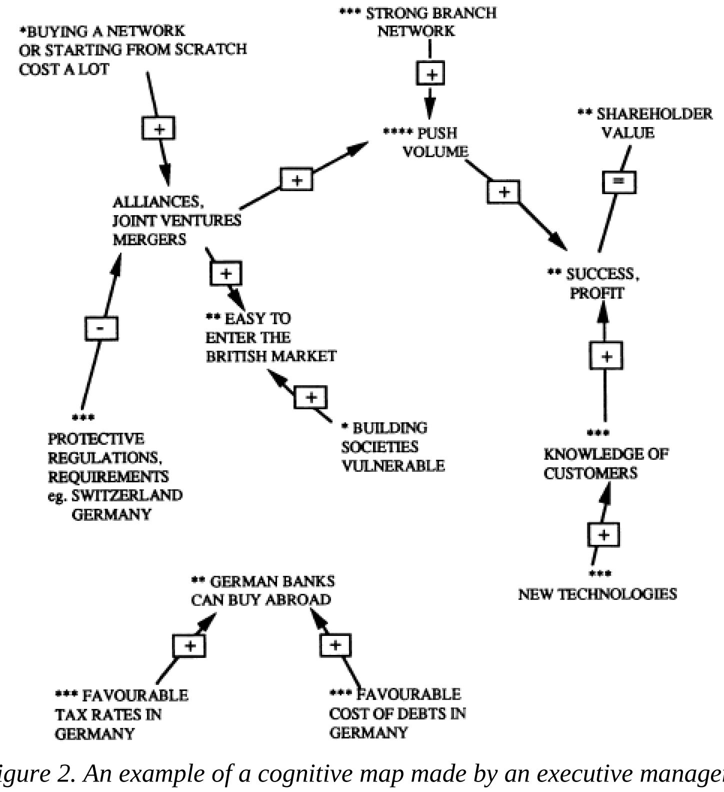



*Figure 2. An example of a cognitive map made by an executive manager*

Causal maps, like these cognitive maps, are often used in consultancy as a measure to map a client’s ill-defined problem, like the failure of a marketing or recruitment campaign.

Almost no research on how experts in the Management Science domain reason while modeling could be found. Willemain (1995) performed an experiment with 12 management experts who were given modeling tasks and were asked to think aloud as they spent 60 minutes on developing a model. Willemain (1995) classified the experts’ modeling under the following heading:

1. The problem context: structuring the problem (15% of the time)
2. The model structure: process of deciding what category of model to use (60% of the time).
3. Model realization: process of parameter estimation for a model and calculation of results (10% of the time).
4. Model assessment: deciding whether the model will be valid, usable and acceptable (15% of the time).
5. Model implementation: working with the client to gain some value from the model (0% of the time).

However, the Operations research tasks given by Willemain (1995) in the experiment mentioned above were highly quantitative. In practice most of the problems that confront managers, for example in Human Resource Management, are ill-defined nowadays.

So, what seems to be missing is descriptive research on the expertise of modelers in ill-defined tasks within the domain of Management Science. To better understand the reasoning processes of modeling experts in ill-defined problem solving, one should compare experts in the management domain with novices. Therefore, an exploratory expert-novice experiment will shed light on the cognitive differences.

The research questions that have to be answered can be stated as following:

*1. What are the reasoning processes exerted by modeling experts in ill-defined tasks within the domain of management science?*
*2. What are the differences between the reasoning processes of modeling experts and novices while modeling an ill-defined task within the domain of management science?*

These questions will be answered by exploring and describing the reasoning processes exerted by management modeling experts when constructing a model and comparing the experts with novices executing the same task in detail. A Human Resource Management case will be used as an instance of an ill-defined problem.

Knowing how experts cognitively deal with ill-defined problems is important in Educational Psychology. As stated in the introduction, computer models have potential to catalyze the novice’s learning process on conceptual domain knowledge, inquiry skills and modeling knowledge. A description of the expert’s reasoning processes gives insight on how novices could deal effectively with modeling tasks. The results of this research could help in the development of scaffolding learning tools and instruction on modeling ill-defined problems.

# 2 Methodology

## 2.1 Participants

Five male experts and five male novices in the Human Resource Management (HRM) domain took part in the experiments. The experts were recruited from consultancy firms and one research institute. They were selected by holding a short unstructured interview, in which the following criteria were assessed:

- At least two years of practical experience in HRM consultancy or research
- A (post-) academic degree in a HRM related study, like Business Administration, Organizational Psychology and Personnel studies.
- At least two years of experience in mapping HRM problems.

More details on the experts can be found in the Table below.

*Table 2. Data about the experts*

| Expert | Educational background | Experience |
|---|---|---|
| A | MSc in Public Administration, Post master courses in HRM | 7 years of experience in HRM consultancy |
| B | MSc in HRM, PhD in HRM | 6 years of experience in HRM research and consultancy |
| C | MSc in Business administration, PhD in HRM | 15 years of experience in HRM research, consultancy and education |
| D | MSc in Business Administration | 6 years of experience in HRM consultancy |
| E | MBA | 12 years of experience in HRM and Change Management consultancy |

These experts were compared with five novices in the HRM domain (Table 3). The novices were Bachelor of Science students in their final year at the school of Management & Governance of the University of Twente, the Netherlands. They had followed a few introduction courses on HRM. This prior knowledge criterion was important for their understanding of the case description.

All subjects participated on a voluntary basis.

*Table 3. Data about the novices*

| Novice | Educational background |
|---|---|
| A | 3rd year of BSc Industrial Engineering & Management |
| B | 3rd year of BSc Industrial Engineering & Management |
| C | 3rd year of BSc Business Administration |
| D | 3rd year of BSc Business Administration |
| E | 3rd year of BSc Business Administration |

## 2.2 Materials

The case was derived from a case study of the Harvard Business Review (Ehrenfeld, 1992). This case sketches the difficulties a large chemical company has with retaining its R&D personnel, although it performed very well financially as well in productivity. The case requested the participant to take the role of an HRM consultant and model the problem situation. The criteria of van Someren et al. (1994) were used to ensure well-verbalized protocols. The most important criteria for the case selection were

- Variables could be easily derived from the case description.

- Only qualitative relations could be derived from the case description.
- The case should be at an appropriate level for both experts and novices.
- The case should be realistic, because task commitment is important.
- The case should be ill-defined.

The case was translated to Dutch and the storyline was shortened without making it unrealistic. Furthermore, the case was reviewed by an HRM expert on the criteria above and tested twice on other experts to test the case on practicality. As a result, some minor changes in the case were made to make the description more clear and realistic. The case can be found in Appendix A.

Participants had the choice to use either the modeling tool in Co-Lab or a whiteboard with an eraser to construct their model. Co-Lab is an environment for inquiry learning that includes a qualitative modeling tool (van Joolingen et al., 2004). The modeling tool in Co-Lab has a graphical representation and works with qualitative, semi-quantitative and quantitative relations. The Co-Lab tool was originally intended to be the main tool for modeling the case, but it proved to be too restrictive in two pilot studies. However, the choice was made to use this tool as training in the experiment, because both pilot studies showed that this exercise made the participant more at ease with modeling.

Two training exercises were developed for the experiment. The first exercise gave the participants instructions on how to think aloud properly. This exercise contained an example video and two small assignments, in which the participants had to solve a block puzzle and come up with five improvements for a kitchen machine, while verbalizing their thoughts. The other exercise made the participant acquainted with modeling as an activity. This exercise consisted of a brief Co-Lab tutorial and two exercises, in which a model of a rabbit and fox population had to be constructed.

## 2.3 Measures and analysis

The participants’ verbalization during the experiment was recorded on a laptop computer and all modeling actions were logged by the experimenter using paper and pencil. At the end of the experiment a photo was taken of the model constructed by the participant, which was used later for the analysis of the model characteristics between the experts and the novices.

The recordings were transcribed to protocols and segmented into episodes based on natural breaks between sentences (van Someren et al., 1994). The length of the episodes varied between 1 and 90 seconds. These segmented protocols were coded with the coding scheme described in the next paragraph. After the coding the percentage of time students spent on each activity and the sequence between episodes were computed by using the protocol analysis program MEPA (Erkens, 1998), Multiple Episode Protocol Analysis. Furthermore, the protocols were analyzed in a spreadsheet to find out the modeling strategies.

Due to the relatively small sample size, non-parametric Mann-Whitney U tests were used to analyze the differences in reasoning processes and the differences in the produced models.

## 2.4 Instrument

The coding scheme of Löhner et al. (2005) was used as an instrument to interpret the activities of the participants during the experiment. Only a few categories and examples for coders were changed to fit ill-structured problems and the Management Science domain particularly. The final coding scheme has two types of categories:

- Scientific reasoning activities

- Other activities

The scientific reasoning activities consisted of five main categories:
- Orientation
- Hypothesizing
- Experimenting
- Model implementation
- Model evaluation
-

The other activities part had the following main categories:
- Actions
- Regulation
- Off task
- Experimenter

A complete list of categories and subcategories can be found in Appendix B.

In order to compute the inter-rater reliability a second coder received a protocol and coded it independently. The analysis program MEPA was used to calculate both the agreement percentage and Cohen’s kappa for the main categories and the subcategories. The inter-rater reliability of the main categories seemed to be acceptable (69.1 agreement percentage and Cohen’s kappa = 0.61).

## *2.5 Procedure*

As it was important for the participants to feel comfortable, the experiment for the experts took place in a quiet conference room at their office. The experiments with the novices took place in a conference room at the University of Twente. The procedure was standardized with an instruction sheet for the experimenter.

First, some general information about the study was given and the procedure of the experiment was explained to the participant. The experiment continued with a think aloud exercise, which lasted for 15 minutes. Second, the participant did an exercise in modeling by using a tutorial and a modeling exercise. This took about 30 minutes. After this exercise there was time for a 5 minute break. The actual experiment started with a sound check and after the sound quality proved to be good the participant could work for 45 minutes on the case using either a whiteboard or the modeling tool in Co-Lab. All participants but expert C used the whiteboard. The experimenter had a passive role during the experiment answering only procedural questions. The experiment concluded with a brief interview about how the participant experienced the experiment. This interview evaluated thinking aloud, use of Co-Lab, planning of the experiment, experimenter’s attitude during the experiment and the level of the case. The experiment total duration was approximately two hours.

# 3 Results

## 3.1 Reasoning processes

Table 4 shows the percentages of time participants spent on each activity. The standard deviations of the categories model implementation, regulation and actions are very high, indicating large differences between individuals. The share of the orientation and the hypothesizing categories together is very large: about 50 percent for the experts and 40 percent for the novices. For example, the orientation process of expert B took 39.2 % of the total time spent, which was the highest share among the participants. The time subjects talked off-task and the time the experimenter intervened are low, indicating that the participants focused strongly on the case.

When looking at the differences between groups, the Mann-Whitney U test (*n = 10*) indicates some significant differences. The percentage of time experts spent on orientating (28.0 %, *SD* = 7.7*)* is significantly higher than the time spent by the novices on this (15.1 %, *SD* = 0.3; *U* = 0, *p* = 0.009). The novices spent significantly more time to implement their model (20.8 %, *SD* = 6.9) than the experts (9.8 %, *SD* = 7.1; *U* = 3, *p = 0.047).* There were no segments coded of expert B's protocol in which he was implementing his model, indicating that he implemented his model silently and parallel to the other processes. The time experts spent on model evaluation (4.6%, *SD* = 1.9) was significantly less than the novices (5.5%, *SD* = 2.5; *U* = 2.5*, p* = 0.036). At last, the time the experimenter intervened to answer procedural questions was longer in the expert's process (4.5%, *SD* = 1.7) than the share of interventions in the novice's process (1.9%, *SD* = 1.9; *U* = 3, *p* = 0.047). In addition to these differences at *p* = 0.05 some smaller differences (*p = 0.1*) were found. The expert's session length (38:12, *SD* = 5:52) is longer than the time novices spent in total (30:10, *SD* = 7:45; *U* = 4, *p* = 0.076). Looking at the time spent on hypothesizing in minutes, experts (8:03) spent the same time as novices do (7:56), but the percentage of hypothesizing is higher at the novices (26.3 %, *SD* = 4.4) than at the experts (21.1 %, *SD* = 4.4*; U = 4, p* = 0.076). There are no significant differences noticed in the categories experimenting, regulation, actions and off-task processes.

*Table 4. Time experts and novices spent on reasoning processes while constructing a model*

| | Experts | Novices | U | p | |
|---|---|---|---|---|---|
| Length (m:s) | 38:12 (5:52) | 30:10 (7:45) | 4.00 | 0.076 | ` |
| Orientation (%) | 28.0 (7.7) | 15.1 (0.3) | 0.00 | 0.009 | * |
| Hypothesizing (%) | 21.1 (4.4) | 26.3 (3.8) | 4.00 | 0.076 | ` |
| Experimenting (%) | 5.8 (2.0) | 3.1 (3.3) | 6.00 | 0.173 | |
| Model implementation (%) | 9.8 (7.1) | 20.8 (6.9) | 3.00 | 0.047 | * |
| Model evaluation (%) | 4.6 (1.9) | 10.8 (5.5) | 2.50 | 0.036 | * |
| Regulation (%) | 16.8 (8.5) | 12.6 (6.4) | 9.00 | 0.465 | |
| Actions (%) | 8.4 (4.3) | 8.6 (6.8) | 11.00 | 0.754 | |
| Off task (%) | 1.0 (1.5) | 0.9 (0.8) | 12.00 | 0.911 | |
| Experimenter (%) | 4.5 (1.7) | 1.9 (1.9) | 3.00 | 0.047 | * |

*U and p scores were obtained with a non-parametric Mann-Whitney U-test*
** p = 0.05*
*` p = 0.1*

When one has a closer look in the orientation process (Table 4), experts spent in total almost 6 percent of their time on prior knowledge and novices almost nothing; 0.1 percent in total. The experts spent more time on both experience knowledge (3.0 %, *SD* = 1.5) and theoretical knowledge (2.8 %, *SD* = 3.3) than novices do (respectively 0.1 %, *SD* = 0.2; *U* = 0.0, *p* = 0.007 and 0.0 %, *SD* = 0; *U* = 2.5, *p* = 0.019). Moreover, experts take significantly (*p* = 0.1) more time to interpret the data in the case description (14.9 %, *SD* = 4.3) compared to the novices (9.3 %, *SD* = 1.4; *U* = 4.0, *p* = 0.076).

*Table 5. The sub processes of the orientation process*

| | *Experts* | *Novices* | *U* | *p* |
|---|---|---|---|---|
| Interpreting (%) | 14.9 (4.3) | 9.3 (1.4) | 4.00 | 0.076 ` |
| Defining variables (%) | 6.2 (4.6) | 3.6 (1.4) | 10.00 | 0.602 |
| Experience knowledge (%) | 3.0 (1.5) | 0.1 (0.2) | 0.00 | 0.007 * |
| Theoretical knowledge (%) | 2.8 (3.3) | 0.0 (0.0) | 2.50 | 0.019 * |
| Refer to instruction (%) | 1.2 (0.9) | 2.0 (1.4) | 6.50 | 0.209 |

*U and p scores were obtained with a non-parametric Mann-Whitney U-test*
** p = 0.05*
*` p = 0.1*

Within the hypothesizing process, no significant differences were found on a lower aggregation level.

## 3.2 Modeling strategy

Sequential analysis revealed that the experts stayed longer within a single process: an expert generates relatively more hypotheses after each other than a novice, for example. No inquiry cycle could be observed for any participant; hypothesis generation → experimenting → model evaluation → hypothesis generation.

When adding a time dimension, one can analyze the modeling strategy in more detail. The five scientific reasoning categories are plotted on the protocol position in percentage (X-axis) and the relative share of the process (Y-axis), assuming that the length of the sub processes change proportionally when scaling up or down between different protocol lengths. The relative share is computed within a window of 15 percent. The x-axis ends at 85%, because after this point no 15% window can be computed. The graphs in Figure 3 and Figure 4 show how the average share of each process for respectively the experts and novices evolve during the construction of their model.

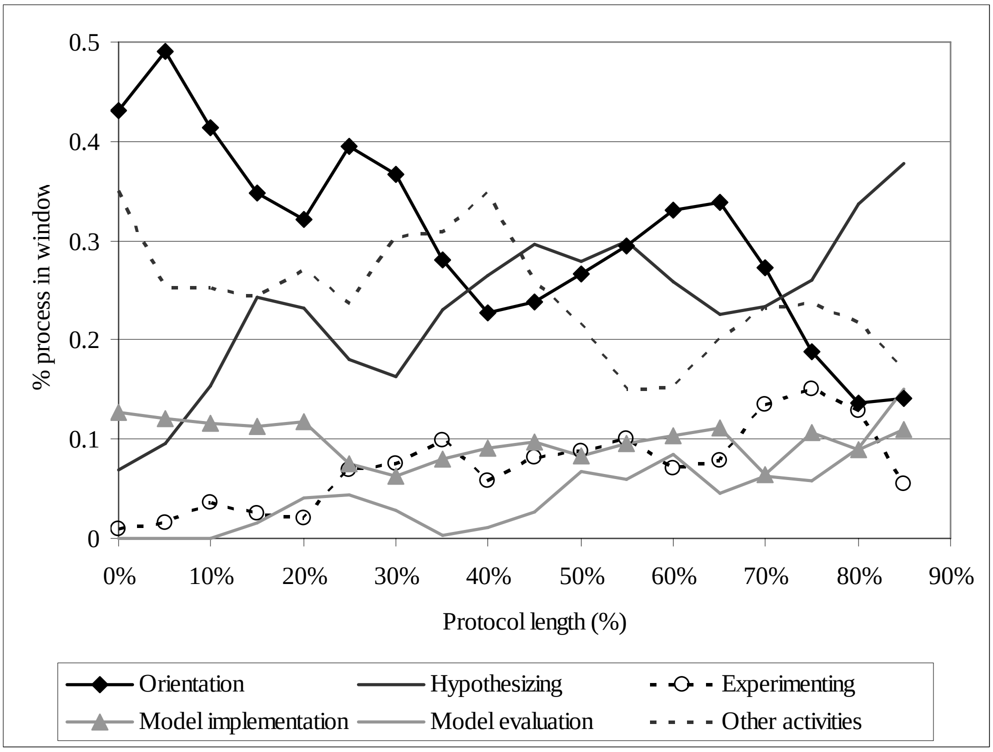



*Figure 3. Process graphs of the expert's scientific reasoning processes*

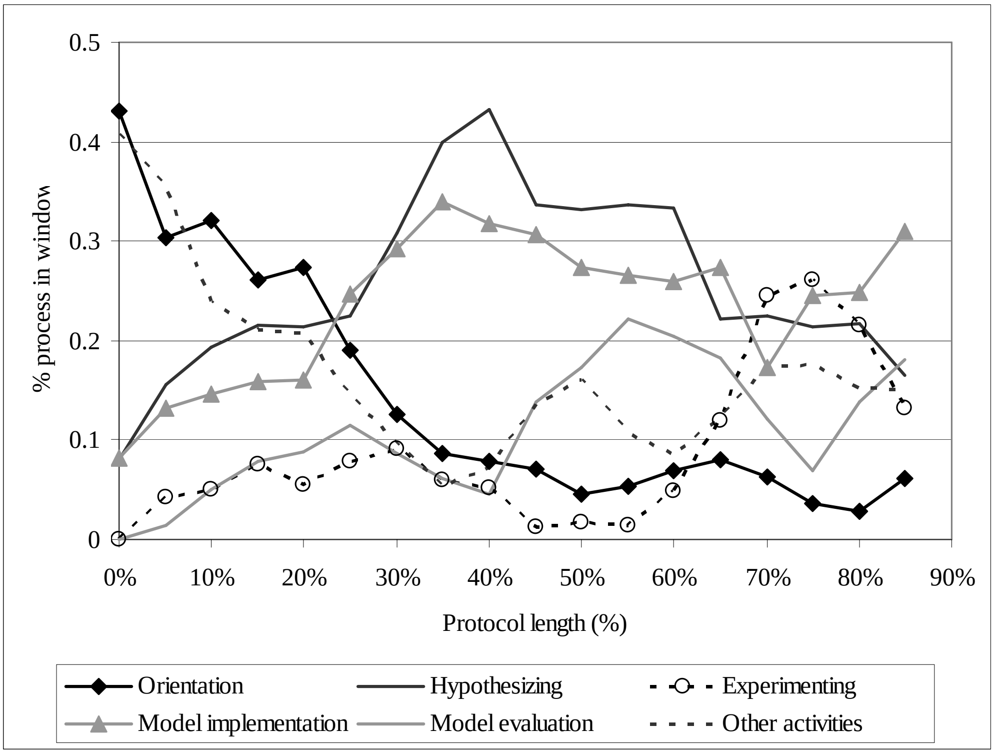



*Figure 4. Process graphs of the novice's scientific reasoning processes*

The differences support the proportion tables 3 and 4. In general, the share of each process during the complete protocol is more stable for the experts, resulting in less steeper lines. For the model implementation phase the deviation between the single processes is high. The expert's orientation process is more dominant during the whole time line compared to the novice's orientation process, mostly orientating in the first part of the protocol. On the other hand, the model implementation process appears to be more central in the novice's protocol. The novice's hypothesizing curve seems to have the same shape as the model implementation curve, suggesting a relation between both activities. Both curves have a strong peak at approximately 35%, indicating that the novice has closed the orientation phase and started mainly with implementing his representation. In addition, it seems that novices build their model while generating hypotheses, instead of experts who separate those processes. In the latter case, the hypothesis process gains share when time passes, reaching its summit at the end of the protocol. In both graphs the model evaluation process has its central point at the end of the protocol, indicating both experts and novices asses their own constructions after hypothesizing and implementing it.

### *3.3 Protocol examples*

Qualitative analysis of the protocols provides some details about how processes differ substantially. The experts demonstrate some schemata in their protocols. Table 6 shows two examples of theoretical knowledge schemata exerted by experts A and C in their protocols. Expert A puts the problem situation in a HRM model he uses in practice. First, he tries to state the model's six factors, except one he forgot. Later in the protocol he remembers the missing factor. Expert C demonstrates a schema about the meaning of HRM. The words "working in teams" elicit connections with other HRM concepts like performance management. Furthermore, he decomposes the HRM domain in seven constituents. These schemata of theoretical knowledge are often used to analyze the case, define variables and evaluate their model.

*Table 6. Examples of theoretical knowledge schemata exerted by expert A and C*

| Expert | Time | Code | Segment |
|---|---|---|---|
| A | 00:04:13 | Theoretical knowledge | *Uuh…from our…from my own practice there is a model that consists of six factors, which are related to each other…uuh…this model represents balance, coherence and heterogeneity, ESH, and these factors are organizational strategy, organizational structure, style of management and governance, personnel, systems and I always forget one..uuh..perss..no, I come back on it later. Those things are such related to each other that if you influence one of the dots presented here, it will directly effect the other dots.* |
| C | 00:35:44 | Theoretical knowledge | *Working in teams...uuh…to be complete, it is about organizing of the work, it is also about things like performance measurement, safe environment, eeh performance measurement. Uuh…measurement, yes. I also would like something about rewards. Rewards...Performance measurement, I will deal with appraisal separately. It is about a HRM manager, isn't it? Yes, it is about HRM. I think HRM deals with leadership, work organization, performance measurement, rewarding, evaluating, career, hop.* |

Some schemata of experiential knowledge occur in the experts' protocols too. Table 7 shows how the case reminds expert B to his own experiences with a similar company in the chemical industry. He transfers his experiences to visualize the work situation for the employees in the case.

*Table 7. Example of experiential knowledge schema exerted by Expert B*

| Time | Code | Segment |
| --- | --- | --- |
| 00:12:04 | Experiential knowledge | *But as I can see the company has 350 employees…uuh…that is quite large in my opinion. So…the company could be compared with a division of DSM. I know DSM reasonably well and that is why I think about it. I can imagine a situation with crackers all over the place, a few people, a lot of…uuh…machines and factory..uuh these are not actually factory halls, but pipelines. People who wear helmets and one office, in which all R&D employees sit together. Something like that appears to me.* |

The novices are more concerned with the construction of their model and how to fit all the variables on the whiteboard. Table 8 shows how novice C works with a numbering system to implement his model more efficiently. He designates referential numbers to variables, like number six for the variable enthusiasm, and makes a legend in which links the numbers to the variables.

*Table 8. Example of novice C's model implementation*

| Time | Code | Segment |
| --- | --- | --- |
| 00:12:39 | Model implementation | *Uuh..wait…there stand and is absence through illness and so…everything from one till five has a negative influence, uuh..number six is enthusiasm...has a negative influence, seven is dissatisfaction, number eight is absence through illness, number nine is the retention and that has of course in its turn an influence on number ten, the total number of personnel.* |

As stated in the previous paragraph, novices make more hypotheses during the implementation of the model. Table 9 shows a few segments of novice E's protocol, in which the novice switches between model implementation and hypothesizing. He is looking at the whiteboard while placing a variable. However, this activity triggers a hypothesis about the variable's relation, which is implemented subsequently. These transitions between hypothesis generation and model implementation repeat during the protocol.

*Table 9. An example of novice E switching between hypothesis generation and model implementation*

| Time | Code | Segment |
| --- | --- | --- |
| 00:09:47 | Model Implementation | *So this should be on this side* |
| 00:09:51 | Hypothesizing | *However, reverse could also be true…a low labor productivity causes illness. So, it is plausible or in any case not ill, but ill for the boss.* |
| 00:10:05 | Model Implementation | *Possibly an arrow to the other side* |
| 00:10:11 | Hypothesizing | *Uuuh…ok..creativity has also to do with…if they are not creative, then they will never be productive* |
| 00:10:22 | Model Implementation | *Also an arrow to this side, reverse is not valid* |
| 00:10:29 | Hypothesizing | *One will not become less creative when not productive. Further, a better job has…uuh…the possibility of getting a better job could influence the absence* |

| | | |
|---|---|---|
| | | *through illness.* |
| 00:14:17 | Hypothesizing | *Uuh..yes…for the reason that if you posted your solicitation letter, you probably do not want to work at the R&D department of this company anymore.* |
| 00:11:00 | Hypothesizing | *Working conditions have an influence on about everything over here.* |
| 00:11:03 | Model Implementation | *I will not even draw arrows for it, because the whiteboard will become a chaos.* |
| 00:11:10 | Hypothesizing | *A conflict has somehow an influence on productivity and absence through illness and probably also on creativity. As soon as a conflict stresses you, you will not act very creatively.* |

Table 10 shows the hypothesizing process of expert A. He stays for a long time hypothesizing without placing the relations directly in the model. The expert forms a chain of causal reasoning starting with the variable "recruitment of HRM personnel" and ending with "Conflicts".

*Table 10. An example of Expert D's hypothesizing*

| **Time** | **Code** | **Segment** |
|---|---|---|
| 00:15:17 | Hypothesizing | *That makes the…eeeh…the incoming personnel available…uuh the incoming personnel consists of eeh..the the the capacity and recruitment's dedication* |
| 00:15:51 | Hypothesizing | *Recruitment of incoming personnel..eeh this input can be important for the effective available human resources eeh....no, for the available human resources uuh, because there is still absence through illness in between* |
| 00:16:08 | Hypothesizing | *So, this leads to…uuh..effective...available and then it will be also…uuh…caused by absence through illness. In that case absence through illness is…absence through illness is determined by…uuh by conflicts too and conflicts have in its turn an effect on the actual effective available eeh…labor hours. If employees have disputes with each other, they will not be able to work together* |

## 3.4 Models

Figure 5 and 6 depict the external representations constructed by expert A and one constructed by novice E respectively. Both models appear to be interpretative stating only constructs and qualitative relations. They can be characterized as causal flow models. These examples exemplify some obvious differences between expert A and novice E. First, the expert's model constitutes more variables and relations and therefore seems to be more complex. The novice used 17 variables in his model from which 12 variables came directly out of the case description. The expert on the other hand used 30 variables of which only 9 were deduced from the case description. Second, experiential and theoretical knowledge are embedded within the expert's model. Examples are the strategic triangle in the left corner of figure 5 and the ESH model, which is located on the right site of the strategic triangle.

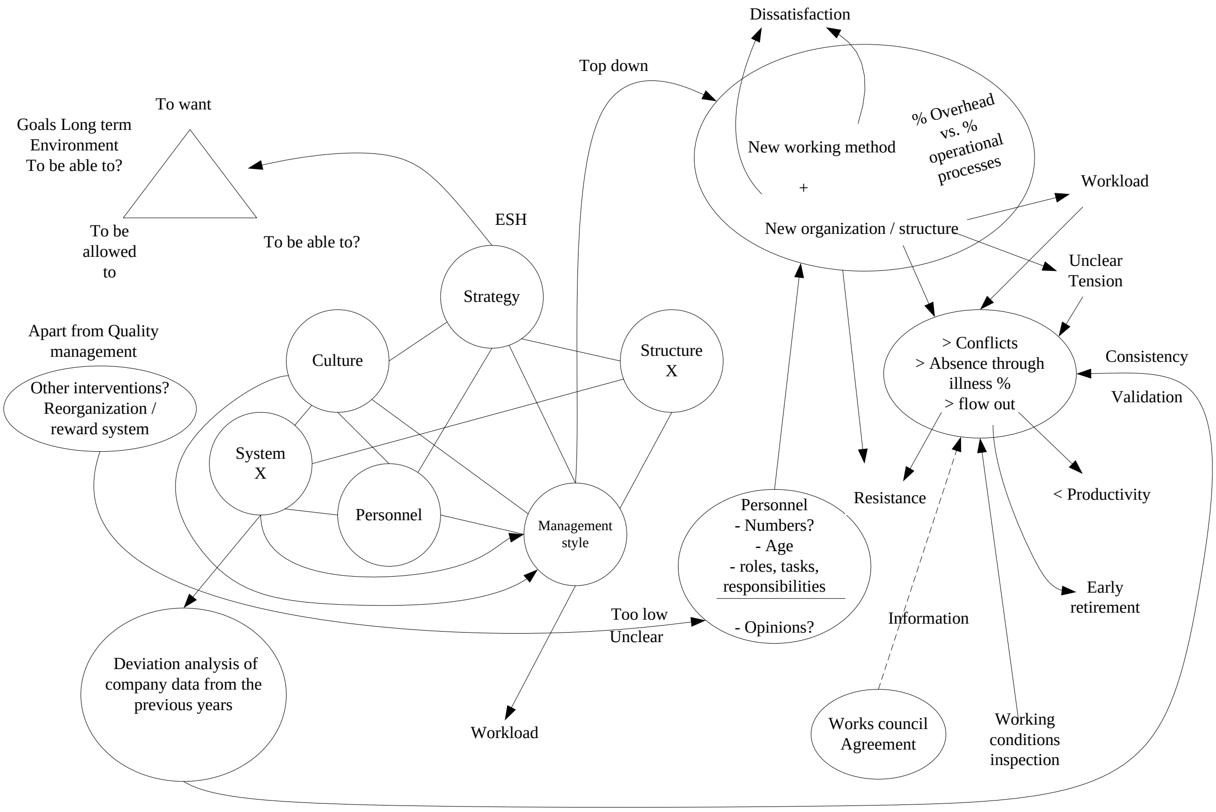



*Figure 5. Model made by expert A*

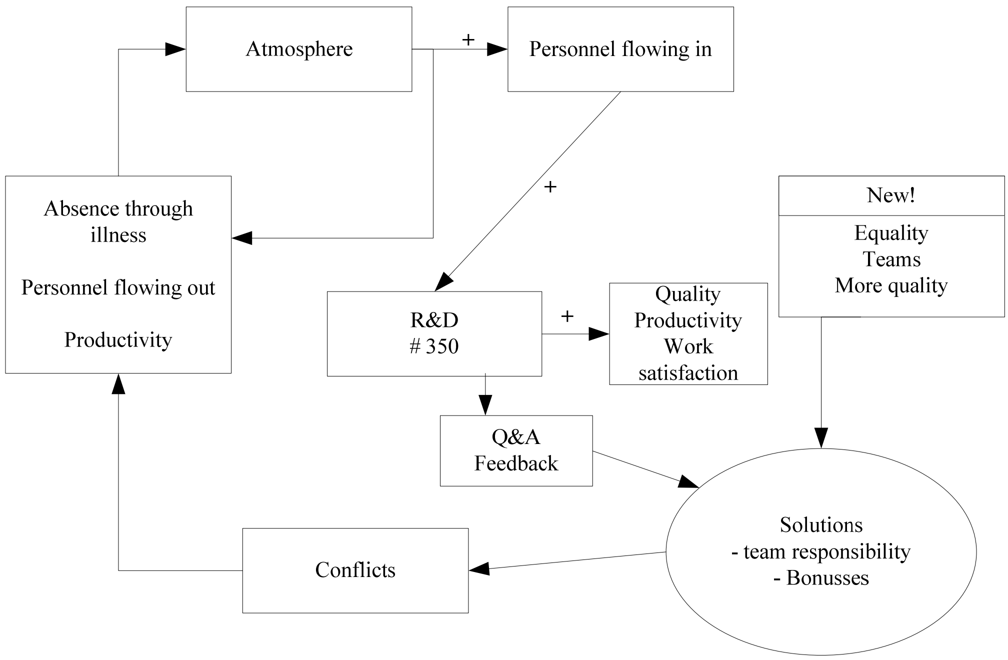



*Figure 6. Model made by novice E*

Table 6 shows the differences between the main characteristics (variables and relations), of the models produced by both groups. In line with the higher orientation activity of experts, the number of new variables experts came up with (12.2, SD = 5.5) is significantly higher than the novices came up with (4.6, SD = 3.4; *U* = 0.50, *p* = 0.012*)*. This comes back in the share of variables that could not deducted from the case, since this is significantly higher in the experts’ models (60.4 %, SD = 9.5 %) than in the novices’ models (33.3 %, SD = 12.9 %; *U* = 1.00, *p* = 0.016). On the other hand the total number of variables, the number of variables extracted from the case description and the total number of relations does not deviate significantly.

*Table 11. Quantitative characteristics of the models*

| | Experts | Novices | U | p |
|---|---|---|---|---|
| Variables | 19.8 (5.5) | 14.2 (4.0) | 6.00 | 0.169 |
| - Variables case | 7.6 (1.7) | 9.6 (3.4) | 8.00 | 0.337 |
| - Variables new | 12.2 (4.7) | 4.6 (1.9) | 0.50 | 0.012 * |
| Relations | 20.8 (8.3) | 16.2 (5.7) | 8.00 | 0.346 |

*U and p scores were obtained with a non-parametric Mann-Whitney U-test*
** p = 0.05*

# 4 Conclusion and discussion

The results show that in modeling an ill-defined task experts spent significantly more time on orientation when compared with novices. The differences in orientation consist of the use of own experiences and theoretical knowledge, like HRM models and theories. This supports the claim that domain knowledge is important for modeling (Sins et al., 2005; Willemain, 1995) and problem solving (van Joolingen & de Jong, 1997), especially solving an ill-defined problem (Voss & Post, 1985). In addition, experts took less time to build on models with more variables and relations. It seems that expert’s familiarity makes the implementation activity more automatic, giving experts the opportunity to delve conceptually deeper in the situation as they have more time for orientation. The new variables novices brought in, suggest that novices have prior knowledge, but do not state it explicitly. On the other hand, novices spent more time on the evaluation of their external representation by comparing it with the data in the case description. This result deviates from Löhner et al.’s (2005) claim that high-performing modelers evaluate their models more, although the domain and the nature of the problem differ between both studies. Future research should compare expert modelers from well- and ill-defined task domains to shed light on the differences between task domains. The share of the regulative process was high for both groups and exceeded the share of regulative processes Löhner et al (2005) found. However, no significant differences between novices and experts could be found, which does not support Alexander’s (2003) and Glaser’s (1996) claim that experts would regulate their work more due to higher meta cognitive capabilities and indicates that apart from regulative processes orientating is important too. It could be that experience in course work made the novices, all in the third year of their Bachelor degree studies, familiar with planning, monitoring and evaluating their performance.

We could not see that subjects from either group sequenced their actions according to a normative inquiry cycle. As the ill-defined task resulted in interpretative models, highly varying in syntaxes, it proved to be hard to formalize these models in a modeling environment like Co-Lab. This confirms the suggestion that ill-defined tasks are hard to quantify in a computer model, because of their interpretative nature (Hicks, 1999). Still, an inquiry cycle requires experimental data, for example from a dynamic model, which serve as input for the hypothesis generation process.

Apart from the reasoning activities, the modeling strategy varied between beginners and experts. They start with orientating, continue with hypothesizing and end with evaluation, but the distribution of the reasoning processes is different: the expert’s orientation process lasts the whole modeling process, whereas novices stop with orientating after they started with the model implementation and hypothesizing. Most experts started later in the process with model implementation, first carefully orientating on the task, which supports Zhang’s et al. (2006) finding that expert modelers wait with constructing the model until the last phase of the modeling process. The hypothesizing and model implementation highly correlated within the novice’s process: they formulate a hypothesis and put it in the model. In contrast, blocks of hypotheses without implementing these were found in the experts’ protocols, indicating that experts can make more reasoning steps at once. These blocks of hypothesizing is similar to the reasoning chains found by Voss et al. (1983). Qualitative support was found for these findings: for example, having a closer look at the experts’ orientation process the use of schemata of prior knowledge can be found as predicted by Chi et al. (1985). The reasoning processes resulted in different models. All of them consisted of only qualitative relations and abstract constructs as variables. Although experts were less busy with building their model, they produced models. The share of variables that were not stated in the case description was significantly higher, indicating that more extensive orientation process exerted by the experts resulted in more new variables. On the other hand experts used less variables stated in the case, which supports the claim that experts should be better in filtering critical information (de Groot, 1965; Reitman, 1965; Simon, 1973)

Research on the differences between experts and novices contributes to the development of modeling instruction, because it identifies key concepts and strategies that students must acquire to function effectively in a particular domain (Bransford & Vye, 1989). As a snapshot of the reasoning processes exerted by both groups, this study defines the goals novices should pursue. The results of this study and prior research (Sins et al., 2005; van Joolingen & de Jong, 1997; Voss & Post, 1985) suggest that the use of (prior) knowledge is important during the whole modeling process. Although this study seems to provide on overview of the symptoms of expertise, it has implications for the development of modeling instruction, especially for ill-defined tasks. These practical implications regard both declarative and procedural knowledge. In the learning process from beginner to expert, the declarative knowledge of novices and intermediaries is passive in the sense that they are not able to spontaneously make transfers as their schema are still context-dependent. They can develop expertise by repeating similar cases, which in the end decontextualize these schemata (Robertson, 2001). Modeling instruction, for example in the form of a learning environment, can possibly catalyze this schema induction for experiential and theoretical knowledge, but should aware of the drawbacks of schematization (Feltovich et al., 1997). The examples in modeling instruction should start with cases novices are already familiar with. Subsequently, the instructional tool should help the novice to make transfers with earlier cases, that share productions with the current case (Anderson & Singley, 1993) by scaffolding on relevant moments. This scaffolding can include active prompting using a database with previously made models, which supports the modeler in recognizing the current exercise. If the modeler starts recognizing elements in the current model that are similar to elements in previous cases, the modeling tool can remind the modeler to these previous cases. When the novice starts making transfers spontaneously, the instructional tool should help to induce more conceptual knowledge from related cases. The learning environment could support the learner in categorizing and developing conceptual knowledge by offering templates. This strategy can be used for enhancing active use of declarative, both experiential and theoretical, and procedural knowledge on modeling. Apart from the use of knowledge, modeling instruction should stimulate learners to form chains of reasoning instead of generate hypotheses after each other.

The post-experiment interview revealed that the experts appreciated a modeling tool such as the modeling tool in Co-Lab for their domain. Such modeling tools can help HRM consultants or experts to assess models made for clients or research projects. However, they stated that due to the restrictive representation and syntaxes of a modeling tool it would be difficult to implement an ill-defined task like the HRM case. A modeling tool suitable for an ill-defined task should work with causal relations and allow variables not to be defined or missing. As the problem gets more defined in practice, a modeling tool like in Co-Lab can be used.

This study was intended to be exploratory. It sheds light on the reasoning process of experts and gives directions for future research on the reasoning process when modeling ill-defined tasks. Some differences between experts and novices are hard to explain. The difference in model evaluation can be caused by the time pressure experts had to deal with, as most of them did their experiment during work time. The deviation in experimenter’s interventions could be the result of the expert’s habit to question, like they should do when working with a client, and the novice’s habit to work on an assessment silently. Other research should validate the findings with a bigger sample. Moreover, as there is a broad spectrum of ill-defined tasks the reasoning process found in this study can not be generalized to all domains. A meta-analysis of several validated studies within different domains is necessary for such an induction. Comparing modeling expertise in ill-defined tasks with well-defined tasks should tell whether this study’s results are characteristic for ill-defined tasks. Future research should investigate whether modeling environments appropriate for ill-defined tasks can transform an ill-defined task into an inquiry task, possibly provoking the inquiry cycle.

# 5 References

Ackoff, R. L. (1979). The future of operational research is past. *Journal of the Operational Research Society, 30*(2), 93-104.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14.

Anderson, J. R., & Singley, M. K. (1993). The identical elements theory of transfer. In J. R. Anderson (Ed.), *Rules of the mind*. Hillsdale: NJ: Lawrence Erlbaum Associates Inc.

Bliss, J. (1994). From mental models to modelling. In H. Mellar, J. Bliss, R. Boohan, J. Ogborn & C. Tampsett (Eds.), *Learning with artificial worlds: Computer based modeling in the curriculum*. London: The Falmer press.

Bransford, J. D., & Vye, N. J. (1989). A perspective on cogntive research and its implications for instruction. In L. B. Resnick, Klopfer, L.E. (Ed.), *Toward the thinking curriculum: Current cognitive research* (pp. 150-173): ASCD publications.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1985). *The nature of expertise* (first edition ed.). Hillsdale: Lawrence Erlbaum.

de Groot, A. (1965). *Thought and choice in chess*. The Hague: The Netherlands: Mouton.

de Jong, T., van Joolingen, W. R., Lazonder, A. W., Ootes, S., Savelsbergh, E. R., & Wilhelm, P. (2002). *Colab specifications; part 1 theoretical background*. Enschede: The Netherlands: University of Twente.

Ehrenfeld, T. (1992). The case of the unpopular pay plan. *Harvard Business Review* (January - February), 14-23.

Erkens, G. (1998). Multiple episode protocol analysis. Utrecht: Department of Educational Sciences, University of Utrecht.

Fayol, H. (1949). *General and industrial management*. London: Pitman Publishing company.

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1997). Issues in expert flexibility in contexts characterized by complexity and change. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context* (pp. 126-146). London: MIT Press.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21*, 99-120.

Glaser, R. (1996). Changing the agency of learning: Acquiring expert performance. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Hicks, M. J. (1999). *Problem solving in business and management* (Third edition ed.). London: Thompson business press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science, 12*, 1-48.

Langfield-Smith, K. (1992). Exploring the need for a shared cognitive map. *Journal of Management Studies, 29*(3).

Löhner, S., van Joolingen, W. R., Savelsbergh, E. R., & van Hout-Wolters, B. (2005). Student's reasoning during modeling in an inquiry learning environment. *Computers in Human Behavior, 21*, 441-461.

Ogborn, J. (1994). Overview: The nature of modelling. In H. Mellar, J. Bliss, R. Boohan, J. Ogborn & C. Tampsett (Eds.), *Learning with artificial worlds: Computer based modeling in the curriculum.* London: The Falmer Press.

Pidd, M. (2005). *Tool for thinking: Modelling in management science* (Second edition ed.). Chichester: Wiley.

Reitman, W. (1965). *Cognition and thought*. New York: Wiley.

Roberston, S. (2001). *Problem solving* (1st edition ed.): Psychology Press.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4*, 181-201.

Sins, P. H. M., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The difficult process of scientific modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education, 27*(14), 1695-1721.

Suthers, D. (2005). Technology affordances for intersubjective learning: A thematic agenda for cscl. In D. S. T. W. C. T. Koschmann (Ed.), *Computer supported collaborative learning 2005: The next 10 years!* (pp. 662-671). Mahwah: Lawrence Erlbaum Associates.

van Joolingen, W. R., & de Jong, T. (1997). An extended dual search space model of learning with computer simulations. *Instructional Science, 25*, 307-346.

van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E. R., & Manlove, S. (2004). Co-lab: Research and development of an online learning environment for collaborative scientific discovery learning, *Computers in Human Behavior* (Vol. 21, pp. 671–688).

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. Amsterdam: Academic press.

Voss, J. F., & Post, T. A. (1985). On the solving of ill-structured problems. In M. T. H. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise*. Hillsdale: Lawrence Erlbaum.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition* (Vol. 1, pp. 205-232). New York: Academic.

Willemain, T. R. (1995). Model formulation: What experts think about and when. *Operations Research, 46*, 87-122.

Zhang, B., Liu, X., & Krajcik, K. S. (2006). Expert models and modeling processes associated with a computer-modeling tool. *Wiley InterScience*.

# 6 Appendix A: case

HRM CASE: SuperChem

Pieter van Nieuwenhuizen, HRM consultant at Personnel Advise B.V., opened his E-mail in the morning. Superchem's vice-president, Bert Poortvliet, sent him a very urgent E-mail, asking Personnel advice to help Superchem. The E-mail explained that the 93 years old chemical company, which had a turnover of 2 billion Euros in 2004, faced a severe threat though it performed financially well.

Superchem had introduced a quality management program called "Quality for everybody" three years ago. The program aimed to improve the quality of Superchem's Research & development (R&D). R&D employees were placed in product oriented teams, in which different organizational layers and functions strived to innovate products together. At first glance resistance from the personnel was expected, but fortunately it did not occur and the enthusiasm for the new teams grew slowly. The quality management program changed the way people worked together at Superchem. The time to market new products was decreased, the product quality was improved and Superchem became more efficient. The company's cultural values were changed as well during the last years, compared to the old values. In the previous situation cultural values like hierarchy, seniority, functions, hours worked, etc. were appreciated.

However, Pieter read in Bert's E-mail that the absence through illness percentage had doubled from 5 to 10 last year and the turnover of R&D personnel grew with 25 percent to 30 employees in 2006. At the moment the E-mail was sent only 350 R&D employees worked at Superchem. The absence through illness reduced productivity. In addition, the number of conflicts among employees increased. On the other hand the incoming personnel from recruitment stayed the same. Unfortunately, Bert did not have the manpower in the field of HRM to solve this problem and therefore he contacted Pieter to fix it.

Pieter studied the problem and decided that in order to solve the problem he needed to map the situation for the upcoming months. Which are the problems' factors and how do these influence each other and the in and outflow of R&D employees?

**Assignment**
Help Pieter to understand the problem by constructing a model. Examples are causal maps, case diagrams, system models, etc. You have 45 minutes to model the situation mentioned above with the Co-Lab modeling environment or on the white board. Extract relevant variables from the case description and your own experiences and try to link these. You will notice that only a few facts about Superchem are available. Therefore you have to make assumptions by using your own experience or / and theoretical knowledge. The assignment's aim is to make a model that is as complete and valid as possible, so try to assess and improve your model during the process.

# Appendix B: coding scheme

| **Scientific reasoning category** |
|---|
| Orientation |
|     Interpreting |
|     Defining variables |
|     Theoretical knowledge |
|     Experiential knowledge |
|     Refering to instruction |
| Hypothesizing |
|     Predicting |
|     Hypothesis generation |
|     Domain talk |
| Experimenting |
| Model implementation |
| Model evaluation |

| **Other category** |
|---|
| Actions |
| Regulation |
| Off task |
| Experimenter |