Evaluating Qualitative Reasoning as a Support Tool for the Transfer of Conceptual Knowledge

Joram Rafalowicz
Supervisor: Dr. Bert Bredeweg

Doctoral Thesis
Social Computer Science - Complex Human Computer Systems

Human Computer Studies Laboratory
Informatics Institute
Faculty of Science
University of Amsterdam

Kruislaan 419, Matrix I
1098 VA Amsterdam, The Netherlands

August 13, 2007

Student Number: 9842837

Abstract

Qualitative reasoning (QR) is the area of Artificial Intelligence (AI) that captures and simulates conceptual knowledge about system behaviour. This thesis explores how QR can support the transfer of conceptual knowledge and how tools that provide that support can be evaluated. The approach in this thesis is to create an evaluation framework. This framework provides structure for exploring and evaluating how a QR software program can support knowledge transfer. The framework contains five types of support QR can provide to knowledge transfer. Dimensions are provided to describe usages of these types and guidance for evaluation those usages are set up. To investigate the usefulness of this framework it applied to Garp3. Garp3 is a software tool with a graphical interface that can be used to build and simulate QR models. Two studies are conducted to evaluate two usages. The first study addresses the acquisition of conceptual knowledge by working with QR and showed that novice users can work and learn with Garp3 with minimal software instructions. The second study addresses the articulation of conceptual knowledge with a recently developed structured approach. This study showed that novice users were able to build models with Garp3 with minimal software instructions. However, instruction and support to the structured approach is vital to model building success. The thesis discusses QR, ways to use and evaluate tools based on QR, and the specifics to the two evaluations studies that have been carried out.

Contents

Chapter 1

Introduction

"Qualitative reasoning (QR) is the area of Artificial Intelligence (AI) which creates representations for continuous aspects of the world, such as space, time, and quantity, which supports automated reasoning with very little information" [Forbus, 1996]. The early origins of QR can be found in the 1977 paper by de Kleer in which he examined multiple knowledge representations, both quantitative and qualitative, which were used to create a problem solver for simple mechanical problems [de Kleer, 1977]. In the years that followed, QR developed into a research area of its own, with interests from both the scientific (e.g. [Salles and Bredeweg, 2005]) and industrial community (e.g. [Struss and Price, 2005]).

QR deals with conceptual knowledge and the description of systems and their behaviours. Qualitative models capture conceptual knowledge about system behaviour. "Qualitative reasoning provides a vocabulary (an ontology if one likes) by which computer programs can reason about the behavior of systems in such a way that these computers can communicate about the behavior of these systems with humans." [Bredeweg and Winkels, 1998]. QR software can serve as a medium for the transfer of conceptual knowledge; articulating knowledge from humans to computers [Salles and Bredeweg, 2002] and acquiring knowledge by humans from computers [Forbus, 1996], [Bredeweg and Winkels, 1998], [Werf, 2003], [Tjaris, 2002]. However, it is not evident how support for knowledge transfer using QR can be evaluated.

This thesis explores how QR software can support the transfer of conceptual knowledge and how such software can be evaluated. The approach described in this thesis is to create a framework to provide a structure to evaluate the support QR software can provide. The framework consists of three parts: knowledge transfer types, dimensions of usage and approaches to evaluation. To investigate the usefulness of this framework, it will be applied in the context of Garp3 in this thesis. Garp3 is a software tool, which offers a graphical interface to build and inspect QR models and run simulations. The Garp3 workbench was developed as part of the Naturnet-Redime[1] project [Bredeweg et al., 2006a] to complement GARP [Bredeweg, 1992], a domain independent qualitative reasoning engine.

Two usages specific to Garp3 are defined in this thesis using the evaluation framework. Two studies will be carried out that evaluate to what extent these usages provide knowledge transfer support. The first study addresses the acquisition of conceptual knowledge by observing qualitative simulations. The second study addresses the articulation of conceptual knowledge with the recently developed structured approach [Bredeweg et al., 2007] and the acquisition of knowledge through constructing a qualitative model. The evaluations of these two usages address three of the five support types.

Beyond evaluation of Garp3, this thesis aims to investigate the usefulness of the framework, and also provides insight in the issues that are encountered when using QR tools for articulation and transfer of conceptual knowledge.

Problem statement

QR appears to be a promising way to support transfer of conceptual knowledge. However, little information is available on how QR tools can support this, and how this support itself can be evaluated. The following questions are therefore investigated:

- How can QR based software support transfer of conceptual knowledge?
- Can a framework be developed to facilitate evaluating to what extent a particular QR software program can support this transfer?

The Garp3 studies are used as cases to provide insight on a number of levels. Using the two studies it investigates the usefulness of Garp3 in knowledge transfer support. The thesis also explores the utility of the framework's approach to evaluating QR software. These insights can be used to improve QR

[1] http://www.naturnet.org/

tools that aim to support knowledge transfer. Beyond improving tools, it can also provide insight in the processes of conceptual knowledge articulation and acquisition.

Thesis outline

Chapter 2 discusses the theoretical background for the framework. The key concepts are explored before the framework is introduced. The framework consists of three parts: knowledge support types, dimensions of usage and guidance to evaluation. Chapter 3 focuses on Garp3. The representation used by Garp3 and its features are described. The structured approach to model building [Bredeweg et al., 2007] is discussed and the framework is applied to Garp3. Following the framework two usages are chosen for evaluation. These two evaluation studies are described in chapters 4 and 5. For both studies the results are analysed and discussed. In the last chapter conclusions are drawn on the support QR can provide for knowledge transfer and the framework presented in this thesis.

Chapter 2

Knowledge transfer using qualitative reasoning

2.1 Introduction

This thesis examines the use of qualitative reasoning (QR), and QR software specifically, for the transfer of conceptual knowledge. This chapter provides background information to this exploration of concepts. First Qualitative Reasoning is described. This is followed by an examination of the transfer of conceptual knowledge. Section 2.3.1 looks at QR techniques to support the transfer of conceptual knowledge. The last section focuses on how to evaluate the support QR can give. In this section an 'evaluation framework' is developed, which aims to provide a structured approach to evaluation of support a QR program can give to conceptual knowledge transfer.

2.2 Background

2.2.1 Qualitative reasoning

The field of QR is concerned with the representation of the physical world and the automation of reasoning about the entities and properties of this world [Forbus, 1996]. QR contrasts with quantitative reasoning in that QR makes it possible to reason about the world without precise and complete numerical information. In this sense, QR resembles human reasoning more closely than quantitative methods. Iwasaki provides an informal example to explain the usefulness of QR [Iwasaki, 1997]: "To get a flavor of what is meant by qualitative reasoning, consider a the following everyday scenario: You are confronted with a waterfilled pan on a lit stove. You can easily predict that the pan will warm up, which will warm the water, the water will start to boil sometime, the pan may eventually become empty, and so on. To make these predictions, I need not tell you the exact values of the variables involved, such as the amount of water, the temperatures of the stove and the water, or the boiling temperature. Neither do you need to know the exact mathematical relations among the variables Because of the lack of precision in the available information, you would not be able to say exactly when the water will start to boil or how long it will take it to evaporate completely. Nevertheless, this type of imprecise prediction suffices in many situations to allow people to react appropriately."

QR deals with conceptual knowledge. QR is concerned with structural and behavioural knowledge of systems. In other words QR models aim to capture the structure of a system and can be used to simulate how the system behaves over time. QR is domain independent, given that the domain contains system behaviour. Domains suitable for qualitative modelling contain for example physics or ecology as opposed to history. In physics for example, causal relations for system behaviour are always the same (e.g. an apple always falls down, due to the force of gravity), whereas in history fixed causal relations are absent (one human behaviour does not always lead to a fixed reaction) and behaviour can not always be derived from physical structure. In order to create a qualitative model a qualitative representation is needed, which allows for description of, and reasoning about the physical world. A qualitative representation is a language to articulate knowledge. Such a language provides for a notation that can be used to describe and reason about continuous properties of the physical world [Forbus, 1996]. Qualitative models capture conceptual knowledge on system structure and behaviour. According to Forbus two key issues to qualitative representations are resolution and compositionality. The resolution of a representation is the level of detail it can represent. Compositionality is the ability to combine representations of aspects of a system into a whole. When developing a qualitative representation tradeoffs associated to each of these two issues have to be considered, e.g. not all models of physical systems require the same level of detail. Consider, for example, a representation of a traffic light system, for a pedestrian or motorist the representation should capture that a green light means go and a red light stop. For a traffic light repair engineer the representation needs a much higher level of detail to be useful.

Qualitative simulation

Qualitative simulation is the prediction of the possible behaviours consistent with incomplete knowledge of the structure of physical system [Kuipers, 1993]. The behaviour of a system is represented in QR as states and transitions between states. Qualitative representations provide a way to denote the causal relationships of a system, besides the structure, which allows the reasoning over time and transitions between states. With a qualitative representation a qualitative model can be created of a physical system, this model contains both the structural and behavioural information of the system. This model can then be reasoned with or simulated to see the behaviour change over time.

2.2.2 Conceptual knowledge

In this thesis a distinction is made between two types of knowledge: *conceptual* and *procedural* knowledge. QR deals with conceptual knowledge. Conceptual knowledge can be defined as "explicit or implicit understanding of the principles that govern a domain and of the interrelations between pieces of knowledge in a domain" [Rittle-Johnson and Alibali, 1999]. Alternative terms for conceptual knowledge that are used are declarative, descriptive and propositional knowledge. Procedural knowledge on the other hand can be defined as "action sequences to solving problems" [Rittle-Johnson and Alibali, 1999]. Conceptual knowledge and procedural knowledge can be seen as 'know that' and 'know how' respectively [McCormick, 1997].

2.2.3 Knowledge transfer

According to psychological research the human mind stores conceptual knowledge in so-called mental models. Mental models are models people use to reason about the physical world. These mental models allow people to reason about, explain and predict the behaviour of the world around them [Greca and Moreira, 2000]. The acquisition of correct mental models, i.e. mental models that correspond to correct conceptual knowledge, is considered important in science education [Greca and Moreira, 2000], [Forbus et al., 2005], [Rittle-Johnson and Alibali, 1999], [Bredeweg and Winkels, 1998]. The acquisition of correct mental models is not straightforward. Behaviour, structure and causal relations of a correct model need to correspond to those of the correct conceptual knowledge about a system. This requires deep understanding of the physical world and its behaviour. Knowledge transfer involves obtaining knowledge, also known as acquiring knowledge or learning, and articulating knowledge. Knowledge transfer requires knowledge to be articulated (articulated, or explicit knowledge). Knowledge needs to be captured in some kind of representation (e.g. text, diagrams or models) in order to make learning possible. A distinction can be made for the way knowledge can be obtained; via empirical or inferential learning. Empirical learning involves observing or interacting with the world in order to obtain knowledge. Inferential learning involves obtaining knowledge through reasoning from facts or theory. Knowledge transfer thus involves:

- Articulating knowledge
- Acquiring knowledge empirically
- Acquiring knowledge through inference

Communicative interaction

Computers can help the transfer of conceptual knowledge (e.g. [Çepni et al., 2006]). Elsom-Cook describes the transfer of knowledge between two 'agents' in his communication model [Elsom-Cook, 2001]. Agents are systems with a mental state and intentions to change the mental state of other systems. This mental state can be seen as the knowledge an agent possesses. The transfer of knowledge between agents takes place through communicative interaction. Interaction between the agents takes place through a medium. To transfer knowledge an agent has to articulate knowledge in a natural or artificial language. This representation of the knowledge can then be sent to another agent, who has to interpret the representation and adapt his mental state. The communication model is a general model for the transfer of knowledge. This can be applied to the transfer of conceptual knowledge and more specifically computer-supported transfer of conceptual knowledge. QR serves as a medium for knowledge representation and communication.

2.2.4 Using QR to transfer knowledge

QR can facilitate articulation and acquisition of knowledge. QR can capture many aspects of mental models [Forbus and Gentner, 1997]. This includes the ability to capture partial information about values, causal and mathematical relationships, assumptions and domain ontologies. Furthermore QR deals with conceptual models, making it very suitable to support the transfer of conceptual knowledge. Besides articulation, QR also lends itself very well to the acquisition of knowledge. Not only can a model be interpreted by inspection, the simulation aspect of QR allows for interaction with qualitative systems in the analogue manner as with physical systems, thus enabling empirical learning. This thesis aims to provide a framework to support evaluation of knowledge transfer using QR and QR based software.

2.2.5 Evaluating knowledge transfer support software

The evaluation of knowledge transfer support software has overlaps with the field of evaluation of educational software. The evaluation of educational software is a complex task [Oliver, 2000], mainly because of the wide variety of software types and a lack of a uniform approach, if such an approach even exists. A lot of evaluation research done in the field of educational software focuses on usability (e.g. [Bednarik et al., 2004]) or the predictive evaluation of software, i.e. evaluating software to determine which is most suitable [Squires and Preece, 1999], [Carney and Wallnau, 1998], [Khalifa et al., 2000]. Research is also being done into evaluations of software support to education in terms of learning effect, but these focus on other kinds of software than QR software; existing internet based communication tools for example [Daradoumis et al., 2006] or visualization tools [Squire et al., 2004].

2.3 Framework

This section describes a framework that provides a structure for evaluating the support QR software can provide to knowledge transfer, both articulation and acquisition of knowledge. The framework contains five types of such support. Instances of these five types can be used to investigate to what extent the software under evaluation facilitates a particular type of support. These instances are called 'usages' from now on. The framework can be used in different ways. First of all, it provides dimensions to define usages of QR programs in knowledge transfer. It can be used to systematically explore what the possible usages of a program are and to what extent usages are comparable. It subsequently facilitates evaluation of whether a QR program supports a particular usage. This also facilitates evaluation of the usefulness of this usage in knowledge transfer.

In this way, the framework's approach aims to facilitate the assessment of the usefulness of a QR program for a particular usage, as well as the utility of the particular usage in knowledge transfer in a particular context. By applying the framework a number of usages can be identified. These usages can be evaluated on their support to knowledge transfer and the best usage can be chosen. By using the dimensions of the framework, the exploration will be structured and the discovered usages will be easily comparable. The framework consists of three parts. The first part consists of knowledge transfer support types; the types of usages of QR for knowledge transfer. These types are explored in section 2.3.1. The second part is the dimensions of usage. These dimensions can be used to describe usages and can be used to explore possible usages of a QR software program. These usage dimensions are described in section 2.3.2. It is important to note here that all usages discussed here are usages where knowledge transfer with QR plays a part. These are all usages wherein QR software tools are used on a computer. The third part of the framework focuses on the evaluation of the QR usages. Every usage can require a different approach to evaluation. This part does not try to dictate how usages should be evaluated. Instead it provides guidance for choosing measurements and focus points for evaluations. Focus is on the knowledge transfer aspects of the usages. The frameworks third part is discussed in section 2.3.3. The framework helps to evaluate how well software supports knowledge transfer of a specific type (a category of knowledge acquisition or articulation) in a specific context (the dimensions of the framework).

2.3.1 Knowledge transfer support types

In this section the QR usages that can be used to support knowledge transfer are explored. By working with QR the user's mental model is either externalised in a qualitative representation (knowledge articulation) or adapted (knowledge acquisition). To explore the ways in which QR can support knowledge transfer, the different support types which can be performed with QR are discussed.

Acquiring conceptual knowledge

The knowledge acquisition part of this framework focuses on empirical learning. Empirical learning can take place through observing or interacting with the world. Interactions of humans with physical systems can be divided in three main categories: (1) controlling and operating, (2) designing and constructing, (3) diagnosing and repairing [Bredeweg and Winkels, 1998]. These categories are all based on the behaviour of the system. By controlling and operating humans interact with a system but don't change it, so that the system performs certain behaviour. By designing and constructing humans create a system that will perform certain behaviour. If a system does not show the desired behaviour, humans need to change the system by diagnosing and repairing. The observation of the world plays big part in each of these categories, it is an important way of gathering information. The fourth category of empirical learning is (4) observing and inspecting.

The same four categories can be applied to learning with QR. Qualitative models are representations of physical systems. If a user is learning about a system, the user develops a mental model of the system in question. In most cases this mental model already contains some of the conceptual knowledge and possibly some knowledge of a different nature (for example numerical knowledge or factual knowledge). The goal of learning conceptual knowledge is to build a mental model that corresponds to that knowledge. This section examines the four ways of empirical learning with QR.

(1) Acquiring knowledge by controlling and operating Controlling and operating requires an existing system, i.e. an existing qualitative model. Users can interact with the model by simulating it with different parameters or initial states and examining the behaviours. This can be used to construct, test and/or adapt knowledge in the minds of users and it can be used to test existing models, i.e. does the model show the expected behaviour.

Controlling and operating qualitative models is achieved by simulation. Users can interact with the simulation in analogue fashion to the real world. So in a model of liquids and container, a user could open a valve for example. The simulator would simulate the results of the action and give the user feedback. This type of interaction leads itself very well to experimentation, mostly because of the immediate feedback. This kind of simulations is seen often in quantitative simulators, but not in qualitative simulators.

(2) Acquiring knowledge by designing and constructing By designing and constructing a qualitative model of a system, the user must explicate the structure and causal relations of the system. The user must think about the system at a conceptual level, the user's mental model must be adapted to correspond with the conceptual model of the system. If the user is to create a correct qualitative model, the user's mental model will correspond to the conceptual model of the system. Using model building to learn from has been shown to be successful. VModel [Forbus et al., 2005] is visual qualitative modelling environment for students in secondary education that has shown positive results [Bredeweg and Forbus, 2005]. Betty's brain [Leelawong et al., 2001] uses a different approach to model building. Users build a qualitative model to 'teach' an agent about a system. The agent can give feedback and ask questions to the user. This approach has also yielded positive results.

(3) Acquiring knowledge by diagnosing and repairing To recognise erroneous behaviour of modelling error requires a good understanding of the system and its behaviour. If the user has no clue to how a system should behave, erroneous behaviour will look the same as correct behaviour. The same goes for the model itself. But that is not to say that users cannot learn form diagnosing and repairing, on the contrary. The process of recognising erroneous behaviour, finding the cause and fixing the problem can lead to better understanding. To be able to do this in the first place the user must posses a mental model, which will allow the recognition of the erroneous behaviour. The process of finding the cause and fixing it can lead to a much better mental model and understanding of the system and its behaviour.

(4) Acquiring knowledge by observing and inspecting By inspecting qualitative models and simulations users can gain insight on the structure and behaviour of a system. Research into learning through inspection of simulation results has shown that guided learning with simulations lead to better understanding [Squire et al., 2004], [Tjaris, 2002]. By observing the behaviour of a system, observing the effects of certain actions, users can learn about the system on a conceptual level.

Articulating conceptual knowledge

The articulation of conceptual knowledge can be used for many purposes. As the previous section showed, for a lot of learning activities correct qualitative models are needed. Furthermore articulated conceptual knowledge can be used to support decision-making, reaching consensus and communication in general. QR is very suitable for conceptual knowledge as stated earlier. The reasoning aspect of QR allows users to test their articulated conceptual knowledge, providing extra control over the correctness of their representation.

(5) Articulating knowledge by designing and constructing The qualitative representation is a language to express conceptual knowledge. For a representation a trade-off can be made between expressiveness and ease of use. A representation with high expressiveness can be very difficult to model with. A simplified representation may have less expressive power, but can be easier to use. QR software can include tools to build models. These tools can support the user in the model building effort (e.g. [Groen, 2003]). The support QR software can give can include help with the syntax of the representation and the model building process. The ability of QR software to simulate the models allows users to observe the behaviour of the their models. This allows users to compare the simulated behaviour with real or expected behaviour. Making it possible to test if the articulated knowledge is correct. This is actually debugging of the articulated knowledge. it is possible that someone has a correct mental model, but that the knowledge is represented correctly.

This support type is different from support type 2, because the knowledge is transferred in the other direction. The knowledge transfer goal of type 2 is to acquire knowledge, the support QR can give is geared to knowledge acquisition. The support to articulation is focussed on other aspects of the QR. For articulation it is important that the user's mental model is captured in a correct qualitative model.

2.3.2 Dimensions of usage

The context of use of a computer system can affects its utility and usability [Preece et al., 2002], [Beyer and Holtzblatt, 1999]. Equally, evaluation of QR software should take into account the context in which it is used. Usages can be distinguished from each other by looking at their contextual properties. [International Standards Organization, 1998] defines context of use as the users, goals, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used. These properties are used as a basis for the dimensions of usage. The domain of the usage is added as a property as it is the subjects of the knowledge transfer usage.

Domain The domain of a usage is the subject of the usage. It is the domain of which knowledge is to be transferred. As QR is suited only for domain with systemic behaviour, the domain of a usage has the same constraints. Domains can include for example physics, ecology and automotive industry.

Users The focus here lies on the level of expertise of the user. This relates mostly to expertise on QR and the QR software program, expertise on the domain and expertise with computers. It is important to take this into account because it can have an effect on many other aspects of the usage.

Environment The environment where the usage is performed. This can be an educational or non-educational environment. For educational environments the following distinction is made:

- Formal education
  - Primary education
  - Secondary education
  - Higher education
- Non-formal education

Non-formal education relates to all forms of education outside of the formal educational system. This could be a workshop or self-education. Non-educational environments can include vocational, scientific and home settings. The environment of a usage can determine how much control there is over the usage; there can be much more control over a usage in a classroom than with a user sitting behind a computer at home. The environment also contains the set-up of the usage, how the users will work (e.g. guided or unguided), how much time the usage should take, how many users there are and how they work together (e.g. synchronous or asynchronous).

Tasks The tasks of a usage are the activities that are performed during a usage. This could be building or testing a model for example.

Equipment This includes the hardware, the software and the materials needed for the usage. The hardware must be able to run the QR software. The software includes the QR software, but can also contain additional software (e.g. tools for communication or collaboration). The material can include domain sources, instructions.

Goals A distinction can be made between two types of goals for a usage: task goals and knowledge transfer goals. Task goals are the goals the user is to achieve through the tasks (e.g. to build a model or complete a number of assignment). The task goals are the goals that are communicated to the users and relate to the tasks of the usage. The knowledge transfer goals refer to the desired knowledge transfer effect. The knowledge transfer goals describe which knowledge should be transferred. The knowledge transfer goals should also take into account the initial and final knowledge states or mental models.

2.3.3 Evaluation

In section 2.3.1 five types of support were identified, four of which fall in the category of acquisition support. The other types provide support for the articulation of conceptual knowledge. The goal of evaluating knowledge transfer with QR is to give some measure of quality, either quantitative or qualitative, to the QR usage with respects to the knowledge transfer. The subject of the evaluation is the usage in which the knowledge transfer takes place. When evaluating a usage the dimensions should be taken into account. Such contextual elements are also taken into account in other frameworks for evaluation of learning support. [Freitas and Oliver, 2006] for example have proposed a framework to evaluate games and simulations in education. Their framework consists of four dimensions: context, learner, mode of representation and process of learning. These are the elements, which need to be considered when evaluating a game or simulation in education. The framework presented here is focussed on the conceptual knowledge QR provides as opposed more to Freitas and Olivers focus on quantitative 'realistic' games and simulations.

Evaluation of acquisition Evaluation of learning conceptual knowledge has to focus on what the users are supposed to learn. [Bouwer et al., 2002] points out the key requirements with respect to what is needed to effectively interact with systems and their behaviour. The central issue is behaviour analysis, or more specifically the pre- and post-diction of behaviour [Forbus, 1984] and the deriving of behaviour form structure. The ability to predict and explain the behaviour of a system is considered a good measure of conceptual knowledge. This can be used to measure the learning effect of a usage and help evaluate it. The learning effect is a key measurement for learning usages. The learning effect is a measure of the amount that is learned by the usage. It is the difference in knowledge the learner has before and after the usage. It can provide a quantitative measure for the quality of the usage. In an evaluation study it can be very useful to compare the learning effects of different usages or methods.

Evaluation of articulation The evaluation of the support to the articulation of conceptual knowledge should focus on the expressiveness and process and the Qualitative representation. The expressiveness of a qualitative representation determines what can be articulated with that representation. Assuming that the conceptual knowledge a user wants to articulate can be captured in the QR representation, the process should be evaluated. What is important here is how the QR software program supports the user in the model building effort, for the actual modelling, debugging and testing with simulation.

Chapter 3

Garp3

3.1 Introduction

The research presented in this thesis focuses on Garp3 [Bredeweg et al., 2006a], a QR software implementation. Garp3 is a visual tool for building and simulating qualitative models with an advanced graphical user interface (see figure 3.1) written in SWI-Prolog[2]. It was developed at HCS lab at the University of Amsterdam and is based on previous tools: model building tool Homer [Jellema, 2000], simulation tool Garp2 [Linnebank, 2004] and simulation inspection tool VisiGarp [Bouwer and Bredeweg, 2001]. Garp3 uses a complete qualitative representation, it uses quantity spaces (intervals) to represent quantities and has a very extensive set of relations. Garp3 allows users to design, build and simulate qualitative models as well as providing means to share models via an online repository and re-use model parts. In this chapter Garp3 is examined, first a short description of the representation used by Garp3 is given followed by an overview of the features of Garp3.

3.2 Representation

A Garp3 model is a representation of a system and its behaviour. A Garp3 model consists of the following ingredients.

Entities and agents The entities are the physical objects in the system. The entities are arranged in an is-a hierarchy. Agents are a special type of entities, they are entities that reside outside of the system. Agents can influence the system in the same way as entities, but cannot be influenced by the system. Agents are also arranged in an is-a hierarchy.

Configurations Configurations are the relations between entities.

Attributes Attributes are properties of entities, which do not change over time.

Quantities and quantity spaces Quantities are properties of entities that can change over time. Each quantity has an associated quantity space. A quantity space is a range of qualitative values the quantity can assume. Each value in a quantity space is a point or an interval.

Assumptions Assumptions are ingredients used to constrain behaviour of the model. Assumptions are arranged in an is-a hierarchy.

Dependencies Garp3 contains a number of dependencies, which can be used to describe the behaviour of the system:

- Inequalities
- Values
- Influences
- Proportionalities
- Correspondences
- Plus/Min

[2] http://www.swi-prolog.org

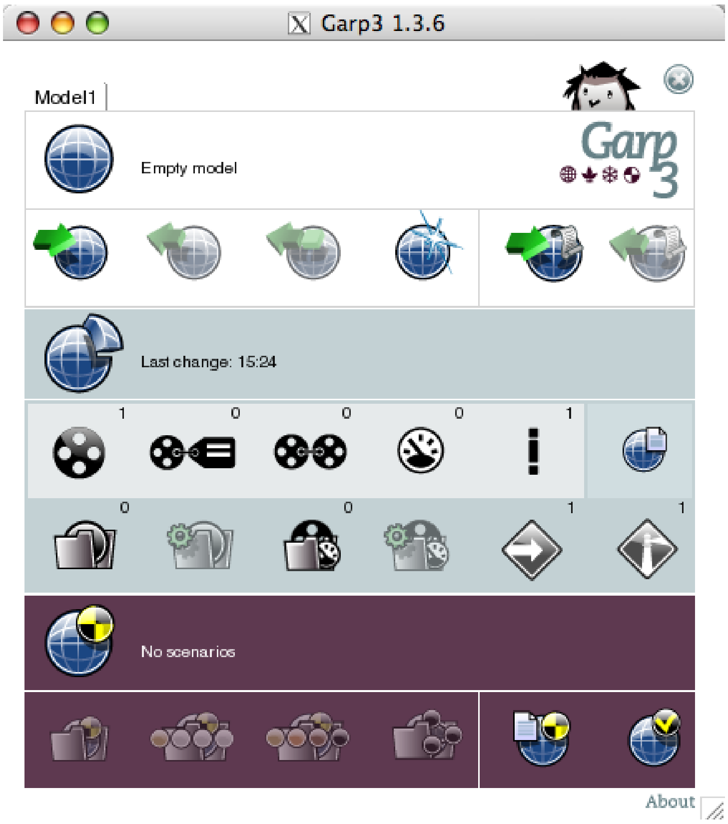

Figure 3.1: Main menu of Garp3.

Model fragments and scenarios Model fragments represent part of the structure and behaviour of the system. The ingredients in a model fragment are either a condition or a consequence. The way model fragments are used in the simulator is of the form IF conditions THEN consequence. There are three types of model fragments: static, process and agent model fragments. Static model fragments represent the structure of the system. Process model fragments describe the processes in the system. Agent model fragments describe the external influences of the system.

Scenarios represent an initial state of the system. A scenario is used as a starting point for the simulator. The simulator tries to match conditions from model fragments to the scenario and executes the consequences. Thus creating new states of the system and simulating the behaviour.

3.3 Features

The Garp3 workbench contains three environments: build, simulate and sketch.

3.3.1 Build environment

The build environment of Garp3 allows users to use a graphical interface to build a model with the representation described in section 3.2. The build environment contains nine editors to create the different model ingredients. With these editors all ingredients described in section 3.2 can be created and inspected. Garp3 also has some extra features to aid the model building process.

Copy-paste

Garp3 allows the user to copy certain elements of a model and paste it in another model, thus providing a way for re-using parts of models.

OWL repository

Garp3 can save models in the OWL format and upload it to an online repository, users can also download models from this repository and load it into Garp3. This feature combined with copy-paste allows users to start re-using parts of models to create a base for their model, saving time and effort.

3.3.2 Simulate environment

The simulate environment of Garp3 allows users to simulate a model within a graphical interface and inspect a variety of elements within the simulation. The simulator can do a full envisionment (all possible states and transitions from the scenario) or can simulate step by step. The simulator takes ambiguity into account, if more than transition is possible, all those transitions will be simulated. The simulate environment allows the user to inspect many aspects of the simulation results. For each state the user can inspect the entities and their relations, the values of the quantities, the active model fragments and the dependencies. The dependencies are visualisation of the entities and the dependency relations between them. For a set of states the user can inspect the history of the transitions, equations and values.

3.3.3 Sketch environment

The sketch environment was created as a set of tools to support the structured approach to model building [Bredeweg et al., 2007]. It contains a number of tools with a graphical interface. The sketch environment and the structured approach are described in the following section.

3.4 Sketch tool and the structured approach

Garp3 supports the structured approach with the sketch environment (see figure 3.2). In the following section the structured approach and the sketch environment are examined with a focus on how they are related. With the sketch environment users can build a concept map, a structural model and a causal model and they can define the processes, the actions and external influences, the scenarios and the behaviour graph. This forces a certain structure upon the user, without the strict syntax of Garp3, which makes it easier to experiment and explore the domain and create a basic structure to start modelling from. The structured approach was designed to support users in two ways. First, the approach structures the model building effort. Second, it supports users in comparing and evaluating the intermediate representations and the final model.

The structured approach is a framework consisting of six steps detailed below (see figure 3.3), the steps represent a waterfall approach. However the actual activities are more akin to a spiral approach [Boehm, 1986]. The set of ingredients is gradually created and refined during the steps, resulting in implemented model ingredients (see figure 3.4 for an overview of the ingredients).

To support the structured approach Garp3 contains a sketch environment. The sketch environment contains tools, which correspond to elements of the structured approach. In the following sections each of the six steps and the corresponding sketch tools are described briefly. In figure 3.3 the six steps are represented in the boxes. For each step the 'deliverables' are shown below the boxes.

3.4.1 Orientation and initial specification

This is a broad exploratory phase, defining the scope and level of detail of the model. The sketch environment supports this step with two tools.

Abstract, Intended audience, Model goals and General remarks This tool provides four separate text fields for description of these elements. The deliverables documentation and model goals can be created with this tool.

Concept map The last deliverable of this step, the concept map, has its own tool in the sketch environment. The concept map is represented as a graph, where the concepts are visualized as nodes and relations as arcs. The concepts defined in the concept map are used to create the rest of the sketch model. Some concepts are re-used, some are refined, some are discarded, but the concept map should provide an information pool for the rest of the sketch model.

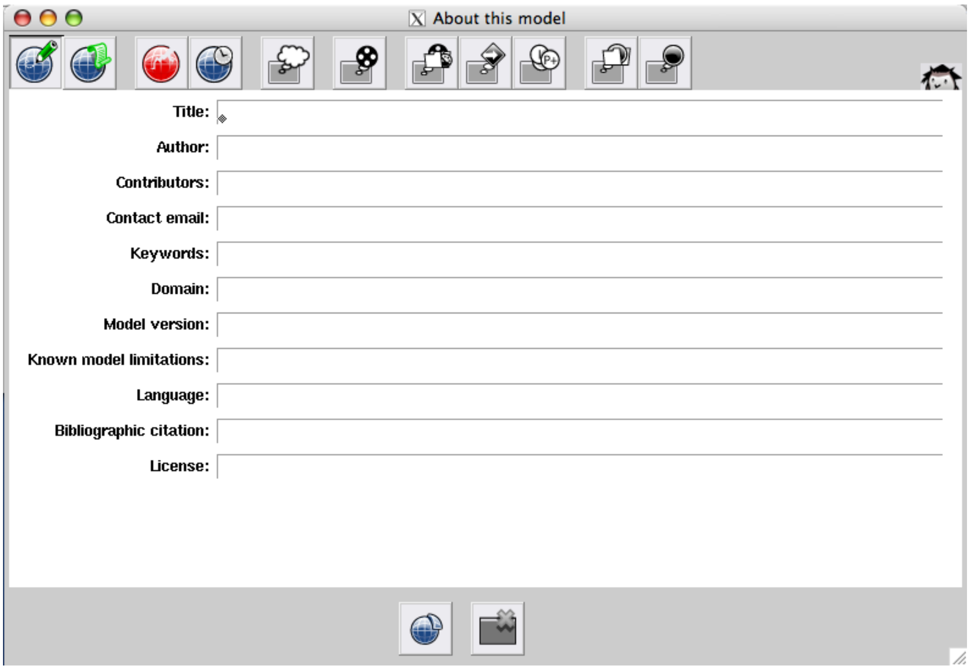

Figure 3.2: Main menu of the sketch environment, the icons represent the tools. From left to right: 'General information', 'Abstract, model goals, intended audience and general remarks', 'Model status and bug reports', 'Model data', 'Concept map', 'Structural model', 'Processes', 'Actions and external influences', 'Causal model', 'Scenarios', 'Behaviour graph'

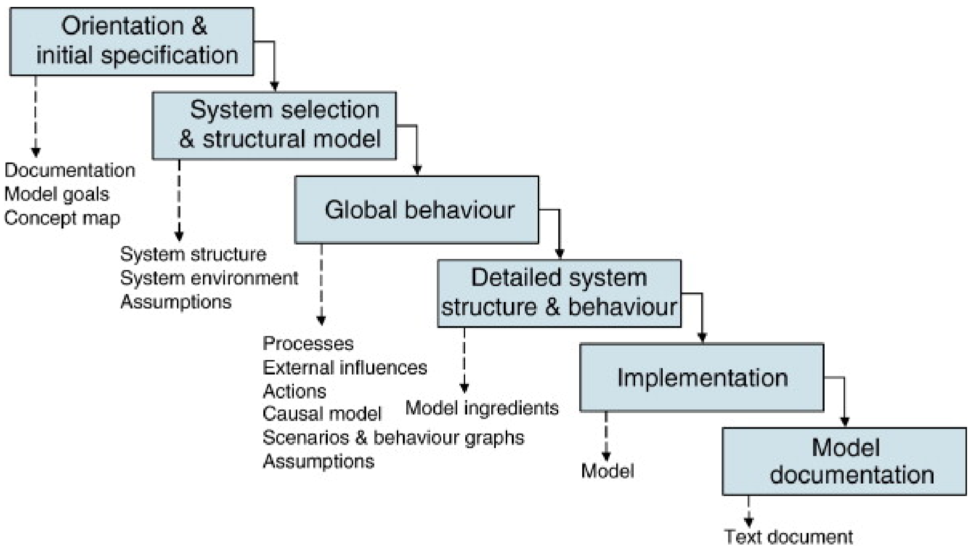

Figure 3.3: Structured approach to building a qualitative model. The corresponding elements are indicated with each step. Taken from [Bredeweg et al., 2007]

3.4.2 System selection & structural model

In this step the focus lies on the physical structure of the system.

Structural model The structural model contains the physical structure of the model (the entities and agents), assumptions and the relations between them all (how they are configured). The structural model is represented in the same vein as the concept map. The structural model can be seen as the backbone of a model, whereas the concept map provided a pool of information to work, the structural model brings structure to it.

3.4.3 Global behaviour

In this step the behaviour of the system is the focal point.

Processes With the processes tool the processes which are active in the model can be described. Unlike the concept map and structural model, the processes tool is text based. For each process the entities and quantities involved are listed, the start- and stop-conditions, the effects and the assumptions of the process are described.

Actions and external influences The actions and external influences are very similar to the processes, the processes have an internal instigator and the actions and external influences have an external instigator. The only difference with the processes tool is that involved agents can also be listed in the actions and external influences tool.

Causal model The causal model contains the dynamic structure of the model, it contains the quantities of the model and the causal relationships between them. It should contain all the dynamics of the defined processes. The causal model is also represented as a graph.

Scenarios The scenarios are the initial states of the model simulation. In the simulation tool the scenarios can be defined in the same vein as the processes.

Behaviour graph With the behaviour graph the expected behaviour of the model can be described. States are defined with quantities and their values and state transitions, how the system will behave.

3.4.4 Detailed system structure & behaviour

This step is the bridge between the representations created in the previous steps and the QR vocabulary (in this case Garp3). The modellers have to specify all the model ingredients in the QR vocabulary. In this step the QR software is not used to give the modellers more freedom. The sketch environment provides no support for this step at the moment.

3.4.5 Implementation

In this step the model is implemented using the QR software. In this step possible errors and inconsistencies can become apparent and have to be fixed. Furthermore simulating the model can result in unexpected behaviour, leading to possible changes to the model. For this the build and simulate environments of Garp3 can be used.

3.4.6 Model documentation

In this step documentation should be created on the model building effort. This documentation should allow the reader to redo the modelling effort and get the same results. The documentation should contain the details of the information captured by the model. Garp3 does not provide support for this step.

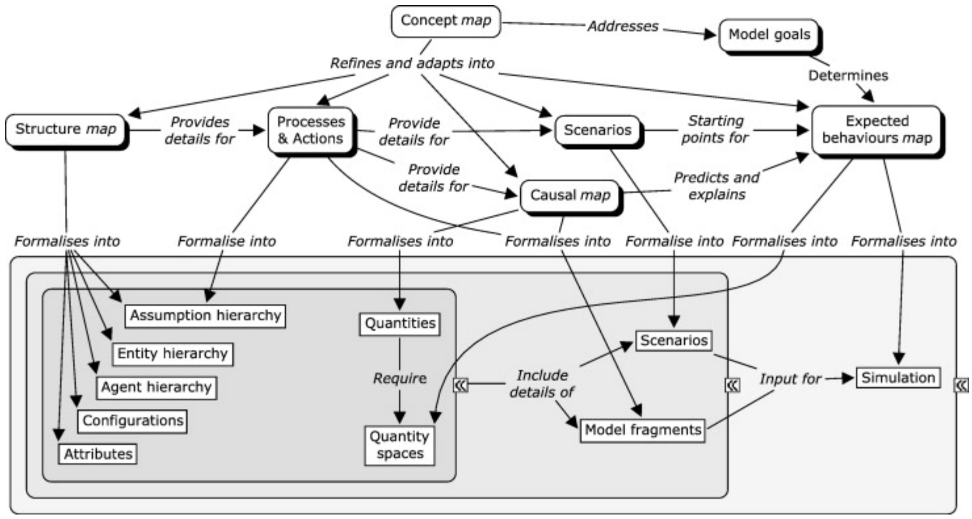

Figure 3.4: Overview of the intermediate representations used in the structured approach. Taken from [Bredeweg et al., 2007]

3.5 Knowledge transfer framework for Garp3

The next three sections detail the application of the framework described in chapter 2 to Garp3. Section 3.5.1 describes how Garp3 can be used for each of the five support types. In section 3.5.2 the dimensions of usage are explored for Garp3. In the last section the evaluation is discussed; this section examines how two usages of Garp3 can be evaluated. These two usages will be evaluated and are described in chapters 4 and 5.

3.5.1 Knowledge transfer types

Knowledge acquisition

(1) Knowledge acquisition by controlling and operating Garp3 has limited support for controlling the simulation. The user can choose the starting scenario of the simulation and choose the type of simulation (e.g. full or step by step). Knowledge acquisition by controlling and operating with Garp3 basically boils down to observing and inspecting simulation results. The user can select or create a starting scenario for the simulation and observe the behaviour.

(2) Knowledge acquisition by designing and constructing The building of a qualitative model is articulating knowledge, so in order to build a model representing a system the builder must have a mental model of that system. This is the basis for knowledge acquisition by model building for it forces the user to build a mental model of the system if it is not available. Garp3 supports model building with it's graphical interface, making it much more usable than text based software. Garp3 also supports the structured approach with the sketch environment. This approach was designed to support the model building effort. When users must create a model of a system they are forced to explicitly think about the structural and causal model of that system, thereby adapting their mental models. The user can use the build and sketch environment to design and construct a qualitative model.

(3) Knowledge acquisition by diagnosing and repairing Garp3 provides no specific functionality for diagnosing and repairing. However the simulate and build environment provides the functionality needed for diagnosing and repairing qualitative models. The simulate environment allows users to inspect the behaviour of the model and the build environment allows the user to diagnose and repair the model.

(4) Knowledge acquisition by observing and inspecting Inspecting a model or observing behaviour can lead to learning if applied correctly. The qualitative model and the simulation can give users insight in the structural and causal relations of the system and its behaviour. Through observing and inspecting the user can create a mental model of the system. As the knowledge captured by the qualitative model is on a conceptual level, the user can create e mental model corresponding to the conceptual model. Thus acquiring conceptual knowledge on the system. The user can observe and inspect a qualitative model, a sketch model and simulation results.

Knowledge articulation

(5) Knowledge articulation by designing and constructing The build and sketch environment of Garp3 provide support for the articulation of conceptual knowledge. The representation of Garp3 (see section 3.2) is very extensive and allows great expressiveness. However building a model with the GARP language can be a complex task. To support this the build environment provides syntax checks and allows visual building of model ingredients. The sketch environment provide tools to structure the model building effort.

3.5.2 Dimensions of usage

Domain Garp3 is domain independent, so as long as the domain is suitable for QR, it is suitable for Garp3.

Users The representation used in Garp3 is fairly complex and building a model is a complex task. Since model building is a complex task it is not reasonable to expect students in primary or secondary education to build working qualitative models. This does not mean that Garp3 cannot be used in primary or secondary education. Garp3's graphical interface allows easy inspection and control of the models. Furthermore the Garp3 workbench could be used in a simplified software tool for use in primary and secondary education. Other features of Garp3 require less expertise, the simulation environment allows inspecting on different levels of detail and tools from the sketch environment can be used separately.

Environment Garp3 can be used in many different environments. For education a distinction can be made between formal and non-formal education. Formal education can be split up into primary, secondary and higher education. Non-formal education is education outside of official education, e.g. self-education or a workshop. Garp3 could be used to support learning in formal education in numerous ways, in primary, secondary and higher education. Garp3 can be used as a stand-alone educational tool, it can be incorporated with other teaching material or the techniques and features of Garp3 can be used to create new educational tools. Garp3 can be used in many different configurations. The sketch environment provides support to discussion and communication with the intermediate representations, making it suitable to be used in collaborative modelling efforts [Salles and Bredeweg, 2002].

Tasks The tasks that can be performed with Garp3 are split over the three environments of Garp3. The build environment allows users to build models, copy-paste ingredients and use the online repository to upload and download models. The simulate environment allows users to simulate models and inspect the simulation results. The sketch environment allows users to apply the structured approach with a number of tools.

Equipment The Garp3 software is platform independent and is freely available for download[3]. It needs the freely available SWI-prolog[4] installed in order to run.

3.5.3 Evaluation

As part of the application of the evaluation framework to Garp3 two usages have been chosen for evaluation. These two usages have been selected to both evaluate Garp3 and to evaluate the usefulness of the framework.

The first usage addresses knowledge acquisition by observing and inspecting. The second usage addresses the articulation of knowledge by designing and constructing. The dimensions of these usages are

[3] http://hcs.science.uva.nl/QRM/software/
[4] http://www.swi-prolog.org

Dimension	Acquisition by observing	Articulation by constructing
Domain	This usages contains two domains: 'Deforestation' and 'Fuel and global warming', both part of sustainable development.	The domains of this usage are development goals of sustainable development.
Users	The users are novices on QR, Garp3 and the domain. They are experienced with the computer.	The users are intermediates on QR and Garp3 and experienced with computers. They are novices on the domain.
Environment	The environment is higher education. The users work individually and are strictly guided by assignments. The users spend approximately one hour and 15 minutes on the usage.	The environment is higher education. The users work individually and are guided by experts. The usage was spread out over 8 weeks.
Tasks	The users have to use the simulate environment of Garp3 to inspect the dependencies and behaviour of the supplied models.	The users have to use the sketch environment and subsequently the build environment to build a working qualitative model. The users also have to fully document the modelling effort.
Equipment	The users use the Garp3 software and a web browser. The materials for this usage are the assignments and instructions.	The users use the Garp3 software. The users have access to documentation on the domain and the structured approach.
Goals	The task goals of this usage are to complete the assignments. The knowledge transfer goals are to learn conceptual knowledge on the domains.	The task goals of this usage are to build and document a working qualitative models. The knowledge transfer goals are to acquire conceptual knowledge on the domain and articulate that knowledge.

Table 3.1: The dimensions of usage for the two usages.

described in table 3.1. In the first usage higher education students use the simulate environment to inspect qualitative models and simulations through guided assignments. In the second usage higher education students build and document a functioning qualitative model with the sketch and build environment following the structured approach.

For evaluation of the first usage the focus lies on learning effect; the difference in knowledge before and after the usage. To measure the knowledge the pre- and post-diction skills of the users is measured. This evaluation addresses support type (4) knowledge acquisition by observing and inspecting.

The evaluation of the second usage is more qualitative in nature than the evaluation of the first usage. The focus is on observing the modelling effort and to investigate the issues encountered when Garp3 is used to support this effort. The users in this case have to convert knowledge represented as text, into a QR model. This usage both addresses support type (5) knowledge articulation by designing and constructing and (2) knowledge acquisition by designing and constructing.

The evaluations of these two usages address three of the five support types in total. In the following two chapters the evaluation studies of the two usages are described and analyzed.

Chapter 4

Evaluation of knowledge acquisition by observing and inspecting

4.1 Introduction

Is it possible to acquire knowledge by working with qualitative simulations? According to [Tjaris, 2002] it is. In her study she showed that participants gained knowledge by doing a treatment with the simulation software. The domain knowledge was tested before and after the treatment with a multiple-choice test. The software that was used for the treatment was VisiGarp, which is now modified and part of Garp3 as the simulate environment.

This research is a follow up to the research of Tjaris, focussed on the learning effect of working with the Garp3 simulate environment. The hypothesis is that knowledge of a particular domain can be taught to people with no experience with Garp3 by doing a treatment with the Garp3 simulate environment.

4.2 Hypotheses and assumptions

Hypotheses:

- h1 The simulate environment of Garp3 can be used by novices with minimal instructions.
- h2 By doing a guided treatment with the simulate environment of Garp3, people can learn about the modelled domain.
- h3 The learning effect will be greater if the model is presented with progressive complexity than when it is presented entirely at once.

Assumptions:

- a1 The models of both domains used in this experiment are of equal complexity.
- a2 The treatment assignments are of equal difficulty for both models.
- a3 The questions in the tests are of equal difficulty for both domains.
- a4 Both forward and backward reasoning is tested with the tests.
- a5 The tests are comparable in difficulty.

4.3 Description of usage

Domain This usage covered two domains: 'deforestation' (DEF) and 'fuel and global warming' (FGW), both domains are related to sustainable development and exhibit system behaviour and are appropriate to use in QR.

Users The users were higher educational students with no prior knowledge of the domain and no prior experience with Garp3. They have the necessary computer skills and might have some prior knowledge of QR.

Tasks The tasks were opening models, running simulations and inspecting the simulation results. The users had to answer questions about the simulation results.

Environment The environment of this usage was formal educational. The users worked individually to complete the tasks, they were be guided by instructions and assignments. The users got immediate feedback on their assignments.

Equipment For this usage, Garp3 ran on Windows XP machines. Besides the Garp3 software, a web browser was needed to view the instructions, the assignments and the tests. The start-up instructions were on paper.

Goals The task goals were to complete the assignments given to the users. The knowledge transfer goal of this usage was for the users to learn about the causal relations in the domains.

4.4 Approach

To test the hypothesis a pre- and pos-test set-up was used with a test group and a control group. In the test group the pre-test was taken before the treatment and the post-test afterwards. The control group did not do a treatment between the tests. The control group is needed to rule out the possibility that the participants 'learn' from the pre-test or on their own between the pre- and post-test, thus making the results of the experiment less reliable. In this experiment the learning effect of conceptual knowledge has to be measured. A good measure for this is the ability for pre- and post-diction or explaining and predicting. The pre- and post-test measure pre- and post-diction on the domains. For this set-up two tests are used (test A and B) which will be used as both pre- and post-test for both groups, the tests are designed to be comparable in difficulty.

4.5 Expectations

According to the hypothesis, the participants in the treatment group will learn about the domains on a conceptual level by doing the treatment. The pre- and post-test measure the ability to do pre- and post-diction on the domain, a good measure of conceptual knowledge. If the participants learn form the treatment, their scores on the post-test will be significantly higher than on the pre-test. The treatment is split in two sections, one for each domain. In the first section the model on DEF is presented with a progression in complexity for each question set, in the second section the entire model on FGW is presented at once, without any model progression. According to learning theories progression will improve the learning effect, so it is expected that the participants will learn more on the first section than the second. So the difference in scores between the pre- and the post-test should be greater for the questions on DEF than those on FGW. Finally it is expected that there are no significant differences between test A and test B, for they are considered to be comparable in difficulty.

4.6 Set-up

The experiment was conducted with twenty-eight participants, divided equally over the test and control group. The participants in the control group were first year Information Science students and the participants in the treatment group were first year Artificial Intelligence students, all at the University of Amsterdam. Both the control and the test group did a pre- and a post-test, consisting of multiple-choice questions on the domains of DEF and FGW. Between the pre- and the post-test the treatment group worked with Garp3 and the control group had a lecture on an unrelated subject (Java programming).

For the pre- and the post-test two tests were created; test A and test B (Appendix A), both tests were designed to be of equal difficulty and content. The tests contained 15 questions each, 9 on the domain of DEF and 6 on the domain of FGW. The distribution of the questions on the two domains was mixed, but identical for both tests. The tests contained two types of questions, pre- and post-diction questions. The post-diction questions required backward reasoning. These are questions of the type "A is observed, what has caused this?" pre-diction question require forward reasoning. These are questions are of the type "A happens, what will be the effect?". The difficulty of a question is determined by its dependancy path, e.g. A effects B, which effects C is a dependency path of length 3. The questions of the tests were of varying difficulty (table 4.1). Test A contained 6 forward (4 on DEF and 2 on FGW) and 9 backward reasoning questions (3 on DEF and 3 on FGW) with an average dependency path length of 2.13 and test B contained 9 forward and 6 backward reasoning questions with an average dependency path length of 2.33 .

		Test A		Test B	
		Nr of quest.	Ave. path	Nr of quest.	Ave. path
Forward	DEF	4	2.5	6	2
	FGW	2	1.5	3	2
	Total	6	2.17	9	2
Backward	DEF	5	2.4	3	3.7
	FGW	4	1.75	3	2
	Total	9	2.1	6	2.8
Total		15	2.13	15	2.3

Table 4.1: Number of forward and backward reasoning questions in Test A and B and the average length of the dependency paths.

The treatment consisted inspecting the simulation results of the models of the domains of DEF and FGW with the simulation environment of Garp3. To guide the participants through the domains they had to answer 12 sets of questions on different subjects related to the domains, with a total of 44 questions (Appendix 4.10). The first 6 sets were on DEF and the last 6 sets were on FGW. To answer the questions the participants had to inspect dependencies of the model (as in figure 4.5) and value histories of quantities (as in figure 4.4). For the DEF domain each question set used a different scenario with increasing complexity, the questions on FGW used one scenario. The questions were created by domain experts.

Figure 4.1: Screenshot of explanation during treatment.

4.6.1 Procedure

Treatment Group

The treatment was conducted in a computer room reserved for the experiment. The participants were placed behind a computer and were handed out instructions and their personal id number. All parts of the treatment experiment were done on-line with the computer, the first part of the experiment was

Figure 4.2: Screenshot of question during treatment.

the pre-test. 50 % of the participants got test A as the pre-test and test B as the post-test, the other 50 % got test B as the pre-test and test A as the post-test. This order was assigned randomly. The participants had 15 minutes for the pre-test. During the treatment, the participants had approximately 1 hour and 15 minutes to work with the software and answer the treatment questions. The web-pages contained detailed instructions (figure 4.1), the treatment questions (figure 4.2) and the answers with a visual explanation (figure 4.3). The answers and the time spend on each question set was logged for analysis. Figure 4.4 and 4.5 are representative screenshots of what the participants had to inspect in order to answer the treatment questions. When the treatment was finished the participants did the post-test, again they had 15 minutes. After the post-test the participants were asked to fill out a questionnaire about their background and an attitude questionnaire on the experiment, both questionnaires were done on the computer.

Control group

The control experiment was conducted during a Java programming course. In a short introduction it was explained that the experiment was part of a master research project and the participants were asked to try their best to answer the questions correctly. They did not get any information about the domain nor about the goal of the experiment. The questionnaires were handed out to the participants on paper, 50 % of the participants got test A as the pre-test and test B as the post-test, the other 50 % got test B as the pre-test and test A as the post-test. This order was assigned randomly. The participants had 15 minutes to complete the first test. After the pre-test the participants attended the regular class for 30 minutes and were then presented with the post-test. Afterwards the participants had the opportunity to ask some questions.

4.7 Results

Everybody in the treatment and control group finished well within the time limits on the pre- and post-tests. Out of the 14 participants in the treatment, 9 finished all the treatment questions within the time limit, 3 almost finished and 2 finished roughly 50 % of the questions, one of those two participant started significantly later than the rest.

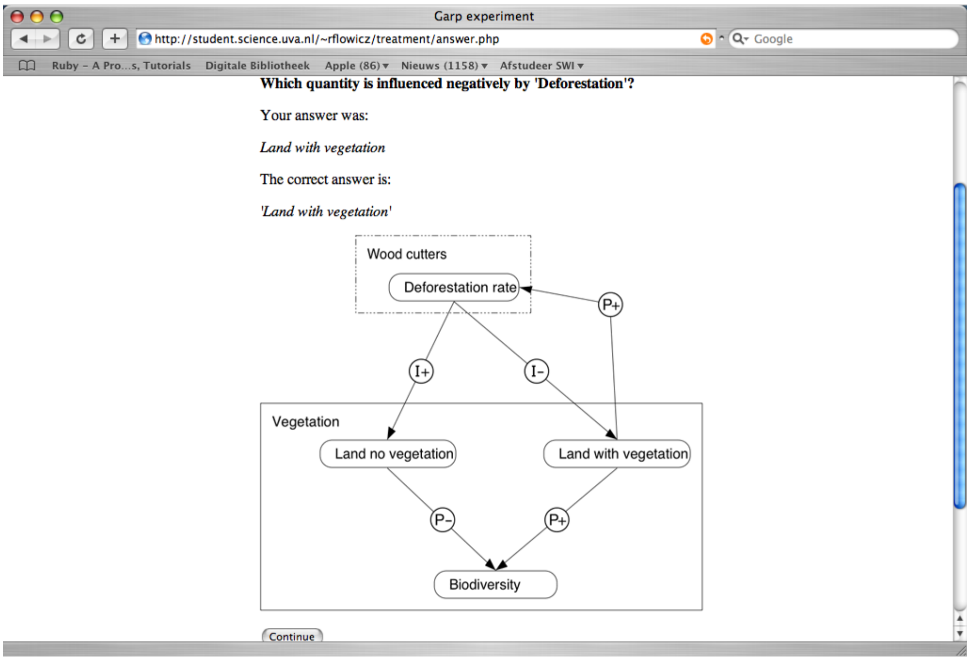

Figure 4.3: Screenshot of answer during treatment.

4.7.1 Analysis

For the analysis of the results two-tied t-tests were used; an independent samples t-test for comparison between the control group and the treatment group (inter) and a paired samples t-test for comparison of the pre- and post-tests within a group (intra). To compute the results SPSS was used and excel was used for further analysis. The scores on the tests are in percentiles (values ranging from 0.00 to 1.00). The mean (m) and standard deviation (sd) of the group scores are of interest. For comparison of group scores the t-value (t) and the significance (sig) are of interest. The results of the t-tests can be interpreted as follows: two sets of values are compared with a t-test, if the significance value is below a threshold, the difference of the values between the two sets is significant; it is not coincidental. For this experiment the threshold is 0.05, roughly meaning a certainty level of 95%.

4.7.2 Results of the pre- and post-tests

The results of the pre- and post-tests are in table 4.2 and the results for the t-tests can be found in table 4.3. The control group scored lower on the post-test (m=0.51) than on the pre-test (m=0.53), but this difference was not significant (t=0.418, sig=0.683). The treatment group scored higher on the post-test (m=0.60) than on the pre-test (m=0.48) and the difference was significant (t=-2.249, sig=0.043). On the pre-test the control group (m=0.53) scored higher than the treatment group (m=0.48), but this difference is not significant (t=0.843, sig=0.407). On the post-test the treatment group (m=0.60) scored higher than the control group (m=0.51), but this difference was also not significant (t=-1.206, sig=0.239). In table 4.3 the results for the t-tests split by domain are also shown. This shows a significant difference between the treatment and control group on the post-tests for the FGW domain. Also no significant difference is measured between the pre- and post-test for the treatment group on the FGW domain.

4.7.3 Further analysis of the test results

Individual performance Table 4.4 and 4.5 show the average scores for the individual participants in the control group and the treatment group. In the treatment group t3, t7 and t40 score lower on the post-test than on the pre-test, the rest improve or equal their scores.

Case	N	Mean	Std. dev.
Pre control	14	0.53	0.20
Post control	14	0.51	0.23
Pre treatment	14	0.48	0.15
Post treatment	14	0.60	0.20

Table 4.2: Average scores for the four cases.

	All	DEF	FGW
Inter			
Pre			
t	0.843	0.871	0.413
sig	0.407	0.392	0.683
Post			
t	-1.206	-0.187	-2.526
sig	0.239	0.853	0.018
Intra			
Control			
t	0.418	-0.688	1.307
sig	0.683	0.503	0.214
Treatment			
t	-2.249	-2.412	-1.389
sig	0.043	0.031	0.188

Table 4.3: T-test scores for inter and intra groups. For all questions and split by domain. For the inter tests negative t-values indicate a higher average for the treatment group. For the intra tests negative t-values indicate a higher average for the post-test.

(a) Pre-test, N=15

ID	Test	Mean	Std. dev.
c1	A	0.80	0.41
c2	B	0.27	0.46
c3	A	0.53	0.52
c4	B	0.27	0.46
c5	A	0.27	0.46
c6	B	0.80	0.41
c7	A	0.73	0.46
c8	B	0.33	0.49
c9	A	0.53	0.52
c10	B	0.67	0.49
c11	A	0.80	0.41
c12	B	0.47	0.52
c13	A	0.53	0.52
c14	B	0.47	0.52

(b) Post-test. N=15

ID	Test	Mean	Std. dev.
c1	B	0.60	0.51
c2	A	0.13	0.35
c3	B	0.20	0.41
c4	A	0.60	0.51
c5	B	0.27	0.46
c6	A	0.87	0.35
c7	B	0.53	0.52
c8	A	0.80	0.41
c9	B	0.33	0.49
c10	A	0.67	0.49
c11	B	0.40	0.51
c12	A	0.80	0.41
c13	B	0.47	0.52
c14	A	0.40	0.51

Table 4.4: Average scores for the participants in the control group. Test A or B indicate which test the participant got as pre-test.

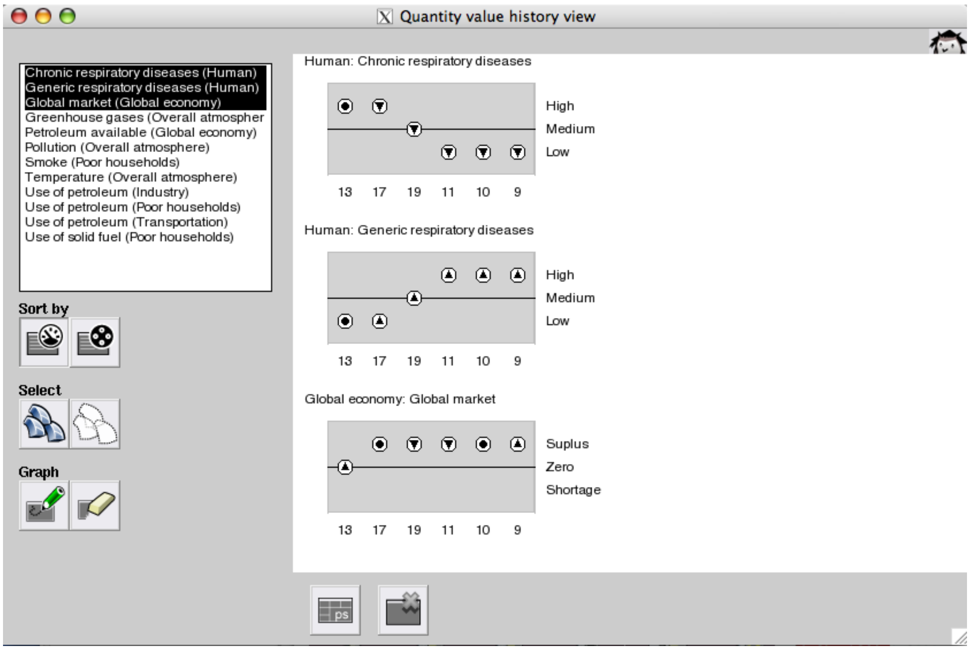

Figure 4.4: Screenshot of Garp3 value history of FGW model.

Test A and B The average scores on test A were higher than test B for the pre- and post-test of the control group and the pre-test of the treatment group, only for the post-test of the treatment group were the scores of test B higher than test A (table 4.6). However none of these differences were significant (table 4.7) supporting the assumption that test A and B are comparable in difficulty.

Domains In the pre- and post-test of both the control and the treatment group the performance on the questions on FGW was lower than on the questions on DEF (table 4.8). This difference was only significant for the post-test of the control group (table 4.9).

4.7.4 Examining the treatment

Scores

The treatment consisted of 12 sets of questions divided over the two domains. The participants had to work with 7 different scenarios, 6 incrementing scenarios on DEF and 1 scenario on FGW. There was a significant difference (t=2.912, sig=0.005) between the questions on DEF (m=0.93) and FGW (m=0.79), set 8 scored especially low.

The averages in table 4.10 are computed over the questions that are answered and not all the participants finished the treatment. Furthermore every set did not have the same amount of questions and some questions had multiple parts, therefore N differs over the sets.

None of the participants scored notably low on the treatment questions they answered (table 4.11), however t3 (N=39) and t40 (N=28) did not get very far, these are also two of the three participants who scored lower on the post-test than on the pre-test.

Time

The average time spent on question sets in the treatment lies between 50 and 150 seconds for most sets (figure 4.6) except for the first and the eighth set, which took much longer. This was expected for the first set, since it included the introduction, so it contains the start-up time, there is no such reason for set 8.

(a) Pre-test, N=15

ID	Test	Mean	Std. dev.
t1	A	0.80	0.41
t2	A	0.27	0.46
t3	A	0.40	0.51
t4	A	0.33	0.49
t5	A	0.60	0.51
t6	A	0.53	0.52
t7	A	0.60	0.51
t21	B	0.47	0.52
t22	B	0.60	0.51
t23	B	0.33	0.49
t24	B	0.27	0.46
t25	B	0.40	0.51
t26	B	0.47	0.52
t40	B	0.60	0.51

(b) Post-test. N=15

ID	Test	Mean	Std. dev.
t1	B	0.80	0.41
t2	B	0.53	0.52
t3	B	0.27	0.46
t4	B	0.33	0.49
t5	B	0.80	0.41
t6	B	0.53	0.52
t7	B	0.53	0.52
t21	A	0.73	0.46
t22	A	0.60	0.51
t23	A	0.87	0.35
t24	A	0.40	0.51
t25	A	0.93	0.26
t26	A	0.60	0.51
t40	A	0.53	0.52

Table 4.5: Average scores for the participants in the treatment group. Test A or B indicate which test the participant got as pre-test.

Case	Test	N	Mean	Std. dev.
Pre control	A	7	0.60	0.19
	B	7	0.47	0.20
Post control	A	7	0.61	0.26
	B	7	0.40	0.14
Pre treatment	A	7	0.50	0.18
	B	7	0.45	0.13
Post treatment	A	7	0.54	0.21
	B	7	0.67	0.19

Table 4.6: Average scores of test A and B for the four cases.

Case	t	sig.
Pre control	1.230	0.242
Post control	1.849	0.089
Pre treatment	0.662	0.521
Post treatment	-1.180	0.261

Table 4.7: T-test results for difference between test A and B for the four cases.

Case	Test	N	Mean	Std. dev.
Pre control	DEF	9	0.57	0.15
	FGW	6	0.48	0.18
Post control	DEF	9	0.61	0.11
	FGW	6	0.35	0.13
Pre treatment	DEF	9	0.50	0.19
	FGW	6	0.44	0.20
Post treatment	DEF	9	0.63	0.19
	FGW	6	0.57	0.12

Table 4.8: Average scores of on the DEF and FGW domains for the four cases.

Case	t	sig.
Pre control	1.113	0.286
Post control	4.271	0.001
Pre treatment	0.564	0.583
Post treatment	0.650	0.527

Table 4.9: T-test results for difference between the DEF and FGW domains for the four cases.

Set	N	Mean	Std. dev.
1	70	0.93	0.26
2	70	0.89	0.32
3	56	0.98	0.13
4	70	0.93	0.26
5	56	0.89	0.31
6	83	0.94	0.24
7	78	0.85	0.36
8	52	0.52	0.50
9	120	0.94	0.24
10	48	0.60	0.49
11	47	0.83	0.38
12	39	0.74	0.44

Table 4.10: Average scores of the question sets in the treatment. N denotes the total number of questions answered by participants for the question set.

ID	N	Mean	Std. dev.
t1	58	0.91	0.28
t2	56	0.96	0.19
t3	39	0.90	0.31
t4	61	0.74	0.44
t5	61	0.80	0.40
t6	61	0.85	0.36
t7	61	0.84	0.37
t21	61	0.93	0.25
t22	61	0.89	0.32
t23	61	0.75	0.43
t24	61	0.87	0.34
t25	61	0.85	0.36
t26	59	0.88	0.33
t40	28	0.89	0.31

Table 4.11: Average individual scores on the treatment questions. N denotes the number of questions answered by the participant.

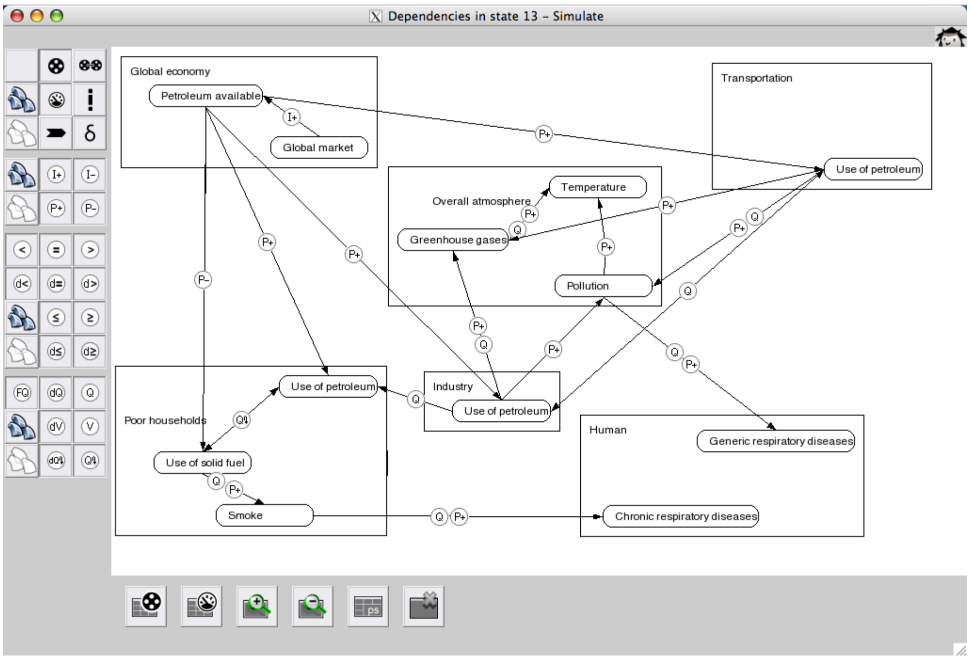

Figure 4.5: Screenshot of Garp3 dependencies of FGW model.

	Mean	Std. def.
Computer experience	5.36	1.15
Ecology expertise	2.64	1.08
Conceptual Modelling expertise	3.93	0.92
Qualitative Reasoning expertise	4.43	1.02

Table 4.12: Average scores of the background questionnaire, N=14. The scores range from none to very high on a scale of 1 to 7.

4.7.5 Results of the background and attitude questionnaires

The questions of the background and attitude questionnaires (Appendix C) were answered on a scale form 1 to 7, 1 being the lowest score and 7 the highest. The participants were all first year Artificial Intelligence students, 12 male and 2 female, with an average age of 19. From the average scores (table 4.12) it appears that the participants were experienced with computers, averagely experienced with qualitative reasoning and conceptual modelling and not very experienced on ecology. This is what is to be expected of a first year Artificial Intelligence student, except for the score on qualitative reasoning. The participants had no instruction on QR yet and it is very unlikely that they had gained experience in QR in another way.

The ease of use and the understandability of the diagrams scored high on the attitude questionnaire (table 4.13), the rest scored approximately average.

	Mean	Std. def.
How much was learned	4.00	1.24
Difficulty of the test questions (1=difficult,7=easy)	4.36	1.60
Ease of use of the software interface	5.07	1.07
How easy did were the diagrams to understand	5.29	1.07
How enjoyable was the session	3.71	1.27
How enjoyable was it using the software	3.93	1.21

Table 4.13: Average scores of the attitude questionnaire, N=14. Unless noted otherwise, scores range from low to high on a 1 to 7 scale.

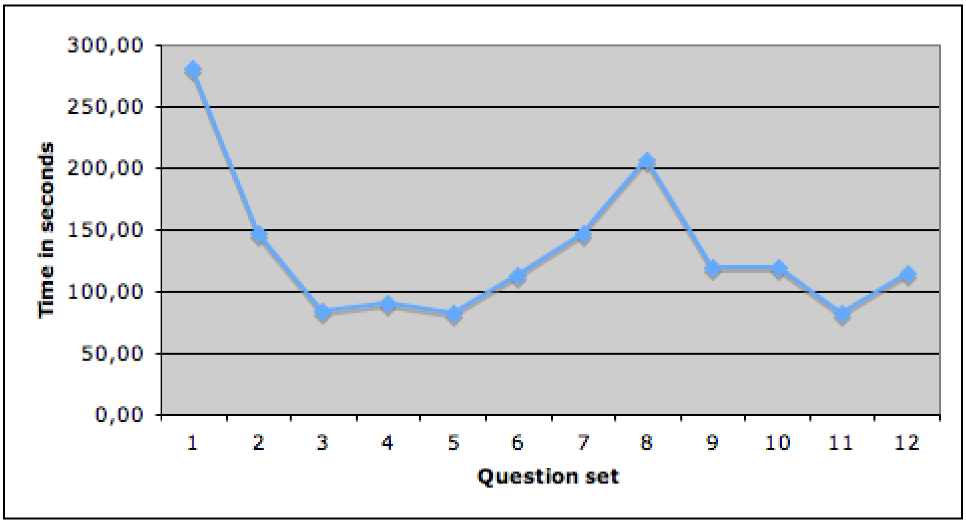

Figure 4.6: Average time spent on the treatment question sets.

4.8 Conclusion

The results of the experiment support the hypothesis that people can learn conceptual knowledge through observing and inspecting qualitative simulations, even within very limited time. There was no significant difference between the pre- and post-test in the control group, thus it is safe to conclude that any difference observed in the treatment group is a direct effect of the treatment itself. Almost all the participants in the treatment group scored higher on the post-test than on the pre-test. Further supporting the conclusion is the observation that the two participants (t3 and t40), who only got about halve way through the treatment, are also two of the three participants that scored lower on the post-test than on the pre-test. Since there are no significant differences between test A and B in any of the four cases (table 4.7), the tests can be considered comparable in difficulty.

According to learning theory the treatment participants should learn more about the domain if the model is worked through progressively, as with the DEF domain, than if the model is presented entirely at once, as with the FGW domain. Overall the scores of the questions on the FGW domain were lower than those on the DEF domain, but the only significant difference between those scores was in the post-test in the control group. In fact, table 4.3 shows that for the treatment group there is no significant learning effect on the FGW domain (t=-1.389, sig=0.188), only on the DEF domain (t=-2.412, sig=0.031). The treatment question scores show a similar result, significant lower scores on the questions on the FGW domain than those on the DEF domain. This all supports the theory that model progression improves learning, if it can be shown that the FGW model is not more complex than the DEF model and the test questions on FGW are not more difficult than those on DEF.

How to view these results? First the test questions, in table 4.1 shows that in both tests the questions on FGW are less difficult than those on DEF (shorter average causal path), for both forward and backward reasoning. Figures 4.7 and 4.8 show the model dependencies for the complete DEF domain and FGW domain respectively, in these figures the entities and the causal relations between the entities are shown. Both models contain 12 entities, the DEF model contains 16 relations and the FGW model contains 14 relations. The longest causal path in the DEF model has a length of 5, the longest path in the FGW model has a length of 4. From this data it is clear that the FGW model is not more complex than the DEF model. It can be concluded that presenting a model progressively improves the learning effect, in fact in this case it was necessary to achieve a significant learning effect.

The time spike for question set 8 of the treatment can be explained by the nature of the questions of set 8, especially the second question. While most questions relationsred the participants to inspect some values and relations, the second question of set 8 asked the participants to compare the behaviour of multiple quantities. This most likely took more time then the other questions, because the participants had to inspect and compare multiple value histories.

The participants of the treatment indicated that they had an average experience with QR. This is

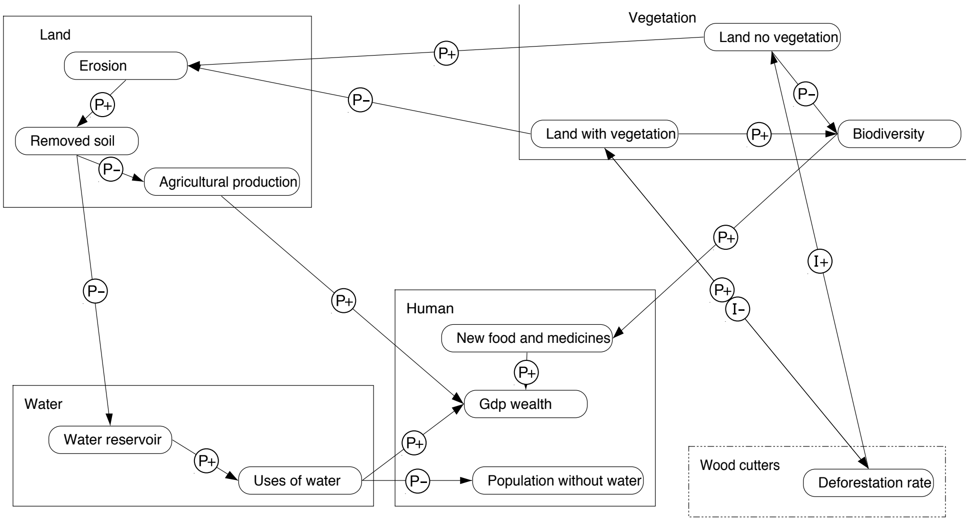

Figure 4.7: Model dependencies for the complete DEF model.

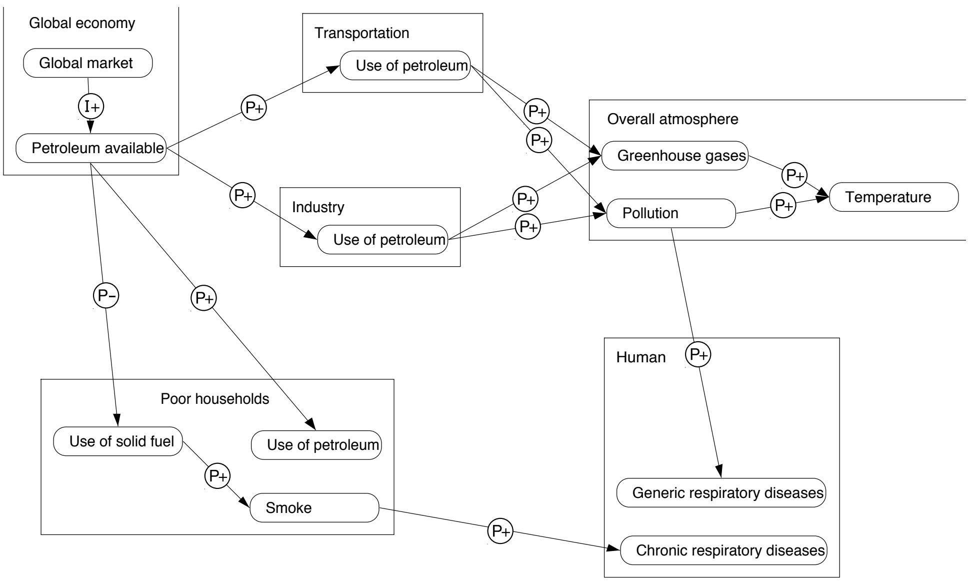

Figure 4.8: Model dependencies for the complete FGW model.

very unlikely, since they had had no prior instruction on QR. We assume that the participants did not understand what QR was or misunderstood the question.

In this experiment the treatment was strictly guided by assignments and the participants got immediate feedback on the assignments. Further research could be done on the effect of guidance and feedback. This can be done by comparing guidance with feedback, guidance without feedback and no guidance groups. The expected outcome of this experiment would be that guidance with feedback will give the best results in terms of knowledge transfer. However it would be very interesting to see how much better the results are compared to the other conditions, is guidance and feedback mandatory for learning or can learning occur without?

Chapter 5

Evaluation of knowledge articulation by designing and constructing

5.1 Introduction

Building a qualitative model is a complex task. To support this a structured approach to building qualitative models was developed [Bredeweg et al., 2006b], [Bredeweg et al., 2007]. This approach is supported by the Garp3 workbench in the form of the sketch environment. The structured approach and the sketch environment have been discussed in section 3.4. To examine how this approach supports the model building effort an exploratory case study has been carried out which is described in this section. This study focuses on actual usage of the structured approach to qualitative modelling using the sketch environment in an educational setting.

5.2 Problem statement

A structured approach has been developed and a tool to support this approach to build models. A case study was carried out where the developers of the structured approach applied the approach successfully in a collaborative modelling effort. Moreover a sketch environment was developed to support the structured approach in a formalized manner. The goal was (1) to determine how working with the sketch environment helped the model building process, (2) to determine how working with the sketch environment supported learning, (3) to map the issues which were encountered by the users of the sketch environment and (4) to create recommendations for guidance of structured approach in an educational setting.

5.3 Method

A case study was chosen as a means to gain deeper insights in applying the structured approach using Garp3. A qualitative in-depth exploratory case study was conducted in an educational setting. Three participants were followed in their model-building efforts following the structured approach. Observations of the model building process, weekly open-ended questionnaires, and analysis of intermediate model representations were combined in this study to gain insight into the process of working with Garp3 and the sketch environment. The participants gave informed consent to being part of this study.

Domain Each participant was presented with one of three modelling problems. These three problems were related to sustainable development. Each of the problems was a target in the Millennium Development Goals (MDG) project of the United Nations[4]. The MDG project is a combined effort to improve the quality of life in third world countries by the year 2015. The MDG project contains 8 goals to be achieved by 2015, each goal has a number of targets with corresponding indicators. The modelling problems for this course were each based on one of the three targets for goal 7: Ensure Environmental Sustainability (see table 5.1). The participants had to build models that would give better insight into how the targets could be achieved. This was done by making the structure and causal relations of the involved systems explicit in a qualitative model.

Users The users were MSc Artificial Intelligence participants at the University of Amsterdam (UvA). They had some experience with model building using Garp3 and QR theory. The users had computer experience, but little prior domain knowledge.

[4]http://www.unmillenniumproject.org

Target 9	Integrate the principles of sustainable development into country policies and programs and reverse the loss of environmental resources.
Indicators	25. Proportion of land area covered by forest (FAO) 26. Ratio of area protected to maintain biological diversity to surface area (UNEP-WCMC) 27. Energy use (kg oil equivalent) per $1 GDP (PPP) (IEA, World Bank) 28. Carbon dioxide emissions per capita (UNFCCC, UNSD) and consumption of ozone-depleting CFCs (ODP tons) (UNEP-Ozone Secretariat) 29. Proportion of population using solid fuels (WHO)
Target 10	Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation.
Indicators	30. Proportion of population with sustainable access to an improved water source, urban and rural (UNICEF-WHO) 31. Proportion of population with access to improved sanitation, urban and rural (UNICEF-WHO)
Target 11	Have achieved by 2020 a significant improvement in the lives of at least 100 million slum dwellers.
Indicators	32. Proportion of households with access to secure tenure (UN-HABITAT)

Table 5.1: Millennium Development Goal 7: Ensure Environmental Sustainability

Environment This case study involved three MSc Artificial Intelligence students at the University of Amsterdam (UvA) enrolled in the course 'Qualitative Reasoning'. The course lasted 16 weeks and was divided in two parts. The first part contained theory and practical exercises on QR and Garp3, in the second part the students had to do a practical project related to Garp3. Three of the five students chose to build a qualitative model as their project. These three students were chosen as the participants of this study. In the 8 weeks of the model building effort the students were supervised by three experts. The students presented and discussed their progress at weekly meetings. During the meetings the experts commented on the models and also provided explanations on the domain and modelling techniques. The students were guided in their modelling effort by weekly assignments, corresponding to the steps of the structured approach as described in section 3.4. The users worked individually on their models. There was no predetermined work environment, the users could work at home or at the computer facilities at the UvA.

Tasks The users had to use the sketch environment to design their model and implement a functioning model with the build and simulate environment.

Equipment The users used Garp3 and worked on their own computers. For this task the users were given literature on the domain [Lee and Ghanime, 2004], [United Nations Development Group, 2001], [Smeets and Weterings, 1999] and modelling the domain [Salles et al., 2005]. The users were also given an article on the structured approach [Bredeweg et al., 2007].

Goals The task goals of this usage were to create a working qualitative model and fully document it. The knowledge transfer goals were to learn more about QR and about the domain.

5.3.1 Observation of modelling process

The weekly meetings were recorded on video for further analysis. The videos were analyzed on for a temporal division. The temporal division is an analysis of proceedings of the meetings. The meetings are divided temporally according to the following categories:

Name
Date
Representation
What did you do?
What were the problems you encountered?
What is the status of the representation?
Remarks
Are you satisfied with the representation?

Table 5.2: Open-ended questionnaire for weekly report by participants.

- Participant: when a Participant is presenting his model, this is usually only in the beginning of their allotted time.
- Comment: most likely given by one of the teachers, it is a comment on what the participant has created, this can lead to further discussion or explanation.
- Discusion: When an issue arises and it is not clear how this should be resolved or someone is not convinced a certain approach is correct it can be discussed.
- Explanation: When more explanation is required, one of the teachers could elaborate on certain subjects. During the explanation the teacher could ask questions to the participant for more interaction.

All meetings were examined extensively and were categorised the category systems described above.

5.3.2 Open-ended questionnaire

The participants were asked to answer a few open-ended questions each week (table 5.2), on what they did and what issues they encountered for the purpose of this study.

5.3.3 Intermediate model representations

To analyse the structured approach the model progress was mapped for the three participants. The structured approach as described in section 3.4 contains a number of assumptions with regards to the re-use, refinements and formalisation of ingredients (see figure 3.4). According to the structured approach the elements should be re-used and refined in following steps rather than creating new elements in each step. The structured approach is a process of refinement and iteration. For analysis all the intermediate representations were examined on content for each week. The representations were compared to the representations of the previous week and other current representations. This analysis focussed on where the elements came from, where they re-used or refined from another representation or were they new. For the implementation of the model only the final model was examined for this analysis. This analysis should give an insight in the model building process and the structured approach.

To analyse the formalisation of the intermediate representation into the actual model the final intermediate representations were compared to the final model. Each of the formalisation steps (see figure 3.4 'formalises into' arcs) were measured for each participant and for all the participants combined. The ratio between the re-used, refined and new elements in the final model is used to compare the different approaches and the different formalisation steps. The goal of this analysis is to see if there is a difference in ratios between the three participants and between formalisation steps. If a certain formalisation step has a lower ratio than other steps it could mean that the participants did not create a correct intermediate representation, did not formalise correctly or that the sketch environment does not support the formalisation correctly.

5.4 The modelling effort

Each week the participants got assignments on what to do. These assignments corresponded with the structured approach. The next sections describe what the participants had to do each week. The participants had to present and discuss their progress each week at the meetings. At the end of each meeting the participants got their next assignments.

Week 1 During the first meeting three modelling problems were presented to the participants, who could choose which one they would like to model. All three problems were targets of the sustainable development project. After choosing which target they had to model, the participants had the rest of the week to study the target documentation, fill in the 'Abstract, model goals, intended audience and general remarks' and create the concept map. This correspondents with the first step of the structured approach: orientation & initial specifications.

Week 2 In the second week the participants had to create the structural model, the processes and the actions & external influences. This corresponds with the second step and part of the third step of the structured approach: System selection & structural model and Global behaviour.

Week 3 In week 4 the participants had to create the causal model and the behaviour graph. This is the rest of the third step of the structured approach: Global behaviour.

Week 4-6 The participants had three weeks to build the actual model. This corresponds with step 5 of the structured approach: Implementation.

Week 7-8 The last two weeks were for writing the report and model documentation. this corresponds with step 6 of the structured approach: Model documentation. The final meeting took place in week 7, in this meeting the participants had to give their final presentation.

5.5 Results

The three participants are labelled P1, P2 and P3 throughout the analysis.

5.5.1 Intermediate representations

The intermediate representations made each week were examined for each participant. To see how the iterative process of the structured approach worked an overview was created for each participant. The results of this analysis can be seen in figures 5.1, 5.2 and 5.3.

The blocks in these figures represent the different sketch representations (e.g. concept map, structural model) and build representation (e.g. quantities and model fragments). In each block the different elements for each representation are denoted and the amount of elements present in that representation (e.g. 28 concepts and 9 relations in concept map in figure 5.1). The vertical lines separate each iteration and the dashed vertical line separates the sketch representation from the build representation. P2 and P3 worked on their sketch representation one week more than P1 and therefore have one more iteration. The lines between elements denote re-use or refinement of sketch elements and formalisation of sketch into build elements. The number above the line represent the number of re-used elements, the number below the line represent the number of refined elements (e.g. 5 concepts are re-used and 3 concepts are refined as entities in the structural model in figure 5.1). For the model fragments block, the instantiations denote the amount of model fragments that are specific variants of another model fragment.

In table 5.3 the overview figures 5.1, 5.2 and 5.3 are summarized. The total amount of elements denoted in the figures are looked at and split by sketch and build model. The total number of re-uses and refinements are counted and the percentile of the total amount of elements calculated.

Figure 5.1 shows less relations than figures 5.2 and 5.3, this indicates less refinement and re-use steps. Table 5.3 shows that P1 created relatively the most new elements during the modelling effort; 58% of all created elements were new, as opposed to 24% and 43% by P2 and P3 respectively. All the participants' final models were of comparable size in terms of number of elements.

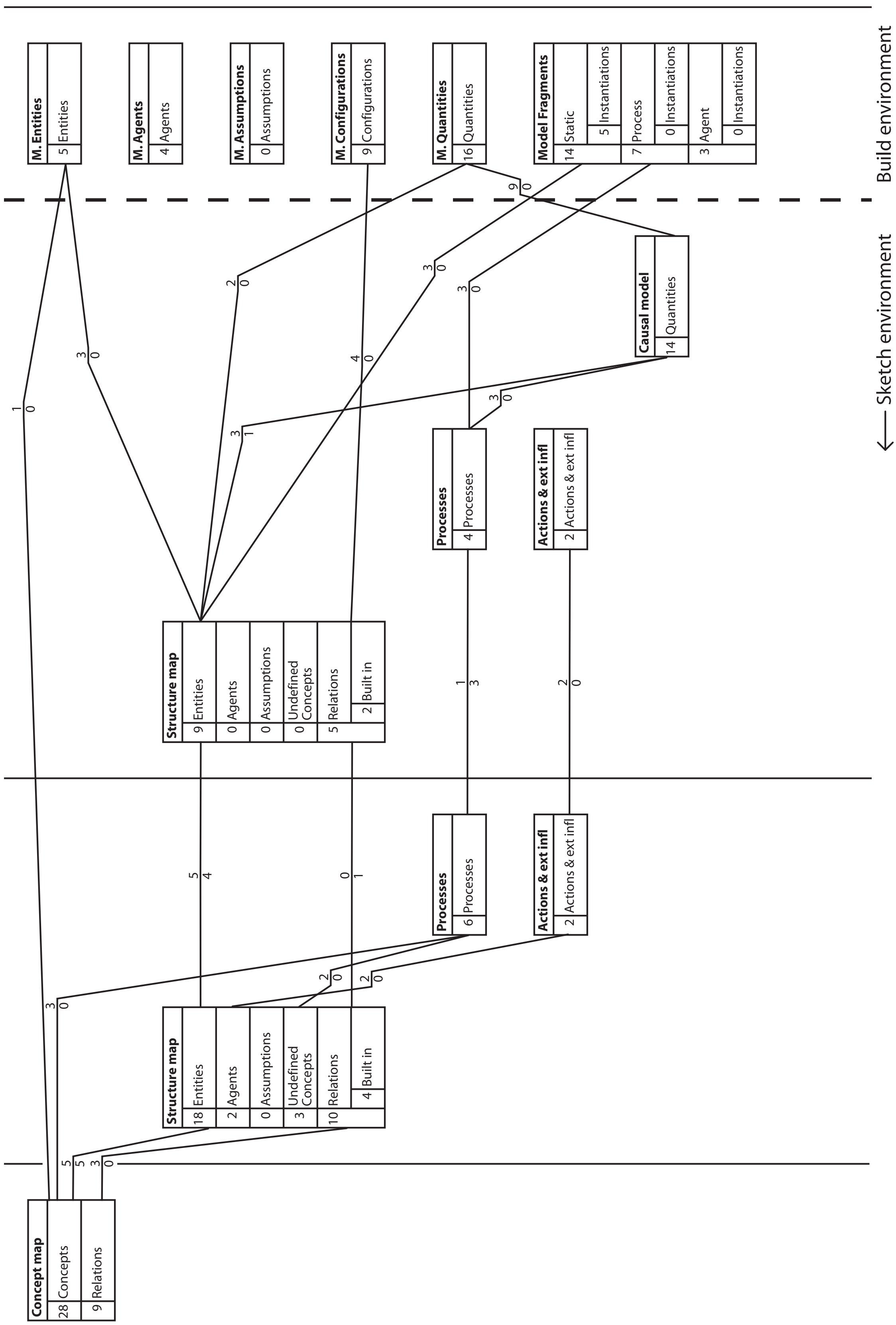

Figure 5.1: Overview of structured approach for P1.

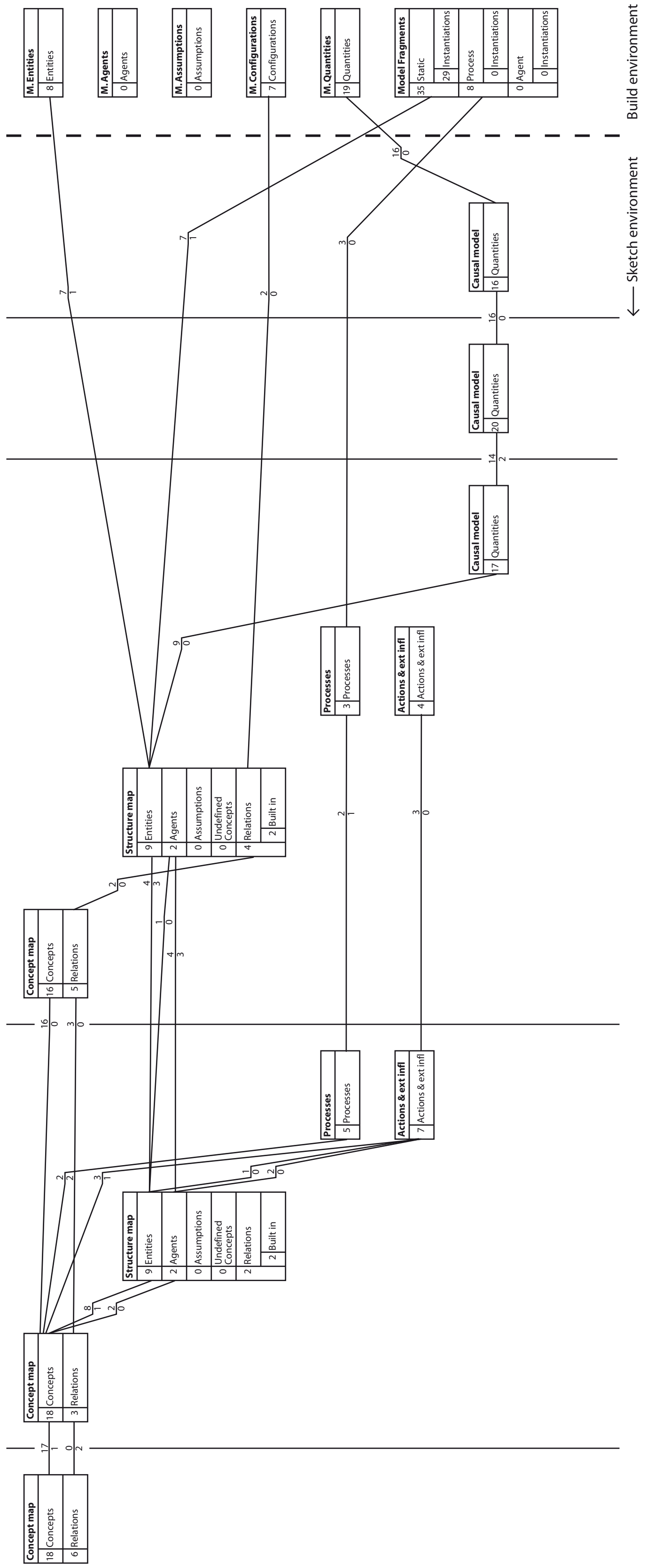

Figure 5.2: Overview of structured approach for P2.

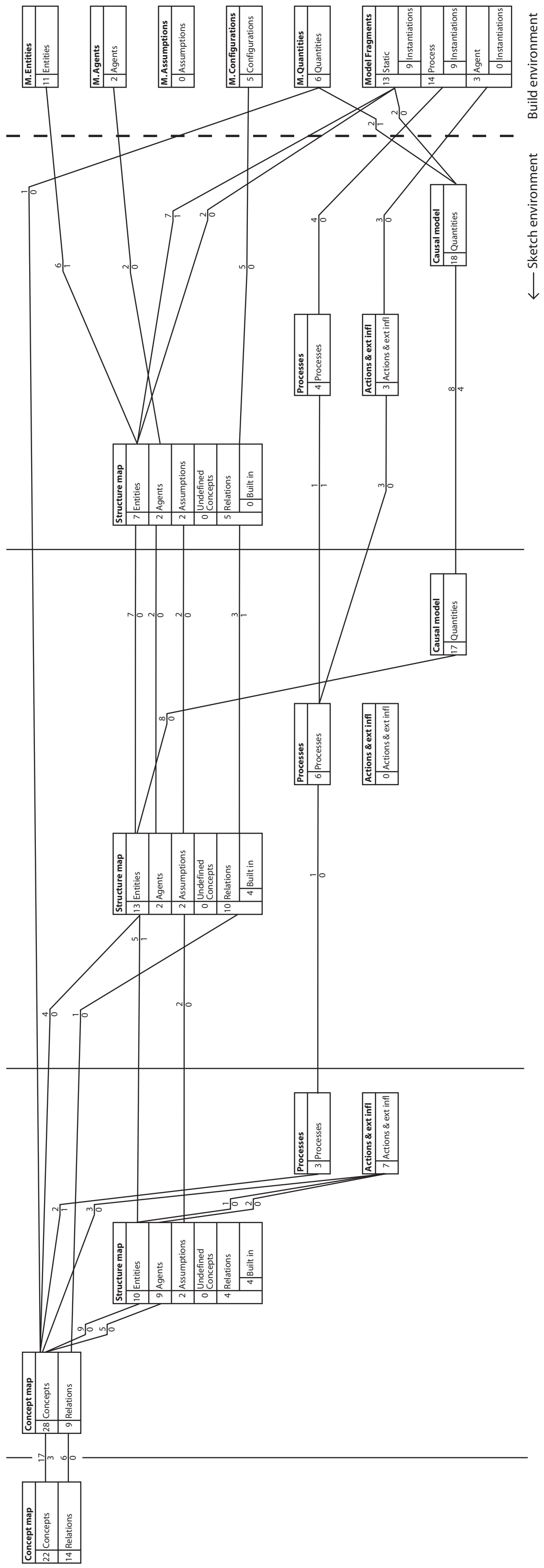

Figure 5.3: Overview of structured approach for P3.

	P1	P2	P3
Total nr of elements	164	214	245
Nr of sketch elements	106	162	191
Nr of model elements	58	52	54
Nr of re-uses	54	144	126
Percentage of re-uses	33%	67%	51%
Nr of refinements	14	18	13
Percentage of refinements	9%	8%	5%
Nr of original elements	96	52	106
Percentage of original elements	58%	24%	43%

Table 5.3: Overview of elements, re-use and refinements in overview figures. Percentage relate to total number of elements.

Explanation subject	**Total time (minutes)**
Behaviour graph	00:30
Causal model	02:30
Concept map	03:00
Course	01:30
Domain	11:00
Model Fragments	03:00
Modelling	18:30
Processes	24:00
Structural model	10:00
Total	74:00

Table 5.4: Total time and subjects for explanations during weekly meetings.

5.5.2 Temporal division

The videos of the meetings were analysed on a temporal basis. In total there where seven meetings, the first meeting was used for introductions and the final meeting was used for the final presentations. Five meetings were used for presentations and discussion. These five meetings were used for this temporal analysis. Each meeting lasted for approximately 105 minutes, the total time of the meetings was 550 minutes. In the first two meetings other students had a half of the meeting to present their work, so the total time available in the meetings for the participants was approximately 450 minutes. In table 5.4 and overview is presented of the time spent explaining different subjects. Table 5.5 shows the times the participants were explaining their progress. Table 5.6 shows the times of the comments during the meetings. Table 5.7 shows the times of discussions during the meetings. All the temporal divisions are divided by subject. The most time was spent on the concept map, the causal model, the processes and modelling. The total time from the four categories is 427 minutes, some time was lost due to set-up of equipment and presentations, so this corresponds to the time available during the meetings.

5.5.3 Formalisation

The formalisation process is the process from the intermediate representations to a final model. To analyse this the final intermediate (sketch) representations were compared to the final model representations. Each of the formalisations from figure 3.4 was measured quantitatively. For each formalisation the amount of elements re-used, refined and new were measured. Figures 5.4, 5.5 and 5.6 contain the ratios between re-used, refined and new elements for the three participants. Figure 5.7 shows these ratios for the three participants combined. P1 shows the lowest ratios and P3 the highest. Overall it stands out that the entities and quantities have a high re-use rate and the model fragments and behaviour have a low re-use rate. Configurations have a medium re-use rate and agents, attributes, assumptions and scenarios have too few occurrences to give a clear view.

Subject	P1	P2	P3	Total
AMgIaGr	00:01:30	00:01:00	00:02:00	0:04:30
Behaviour graph	00:02:30			0:02:30
Causal model	00:09:00	00:00:30	00:01:30	0:11:00
Concept map	00:09:30	00:15:00	00:05:30	0:30:00
Entities			00:00:30	0:00:30
Model Fragments	00:05:00	00:10:00	00:02:00	0:17:00
Processes	00:07:00	00:02:00		0:09:00
Scenario		00:01:30		0:01:30
Scope		00:01:30		0:01:30
Simulation	00:13:00	00:01:00	00:08:00	0:22:00
Structural model	00:04:00	00:03:30	00:01:30	0:09:00
Total	0:51:30	0:36:00	0:21:00	1:48:30

Table 5.5: Overview of participants speaking during meetings by subject for all three participants. AMgIaGr stands for Abstract, Model goals, Intended audience and General remarks.

Subject	P1	P2	P3	Total
A&EI		00:04:30		0:04:30
Behaviour graph			00:07:00	0:07:00
Causal model	00:11:00	00:06:30	00:07:30	0:25:00
Concept map	00:04:30	00:05:30	00:07:00	0:17:00
Model Fragments		00:07:30		0:07:30
Model help	00:32:00	00:21:00	00:27:30	1:00:30
Processes		00:03:30		0:03:30
Structural model	0:06:30	00:11:30	00:03:30	0:21:30
Total	0:54:00	0:60:00	0:52:30	2:46:30

Table 5.6: Overview of comments during weekly meetings by subject for all three participants.

Subject	P1	P2	P3	Total
Approach		00:01:00		0:01:00
Causal model	00:04:00	00:17:00	00:03:00	0:24:00
Concept map		00:02:30	00:05:30	0:08:00
Domain		00:07:30	00:02:00	0:09:30
Modelling	00:04:00	00:14:30	00:07:00	0:25:30
Processes	00:04:00	00:02:00		0:06:00
Scope	00:04:30			0:04:30
Simulation			00:03:00	0:03:00
Structural model	00:05:30		00:01:30	0:07:00
Total	0:22:00	0:44:30	0:22:00	1:18:30

Table 5.7: Overview of discussions during weekly meetings by subject for all three participants.

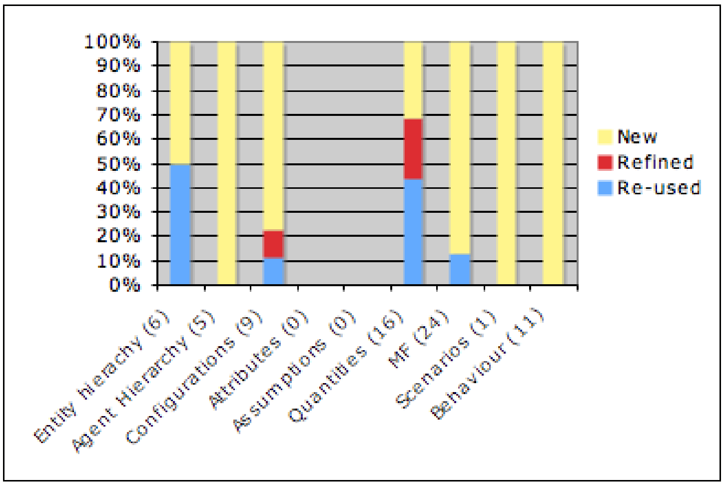

Figure 5.4: Overview of formalisation process for P1.

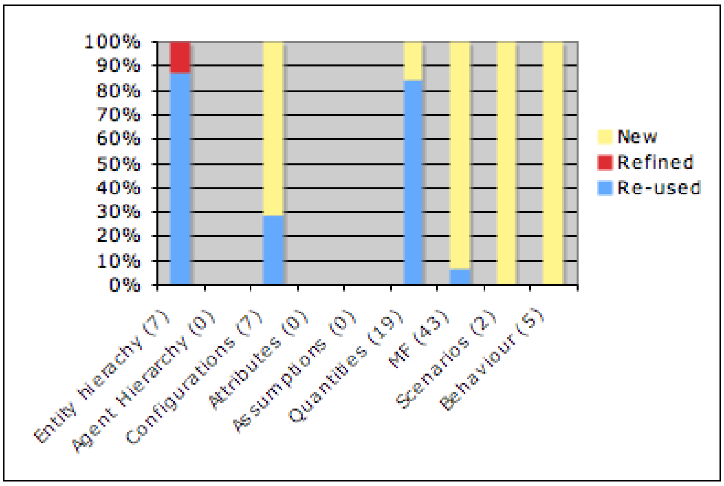

Figure 5.5: Overview of formalisation process for P2.

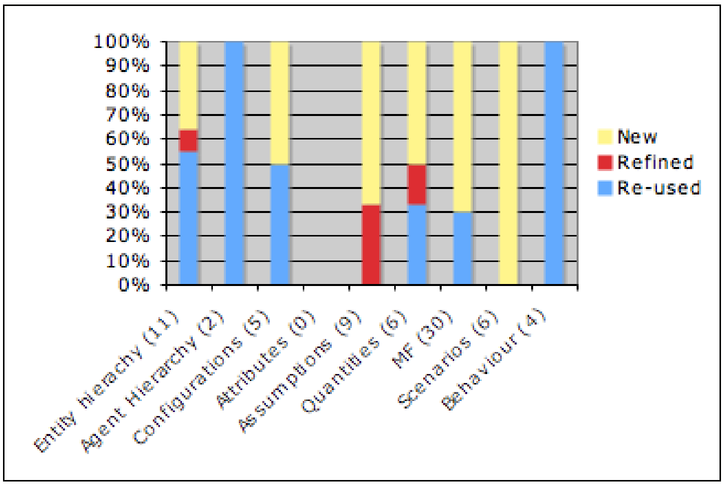

Figure 5.6: Overview of formalisation process for P3.

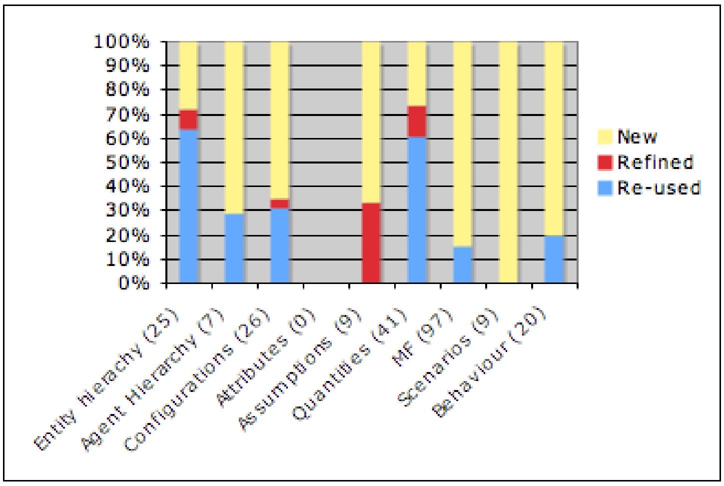

Figure 5.7: Overview of formalisation process for all participants combined.

5.6 Conclusion

Concept map The concept map did not prove to be a problem. The three participants did create different concept maps.

Structure model The structure model provided some problems. Mainly because the participants did not clearly understand what needed to be represented in the structure model and what it's use is. Table 5.4 shows that the structure model needed the second most explanation of the sketch tools during the meetings. Figures 5.1, 5.2 and 5.3 show that the structure model was refined and adapted the most together with the processes and actions & external influences. This all suggests that the structural model may need more explanation than the other models in the sketch environment.

Processes The processes proved to be the most problematic of the sketch tools. It needed the most explanations during the meetings, all participants had to refine and adapt their processes.

Actions & external influences Since actions & external influences have a lot in common with the processes (see section 3.4), the explanations on processes also applied on the actions & external influences. There were no further notable observations.

Causal model During the meetings a lot of time was spend on the causal models, especially comments and discussion. This had to do with the modelling of the processes. The difficulties the participants had with the processes were reflected in the causal map.

Behaviour graph The participants found it cumbersome to create the behaviour graph in the sketch environment. The expected behaviour of P1 and P2 did not correspond to final model behaviour. The behaviour graph of P3 corresponded completely to the final model behaviour.

Modelling During the implementation of the models the participants had a lot of modelling problems. The problems the participants encountered were mostly 'model problems' and 'model errors'.

Overall

This study has shown that is possible to gain a deep conceptual understanding of a domain by using the structured approach to qualitative model building. Even though this understanding has not been tested explicitly, the domain experts agreed that the participants had shown their knowledge of the domain through their work. The participants were graded on their final report and their participation. All three participants passed the course. The next step is to compare this approach to other approaches to learn domain knowledge.

The participants in this experiment showed a different approach to model building, even within the structured approach. From the analysis of the modelling effort it can be concluded that P1 re-used less of the model elements than P2 and P3. P1 also did not refine earlier sketch representations in later steps, whereas P2 and P3 did. This does not mean that P1 had a wrong approach or that the structured approach was not useful. As mentioned in section 3.4 the structured approach serves two purposes, namely to support the model building effort and to externalise intermediate steps. The structured approach supports the model building effort by 'forcing' different viewpoint on the modeller. In each step the modeller has to focus on a particular aspect of the model, the modeller must look at the problem from a certain angle. This helps the modeller in the model building process and through the steps the modelling problem is tackled. This applies to all the participants, by working through the steps different problems could be discussed and explained in a structured fashion.

The second purpose, the externalisation of intermediate representations, is very useful to get insight in the modelling effort. When multiple modellers are collaborating in a model building effort intermediate representations can be very helpful in comparing and discussing the effort. Although the participants did not collaborate on the same model in this study, they did work with overlapping vocabulary. Furthermore the representations can also be very insightful for the experts who guide the modellers. When earlier representations are changed it reflects that the modeller has adapted his view of the model. In an educational situation, like the one in this study, it can help the participants and the experts to communicate and discuss the progress. As P1 did not refine earlier representations to reflect changes in approach, the changes had to be explained verbally or not at all. This caused some confusion during the meetings.

For P1 this meant that the changes made to his approach did not reflect in the representations that were created before the changes. The changes had to be explained verbally or not at all.

In conclusion it can be said that the structured approach and the sketch helped the participants in their modelling effort by providing focus on the important viewpoints. It also helped the communication and discussion of the effort through the intermediate representations. Although it is not essential to re-use and refine a lot of the intermediate elements for the individual, it can hinder communication.

The biggest issues in the modelling effort were the structure model, the processes and the representation thereof in the causal model and the actual creating of the model in the build environment. The issues with the structure model can be attributed to the participants understanding of its use. The participants had some trouble with the concept of the structure model, but when this was explained the issues were resolved. The processes and how to represent them in the causal map also needed some extra explanation before the issues were resolved. The modelling difficulties were mostly of a technical nature, where the model did not behave the way it was expected. These issues were resolved with help from the experts. Some model difficulties can be attributed to the lack of support the sketch environment provides to step 4 of the structured approach. In this step all the model ingredients should be created without using the actual model building tool. This step should provide a clear 'blueprint' of all the ingredients and their role. This is especially useful for the model fragments, the ingredients that contain partial structural and behaviour patterns of the system. Due to this lack the transition from sketch environment to build environment may have provided some problems for the model fragments. Figure 5.7 indicates that the formalisation of model fragments was far less than the formalisation of other ingredients that were used in large numbers.

The issues encountered during the modelling effort indicate problems the participants had. These were subjects the participants did not fully understand beforehand. Through exercise and explanation these the participants had learned and the issues were resolved. The structured approach provided focus on different aspects of the modelling effort, thus allowing the participants to learn about the different aspects of qualitative modelling. In this way the structured approach supports learning about modelling.

In similar usages as this it can be useful to explain the structure model and the processes beforehand to avoid the issues encountered in this study. Furthermore the addition of support to step 4 of the structured model could improve the sketch environment in its support to model building, but this should be tested in further research. In this experiment only three participants were observed, the following step is to expand this experiment. With more participants quantitative data can be collected on the support structured approach and the sketch environment can give to knowledge transfer. It is also interesting to compare participants working with and without the structured approach. This would shed more light on the actual impact on working with the structured approach.

Chapter 6

Conclusion and discussion

This thesis discussed how Qualitative Reasoning (QR) can support the transfer of conceptual knowledge and how that support can be evaluated. The support QR can give depends on how it is used, this is called the usage. Each usage of QR can possibly support the transfer of conceptual knowledge. In order to determine how QR can support knowledge transfer, the usages must be tested. To structure the process of evaluation a framework was constructed. With this framework the possible support for knowledge transfer of a QR program can be explored systematically. The framework also provides dimensions to define usages, facilitating the comparison and evaluation of usages.

The transfer of conceptual knowledge consists of acquisition and articulation of knowledge. The acquisition of knowledge, or learning, can be supported in four ways. These ways are analogous to empirical learning, with QR representing the physical world. Knowledge can be acquired by (1) controlling and operating qualitative simulations, by (2) designing and constructing qualitative models, by (3) diagnosing and repairing qualitative models and by (4) observing and inspecting qualitative models and simulations. The articulation of knowledge can be supported by (5) designing and constructing qualitative models. These five support types can be tested for QR software, however each type can contain multiple usages. To test these, the usages for these types must be tested. The usages can be defined using the usage dimensions (domain, users, environment, tasks, equipment and goals). With these dimensions the scope of the usages can be determined. These usages can then be tested to evaluate the QR software.

The framework has been applied in the context of Garp3, a software program for building and simulating qualitative models with a graphical user interface. By applying the framework possible usages of Grap3 were identified, two of these usages were evaluated. The first evaluation study focussed on learning through observing and inspecting. The second study looked articulating by designing and constructing and learning through designing and constructing of qualitative models.

The first study was a between-subject experiment. The participants did a treatment with the simulate environment of Garp3. The participants were guided through a number of assignments on the computer, which required them to use Garp3 and inspect simulation results. A control group was also tested, the control participants did not do the treatment. To test the participants, they had to take a pre- and post-test. Both test measured the participants' ability to do pre- and post-diction on the domain, which is a good measure for conceptual knowledge. The study showed a significant learning effect on conceptual knowledge in the treatment group. The study also showed that novice users were able to use Garp3 with minimal instructions.

The second study was an exploratory case study. Three participants were observed in their modelling effort. The participants followed the structured approach to model building and used the sketch environment of Garp3, as well as the build and simulate environment. The participants were able to build a functioning model using the structured approach. Even though the participants had their own working style, the structured approach brought focus to different aspects of the modelling effort. This focus facilitated discussion and problem solving during the modelling effort. This study also showed that the participants acquired knowledge on the domain on a conceptual level by building the model.

These two studies were conducted as part of the evaluation of Garp3 following the framework proposed in this thesis. For these two studies the framework has been successful, however more experiments to evaluate Garp3 should be done to determine the usefulness of the framework to its fullest extend. The studies performed in the context of this thesis addressed three of the five support types from the framework and looked at two usages by making choices for each of the usage dimensions. Further studies could address all support types and look at different usages. One step further would be to look at real world usages and applications like in the Naturenet-Redime project. For example the evaluation of Garp3 in a real curriculum, professional collaborative model building or an interactive learning application based on Garp3.

To further explore the support QR can provide for knowledge transfer other QR software should be evaluated as well. Different software can be compared, as a package or on individual usages. Not only can QR software be compared with each other, they can also be compared with other software and methods for knowledge transfer. QR software can be compared to quantitative simulations in educational context to see where the differences lie. QR software can also be compared to more traditional existing learning methods. This will lead to a better understanding of how transfer of conceptual knowledge can be supported.

Other areas that can be researched for QR are how well it is suited for different learning styles. As mentioned in section 2.3.3 it is important to look at the learning style of a usage. However research could be done to compare the different learning styles and methods with QR as a support tool, e.g. discovery based learning or guided linear learning. The differences in impact on knowledge transfer between styles can be used to create better interactive learning environments based on QR. In the future principles from HCI research and education research should be integrated in development and research of usable and useful tools to support the transfer of conceptual knowledge.

References

[Bednarik et al., 2004] Bednarik, R.; Gerdt, P.; Miraftabi, R.; and Tukiainen, M. (2004). "Development of the TUP Model - Evaluating Educational Software". In *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT04).*

[Beyer and Holtzblatt, 1999] Beyer, H. and Holtzblatt, K. (1999). "Contextual design". *Interactions*, 6(1), pp. 32–42.

[Boehm, 1986] Boehm, B. (1986). "A spiral model of software development and enhancement". *SIGSOFT Softw. Eng. Notes*, 11(4), pp. 14–24.

[Bouwer et al., 2002] Bouwer, A.; Bessa Machado, V.; and Bredeweg, B. (2002). "Interactive Model Building Environments". *The Role of Communication in Learning to Model*, pages 155–182.

[Bouwer and Bredeweg, 2001] Bouwer, A. and Bredeweg, B. (2001). *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future*, chapter VisiGarp: Graphical Representation of Qualitative Simulation Models, pages 294–305. IOS-Press/Ohmsha, Japan, Osaka.

[Bredeweg, 1992] Bredeweg, B. (1992). *Expertise in Qualitative Prediction of Behaviour.* PhD thesis, University of Amsterdam.

[Bredeweg et al., 2006a] Bredeweg, B.; Bouwer, A.; Jellema, J.; Bertels, D.; Linnebank, F.; and Liem, J. (2006a). "Garp3 A new Workbench for Qualitative Reasoning and Modelling". In *Proceedings of the 20th International Workshop on Qualitative Reasoning (QR'06)*, pages 21–28.

[Bredeweg and Forbus, 2005] Bredeweg, B. and Forbus, K. (2005). "Qualitative Modelling in Education". *AI Magazine*, 26(4), pp. 35–46.

[Bredeweg et al., 2006b] Bredeweg, B.; Salles, P.; Bouwer, A.; and Liem, J. (2006b). "Towards a Structured Approach to Qualitative Modelling". Technical report, University of Amsterdam.

[Bredeweg et al., 2007] Bredeweg, B.; Salles, P.; Bouwer, A.; Liem, J.; Nuttle, T.; Cioaca, E.; Nakova, E.; Noble, R.; Caldas, A. L. R.; Uzunov, Y.; Varadinova, E.; and Zitek, A. (2007). "Towards a structured approach to building qualitative reasoning models and simulations". *Ecological Informatics*, page In print.

[Bredeweg and Winkels, 1998] Bredeweg, B. and Winkels, R. (1998). "Qualitative Models in Interactive Learning Environments: An Introduction". *Interactive Learning Environments*, 5(1-2), pp. 1–18.

[Carney and Wallnau, 1998] Carney, D. and Wallnau, K. (1998). "A basis for evaluation of commercial software". *Information and Software Technology*, 40, pp. 851–860.

[Çepni et al., 2006] Çepni, S.; Taş, E.; and Köse, S. (2006). "The effects of computer-assisted material on students cognitive levels, misconceptions and attitudes towards science". *Computers & Education*, 46(2), pp. 192–205.

[Daradoumis et al., 2006] Daradoumis, T.; Martnez-Mons, A.; and Xhafa, F. (2006). "A layered framework for evaluating on-line collaborative learning interactions". *International Journal of Human-Computer Studies*, 64(7), pp. 622–635.

[de Kleer, 1977] de Kleer, J. (1977). "Multiple Representations of Knowledge in a Mechanics Problem-Solver". In *International Joint Conferences on Artificial Intelligence (IJCAI)*, volume 1.

[Elsom-Cook, 2001] Elsom-Cook, M. (2001). *Principles of Interactive Multimedia.* McGraw-Hill.

[Forbus, 1984] Forbus, K. (1984). "Qualitative process theory". *Artificial Intelligence*, 24, pp. 85–168.

[Forbus, 1996] Forbus, K. D. (1996). "Qualitative reasoning". In *The Computer Science and Engineering Handbook*, chapter Qualitative Reasoning, pages 715–733. CRC.

[Forbus et al., 2005] Forbus, K. D.; Camey, K.; Sherin, B. L.; and Ureel, L. C. (2005). "VModel: A visual qualitative modeling environment for middle-school students". *The AI magazine*, 26(3), pp. 63–72.

[Forbus and Gentner, 1997] Forbus, K. D. and Gentner, D. (1997). "Qualitative Mental Models: Simulations or Memories?". In *Proceedings of the Eleventh International Workshop on Qualitative Reasoning*, Cortona, Italy.

[Freitas and Oliver, 2006] Freitas, S. d. and Oliver, M. (2006). "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?". *Computers & Education*, 46(3), pp. 249–264.

[Greca and Moreira, 2000] Greca, I. M. and Moreira, M. A. (2000). "Mental models, conceptual models, and modelling". *International Journal of Science Education*, 22(1), pp. 1–11.

[Groen, 2003] Groen, R. M. S. (2003). "Supporting model building". Master's thesis, University of Amsterdam (UvA).

[International Standards Organization, 1998] International Standards Organization (1998). "ISO 9241-11:Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability".

[Iwasaki, 1997] Iwasaki, Y. (1997). "Real-world applications of qualitative reasoning". *IEEE Expert: Intelligence Systems*, 12(3), pp. 16–21.

[Jellema, 2000] Jellema, J. (2000). "Ontwerpen voor ondersteuning". Master's thesis, University of Amsterdam (UvA).

[Khalifa et al., 2000] Khalifa, S.; Bloor, C. a.; W.; and Jones, C. (2000). "Educational computer software, technical, criteria, and Quality". In *Proceedings of the Information Systems Education Conference.*

[Kuipers, 1993] Kuipers, B. (1993). "Qualitative Simulation: Then and Now". *Artificial Intelligence*, 59, pp. 133–140.

[Lee and Ghanime, 2004] Lee, L. and Ghanime, L. (2004). "Millenium Development Goal 7 Summary Review: 67 Country MDG Reports". Bureau for Development Policy Energy and Environment Group.

[Leelawong et al., 2001] Leelawong, K.; Wang, Y.; Biswas, G.; Vye, N.; and Bransford, J. (2001). "Qualitative Reasoning techniques to support Learning by Teaching: The Teachable Agents Project". In *Proceedings of the AAAI Qualitative Reasoning Workshop.*

[Linnebank, 2004] Linnebank, F. (2004). "Common Sense Reasoning - Towards Mature Qualitative Reasoning Engines". Master's thesis, University of Amsterdam.

[McCormick, 1997] McCormick, R. (1997). "Conceptual and procedural knowledge". *International Journal of Technology and Design Education*, 7, pp. 141–159.

[Oliver, 2000] Oliver, M. (2000). "An introduction to the Evaluation of Learning Technology". *Educational Technology & Society*, 3(4), pp. 20–30.

[Preece et al., 2002] Preece, J.; Rogers, Y.; and Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction.* John Wiley & Sons.

[Rittle-Johnson and Alibali, 1999] Rittle-Johnson, B. and Alibali, M. W. (1999). "Concpeptual and procedural knowledge of mathematics: does one lead to the other?". *Journal of educational psychology*, 1, pp. 175–189.

[Salles and Bredeweg, 2002] Salles, P. and Bredeweg, B. (2002). "A case study of collaborative modelling: building qualitative models in ecology". In *Proceedings of the International workshop on Model-based Systems and Qualitative Reasoning for Intelligent Tutoring Systems*, pages 75–84.

[Salles and Bredeweg, 2005] Salles, P. and Bredeweg, B. (2005). "Qualitative Reasoning about Population and Community Ecology". *AI Magazine*, 24(4), pp. 77–90.

[Salles et al., 2005] Salles, P.; Bredeweg, B.; and Nuttle, T. (2005). "Qualitative Models of Indicators for Environmental Sustainability Related to the Millennium Development Goals". In *Proceedings of 2nd MONET Workshop on Model-Based Systems (MONET 05) at the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*, pages 66–72.

[Smeets and Weterings, 1999] Smeets, E. and Weterings, R. (1999). "Environmental indicators: Typology and overview". Technical report, TNO Centre for Strategy, Technology and Policy, The Netherlands.

[Squire et al., 2004] Squire, K.; Barnett, M.; Grant, J. M.; and Higginbotham, T. (2004). "Electromagnetism Supercharged! Learning Physics with Digital Simulation Games". In *Proceedings of the 6th international conference on Learning sciences*, pages 513–520.

[Squires and Preece, 1999] Squires, D. and Preece, J. (1999). "Predicting quality in educational software: Evaluating for learning, usability and the synergy between them". *Interacting with Computers*, 11(5), pp. 467–483.

[Struss and Price, 2005] Struss, P. and Price, C. (2005). "Model-based systems in the automotive industry". *The AI magazine*, 24(4), pp. 17–34.

[Tjaris, 2002] Tjaris, P. (2002). "Kwalitatieve simulaties als middel tot kennisoverdracht". Master's thesis, University of Amsterdam.

[United Nations Development Group, 2001] United Nations Development Group (2001). "Reporting on the Millennium Development Goals at the Country Level". Guidance Note.

[Werf, 2003] Werf, R. v. d. (2003). "The Use of Qualitative Modelling and Simulation Tools in High School Education: an Engineering Study". Master's thesis, University of Amsterdam.

Acknowledgements

I would like to thank the following people for their support and encouragement.

Bert Bredeweg, Hannah Kay Piché, Amnon Rafalowicz, Hilde Danner, Niels Monshouwer, Daan Vreeswijk, Aron Abbo and Henriette Cramer.

Appendix A

Tests for evaluation of knowledge acquisition by observing and inspecting

Tests used in as the pre and post tests in the 'Learning through simulation' experiment, the correct answers are italic.

Test A

Sustainable development

1. When deforestation happens,
   a. *land with no vegetation increases.*
   b. water reservoir increases.
   c. production of new food and medicines increases.
   d. GDP (wealth) increases.
2. When the production of smoke in poor households increases,
   a. the production of greenhouse gases increases.
   b. the use of petroleum for industry increases.
   c. the petroleum available in global economy increases.
   d. the occurrence of chronic respiratory diseases increases.
3. When land with vegetation decreases,
   a. production of new food and medicines increases.
   b. erosion decreases.
   c. GDP (wealth) decreases.
   d. water reservoir increases.
4. When biodiversity increases,
   a. GDP (wealth) decreases.
   b. production of new food and medicines increases.
   c. erosion increases.
   d. agricultural production decreases.
5. The occurrence of chronic respiratory diseases increases because
   a. the petroleum available in the global economy increases.
   b. the use of petroleum for transportation increases.
   c. the use of solid fuel in poor households increases.
   d. there is a surplus of petroleum in the global economy.
6. New food and medicines increases because
   a. biodiversity increases.
   b. GDP (wealth) increases.
   c. agricultural production decreases.
   d. deforestation is active.

7. When erosion increases,
   a. uses of water increase.
   b. (GDP) wealth increases.
   c. amount of removed soil decreases.
   d. population without water increases.

8. When the use of petroleum for industry increases,
   a. the overall atmosphere temperature decreases.
   b. the occurrence of generic respiratory diseases increases.
   c. the use of solid fuel in poor households increases.
   d. the use of petroleum for transportation decreases.

9. Agricultural production decreases because
   a. erosion decreases.
   b. land with no vegetation decreases.
   c. deforestation stopped.
   d. removed soil increases.

10. The use of petroleum in poor households decreases because
    a. the overall atmosphere temperature increases.
    b. the use of petroleum for transportation increases.
    c. there is a shortage of petroleum in the global economy.
    d. the occurrence of generic respiratory diseases increases.

11. The water reservoir decreases because
    a. erosion increases.
    b. deforestation stopped.
    c. amount of removed soil decreases.
    d. land with no vegetation decreases.

12. The overall atmosphere temperature increases because
    a. the use of petroleum for industry decreases.
    b. there is a surplus of petroleum in the global market.
    c. the occurrence of chronic respiratory diseases increases.
    d. the use of solid fuel in poor households increases.

13. Population without water increases because
    a. land with no vegetation decreases.
    b. uses of water increase.
    c. erosion increases.
    d. amount of removed soil decreases

14. GDP (wealth) decreases because
    a. uses of water increase.
    b. biodiversity increases.
    c. erosion is decreasing
    d. land with no vegetation increases.

15. The use of solid fuel in poor households decreases because
    a. the petroleum available in the global economy increases.
    b. the occurrence of chronic respiratory diseases increases.
    c. the overall atmosphere temperature decreases.
    d. the use of petroleum for industry decreases.

Test B

Sustainable development

1. When deforestation stops,
   a. agricultural production increases.
   b. biodiversity decreases.
   c. erosion increases.
   d. uses of water decrease.

2. The occurrence of generic respiratory diseases decreases because
   a. the production of greenhouse gases increases.
   b. the overall atmosphere temperature decreases.
   c. the production of smoke in poor households decreases.
   d. the use of petroleum for transportation increases.

3. When land with no vegetation increases,
   a. uses of water increase.
   b. removed soil increases.
   c. new food and medicines decrease.
   d. wealth (GDP) increases.

4. When biodiversity decreases
   a. soil erosion decreases.
   b. new food and medicines decreases.
   c. population without water decreases.
   d. uses of water increase.

5. The occurrence of chronic respiratory diseases decreases because
   a. the petroleum available in the global economy increases.
   b. the overall atmosphere pollution decreases.
   c. the occurrence of generic respiratory diseases decreases.
   d. the use of petroleum for industry increases.

6. When new food and medicines decreases,
   a. biodiversity increases.
   b. GDP (wealth) increases.
   c. deforestation is active.
   d. agricultural production increases.

7. When erosion decreases,
   a. amount of removed soil increases.
   b. uses of water decrease.
   c. water reservoir decreases.
   d. agricultural production increases.

8. The production of smoke in poor households decreases because
   a. the use of petroleum for transportation decreases.
   b. the use of solid fuel in poor households increases.
   c. the occurrence of chronic respiratory diseases increases.
   d. the overall atmosphere temperature decreases.

9. When removed soil decreases,
   a. GDP (wealth) decreases.
   b. agricultural production increases.
   c. uses of water increase.
   d. erosion decreases.

10. When the petroleum available in the global economy increases,
    a. the overall atmosphere temperature increases.
    b. the use of petroleum for transportation decreases.
    c. the use of solid fuel in poor households increases.
    d. the production of smoke in poor households increases.

11. When uses of water decrease,
    a. erosion decreases,
    b. agricultural production increases.
    c. GDP (wealth) decreases.
    d. new food and medicines decreases.

12. When the use of petroleum for industry decreases,
    a. the production of smoke in poor households decreases.
    b. the production of greenhouse gases decreases.
    c. the use of petroleum for transportation increases.
    d. the occurrence of chronic respiratory diseases decreases.

13. When population without water decreases,
    a. agricultural production decreases.
    b. biodiversity decreases.
    c. land with no vegetation decreases.
    d. GDP (wealth) decreases.

14. GDP (wealth) increases because
    a. water reservoir decreases.
    b. uses of water decrease.
    c. new food and medicines decreases.
    d. erosion decreases.

15. When the use of petroleum in poor households increases,
    a. the occurrence of chronic respiratory diseases decreases.
    b. the production of greenhouse gases decreases.
    c. the use of solid fuel in poor households increases.
    d. the use of petroleum for industry decreases.

Appendix B

Treatment for evaluation of knowledge acquisition by observing and inspecting

The questions with the answers of the treatment of the 'learning through simulation' experiment.

1. Impact on Vegetation

1. Which quantity is influenced negatively by deforestation?
   Land with vegetation

2. If land with vegetation decreases, what will happen to biodiversity?
   Decreases

3. What is the value of biodiversity in state 1 and 4?
   Large and zero

4. Which quantity is increasing?
   Land without vegetation

2. Impact on Food and Medicines

1. The indirect effect of deforestation on biodiversity via land with vegetation is negative. Is the effect via land without vegetation also negative, or is it positive?
   Also negative

2. Will the production of new food and medicine increase or decrease because of deforestation?
   Decrease

3. In which state is the value of biodiversity equal to medium?
   State 2

4. What is the value of food and medicine in that state? And deforestation?
   Food and medicine = medium, Deforestation = plus

3. Impact on Land

1. If there is deforestation, will erosion increase or decrease?
   Increase

2. Does agricultural production increase or decrease because of erosion?
   Decrease

3. In which state is deforestation equal to zero?
   State 4

4. What is the value of land with vegetation in that state?
   Zero

4. Impact on Land and Water

1. Via which two quantities does erosion affect uses of water?
   Removed soil and water reservoirs

2. Does agricultural production have an effect on uses of water according to this model?
   No, not according to this model

3. When removed soil = medium, is agricultural production increasing or decreasing?
   Decreasing

4. And when removed soil = max, is agricultural production increasing or decreasing?
   Steady

5. Impact on Land, Water & Human

1. How should water reservoirs change to make the population without (access to) water decrease?
   Water reservoirs has to increase

2. If there is erosion, what will happen to water reservoirs?
   Decrease

3. When removed soil is small, what is the value of water reservoirs?
   Large

4. Is it true that population without water changes in the same direction as water reservoirs?
   No

6. Impact on GDP (Wealth)

1. What are the three quantities affecting GDP (human wealth) in this model?
   New food and medicine, agricultural production, and uses of water

2. Are these quantities influenced by deforestation, or by erosion?
   Food & med, agricultural production and uses of water are all influenced by deforestation

3. When biodiversity decreases, what happens to the production of new food and medicine?
   Decrease

4. What is the value of GDP (human wealth) in the end?
   Zero

7. Global Economy

1. What are the entities included in the model?
   Global economy, Transportation, Industry, Households, Atmosphere and Human

2. What type of relation is established between 'global market' and 'petroleum available'?
   A direct influence (I+)

3. What are the values of 'global market' in states [19, 16, 17]?
   (surplus, decreasing); (shortage, stable) and (surplus, stable), respectively

4. How can the changes in the values of 'global market' be described ?
   A cyclic behaviour, oscillating between 'surplus' and 'shortage'

8. Uses of petroleum

1. Which quantities are positively influenced by 'petroleum available'?
   'Use of petroleum' for industry, for transportation, and in households

2. Compare the behaviour of the 'global market' and the 'use of petroleum' for industry, transportation, and households. Are they the same or different?
   The four quantities show the same cyclic behaviour

9. Effects on the atmosphere

1. Via which two quantities does 'petroleum available' affect 'greenhouse gases'? And 'pollution'?
   Both 'greenhouse gases' and 'pollution' are affected via 'use of petroleum' for industry and for transportation

2. What are the values of the quantities 'use of petroleum' for the industry and 'greenhouse gases' in states [16, 21, 12]?
   High, medium, low, respectively, for the two quantities

3. Which quantities influence the overall atmosphere 'temperature'?
   'Greenhouse gases' and 'pollution'

4. In which states is the overall atmosphere 'temperature' in the zone of global warming?
   11, 14, 16

10. Effects on households

1. Which quantity is negatively influenced by 'petroleum available'?
   'Use of solid fuel'

2. What is the relation between the values of quantities 'use of petroleum' in households and 'use of solid fuel'?
   They have exactly opposite values (inverse correspondence)

3. What happens with 'smoke' in states [16, 21, 12]? And how does it compare to the behaviour of 'global market' in the global economy?
   'Smoke' increases from low to high, while the 'global market' shows a shortage of petroleum in the global economy

11. Effects on human health

1. What happens with 'chronic respiratory diseases' when 'smoke' decreases?
   'Chronic respiratory diseases' also decreases

2. How do the values of 'chronic respiratory diseases' relate to 'use of solid fuel'?
   These quantities have the same values in every state

3. If 'generic respiratory diseases' increases, what is the reason for that behaviour?
   'Generic respiratory diseases' increases because atmospheric 'pollution' increases

4. How do the behaviour of 'chronic respiratory diseases' and 'generic respiratory diseases' compare in the selected behaviour path?
   These quantities present opposite behaviours: when 'chronic respiratory diseases' increases, 'generic respiratory diseases' decreases

12. Global effects of the petroleum use

1. According to this model: is smoke produced in households related to global warming?
   No, there is no relation between 'smoke' and the overall atmosphere 'temperature' in this model

2. What happens to the overall atmosphere temperature when petroleum available increases, remains stable and decreasing?
   'Temperature' changes in the same direction, that is, it also increases, remains stable and decreases

3. What happens to generic respiratory diseases and to chronic respiratory diseases when petroleum available increases, remains stable and decreases?
   'Generic respiratory diseases' changes in the same direction as 'petroleum available'; 'chronic respiratory diseases' changes in opposite direction when compared to 'petroleum available'

Appendix C

Questionnaires for evaluation of knowledge acquisition by observing and inspecting

The questionnaires used in the 'Learning through simulation' experiment.

PARTICIPANT DATA

Please fill in or circle the relevant data

1 Participant code:

2 Male / Female

3 Age:

4 Topic of Study:

5 Year of Study:

1st / 2nd / 3rd / 4th / ...th

6 Amount of computer experience:

1	2	3	4	5	6	7
none						much

7 Amount of expertise about Ecology:

1	2	3	4	5	6	7
none						much

8 Amount of expertise about Conceptual Modelling:

1	2	3	4	5	6	7
none						much

9 Amount of expertise about Qualitative Reasoning:

1	2	3	4	5	6	7
none						much

PARTICIPANT FEEDBACK

1 How much do you feel you have learned during the session?

1	2	3	4	5	6	7
none						much

2 Did you find it difficult or easy to answer the test questions?

1	2	3	4	5	6	7
very easy						very difficult

3 Did you find it difficult or easy to use the software interface?

1	2	3	4	5	6	7
very easy						very difficult

4 Did you find the diagrams difficult or easy to understand?

1	2	3	4	5	6	7
very easy						very difficult

5 Did you enjoy today's session?

1	2	3	4	5	6	7
not at all						very much

6 Did you enjoy using the software?

1	2	3	4	5	6	7
not at all						very much

7 If you want, you can leave remarks about today's session: