



Automatically Generating Tutoring Questions for Qualitative Simulations

Floor Goddijn, Anders Bouwer and Bert Bredeweg

University of Amsterdam
Department of Social Science Informatics
Roetersstraat 15, 1018 WB Amsterdam, The Netherlands
E-mail: { anders, bert } @swi.psy.uva.nl

Abstract

Learning by interacting with simulations must be supported in order for the learning experience to be effective. Questions play an important role in this respect. In this paper we discuss QUAGS, a program that automatically generates tutoring questions for qualitative simulation models. QUAGS is domain-independent and is designed for embedding in an interactive learning environment. An important problem with automated question generation is the huge amount of questions that in general can be asked. This problem is solved by five restriction methods that can be controlled externally or are executed internally by QUAGS when external control is omitted. The amount of questions is further restricted by heuristics working during the construction and selection phase of the generation process.

Introduction

Qualitative simulations capture the knowledge necessary to teach students about many aspects of systems and their behaviour. However, such simulations are complex and putting their potential to use requires intelligent user support. There are various forms of support, or *cognitive tools* (e.g., sequencing curriculum goals, visualisation tools, tutoring dialogue interaction) (vanJoolingen, 1999). After previous work on curriculum planning (Salles & Bredeweg, 2001) and visualisation (Bouwer & Bredeweg, 2001), in this paper we focus on a particular form of support towards the creation of tutorial dialogue: the use of automatically generated questions.

Much research has been done on dealing with user (student) questions: how to enable them and how to answer them (e.g., Acker, Lester, Souther & Porter, 1991; Cawsey, 1992; Suthers, 1992; Moore, 1995; Metzler & Martincic, 1998). Much less attention has been paid to didactic *tutoring* questions from the teacher (or the system) for the student to answer, however, despite their important role in teaching interactions (e.g., see Sinclair & Coulthard, 1975; Collins & Stevens, 1983; Pilkington, 1997). Inquiry teaching turns learning into problem solving, which motivates students and keeps them on track (Collins & Stevens, 1983). Questions structure the interaction, as they introduce a specific topic, set a goal (to find an answer), and elicit a response (Pilkington, 1997). They can serve to train, test, evaluate, or to diagnose.

The work presented in this paper investigates the kinds of questions that can be asked about the behaviour of systems as modelled in qualitative simulations. A generic taxonomy of question types has been devised for this purpose, which takes into account factors such as the topic of the question, the simulation context, and the kind of reasoning required from the student. When applying this taxonomy to simulations of moderate to high complexity, this results in an enormous number of possible questions. To be of any practical use, selection and ordering mechanisms are necessary to ensure that only the most relevant questions remain. To test the feasibility of our approach, a prototype system, called QUAGS, has been implemented. QUAGS can generate questions for any GARP simulation, in any domain. When given no further constraints, QUAGS searches for the most notable developments in the simulation and generates a set of questions about them. In this case, the amount of questions generated is reduced by heuristics to favour the most interesting ones, fitting them together where possible, without too much overlap between questions. QUAGS can also be given more specific constraints (which could be the output of a curriculum planner, if available), such as a specific state, or part of the system, or a particular type of question. This will generate, if possible, a (set of) question(s) specifically targeted at the constraints.

Qualitative Simulations and Question types

In our work we use GARP (Bredeweg, 1992), a qualitative reasoning engine that follows a compositional approach (Falkenhainer & Forbus, 1991) and implements the important characteristics of the qualitative reasoning technology as originally presented in deKleer & Brown (1984) and Forbus (1984). The engine works on the basis of a number of constructs, including scenarios, model fragments and transition rules. *Scenarios* specify initial situations for the simulator to start a behaviour prediction. *Model fragments* capture knowledge about the structure and behaviour of (partial) systems and are used to assemble states of behaviour. *Assumptions* may be used to further detail the applicability of model fragments. *Transition rules* determine valid transitions between states of behaviour. The basic idea is depicted in Figure 1.

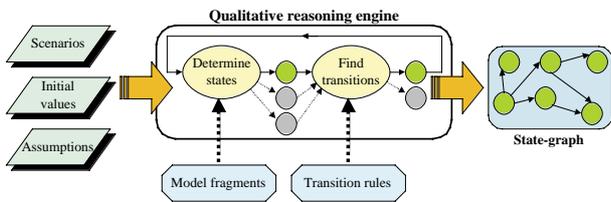


Figure 1: Qualitative simulation

An important output of a qualitative simulator is the *state-graph*, which describes all possible behaviours of a system. Each state in such a graph refers to a qualitative unique behaviour that the simulated system may manifest. The following concepts are relevant for understanding and inspecting the contents of a state-graph:

- States and state transitions: Represent the qualitative distinct behaviours of a system and how distinct behaviours succeed each other.
- Entities: Physical objects or abstract concepts that specify the structure of a system.
- Quantities: Changeable features of entities.
- Quantity spaces: A range of values for each quantity.
- Values and derivatives: The specific qualitative value a quantity has (in a certain state) and how it changes.
- Influences (I+ & I-): Indicate the start of a process, positive or negative.
- Proportionalities (P+ & P-): Propagate the effects of a process through the system.
- Correspondences (V, V[^], Q, & Q[^]): Value correspondences (V) and quantity space correspondences (Q), both undirected and directed ([^]), indicate that certain values for different quantities co-exist.
- Inequalities (<, >, ≤, ≥, & =): Indicate the relative sizes of values or derivatives of quantities.

The concepts central to the notion of a state-graph form the basis of the question types and their generation in QUAGS. The question generation starts by instantiating question type templates. In order to control which question types are used we have developed five criteria to characterise them:

- Perspective (P): What kind of information will be used
 - Simulation run-through (sr) - *changing values*
 - Causal model (cm) – *reasons of change*
- Concept (C): Domain-independent subject
 - Causal relation (cr)
 - Inequality (ie)
 - Correspondence (cp)
 - Value (vl)
 - Derivative (dr)
 - Calculation (ca)
- Behaviour (B): Amount of contribution to the behaviour of the system
 - Real (re) – *representing the behaviour*
 - All (al) – *makes no distinction on this subject*
 - Submissive (su) – *dominated by other effects, therefore not contributing*
- InfoState (S): Whether the needed information is available in the present state or not

- Present (pr)
- Other (ot)
- Answer method (A): Learner’s strategy to find the answer, from easy to hard:
 - Report (rp) – *question indicates place of answer*
 - Search singular (ss) – *answer is to be found near the references in the question*
 - Search plural (sp) – *answer must be found somewhere*
 - Explain singular (es) – *learner must reason with uniform information to answer*
 - Explain plural (ep) – *learner must reason with pluriform information*

Based on these criteria a set of question types can be defined. The generic question type formats are enumerated in the three tables below, using example questions from an imaginary simulation. Figure 2 shows the dependencies between the five quantities in the causal model of this simulation. The information from figure 2 is shown in *italics> in the tables.*

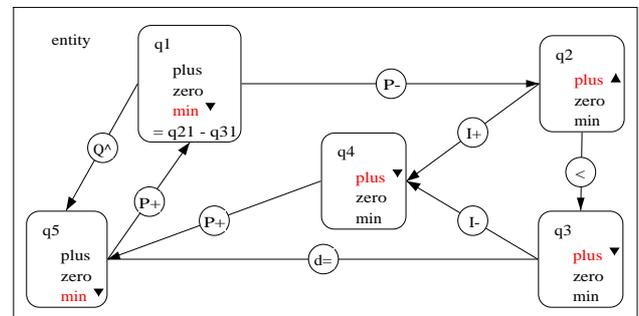


Figure 2: Dependencies in an imaginary simulation

The questions in Table 1 have the perspective criterion *causal model*. Because the questions of this perspective criterion are about dependencies and not about changes over states, all the information needed can be found in one state. Therefore, the *infostate* criterion is *present* for all these questions. The other criteria differ for each question type. For the meaning of C, B, and A (in Table 1) see the criteria and abbreviations discussed above. The last column (covers) shows the dependencies used by the question type.

Nr	C	B	A	Covers
1	ie	re	rp	< > = ≤ ≥
Q: Which quantity is greater, <i>Q2</i> or <i>Q3</i> ? A: <i>Q3</i>				
2	cr	al	rp	I+ I- P+ P-
Q: Describe the influence of <i>Q3</i> on <i>Q4</i> . A: <i>Q3</i> has a <i>negative influence</i> on <i>Q4</i> .				
3	cp	al	rp	Q Q [^] V V [^]
Q: What kind of connection is there between the values of <i>Q5</i> and <i>Q1</i> ? A: <i>All their values correspond, they will always have the same value.</i>				
4	cr	al	rp	I+ I- P+ P-
Q: Does <i>Q3</i> influence <i>Q4</i> directly? A: <i>yes.</i>				
5	cr	al	ss	I+ I- P+ P-
Q: Which quantities have a direct influence on <i>Q4</i> ?				

A: Q2 and Q3.				
6	cr / ca	re	sp	chains: I+ I- P+ P-
Q: How does Q3 influence Q1? A: Q3 has a negative influence on Q4, which propagates its change to Q5, which propagates its change to Q1.				
7	cr	al	sp	chains: I+ I- P+ P-
Q: Does Q5 influence Q2? A: Yes.				
8	cr	al	sp	chains: I+ I- P+ P-
Q: We saw that Q5 influences Q2 and vice versa. Does Q5 influence itself via Q2? A: No, the end of the path from Q5 to Q2 can't be joined to the start of the path from Q2 to Q5				
9	cr	al	es	I+ I- P+ P-
Q: Is the influence from Q2 on Q4 effective? A: No, it is submissive				
10	cp	re	es	values
Q: Why does Q5 have value min? A: Q1 has value min and there is a quantity correspondence between Q5 and Q1.				
11	ca	re	es	values
Q: Why does Q1 have value min? A: Q1 is calculated by Q2 min Q3 and Q2 is smaller than Q3				
12	cr / ie	su	Ep	I+ I- P+ P- <>
Q: Why does Q4 decrease, although Q2 is positive and Q2 has a positive influence on Q4? A: Q3 is positive and Q3 has a negative influence on Q4 and Q3 is greater than Q2.				
13	cr	al	Ep	I+ I- P+ P-
Q: Describe the influence of Q3 on Q2. A: Q3 has a positive influence on Q2.				

Table 1: Questions with perspective = causal model

Table 2 shows the questions of perspective criterion *simulation run-through*. These questions are about changing values over states, the *infostate* criterion is therefore *other* for all these questions.

Nr	C	B	A
14	vl	Re	rp
Q: What will be the value of Q5 in the next state? A: Q5's value will rise from min to zero from this state to state S1.			
15	x	Re	rp
Q: Q4 changed from the state S to this state: its value decreased from plus to zero, what else could have happened? A: Q4 could have stayed plus going to state S1.			
16	vl	Re	rp
Q: What can possibly be the value of Q4 in the next state? A: Q4 can decrease from plus to zero, going to state(s) S1, or stay plus going to state(s) S2.			
17	x	Re	rp
Q: Which quantity will be changed in the next state: Q4 or Q3? A: Just Q4, going to state(s) S1.			
18	cr	Re	es
Q: Q4 changed from state S to this state; namely its derivative rose from min to zero, what is the effect of this change on Q1? A: Q1's derivative rose from min to zero.			
19	x	re	ep
Q: What is going to happen with Q1 during simulation? A: Q1 will decrease and reach minimum value and then			

<i>stay steady ending in state S via states S1 and S2</i>				
---	--	--	--	--

Table 2: Questions with perspective = simulation run-through

Table 3 shows the questions that are applicable for both perspectives. However, they can be seen as causal model questions, because all necessary information can be found in the dependencies (as shown in Figure 2), but because they are about values, they are useful in sets together with the simulation run-through questions.

Nr	C	B	A	Covers
20	vl	re	rp	Values
Q: What is the value of Q5? A: The value of Q5 is min.				
21	vl	al	rp	quantity spaces
Q: Which values can Q5 adopt? A: Min, point (zero), and plus.				
22	cr / dr	re	es	I+ I- P+ P-
Q: Why does Q4 decrease? A: Q3 is positive and Q3 has a negative influence on Q4.				
23	dr	re	es	d=
Q: Why does Q5 decrease? A: Because the derivative of Q5 is equal to the derivative of Q3				

Table 3: Question types applicable to both perspectives.

Question Generation

The generation of the questions proceeds in five steps. First, the given restrictions are analysed. Second, the possible questions are constructed as templates. Third, the most interesting questions are selected. Finally, the last two steps translate the selected questions into 'real' questions (using natural language templates) and order them.

Restriction analysis

One of the most important problems with automatic question generation is the large amount of possible questions. The reasons for this large amount lay in the detailed character of our question types: they are about small aspects of the simulation, which occur many times. To control the amount of questions, we adopt two approaches. The first approach is to enable selection of a certain part of the simulation as subject for the questions and a certain part of the question types, by placing five *restriction methods* on the generation process. The other approach is to make use of heuristics to construct and select the most interesting questions. QUAGS is organised such that it automatically starts to apply heuristics when insufficient restrictions are given and the amount of questions becomes too large. The restriction analysis step consists of applying the restriction methods.

Restriction method 1: Question type criteria

If all question type criteria are undefined, QUAGS takes only the question types that have as *behaviour* criterion the value *real*. This yields the most general questions types.

Restriction method 2: Selection of states

The entered states define, together with the *perspective* criterion, the states for which questions will be generated.

Perspective criterion: *simulation run-through*:

The questions of this perspective criterion are about changes over states that are directly connected. Therefore there must be some sort of awareness which states succeed each other. If the entered states do not form a connected graph, intermediate states are added to make the graph connected. If no states are entered at all, the shortest path(s) between *start* and *end* state(s) are chosen.

Perspective criterion: *causal model*:

Here, no graph is needed because the information for these questions can be found in a single state. Therefore only the entered states are used. Just in case no states were entered, the shortest path(s) between *start* and *end* state(s) are chosen.

Perspective criterion is not defined:

Now, both causal model questions and simulation run-through questions must be possible and therefore a graph is constructed as with the simulation run-through perspective.

If no states are entered, QUAGS uses the start and end states and the shortest path(s) that connect(s) these states. When searching for a path the smallest/shortest one is preferred because of two reasons. First, the shortest path contains all changes needed to transform the start state into the end state. Changes that are not on this shortest path do not contribute to the final situation and can therefore be considered as 'superfluous' detail. Secondly, a short path minimises the amount of repeated information and as a consequence the amount of questions.

Restriction method 3: *System scope*

The system scope defines the part of the system for which questions may be generated. By defining a scope, separate entities can be investigated, but also specific processes. Processes tend to deal with a few quantities from several entities. Therefore, the scope may be defined in terms of entities or quantities. A curriculum planner can use this method to focus QUAGS' question generation process. If no scope is entered, QUAGS considers the whole system.

Restriction method 4: *Subject quantities*

This is an important restriction method. First, because selecting a small set of subject quantities excludes a large number of questions. Second, because all questions are about quantities, as a consequence, the selection of the subject quantities determines for a large part which questions will be posed.

Subject quantities may also be defined as input to QUAGS (e.g., by a curriculum planner). All questions must then be about at least one of these quantities. If no quantities are defined, QUAGS uses two heuristics for determining the subject quantities. The most significant heuristic is that quantities that change *most* during the selected part of the simulation (the selected states) are considered important. 'Changing' can be defined in several ways. One way is to count the number of states where the quantity's derivative has a non-zero value. But this is not really in accordance with the qualitative character of the simulation, which explicitly shows the relevant changes. Another way is to look at the actual

qualitative value and inequality statement changes. In addition to this heuristic, QUAGS considers *start quantities* as important, because they are the causes of the described behaviour in the simulation. A start quantity is not influenced by the system itself and is the instigator of some process in the system.

Restriction method 5: *Forbidden subject quantities*

If some quantities fall under the system scope but may not be explicitly named in the questions or answers, they can be indicated as forbidden. This is a method for a curriculum planner. If no forbidden subject quantities are entered, QUAGS leaves this restriction empty.

Construction of the questions and answers

Construction occurs for each question type selected during the restriction analysis. Each question type has some constraints associated with it. They prescribe which 'situations' may deliver a possible interesting question. By means of these constraints the amount of possible questions is decreased. For example, a constraint may be that a quantity must have a non-zero derivative (so it is changing), in order to qualify for a question about change. Constraints may also be about the state-graph. For example, demanding that a state has at least two successor states when a question is about branching. The questions for which all constraints are fulfilled are constructed by 'finding' the necessary information in the simulation (such as values or dependencies).

Selection of the questions

In the selection phase, the full set of questions generated so far is considered: questions will only be dropped when some other question is better suited, because the goal is to make a diverse set with just a few questions of each type. One general mechanism is that questions about quantities with a high grade of change (investigated at restriction method 4) are considered more important than questions about quantities with a lower grade. Another mechanism used is that some questions have more value if they are combined with other questions, possibly from other types. For example, the question about a quantity is relevant when there is another question that asks why the quantity has this value, or what is going to happen with this value during the simulation. Besides deciding which questions should be selected, QUAGS must also determine when the questions should be posed, because most questions are available in several states. Some questions trigger information that can be valuable in later stages; therefore these questions are posed in the first state possible.

Ordering and verbalisation

The selected questions will now be ordered such that they lead the learner from the start to the end of the simulation. Then the questions, which are still abstract structures, are verbalised. This happens by filling the templates of the question types with the information from the states in the simulation.

Generating Questions for a Complex Simulation: an Example

The Cerrado Succession Hypothesis (CSH) simulation is a complex simulation (Salles & Bredeweg, 1997; 2001). To illustrate the working of QUAGS, this section discusses the question generation for this model. Two extreme cases are considered. First, the focus is on how QUAGS behaves in the absence of restrictions. QUAGS will have to reduce the amount of questions using its internal restriction methods. In the second example QUAGS is given specific restrictions so that questions are generated in the context of a particular focus. Below we first describe some typical features of the CSH simulation.

The CSH simulation shows the behaviour of three populations (grass, shrub and tree) under the influence of fire frequency, controlled by a manager. Due to the size of the causal model in the CSH simulation only parts can be shown here. Figure 3 shows how the fire frequency influences moisture, nutrient, light and soil temperature. These in turn influence the born and dead rates of the populations (not shown). Figure 4 shows how these basic processes affect the (grass) population size (number_of).

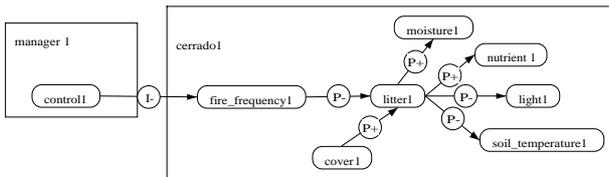


Figure 3: Dependencies in a part of the CSH simulation

In the simulation we use, the number of grass starts at its maximum value, while the shrub and tree populations are non-existing at first (see also Figure 5). During the simulation the colonisation process causes the shrub and tree populations to grow. Figure 6 shows the state-graph of this simulation. The most important issues in this simulation are the changing population sizes, and the origin of the shrub and tree populations. These issues should therefore be considered in the questions generated by QUAGS (when no restrictions are given).

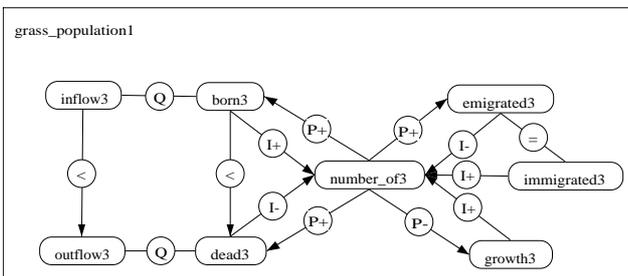


Figure 4: Dependencies for the (grass) population

Generating without restrictions

To illustrate the complexity of generating questions for such a complex simulation (without restrictions), consider the following calculation (Q stands for the amount of quantities, S for the amount of states, and R for the amount of causal relations):

$$\text{Number of questions} = 8QS + 8RS + 3SQ^2$$

In the CSH simulation used here there are 32 quantities, 19 states, and 56 causal relations. Thus, roughly 72000 questions are possible! It is QUAGS' task to generate just a small number of these questions. In this section, we show how QUAGS applies heuristics and generates only the 16 most important questions.

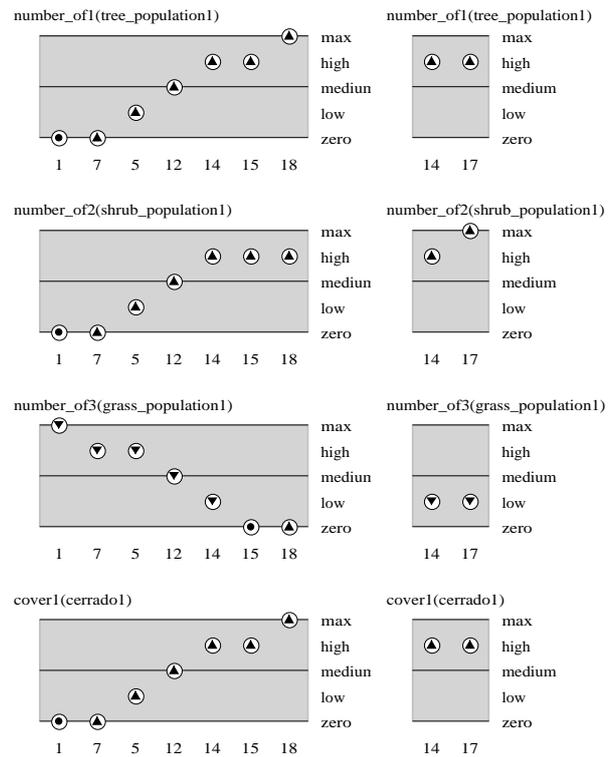


Figure 5: Value history for main quantities in CSH simulation

Restriction analyses

The restriction methods 1, 2 and 4 are applied by QUAGS. Notice that restriction methods 3 and 5 only apply when there is some restriction input, which is not the case here.

Restriction method 1: *question type criteria*

As no constraints have been given, QUAGS assumes that the main behaviour must be communicated. It therefore takes all the question types that have as behaviour criterion the value *real*. This results in the following question types: [1, 6, 10, 11, 14, 15, 16, 17, 18, 19, 20, 22, 23]. This selection reduces the amount of possible question with about 65% from 72000 to 26000.

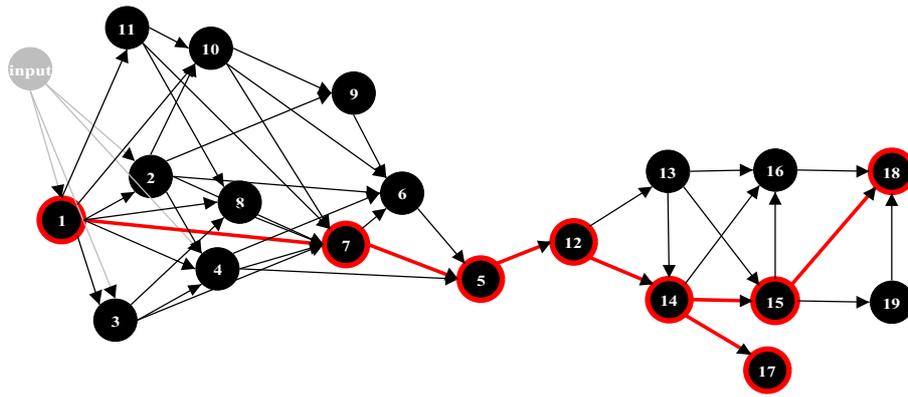


Figure 6: The state-graph of the CSH simulation

Restriction method 2: *selection of states*

QUAGS searches for the smallest connected state-graph that contains the start and end states, because it assumes that the important behaviours are captured within this set of states. The start state is 1; end states are 17 and 18 (Figure 6). The graph found contains the states [1, 7, 5, 12, 14, 15, 17, 18]. Not using all states reduces the amount of possible questions with about 60% from 26000 to 10000.

Restriction method 4: *subject quantities*

QUAGS searches the selected state-graph and examines the changes of the quantities. These changes are stored as degrees of importance of the quantities and are used to select which quantities will be the subjects of the questions. Start quantities, that is quantities at the beginning of a causal path, get the highest degree. In this case, QUAGS selects the quantities *number_of1*, *number_of2*, and *number_of3* belonging to the tree, shrub and grass population, respectively. These quantities have the degree 5. *Cover1* (of the Cerrado) has degree 4 and is also selected. *Control1* (of the manager) is added because it is a start quantity. This selection reduces the amount of possible questions with 95% from 10000 to 500.

Generating the questions

In the construction phase, the amount of questions is further reduced due to constraints (see above), leaving 110 questions: a reduction of 78%. Next, not all these question types actually produce questions, because there are only few direct dependencies between the selected quantities, and they can also not be connected as causal chains. Because of this, no questions are constructed about causal relations. Other question types also produce fewer questions than possible because constraints were able to pick out the most interesting ones. Finally, 16 questions are selected, a reduction of 85%. This is gained by choosing between questions of the same type, making enhancing combinations of questions, and filtering details when overlap exists. Table 4 shows the remaining questions that cover the important issues of the CSH simulation. ‘*Selection*’ gives an indication of why these questions are selected.

In state 1:

Nr 20	<i>Selection:</i> Combination with question of type 19
Q:	What is the value of <i>number_of3</i> of entity <i>grass_population1</i> ?
A:	<i>The value of number_of3 is max(number_of3).</i>
Nr 20	<i>Selection:</i> Combination with question of type 19
Q:	What is the value of <i>number_of3</i> of entity <i>grass_population1</i> ?
A:	<i>The value of number_of3 is max(number_of3).</i>
Nr 19	<i>Selection:</i> High relevancy rate
Q:	What is going to happen with <i>number_of3</i> of entity <i>grass_population1</i> during simulation?
A:	<i>Number_of3 can decrease without reaching its minimum value ending in state 17 via states 14, 12, 5, 7 ...or decrease and reach its minimum value and then stay steady and then rise ending in state 18 via states 15, 14, 12, 5, 7.</i>
Nr 20	<i>Selection:</i> Combination with question of type 10
Q:	Why does <i>cover1</i> of entity <i>cerrado1</i> have value zero?
A:	<i>Number_of1 of entity <i>tree_population1</i> has value zero and there is a quantity correspondence between <i>cover1</i> and <i>number_of1</i>.</i>
Nr 22	<i>Selection:</i> Combination with question of type 19
Q:	Why does <i>number_of3</i> of entity <i>grass_population1</i> decrease?
A:	<i>Dead3 of entity <i>grass_population1</i> is positive and <i>dead3</i> has a negative influence on <i>number_of3</i>, and <i>growth3</i> of entity <i>grass_population1</i> is negative and <i>growth3</i> has a positive influence on <i>number_of3</i>.</i>

In state 7:

Nr 20	<i>Selection:</i> Combination with question of type 19
Q:	What is the value of <i>number_of1</i> of entity <i>tree_population1</i> ?
A:	<i>The value of number_of1 is zero.</i>
Nr 19	<i>Selection:</i> High relevancy rate
Q:	What is going to happen with <i>number_of1</i> of entity <i>tree_population1</i> during simulation?
A:	<i>Number_of1 can rise without reaching its maximum value ending in state 17 via states 14, 12, 5 ...or rise and reach its maximum value ending in state 18 via states 15, 14, 12, 5.</i>
Nr 20	<i>(similar to 20 above but now for <i>shrub_population1</i>)</i>
Nr 19	<i>(similar to 19 above but now for <i>shrub_population1</i>)</i>
Nr 20	<i>Selection:</i> Combination with question of type 14
Q:	What is the value of <i>cover1</i> of entity <i>cerrado1</i> ?
A:	<i>The value of cover1 is zero.</i>
Nr 14	<i>Selection:</i> High relevancy rate
Q:	What will be the value of <i>cover1</i> of entity <i>cerrado1</i> in the

next state? A: <i>Cover1's value will rise from zero to low from this state to state 5.</i>	
Nr 22	Selection: Combination with question of type 19
Q: Why does number_of1 of entity tree_population1 increase? A: <i>Immigrated1 of entity tree_population1 is positive and immigrated1 has a positive influence on number_of1.</i>	
Nr 22	<i>(similar to 22 above but now for shrub_population1)</i>
Nr 22	Selection: Combination with question of type 14
Q: Why does cover1 of entity cerrado1 increase? A: <i>Number_of1 of entity tree_population1 increases and changes in number_of1 are followed by changes in cover1.</i>	

In state 15:

Nr 15	Selection: No competing questions
Q: Number_of3 of entity grass_population1 changed from state 14 to this state: its value decreased from low to zero, what else could have happened? A: <i>Value could have stayed low going to state(s) 17.</i>	

In state 17:

Nr 15	Selection: No competing questions
Q: Number_of2 of entity shrub_population1 changed from state 14 to this state: its value rose from high to max(number_of2), what else could have happened? A: <i>Value could have stayed high going to state(s) 15.</i>	

Table 4: Selected and ordered questions for the CSH simulation

Generating with restrictions as input

In the example above, no restrictions were given as input. Because of the character of the simulation, this resulted in a set of questions almost without any questions about the causal model. In order to get a set that focuses on the causal model for one of the populations, we can place some restrictions:

- Question type criteria: *perspective* criterion is *causal model* (other criteria not entered)
- States: 5
- System scope: Grass population
- Subject quantities: Not entered
- Forbidden subject quantities: Not entered

Now the possible questions amount to about 775 because the number of states, quantities and causal relations is decreased. With restriction method 1, the question type criteria, a reduction of 70% is gained, because only question types of the perspective criterion causal model may be used. Further, with restriction method 4, QUAGS chooses the quantities number_of3 and immigrated3. This reduces the amount of questions with 90% to 26 questions. In the construction phase, the constraints allow just 10 questions, and in the selection phase, QUAGS chooses the following 6 questions:

In state 5:

Nr 5	Q: Which quantities have a direct influence on number_of3? A: <i>Born3 and dead3 and growth3 and immigrated3 and emigrated3.</i>
Nr 9	Q: Is the influence from immigrated3 of entity grass_population1 on number_of3 of entity grass_population1 effective? A: <i>Yes but it is balanced by an other influence</i>

Nr 12	Q: Why does number_of3 of entity grass_population1 decrease, although born3 of entity grass_population1 is positive and born3 has a positive influence on number_of3? A: <i>Dead3 of entity grass_population1 is positive and dead3 has a negative influence on number_of3 and dead3 is greater than born3</i>
Nr 22	Q: Why does number_of3 of entity grass_population1 decrease? A: <i>Dead3 of entity grass_population1 is positive and dead3 has a negative influence on number_of3, and growth3 of entity grass_population1 is negative and growth3 has a positive influence on number_of3</i>
Nr 4	Q: Does number_of3 of entity grass_population1 influence immigrated3 of entity grass_population1 directly? A: <i>No</i>
Nr 4	Q: Does immigrated3 of entity grass_population1 influence number_of3 of entity grass_population1 directly? A: <i>Yes</i>

Table 5: Selected and ordered questions, following restrictions

Four questions were dropped in the selection phase, because they were introduction questions about the values of the chosen quantities. Because there were no other questions about the values, there was an indication that the values were not so important in this situation and therefore the introduction questions were dropped. This is exactly what we had in mind: a set of questions that focuses on the causal model in one of the populations, and not on the changing values.

Evaluation

QUAGS has been evaluated by domain experts to find out whether the questions suggested by QUAGS are indeed the most important ones. We consulted three experts who constructed complex models using GARP. The evaluation consisted of two parts: a questionnaire and an evaluation form. The questionnaire consisted of three parts focussing on the simulations created by the experts. The first part was about the full simulation, whereas the second and third part focussed on specific aspects of the state-graph. In all parts, the experts were asked to develop an ordered set of tutoring questions about their simulations and to indicate the most important quantities. They were also asked to explain their answers.

The experts indicate in the evaluation form that they found the set of question types largely complete, considering the goals that QUAGS aims to cover. This is confirmed by the analyses of their questions. 50% of the questions posed by the experts are available in QUAGS and automatically generated when no additional constraints are given. Another 30% is selected by QUAGS if it is given the same quantities as 'subject' as the experts did. This difference originates from the fact that the heuristics in QUAGS are based on 'changing values' (which the experts agreed with), but the experts had also reasons concerning 'causality' to select quantities. Finally, 20% of the expert questions are not generated by QUAGS. Most of these questions concern

situations in which 'nothing happens'. *E.g.*, a question type such as: Why does the value of quantity Q not change? These situations can indeed be significant. The problem is that the reasons the experts give why these situations are interesting are domain-dependent and can therefore not be implemented as a heuristic in a domain-independent program.

QUAGS has quite a lot of question types which are *not* mentioned at all by the experts. These question types can be divided into two groups. In the first group we find questions with the answer method criterion 'report'. These questions are primarily the basic ones whereas the experts tend to pose more complex questions, possibly assuming that the basic facts are already known. The second group of questions is about the status of causal relations, for example the question why some relation does not have the expected effect. These statuses are not explicitly mentioned in the simulation, which may be the reason why they are not mentioned by the experts.

Conclusion

This paper discusses automatic question generation based on qualitative simulations, with the goal of using these questions as a means to support learners in the process of understanding the details captured in a complex simulation. A computer program, QUAGS, has been implemented that successfully generates tutoring questions automatically. An important problem in automatic question generation is how to restrict the amount of questions. In QUAGS this problem is solved by five restriction methods that can be controlled externally or are executed internally (when external control is omitted). The amount of questions is further restricted by heuristics working during the construction and selection phase of the generation.

The evaluation shows that QUAGS covers a reasonable part of the questions experts would like to pose about the simulations they created. QUAGS heuristic knowledge largely focuses on changing quantities values. The evaluation shows that other aspects, such as the structure of the causal model may also be exploited. Future work should also focus on embedding QUAGS in a fully operational interactive learning environment based on qualitative simulations¹.

References

- Acker, L. H.; Lester, J. C.; Souther, A. F. & Porter, B. W. (1991). Generating coherent explanations to answer students' questions. In H. Burns; J. Parlett & C. Redfield (Eds.) *Intelligent Tutoring Systems: Evolutions in Design*, pages 151- 176, Lawrence Earlbaum, Hillsdale, New Jersey.
- Bouwer, A. & Bredeweg, B. (2001). VisiGarp: Graphical representation of qualitative simulation models. In J.D. Moore, G. Luckhardt Redfield, and J.L. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future*, pages 294-305, IOS-Press/Ohmsha, Osaka, Japan.
- Bredeweg, B. (1992). *Expertise in Qualitative Prediction of Behaviour*. PhD thesis. University of Amsterdam, Amsterdam, The Netherlands
- Cawsey, A. (1992). *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*. MIT Press.
- Collins, A. & Stevens, A. L. (1983). A Cognitive Theory of Inquiry Teaching. In C. M. Reigeluth (Ed.), *Instructional design theories and models: An overview*, pages 247-278, Lawrence Erlbaum, Hillsdale, New Jersey.
- Falkenhainer, B. & Forbus, K. (1991). Compositional modeling: finding the right model for the job. *Artificial Intelligence*, volume 51, number 1-3, pages 95-143.
- Forbus, K.D. (1984). Qualitative process theory. *Artificial Intelligence*, volume 24, number 1-3, pages 85-168.
- vanJoolingen, W. (1999). Cognitive Tools for discovery learning. *International Journal of Artificial Intelligence and Education*, number 10, pages 385-397.
- deKleer, J. & Brown, J.S. (1984). A qualitative physics based on confluences. *Artificial Intelligence*, volume 24, number 1-3, pages 7-83.
- Moore, J.D. (1995). *Participating in Explanatory Dialogues. Interpreting and Responding to Questions in Context*. Cambridge, Massachusetts / London, England: MIT Press.
- Metzler, D.P. & Martincic, C.J. (1998). *Explanatory Mechanisms for Intelligent Tutoring Systems. Proceedings of the international conference on Intelligent Tutoring Systems*, pages 136-145, Spinger-Verlag, volume 1452.
- Pilkington, R. M. (1997). *Analyzing educational discourse: The DISCOUNT scheme. Version 2, march 1997*. Technical report, Computer-based Learning Unit, University of Leeds.
- Salles, P. & Bredeweg, B. (1997). *Building Qualitative Models in Ecology*. In Ironi, L. (ed.) *Proceedings of the 11th. International Workshop on Qualitative Reasoning (QR'97)*. Instituto di Analisi Numerica C.N.R., Pubblicazioni no. 1036, Pavia, Italy.
- Salles, P. & Bredeweg, B. (2001). *Constructing progressive learning routes through qualitative simulation models in ecology*. In: G. Biswas (ed.) *Proceedings of the 15th. International Workshop on Qualitative Reasoning (QR'01)*, pages 82-89, Saint Mary's University, San Antonio, TX, USA.
- Sinclair, J. McH. & R. M. Coulthard (1975). *Towards an Analysis of Discourse: The English used by teachers and pupils*. Oxford University Press, London.
- Suthers, D.D. (1992). *Answering Student Queries: Functionality and Mechanisms*. *Proceedings of the international conference on Intelligent Tutoring Systems*, pages 191-198, Springer-Verlag, volume 608.

¹ QUAGS is currently available as part of VISIGARP and can be downloaded from:
<http://www.swi.psy.uva.nl/projects/GARP/>