Deliverable number:	D7.2.2
Deliverable title:	University of Hull Evaluation of DynaLearn Prototype

Delivery date:	24/12/2010
Submission date:	24/12/2010
Leading beneficiary:	University of Hull (UH)
Status:	Version 05 (final)
Dissemination level:	PU (public)
Authors:	Richard Noble

Project number:	231526
Project acronym:	DynaLearn
Project title:	DynaLearn - Engaging and informed tools for learning conceptual system knowledge
Starting date:	February 1st, 2009
Duration:	36 Months
Call identifier:	FP7-ICT-2007-3
Funding scheme:	Collaborative project (STREP)

Abstract

This deliverable reports on four evaluation experiments undertaken at The University of Hull. These comprised a small-scale pilot study using volunteer biology students and three experiments undertaken within the PGCE Science Education programme. These evaluation studies utilised prototypes of the DynaLearn software (0.6.4 in the pilot study and 0.6.16 in the later studies). The prototype versions contained limited implementations of the conceptual modelling, semantic and virtual character technologies being developed. As such the early case study focused purely on conceptual modelling and the new learning spaces in the software. The later sessions covered semantic grounding, the teachable agent mode of the virtual characters and conceptual modelling.

Results of these studies are analysed within the context of their settings and inferences are drawn to make recommendations for further development of the software and the modelling curriculum/pedagogic approach to learning by conceptual modelling.

Internal Review

- Michael Wißner and René Bühling (UAU), Human Centred Multimedia, University of Augsburg, Germany.
- Petya Borisova and Jordan Uzunov, IBER, Bulgaria.

Acknowledgements

The author would like to thank partners in WP7 for their help in developing the evaluation framework and approaches used. Thanks must also go to all the students who either volunteered or contributed as part of their formal course to the evaluation activities and consented to be research subjects. Thanks must also go to Dr Anne Bore, Dr Janet Gibbs and Richard Davies from the University of Hull Centre for Educational Studies for allowing me access to their PGCE secondary science class. Thanks must also go to Dr Sandra Jones (Ethics Officer, Department of Biological Sciences) and the Faculty Ethics Committee for approving my research with human subjects.

The author also thanks the colleagues at UAU and IBER for their internal review of the deliverable.

Document History

Version	Modification(s)	Date	Author(s)
01	Draft for review	10/12/2010	Richard Noble
02	Sections filled in	17/12/2010	Richard Noble
03	Post review edit and final formatting	22/12/2010	Richard Noble

Contents

1. Introduction

1.1. General contents

This report covers the work undertaken by the University of Hull partners towards the evaluation of the DynaLearn prototype. As such it covers a pilot study (Section 2) undertaken with an early version of the multi learning space conceptual modelling environment (Section 1.2) and later evaluation activities undertaken with trainee teachers using a more advanced prototype (Sections 3 to 5). The activities undertaken focussed on the conceptual modelling environment but also tested prototypes of the virtual character and semantic technologies (Section 1.2). These activities focussed on hypotheses and research questions identified within the general evaluation framework for the overall project and work package seven (Section 1.3).

The results obtained from the prototype evaluations are discussed within the context of the evaluation framework and the general pedagogy of DynaLearn and learning by conceptual modelling in general (Section 6). These studies provide general insights into what is required to advance the software, the curricula and expert modelling of content (WP6, task 6.3/6.4) and the general pedagogy of learning activities within the software (WP7). The deliverable also highlights future plans and evaluation opportunities at the University of Hull (Section 7).

1.2. DynaLearn components

The DynaLearn approach and paradigm are grounded in the learning-by-modelling pedagogical approach (Bredeweg & Forbus, 2003; Mioduser *et al.* 2010). In this approach, students are involved in interpreting, manipulating and constructing models of systems, fostering both conceptual understanding (of subject related content knowledge) and the gradual development of robust scientific reasoning skills (Mioduser *et al.* 2010).

The DynaLearn software comprises six learning spaces (Bredeweg 2009b), Liem *et al.* 2010) that enable students to diagrammatically build conceptual models of varying complexity (from simple concept maps, through basic causal models to full qualitative reasoning models). The qualitative conceptual modelling approach is based on qualitative systems dynamics and qualitative process theory (Bredeweg *et al.*, 2009a; Forbus, 1984). Allied to this conceptual modelling technology are the technologies of the virtual characters and the semantically operated ontology-based feedback (D2.1, Bredeweg et al. 2009). The characters (Wißner *et al.*, 2010b; Andre *et al.* 2009) and feedback (Wißner *et al.*, 2010a,b; Lozano *et al.* 2010) are designed to structure modelling activities and enhance both learning and motivation.

The semantic technology and ontology-based feedback (OBF) operates through a repository of expert and teachers' models of domain knowledge (Gracia *et al.* 2010, Salles *et al.* 2009) providing feedback to students on their modelling efforts, either directly against a single model ort against all of the relevant concepts captured in the repository (Gracia *et al.* 2010). This repository will collate and formalise the models and curricula created by WP6 focussing on environmental science (Salles *et al.* 2009). The evaluation activities presented here are formulated from the initial ideas for learning activities and lesson plans presented in D7.1 (Mioduser *et al.* 2009) and the curricula and content developed in D6.2.4 (Noble, 2010).

1.3. General evaluation framework

The overall aim of Work Package 7 in the DynaLearn project is to assess the contribution of learning by conceptual modelling with the DynaLearn software on students' understanding of environmental sciences. In specific terms, the project aims to assess the effect of DynaLearn's key features and the process of modelling on students':

a. Conceptual understanding (CU) - their learning of content knowledge related to the topics and concepts in environmental science.

b. Scientific reasoning and Modelling Skills through qualitative reasoning (SR/M) - their acquisition of scientific reasoning skills and ability to cope with complexity, through the QR approach and language.

c. Motivation and attitudes (M/A) towards learning science and learning by modelling.

The evaluation of the above, are as function of learning in the DynaLearn environment encompassing:

1. Conceptual Modelling (CM) - in terms of DynaLearn's specific modelling language, modelling process and 6 modelling levels. - the Learning Spaces.
2. Conversational agents (VC) - these act in various functions and roles while interacting with the learner, especially for structuring learning activities and communicating feedback.
3. Semantic Technologies (ST) - individualization of learning via DynaLearn tools for ontology mapping, diagnostic procedures, and the semantic repository.

These six variables form a framework for defining an inquiry space in which individual research questions can be set to evaluate the effects of either individual or combined components of the software on one or more components of learning. The project description of work (Bredeweg *et al.* 2008a) defines seven questions that are to be evaluated. These questions and the components of the inquiry space which they relate to are:

1. Does the diagrammatic approach (as organised in the DynaLearn setting) actually allow learners to address more complex problems? [1a, 1b]

2. Does the meta-vocabulary from which a conceptual interpretation is built, provide learners a domain independent analytic instrument that enables them to construct more fine grained and thorough analyses of how systems work? [3a, 3b]

3. Do the embodied conversational agents establish the ‘involvement momentum’ required for learners to actually benefit from the added value provided by the software for handling conceptual knowledge? Which agents work best? [2a, 2b, 2c]

4. Do the instruments to individualise learning (ontology mapping, diagnostic procedures, and semantic repository) adequately steer learners in acquiring the target subject matter? [3a]

5. Does the personal autonomy cause learners to be more motivated? [General - c]

6. Do learners actually learn better when using the full set of DynaLearn results? [General]

7. And are students more motivated to take on science curricula? [1c, 2c, 3c]

The individual hypotheses and research questions of the activities undertaken for this deliverable are given in each subsection.

2. Evaluation Activity 1 – Pilot study

2.1. Introduction

Development of implementation of modelling functionality in the DynaLearn integrated learning environment will require a good understanding of how students interact with the software. Understanding the approaches they take and the problems students face when learning the approach will contribute to improved software functionality and the development of suitable pedagogical strategies, use-case and modelling activities for the DynaLearn software and learning by conceptual modelling in general. Therefore, a pilot study was undertaken using an early prototype of the DynaLearn conceptual modelling software to evaluate how students worked with the tool, their experiences and whether the approach contributed to their learning.

The prototype software differed from its predecessor (Garp3, Bredeweg *et al.* 2009a) in that it comprised six different learning spaces (Liem *et al.* 2010). Each learning space offers a different approach, focus or complexity to the modelling activity. Therefore, the pilot study focussed on learning activities in the different learning spaces and how these might contribute to learning activities. In particular focus was given to the difference between informal, static, conceptual modelling (traditional concept mapping) versus formalised dynamic conceptual model building (basic causal models in Learning Space 2 or causal differentiation models in Learning Space 4). The use of learning spaces two and four in the pilot study also gave the students the opportunity to work at two levels of complexity in terms describing causality within the DynaLearn modelling language.

Evaluation hypotheses

1) Learning by building and exploring dynamic models created in a formalised structure gives students a better understanding of the system than by building static concept maps.
2) Learning by building and exploring dynamic models created in a formalised structure increases students’ ability to write clear and scientifically correct causal arguments and explanations.

Evaluation questions

1) Do students understand easily the DynaLearn modelling language and approach?
2) Do students build better models when working in the formalised dynamic learning spaces?
3) Does working with dynamic causal models enable students to write clearer scientifically correct causal arguments?
4) Do students find it easy to work with the DynaLearn software?

In addition to this the pilot study was also used to test the beta prototype of the DynaLearn software (versions 0.6.3 and 0.6.4) and provide feedbacks on bugs and functionality issues. Additionally the pilot was used to test teaching approaches and evaluation instruments for use in later evaluation of the complete software.

2.2. Methods

2.2.1. Population/participants

A group of volunteer students were sought from undergraduates studying in the Department of Biological Sciences at The University of Hull. The voluntary activity was open to all undergraduates covering all courses including foundation biology, ecology and biomedical courses. Volunteers were incentivised by the offer of book tokens as a reward for completing the four sessions scheduled outside of normal teaching time. Unfortunately the response of volunteers was severely limited and the original target of 30 students was not met.

Seven volunteers started the series of workshops but only four attended all sessions. The students who dropped out did so because of other academic commitments that formed part of their formal assessments for their degree.

2.2.2. Implementation procedure

The pilot study was conducted over four 3hr session held outside of normal teaching hours in April and May 2010, in the last four weeks of the academic year. The four sessions consisted of:

1) Introduction to the pilot study domain topic and concept mapping activity (Learning Space 1).
    a) Students given stimulus materials (extracts from IPCC 1996 and IPCC 1997; text book diagram of atmospheric heat balance model) and asked to synthesise the key concepts in a concept map.
2) Introduction to conceptual modelling in DynaLearn and basic causal models (Learning Space 2).
    a) Students implement a basic causal model about photosynthesis based on a concept map.
3) Application of basic causal modelling (Learning Space 2) to pilot study domain topic.
    a) Students were given a norm concept map (Appendix B) and asked to develop a basic causal model from it.
4) Introduction to causal differentiation in DynaLearn (Learning Space 4) and application to the pilot study domain topic.
    a) Students were given a norm basic causal model (Appendix C) and asked to implement it in Learning Space 4, differentiating each of the given relations into influences (I’s) or proportionalities (P’s).

In each session a short PowerPoint presentation was delivered covering an introduction to the domain topic (Sessions 1, 3 and 4 covered Climate Change; Session 2 covered photosynthesis), an introduction and explanation of the modelling approach being used in the session and an introduction to the session’s activities.

Sessions 1, 3 and 4 were used as evaluation activities whilst session two was used purely as a means to introduce students to conceptual modelling and to work through with them the approach to building basic causal models in DynaLearn and to introduce the to the domain of qualitative conceptual modelling.

At the end of each session the students were asked to complete a written assignment (Section 2.2.3, Appendix A).

All modelling activities were undertaken in the DynaLearn prototype version 0.6.4.

2.2.3. Design of data collection instruments

The activities of the students during the pilot study were assessed using 5 main instruments: models, written texts, a questionnaire, a log and by observation.

Model instruments

The model files generated by the students (Activity 1, concept map; Activity 2, basic causal model; Activity 3, causal differentiation model) were retained as evidence of what they had implemented in the software during the class sessions. The model instruments were used to assess their domain knowledge and their understanding of the diagrammatic modelling language in the DynaLearn software.

Written texts

After each of the modelling activities the students were given a written assignment comprising four questions (Appendix A). The written assignment was designed to assess students' understanding of the domain concepts and their appreciation of, and ability to, causality and causal reasoning. The questions also aimed to assess the students understanding of some of the reasoning terminology used in the DynaLearn conceptual modelling approach. The four questions focused on identification and description of important domain concepts, identification of key processes involved in the topic, prediction and causal reasoning in given scenarios and identification of conditions that may affect the possible outcomes of scenarios.

The same four questions were administered after each of the three sessions to assess whether the three sequential modelling activities influenced their understanding of the domain or their interpretation of the questions.

Questionnaire

A short questionnaire comprising 20 questions (Appendix D) was administered at the pilot study to assess the experience of the students in relation to their experience of different components of the modelling activities.

Log of experiences

Students were asked to keep a written log of their experiences during each session and submit it at the end of the course. Students were given a list of questions (Appendix E) as stimulus for the type of information they should record. These included previous experience of similar activities, levels of conceptual understanding and usability of the software.

In-class observation

Students were observed during the model activities collecting information regarding common problems, implementation issues, common questions and also any bugs in the prototype software.

2.2.4. Data analysis and variables measured

Concept maps

Student concept maps were analysed for structure, complexity, completeness and for evidence of implementation errors using the list of variables defined in Appendix F. The variables considered the number of elements introduced into the concept maps in the form of propositions consisting of “concepts” (the blocks) and “relationships” (the links between blocks). Maps were analysed to identify the number of items used and whether they related to system structure, processes in the system, aspects of change in the system or whether items contained integrated statements (that is complete or partial propositions within one item).

The concept maps were also marked against a predefined norm list of concepts (Appendix F) that would form an important part of the written causal explanation of the problem defined in the activity (Appendix A). Concepts were marked for their occurrence and completeness/clarity. The number of concepts that were considered superfluous to the explanation was also counted.

Basic Causal Models (LS2)

Student causal models were analysed for structure, complexity, completeness and for evidence of implementation errors using the list of variables defined in Appendix F. The variables considered the number of elements correctly introduced into the models in the form of entities, quantities, configurations and causal dependencies, the correct assignment of quantities to the appropriate entity, the use of state and process quantities (where a process quantity would best be described as a rate and associated with direct influences at LS4 and above) and the scientific correctness of the concepts.

The models were also marked against a predefined norm list of concepts (Appendix F) that would form an important part of the written causal explanation of the problem defined in the activity (Appendix F). Concepts were marked for their occurrence and whether the concept was included causally within the model. The number of concepts that were considered superfluous to the explanation was also counted.

Written texts

The written texts were analysed for their complexity and accuracy in identification of the key concepts (Q1, against a predefined list, Appendix F), named processes involved (Q2), causal concepts and hypotheses (Q3) and conditions affecting the behaviour and possible alternative outcomes (Q4). Complexity was measured by the number of individual statements made and accuracy by the number and proportion of valid statements made. In addition to checking for validity causal arguments were also defined by whether they were complete (if, then, because aspects all included) or incomplete (because missing or the argument was non-sequential and missed an important concept out) (Appendix F). Changes in the scores of individual students were assessed over the three assignments.

2.3. Results

2.3.1. Model instrument analysis

Concept mapping activity

The seven students undertaking the concept mapping task identified between 35% and 65% of the correct concepts from the stimulus materials. Of these, students implemented between 5% and 35% of these key concepts clearly. Most students also introduced around five concepts to their concept maps which were considered to be superfluous to the description of the problem presented to them (Table 1). Of the seven students, three created concept maps in a hierarchical fashion and four in a web design.

Table 1 Implementation and clarity of key domain concepts in the student concept maps (S1, S2, S3, S4, S5, S6 and S7).

	S1	S2	S3	S4	S5	S6	S7
n correct concepts	10	10	11	9	6	8	6
% correct concepts	59	59	65	53	35	47	35
n clear concepts	5	6	5	2	1	3	4
% clear concepts	29	35	29	11	5	17	23
n superfluous concepts	5	6	5	6	6	3	5

There was considerable variation in the complexity of the concept maps produced by the seven students. This was both in terms of the number of concepts and especially in the number of relationships used (Figure 1). Students implemented between 13 and 20 concepts and between 9 and 45 relationships. Students implemented different types of domain concepts in both map concepts and relationships although concepts of change and process were most often implemented in the relationships (Figure 1). Some students implemented multiple domain concepts within the same mapping ingredient, mixing both structural and behaviour information.

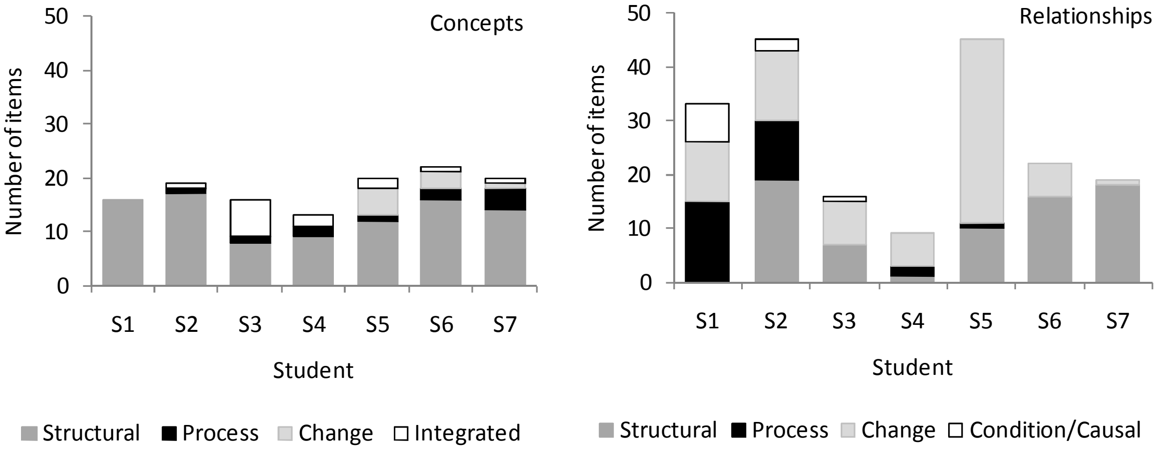


Figure 1 Implementation of domain knowledge within the “concept” and “relationship” components of the student concept maps.

Basic Causal Model Activity

The four students undertaking the basic causal modelling activity implemented quite different numbers of concepts in their models despite all being given the same concept map from which to evolve their models (Table 2). There was also a marked difference in the ability of students to implement the key concepts in a causal manner. Two of the students only managed to implement three domain concepts causally. Despite being given the concept map as a focus all of the students still included at least one superfluous domain concept from their own concept map and the stimulus materials used in the previous week's activity.

Table 2 Implementation of key domain concepts in the basic causal models built by students (S1, S2, S3 and S4).

	S1	S2	S3	S4
n correct concepts	16	9	11	8
% correct concepts	94	53	65	47
n causal concepts	13	7	3	3
% causal concepts	76	41	18	18
n superfluous concepts	1	1	2	2

The student models were relatively similar in complexity in terms of the number of entities that they employed but were markedly different in the numbers of configurations, quantities and causal dependencies they used (Figure 2). Most students exhibited similar numbers of mistakes in terms of conceptually incorrect entities, quantities and configurations. However, students appeared to show marked differences in their understanding and implementation of causal dependencies.

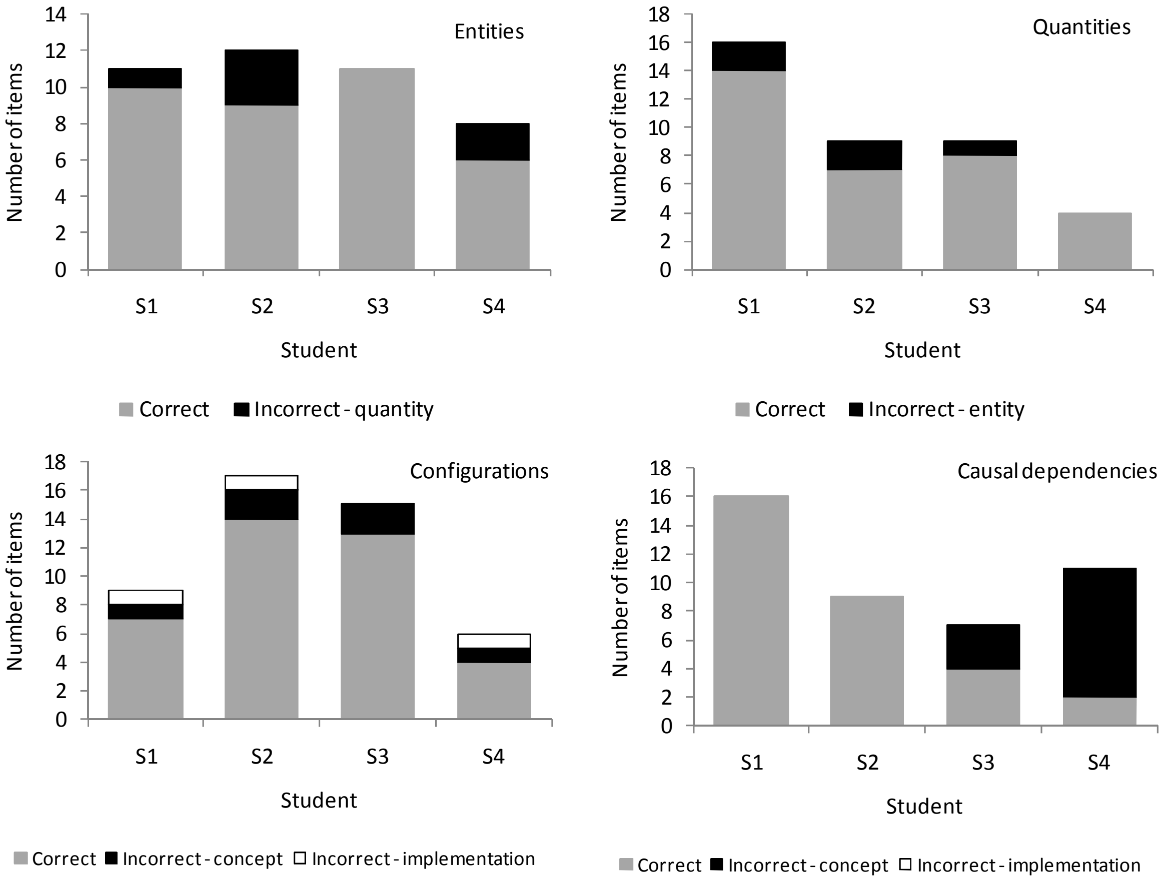


Figure 2 The use of model building blocks by students (S1, S2, S3 and S4) building a basic causal model in Learning Space 2.

2.3.2. Text instrument analysis

The four students showed no discernable pattern in their answers to question one of the written assignments (Figure 3). Student S3 was the only one to exhibit a small increase in complexity and correctness of their answers. Student S4 showed a decline in identifying the key concepts across the three weeks. The answers given by S2 in the final written assignment showed a clear confusion between the key domain concepts and the entities/configurations used in the final model, hence their score declined in both complexity and correctness having improved markedly after the basic causal model exercise.

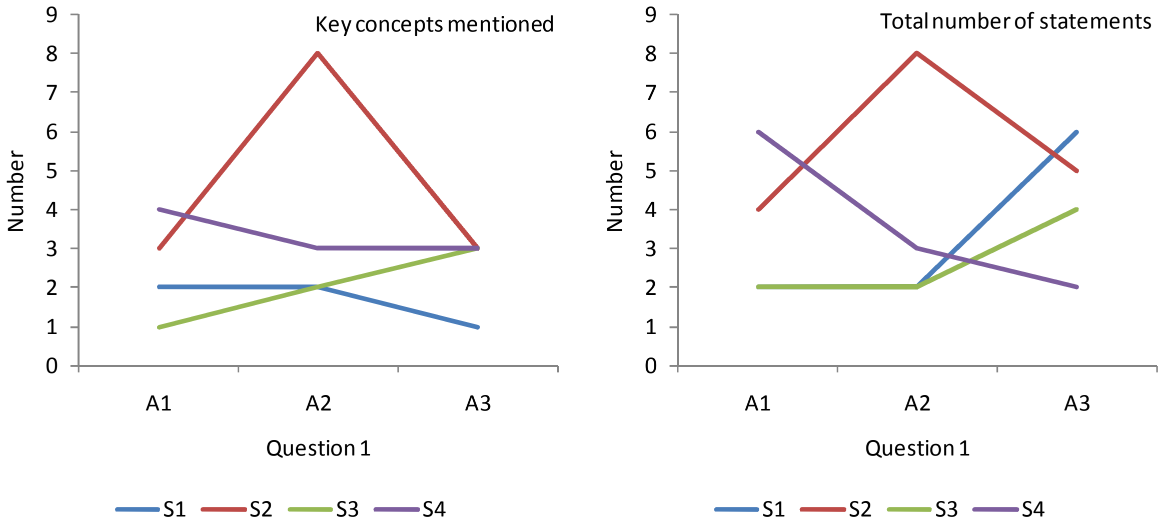


Figure 3 Summary of student responses to Question 1 of the written assignment dealing with the identification of key domain concepts. Answers were analysed as the number of statements made in the answers after the three different activities A1, A2 and A3.

Students showed no discernable pattern in their ability to identify named processes important to the domain concept and the behaviour considered (Figure 4). All their answers were broadly similar in terms of number of processes correctly identified. In general students identified concepts of change but not the processes that were involved.

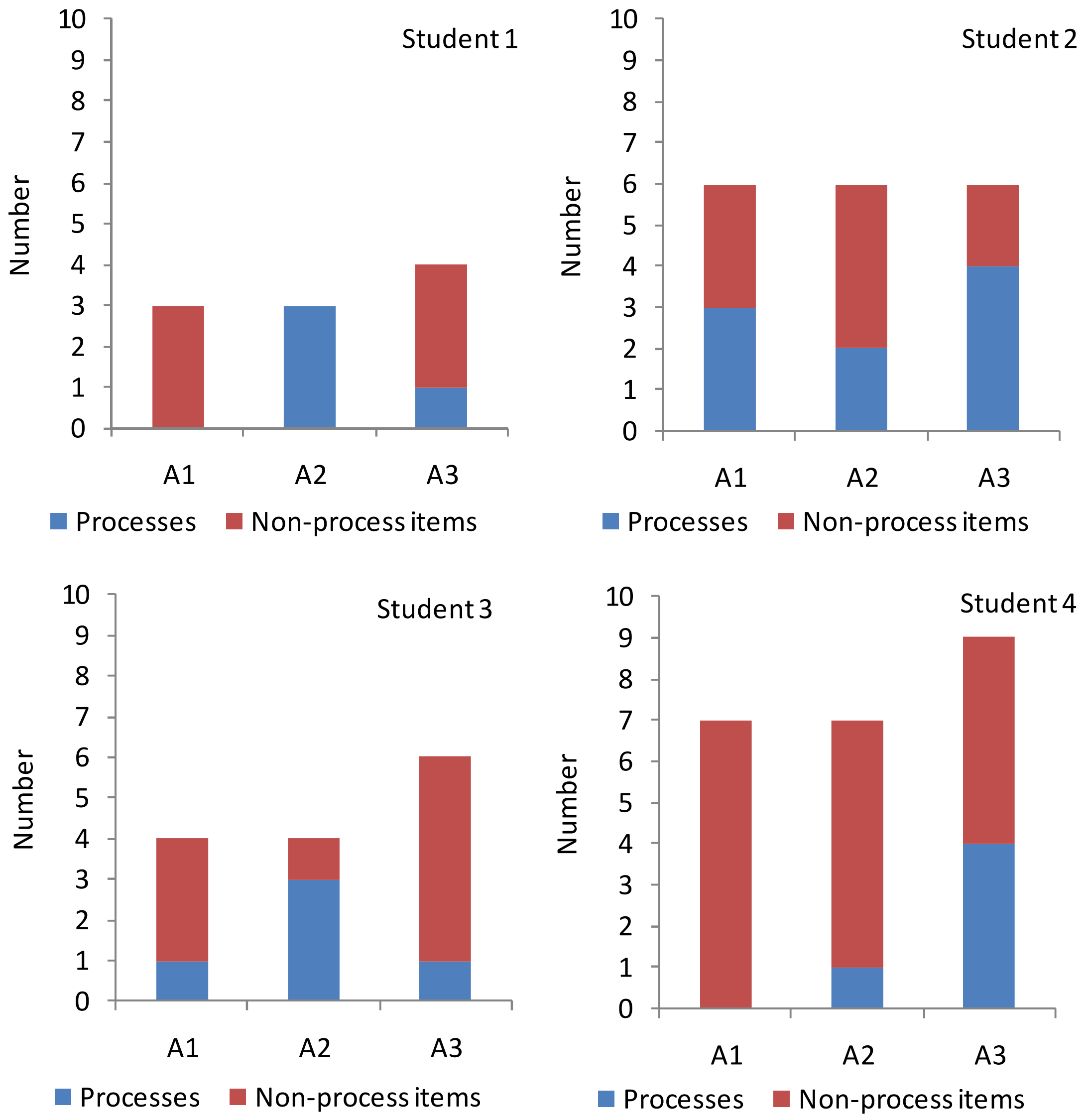


Figure 4 Summary of student responses to Question 2 of the written assignment dealing with the identification of processes. Answers were analysed as the number of statements made in the answers after the three different activities A1, A2 and A3.

Students showed no discernable pattern in their ability to describe causality within the system across the three written assignments following concept mapping, basic causal modelling and causal differentiation of a model. Whilst students could identify causal concepts and relationships between things that were changing most exhibited a mix of true causal statements and causal statements that were either undefined or non-sequential. That is the statements linked two concepts that were changing and could be seen to be related but the student did not define why or missed out an important linking concept between the equivalent of the “if” and “then” components of their statements (Figure 5).

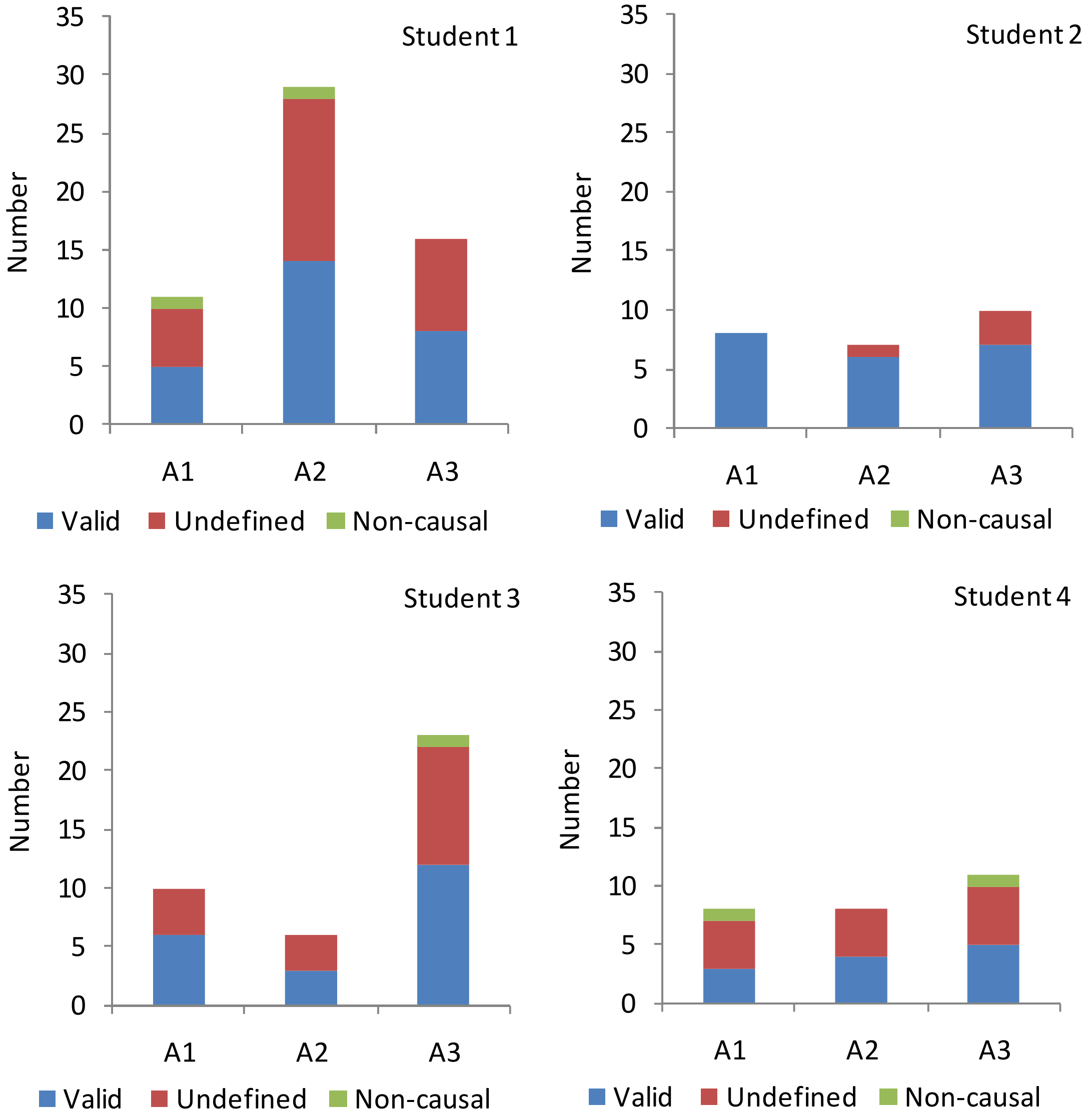


Figure 5 Summary of student responses to Question 3 of the written assignment dealing with causal hypotheses and explanations of behaviours. Answers were analysed as the number of statements made in the answers after the three different activities A1, A2 and A3.

The students showed marked differences in their ability to identify conditional information within their work, or to identify alternative hypotheses and outcomes. Also they did not show any discernable changes with time that could be attributed to an improvement in the understanding of the concept of enabling conditions in the conceptual modelling context (Figure 6).

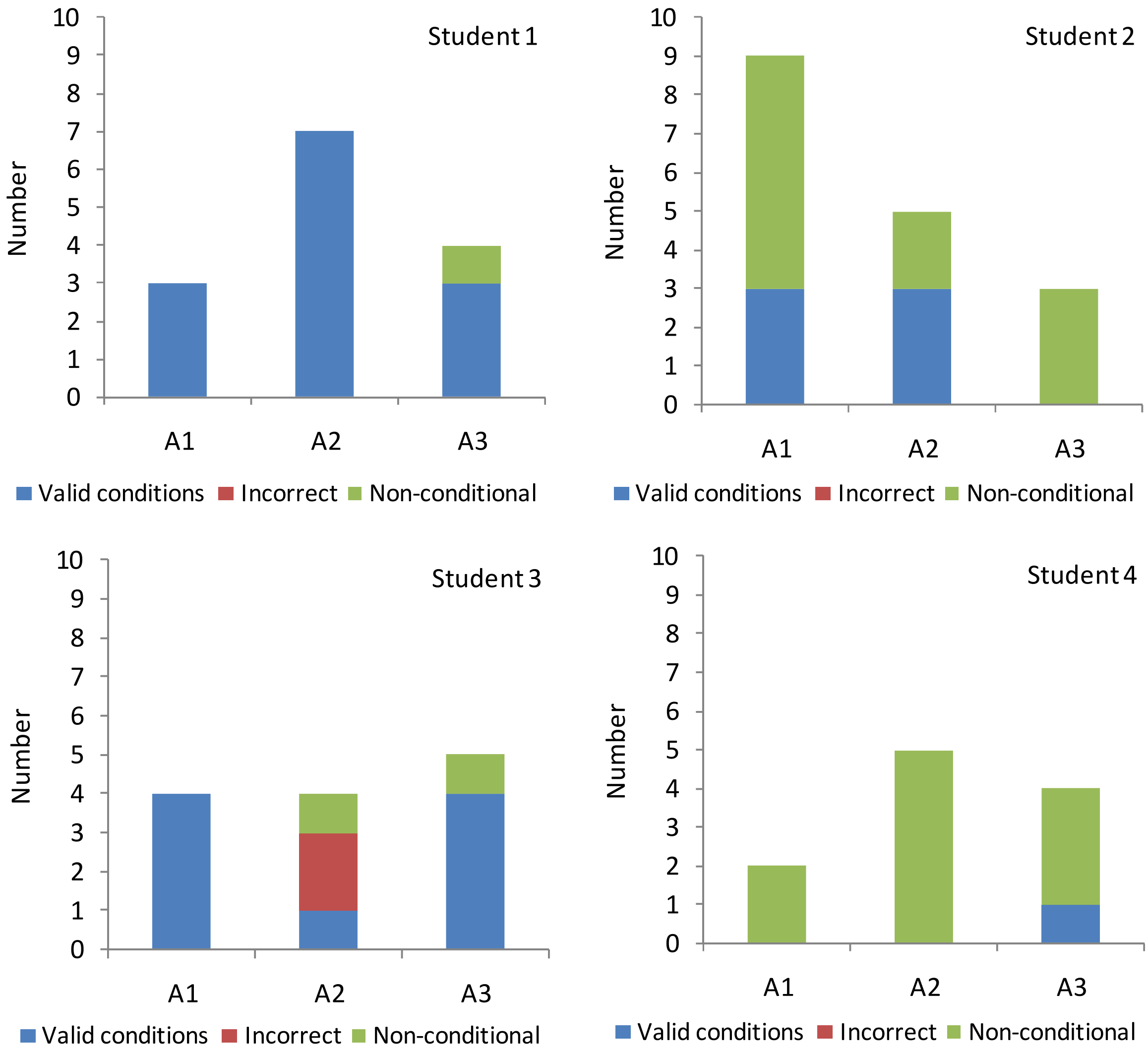


Figure 6 Summary of student responses to Question 4 of the written assignment dealing with causal hypotheses and explanations of behaviours. Answers were analysed as the number of statements made in the answers after the three different activities A1, A2 and A3.

2.3.3. Questionnaire and log responses

The responses of the four students to the questions posed in the questionnaire are summarised in Appendix D. The low number of students completing the course limits the value of the results obtained and restricts any inferences that can be drawn from them. However, a number of features can be noted from the responses. Conceptual modelling was new to all the students (Q1), they felt it contributed to their understanding of systems behaviour (Q4), they found the approach interesting (Q8) and some found it motivating (Q5). Additionally the all felt that they understood the modelling tasks they were given (Q9).

In general students said they experienced some level of difficulty in the modelling tasks specifically related to the new modelling approach and QR language. These included defining entities and quantities (Q14) and defining causal relations (Q15) especially differentiating influences and proportionalities (Q17). Three of the students also expressed difficulty in defining qualitative quantity spaces (Q18).

The answers given in the individual student logs closely mirrored that given in the questionnaire answers and did not provide any new insight into the students' experience during the sessions.

2.4. Discussion

Q1: Do students understand easily the DynaLearn modelling language and approach?

The results of the pilot study indicate that there is a considerable overhead for students to pick up and understand the specifics of the DynaLearn modelling language. This is specifically in terms of the definitions of the model ingredients and the representation of causality through the use of causal dependencies. Students specifically identified understanding the distinction between direct influences and proportionalities as a difficult task. Therefore, learning activities need to be specifically devised to overcome this overhead when learning domain specific knowledge.

Q2: Do students build better models when working in the formalised dynamic learning spaces?

The results of this study gave no indication as to whether students were able to build better models in the formalised conceptual modelling language used in basic causal and causal differentiation models than in the less formal approach using concept maps. This may be confounded by all of the students claiming that they had never previously worked using concept maps. Ideally the formalised approach and framework for modelling of problems available in DynaLearn conceptual modelling should make it easier for students to build models that represent systems in a clear and consistent way. It was clear from the concept maps that students were inconsistent in how the implemented structural and behavioural concepts within a concept map, which made them quite difficult to read and interpret. The formalised approach of DynaLearn to presenting structural and behavioural information should therefore provide an advantage for students organising and communicating their thoughts.

Q3: Does working with dynamic causal models enable students to write clearer scientifically correct causal arguments?

The written assignments used gave no indication of improvement in students writing of causal explanations. This may be because of the time available for students to learn and understand the qualitative process theory that underlies the implementation of causality in DynaLearn. It was apparent in the modelling that most issues they faced related to the implementation of causal arguments in their models and from the distinction between the different types of causal dependencies used in the logical reasoning. Therefore, it could be expected that the use of the qualitative reasoning modelling would only start to influence their written work once they had fully understood how the software reasons during simulations. This is an area where interaction and feedback from virtual characters, such as the questions and answers used in the teachable agent mode, may help students understand how to reason and how to frame their causal arguments and explanations.

Q4: Do students find it easy to work with the DynaLearn software?

Overall most students indicated that once they had identified the appropriate icons that the software was fairly easy to use. Only two key usability issues were identified by the students. Firstly, most students asked for an "undo" button, as found in most other software packages. Secondly, the implementation of the "add quantity" and "add configuration" functions should not automatically set you up to edit an existing model ingredient. On numerous occasions students found themselves inadvertently overwriting an ingredient they had previously added.

3. Evaluation Activity 2 – Grounding of terms in models

3.1. Introduction

Grounding terms is an essential step in the generation of personalised ontology-based feedback based on the semantic technology and model library in the DynaLearn repository. Therefore, this becomes one of the key model building steps for the DynaLearn approach and development of pedagogical and feedback strategies. In addition to this grounding can be a useful step in its own right requiring students to think carefully about the meaning of the domain terms they introduce to a model. Therefore, it is essential that the grounding functionality is fully tested and its capacity to provide cover for all domain terms as defined by students is evaluated.

Evaluation hypothesis

The process of grounding terms, whilst essential for generating personalised feedback, also contributes to conceptual understanding through requiring the student to give careful consideration to the terms they are using.

Evaluation questions

1) Do the current repository, DBpedia and grounding facility enable students to ground all terms?
2) How many terms would currently require the creation of anchor terms in the repository?
3) What type of terms require the generation of anchor terms in the repository?
4) How variable and correct are definitions provided by a range of students?
5) Does grounding (choosing from a predefined list) give students a better conceptual understanding of terms than requiring them to provide their own definition?

3.2. Methods

3.2.1. Population/participants

Evaluation activity 2 was undertaken as a workshop within the normal time-tabled activities of students studying for a Post Graduate Certificate in Education (teacher training). The potential class comprised 37 students studying science education at secondary schools and they came from a back ground of undergraduate degrees in biology, chemistry and physics. Despite the workshop falling within their usual time-tabled sessions, attendance wasn’t complete and the study involved 23 students. The content focus aimed at GCSE level because this was the level at which the students were currently focussing.

3.2.2. Implementation procedure

This 3hr learning by modelling workshop was held in late October 2010. After a short PowerPoint presentation on conceptual modelling the students were set a problem based on a real-life GCSE exam question. The students were then given a printed version of a norm model covering the topic from the exam question to implement in learning space two. Implementing the model manually rather than giving students a version of the model file was chosen so as to give them experience of the functionality of the software prior to later modelling activities (Sections 4 and 5). The class was randomly split into two groups:

1) Experimental group – building the model and defining the modelling terms using the grounding facility in the DynaLearn software and repository (DBpedia functionality).
2) Control group – building the model and defining the modelling terms using the free-text comment boxes in DynaLearn.

At the end of the modelling session all the students were asked to complete a written assignment based on the GCSE exam question.

All modelling activities were undertaken using DynaLearn beta prototype version 0.6.16.

3.2.3. Design of data collection instruments

Model grounding instruments

Students were asked to submit their basic causal model at the end of the session. The model comprised three types of information:

- The completeness of the model in terms of content.
- The groundings used by the experimental group.
- The open definitions created by the control group (and by the experimental group where they were unable to find acceptable groundings).

Written texts

At the end of the session the students were given a short written assignment based on a GCSE exam question (Appendix G) that would require them to generate a causal description of an ecological phenomenon. This question aimed to test their ability to describe a causal argument and also to identify all the important concepts from the model to explain the phenomena.

3.2.4. Data analysis and variables measured

Model grounding

Models submitted by the experimental group were due to be analysed for the number of entities, quantities and configurations successfully grounded. The models were also going to identify the number and types of terms that could be successfully grounded in DBpedia, the number and types of terms for which grounding was not available and whether different students chose alternative groundings for the same terms.

Written texts

The written texts were analysed for their complexity and accuracy in identification of the key causal concepts and arguments explaining the phenomena in the question (Appendix H). Complexity was measured by the number of individual statements made and accuracy by the number and proportion of valid statements made (compared with a predefined list of causal concepts). In addition to checking for validity causal arguments were also defined by whether they were complete (if, then, because aspects all included) or incomplete (because missing or the argument was non-sequential and missed an important concept out) (Appendix I).

3.3. Results

3.3.1. Grounding analysis

Use of DBpedia groundings

Unfortunately during the experiment the students were unable to complete their grounding activity due to instability in the repository server, which shut them out for large periods of the time available to them. Therefore, no data regarding the use and suitability of DBpedia terms could be gathered from this experiment.

Control group definitions

In light of the failure of the repository during the experimental activity data from the control group are to be analysed to gain value added data towards evaluation. The free text definition of terms can be used to look at the potential student behaviour and ability when defining Anchor terms within the DynaLearn repository. These analyses are ongoing and will be fed back internally when available and reported in later deliverables.

3.3.2. Text instrument analysis

Given that the grounding evaluation activity failed due to problems with the stability of the repository server the text-based instruments could not be analysed in the intended way. Data from the written tests are yet to be fully analysed but may prove useful as a control measure for further analysis of any differences between the two groups used in Evaluation activity 3. These data are currently not ready for use in this report.

3.4. Discussion

Given that the repository connection failed during this experiment it is currently impossible to make anything out of the data collected from the control group. However, grounding is one of the critical steps for ontology based feedback in modelling. Therefore, this experiment should be repeated at an appropriate time. Future analysis of the control group definitions will be used to inform the development of technology and handling of anchor terms in the repository.

4. Evaluation Activity 3a – Open-ended modelling in Learning Space 2

4.1. Introduction

Conceptual modelling will be a new activity to most students and as such there will be an educational overhead due to the requirement to learn the modelling approach and language. As such naive modellers will attempt to introduce their knowledge and concepts into models in sub-optimal ways. The development of pedagogical strategies to direct students in good modelling practice, by both teachers and the software's built in help, feedback and diagnosis from the semantic technology and virtual characters, requires that characteristics of naive modelling activities are identified.

Evaluation hypothesis

1) Students will initially face an overhead of learning the modelling approach and language that may initially impede them from formalising their knowledge in a conceptual model. Therefore, early modelling attempts will contain errors in representation. These errors may be predictable and therefore in the future could be addressed using appropriate automated ontology based feedback and diagnosis.

2) One of the key steps in building a conceptual model is defining the system and the important concepts to be modelled (Bredeweg *et al*. 2008b). Students often face difficulties in open-ended tasks to identify what is important or required for the model. Whilst this is not an issue specific to the DynaLearn approach, it needs to be considered to develop appropriate pedagogical approaches for DynaLearn and useful model feedback tools.

3) Naive modellers will learn the modelling language and interface at different rates; this will have implications for pedagogic strategies and activities within the classroom.

Evaluation questions

1) How often do naive modellers make errors in implementation of their model ingredients?

2) How easily do naive modellers identify the important components that need to be included in a causal model to represent a scientifically accurate causal argument?

3) How variable is model complexity between students in terms of number of different ingredients?

4) Do naive modellers work at appreciably different rates during a modelling session?

4.2. Methods

4.2.1. Population/participants

Evaluation activity 3a was undertaken as a workshop within the normal time-tabled activities of students studying for a Post Graduate Certificate in Education (teacher training). The potential class comprised 37 students studying science education at secondary schools and they came from a back ground of undergraduate degrees in biology, chemistry and physics. Despite the workshop falling within their usual time-tabled sessions, attendance wasn't complete and the study involved 18 students. The content focus aimed at GCSE level because this was the level at which the students were currently focussing.

4.2.2. Implementation procedure

This activity was part of 3hr learning by modelling workshop held in mid-November 2010. Students were given a brief PowerPoint presentation giving them a re-cap on conceptual modelling and then an introduction to the concepts to be modelled during the session. At the start of the session students were asked to complete a pre-test to show their existing understanding of the concepts and their ability to write a coherent causal argument.

Students were then given 1hr to build a basic causal model in Learning Space 2 that could be used to explain the phenomenon in the written test. The students were given no guidance on what to include other than the diagram that came with the written test and a text book diagram of osmosis used during the PowerPoint introduction.

All modelling activities were undertaken using DynaLearn beta prototype version 0.6.16.

4.2.3. Design of data collection instruments

Written pre/post test

The students were given a short-answer question about osmosis and diffusion from a GCSE biology exam (Appendix J). The exam question required them to describe and explain causally the reason for the observed phenomena. The same question was administered as a pre and post test to identify if the students changed either the content or style of their answers. This instrument is covered fully in section 5.

Model instruments

The models generated by each student after the free modelling session (1hr) were collected as evidence of what they had managed to implement in the time and also how they had implemented it.

4.2.4. Data analysis and variables measured

Basic causal model LS2

Student causal models were analysed for structure, complexity, completeness and for evidence of implementation errors using the list of variables defined in Appendix I. The variables considered the number of elements correctly introduced into the models in the form of entities, quantities, configurations and causal dependencies, the correct assignment of quantities to the appropriate entity, the use of state and process quantities (where a process quantity would best be described as a rate and associated with direct influences at LS4 and above) and the scientific correctness of the concepts.

The models were also marked against a predefined norm list of concepts (Appendix I) that would form an important part of the written causal explanation of the problem defined in the activity (Appendix I). Concepts were marked for their occurrence and whether the concept was included causally within the model. The number of concepts that were considered superfluous to the explanation was also counted.

4.3. Results

4.3.1. Model analysis

The students' models generally presented quite similar complexity in terms of the number of key concepts (Appendix I) that they managed to implement as naive modellers into a basic causal model (generally around 4-6). However, they generally missed many of the key concepts and in nearly all cases they identified less than half of the list. Only three of the 18 students managed to implement some of their concepts correctly with suitable causal dependencies describing the cause/effect of the concept (Figure 7). In most cases causal dependencies were missing or wrong, or the student implemented notions of change within the configuration statements or the comments boxes of the ingredients.

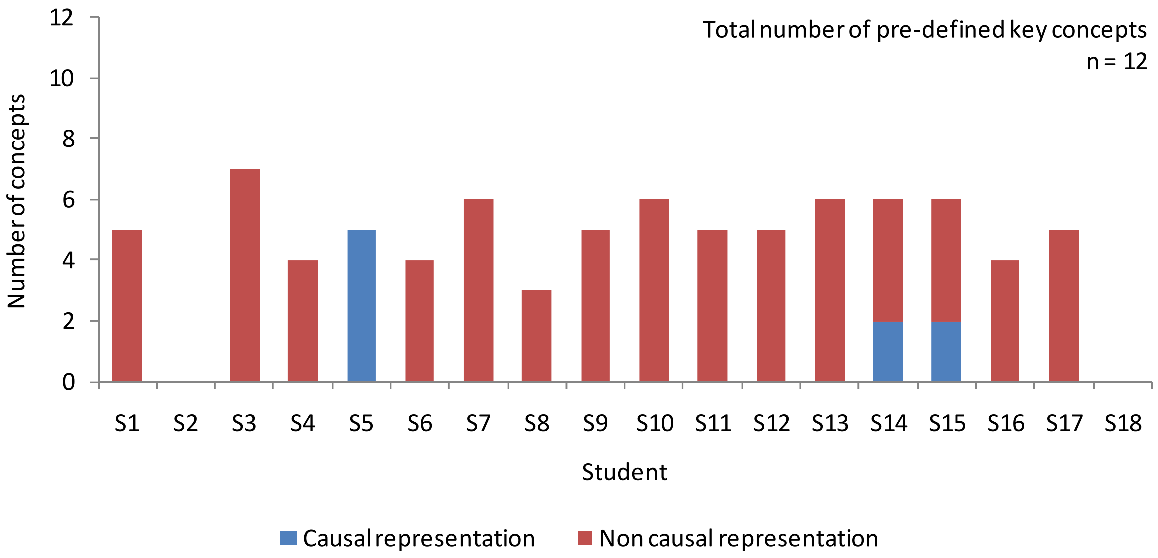


Figure 7 Numbers of key concept (pre-defined from syllabus) included by students whilst building an open ended basic causal model as an aid to answering an exam question based on cellular osmosis.

Very few students exhibited any errors in the implementation of entities and quantities and generally added 4-6 entities and 4-5 quantities to their models (Figure 8). The greatest variability and largest source of error in the student models were from their lack or incorrect use of configurations and causal dependencies (Figure 8). Many students failed to add any correct causal dependencies with large numbers either conceptually incorrect or incorrectly implemented.

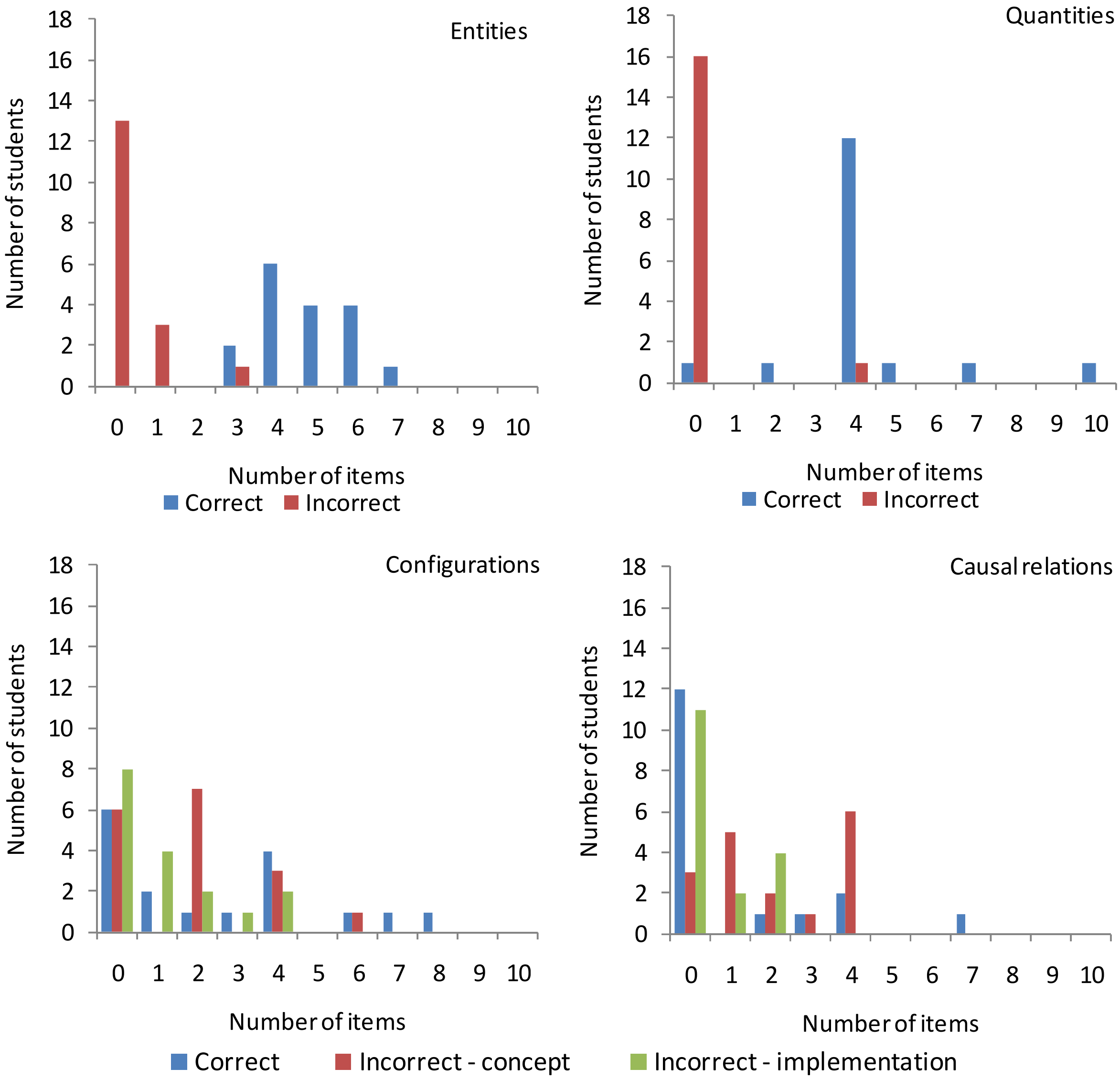


Figure 8 Summary of the implementation of model items by students during the open-ended modelling session. The number of students entering correctly/incorrectly numbers of entities, quantities, configurations and causal dependencies is shown.

One aspect of note from the naive student models was that they all generally used similar amounts of quantities in their models and that nearly all of these related to state variables (Figure 9). Only two of the students used any quantities that could be related to processes. The lack of process quantities probably influenced their issues with implementation of conceptually correct causal dependencies.

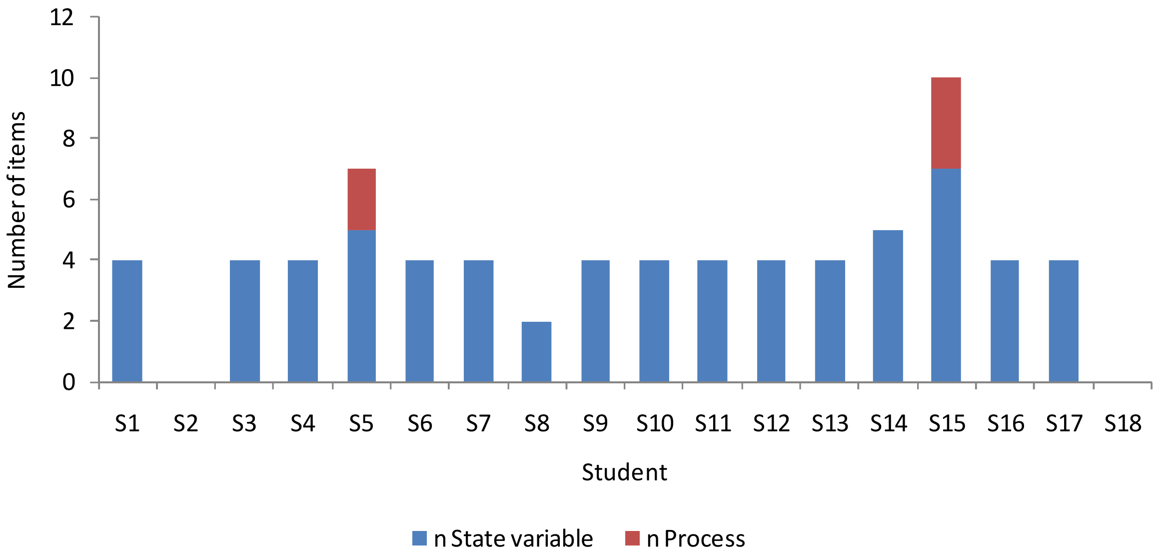


Figure 9 Implementation of state and process concepts as quantities in an open-ended modelling activity. Students S2 and S18 failed to implement any quantities in their models.

4.4. Discussion

Q1: How often do naive modellers make errors in implementation of their model ingredients?

The results shown here indicate that the vast majority of naive modellers quickly pick up on the structural aspects of implementing a model. Few made errors in their implementation of entities and quantities. The errors that occurred with configurations generally occurred due to the attempted inclusion of behavioural concepts in the configurations. This may however be a hangover from students having previous worked with concept mapping and resulting in confusing configurations with concept map relationships. This is an issue that can quickly be solved once students clearly understand how behaviour is represented with qualitative conceptual models. However, it is apparent that naive modellers struggle to understand the representation of causality in learning space 2. The majority of students failed to implement any concepts correctly with causality within their models.

Although a fair amount of the errors in causal implementation were due to the students naivety in modelling, some issues remain with the implementation of causality in learning space two. Although the positive and negative relationships available are meant to be general at this level, they have very specific meanings (that of proportionality) when it comes to simulation and causal explanation of the model. Given this care must be taken when building and simulating basic causal models (LS2 and LS3) such as not to introduce inconsistencies in causal explanations. However, this also provides an opportunity for learning activities to introduce and explain causal differentiation to students. The current implementation may not necessarily allow for simulation approaches as previously used in qualitative reasoning but it provides ample opportunity for development of diagnosis and step by step causal explanations from the virtual characters. Use cases and learning strategies should be developed to exploit this.

Q2: How easily do naive modellers identify the important components that need to be included in a causal model to represent a scientifically accurate causal argument?

It is apparent from this study (and the pilot study) that the students did not identify, or where not able to implement, all the key domain concepts required for a full causal explanation in their models. Whilst this is not a problem that is unique to learning by modelling it is an important step in the evolution of a qualitative model (Bredeweg et al. 2008b). This is especially true where students are working fully in a modelling building mode. Whilst DynaLearn does not necessarily need to identify a solution to this there are aspects of the software that can be developed to address this need. Firstly, the pedagogical strategies, curricula and modelling modes can be developed such that modes such as model evolution and model exploration are used where students need to have the key concepts highlighted to them. Secondly, the personified ontology based feedback could be used to direct students about concepts they should be considering when building a model on a particular topic. These technologies and potential strategies are still under development.

Q3: How variable is model complexity between students in terms of number of different ingredients?

The models developed by the naive modellers showed great levels of variability in terms of complexity. However, in this task the majority of this related to the numbers of configurations and causal relations used rather than the number of entities and quantities used. In the pilot study students exhibited a large variation in the number of quantities used so it may be that complexity in terms of structure will be context and domain dependent. The level of variability in the implementation of causality probably relates in this case to the different students understanding of what they were required to do and how causal representation works within DynaLearn. Therefore, more confident students who gain an understanding quickly of how the system works are more likely to have greater time within a modelling session to enter more complexity into their models. It is unlikely that the models here represent how much the student actually understands of the domain; rather that it represents how much they understand of how they can represent it in the software. To be able to adequately measure an individual’s propensity for developing complexity in a model would require that all the subjects had an equal understanding of the modelling approach. However, it is clear that some students can easily prepare more complex models than others (even if much of what they enter into the model is incorrect). Therefore, teaching strategies and activities (plus inbuilt functionality like help/diagnosis) need to be developed to cope with this.

Q4: Do naive modellers work at appreciably different rates during a modelling session?

It was apparent from verbal communication and observation of the students that they worked at vastly different rates. Amongst the naive modellers four groups could be indentified: those that carefully read the instructions and carefully went about model building; those that skipped the instructions and quickly intuitively started to model; those that skipped the instructions and struggled to model and those that read the instructions and still struggled to model. As such by the time some students had just started to add their first entities to the model others had almost completed their model structure. Whilst this again is not unique to learning by modelling or the DynaLearn system it does pose significant issues to the development of learning activities and lesson planning. The early stages of learning by modelling carry a large overhead for a teacher providing feedback and support to a student; encouraging and facilitating the slower students whilst identifying new activities to keep the faster students motivated and engaged. This is an aspect of learning that could be vastly supported by the use of the virtual characters and the internal help and feedback support. However, suitable teaching activities must be developed to make best use of the tools.

5. Evaluation Activity 3b – Influence of Teachable Agent mode

5.1. Introduction

The teachable agent (TA) mode is one of the strategies developed using the interactive virtual characters (VCs) to focus, direct and motivate a student's modelling activities. Through the use of a teachers norm model in the background and interactive feedback generated in a question/answer style the TA mode is able to direct students in the model ingredients to use and the simulation of behaviours they generate in the given system. The use of the VCs in this mode aims to be beneficial to both the motivation of the students and their understanding of the model behaviour by directing them through sequence of "what if?" style model simulation questions.

Evaluation hypothesis

The TA mode promotes better model building, conceptual understanding and causal explanations than self-directed model simulation by directing the model exploration and simulation using an "if then" question approach.

Evaluation questions

1) Does the TA mode give students a better understanding of the system behaviour?
2) Does the "If, Then" question style of the TA contribute to the way students build causal arguments in the written text?
3) Does the TA mode contribute more to students building better, more correct, models than just through pre-defining the ingredients that should be included in the model?

5.2. Methods

5.2.1. Population/participants

Evaluation activity 3b was undertaken as a workshop within the normal time-tabled activities of students studying for a Post Graduate Certificate in Education (teacher training). The potential class comprised 37 students studying science education at secondary schools and they came from a back ground of undergraduate degrees in biology, chemistry and physics. Despite the workshop falling within their usual time-tabled sessions, attendance wasn't complete and the study involved 18 students. The content focus aimed at GCSE level because this was the level at which the students were currently focussing.

5.2.2. Implementation procedure

This activity was part of a 3hr learning by modelling workshop held in mid-November 2010. The class of students was randomly split into two groups for the evaluation activity.

Experimental group

The experimental group were asked to complete a modelling exercise using the TA mode. A TA mode model file was provided to them with instructions on how to use it. Students were asked to save and submit a number of different versions of their model file at different stages in the activity:

1) The first version of the model built using the pre-defined ingredients, prior to exploring the model behaviour by asking the VC questions (Model B1).

2) If the student made changes to the model after asking the student character questions about the model then the changes were saved and submitted as a new model (Model B2).

3) The final version of the model including any changes made after the student VC has undertaken the challenge(s) from the quiz master VC (Model B3).

Control group

The control group were given a printed list of modelling ingredients (identical to those in the TA model file) and asked to use them to build a model and simulate it using the standard Learning Space 2 scenario/simulation options. The control group were asked to submit models at two points:

1) The initial model produced prior to simulation (Model A1).

2) Any changes made to the model after simulation (Model A2).

After the modelling activity students were asked to repeat the written test.

All modelling activities were undertaken using DynaLearn beta prototype version 0.6.16.

5.2.3. Design of data collection instruments

Written test

The students were given a short answer question about osmosis and diffusion from a GCSE biology exam (Appendix J). The exam question required them to describe and explain causally the reason for the observed phenomena. The same question was administered as a pre and post test to identify if the students changed either the content or style of their answers.

Model instruments

The Learning Space 2 models created by the students were collected as an instrument to measure what they had management to implement in a model within the allotted time. The different versions of the models, collected at different times during the modelling activity were required to enable an assessment of how easily students were able to build a model when given pre-defined ingredients and whether (and how much) simulating the model caused them to identify errors and make changes.

5.2.4. Data analysis and variables measured

Written texts

The written texts were analysed for their complexity and accuracy in identification of the key causal concepts and arguments explaining the phenomena in the question. Complexity was measured by the number of individual statements made and accuracy by the number and proportion of valid statements made (compared with a predefined list of causal concepts). In addition to checking for validity causal arguments were also defined by whether they were complete (if, then, because aspects all included) or incomplete (because missing or the argument was non-sequential and missed an important concept out) (Appendix L).

Model analysis

Student models were assessed for completeness and accuracy in terms of the correct implementation of all the structural ingredients (entities, quantities and configurations) and the causal dependencies. Models at each stage of the activity were compared within and between experimental groups.

5.3. Results

5.3.1. Model instrument analysis

The models initially built by students from a predefined set of ingredients in learning space 2 (TA mode or control group) all correctly implemented the entities and configurations available to them. However, many students decided not to use all the quantities available to them. Generally, this included using either “*Amount of chloride ions*” OR “*Concentration of chloride ions*” but not both as used in the teacher’s model in TA mode. Three of the students also incorrectly assigned some of their quantities to the wrong entity (although this currently has no influence over the TA mode).

With ten of the students missing some of the quantities out this meant that many of the required causal dependencies were missing in their models (Figure 10). The majority of errors by students in causal dependencies were due to them not being implemented, either through missing the quantity out in the first place or through just missing out the correct dependency. Additional errors came from implementing either conceptually incorrect dependencies or dependencies that although correct at a certain level of granularity would be considered non-sequential in the teacher’s model (linked to questions asking whether two quantities have a direct relationship).

Only student S10 implemented their model completely correctly prior to using the questions/simulation of a model. Although there were differences between students it would appear that students using the TA made marginally more errors in the implementation of their first model (Figure 10). However, only three out of seven of the TA mode students made non-sequential causal dependency errors compared with five out of six of the control group.

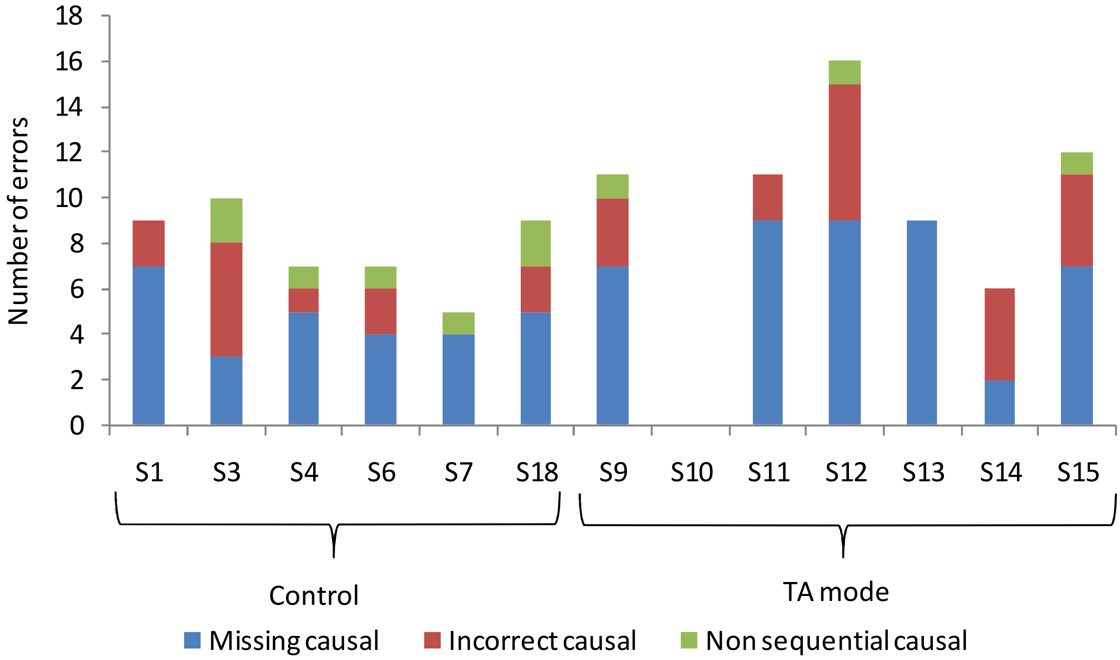


Figure 10 sources of errors in use of causal dependencies by students when given a pre-defined list of model ingredients. Errors assessed prior to the students undertaking model simulation/behaviour exploration in either teachable agent mode or using standard learning space 2 simulation (control). Note the model created by S10 was correct prior to simulation.

Of the seven students who submitted before and after models for the TA mode experiment there were no discernable patterns in terms of them improving their models. S10 did not need to make any changes as their model was initially correct whilst S12 and S13 did not make any changes to their models (Figure 11). Indeed S11 introduced more causal errors in their altered model.

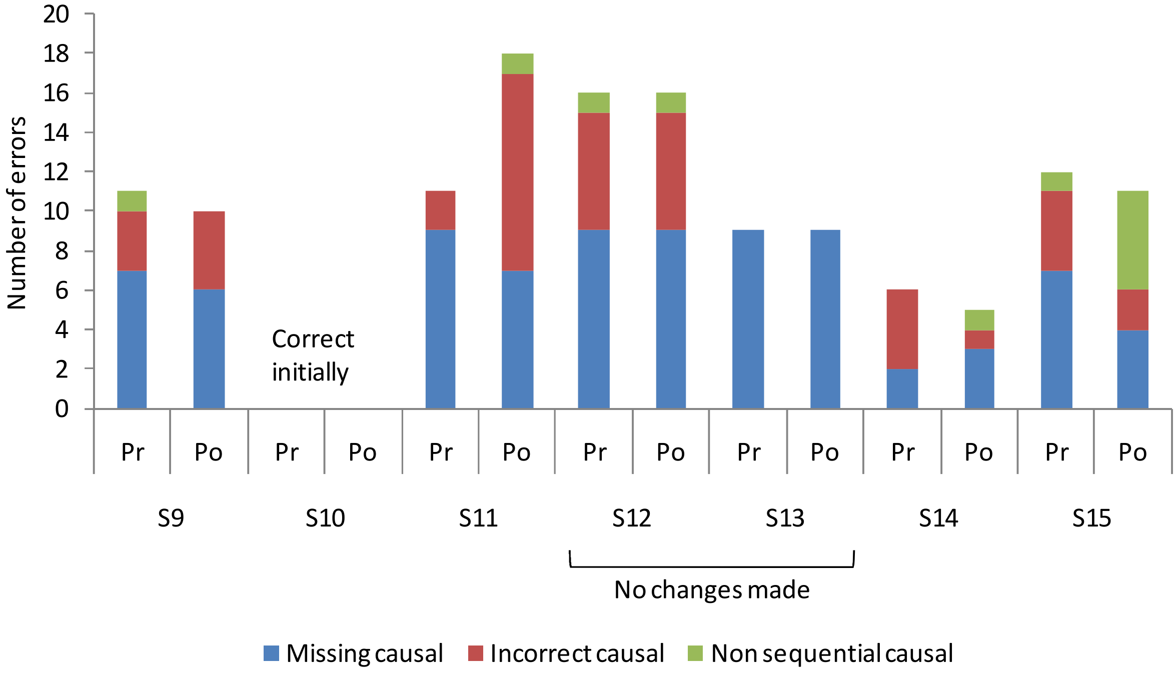


Figure 11 Analysis of the models of the seven students who submitted pre- and post- models for the teachable agent assignment looking at the sources of errors made in implementing causal dependencies. S10's model was complete and correct prior to asking the agent questions and sending the agent to the challenge. S12 and S13 did not make any changes to their model after the challenge.

5.3.2. Text instrument analysis

The written answers provided by students were generally short and lacked structure and clarity. In general the students failed to identify a large number of the key concepts or causal steps that were predefined as being required for a complete explanation. The majority of students mentioned either the amount of chloride ions or the concentration but rarely both. Again, students generally mentioned the solute concentration gradient or a concentration gradient of water but rarely both, and only occasionally was the concentration gradient of water related to the concept of water potential. Causal explanations were generally vague and incomplete with most causal statements being undefined or non-sequential.

Although students in the control group appeared to include a few more concepts (Figure 12) and statements (Figure 13) in their post test than their pre test, there were no significant differences between the pre/post tests or the treatments. However, it must be noted that this was based on only 7 students for the control group and 6 students for the TA mode treatment (Table 3).

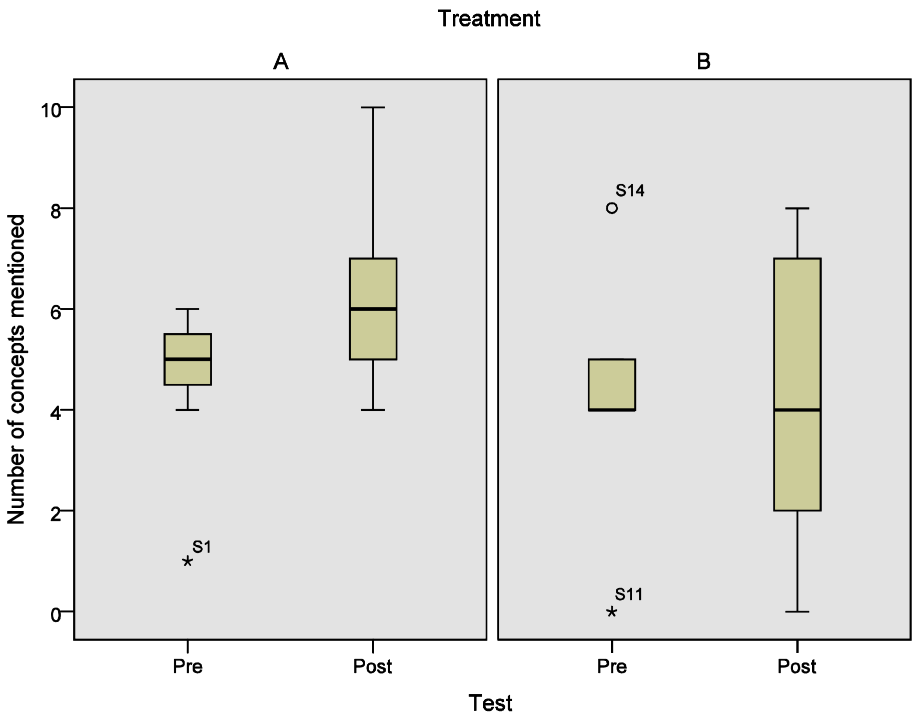


Figure 12 Box plot of the total number of key concepts mentioned by students in the pre and post written tests for the control group (A) and the students using the Teachable Agent mode (B).

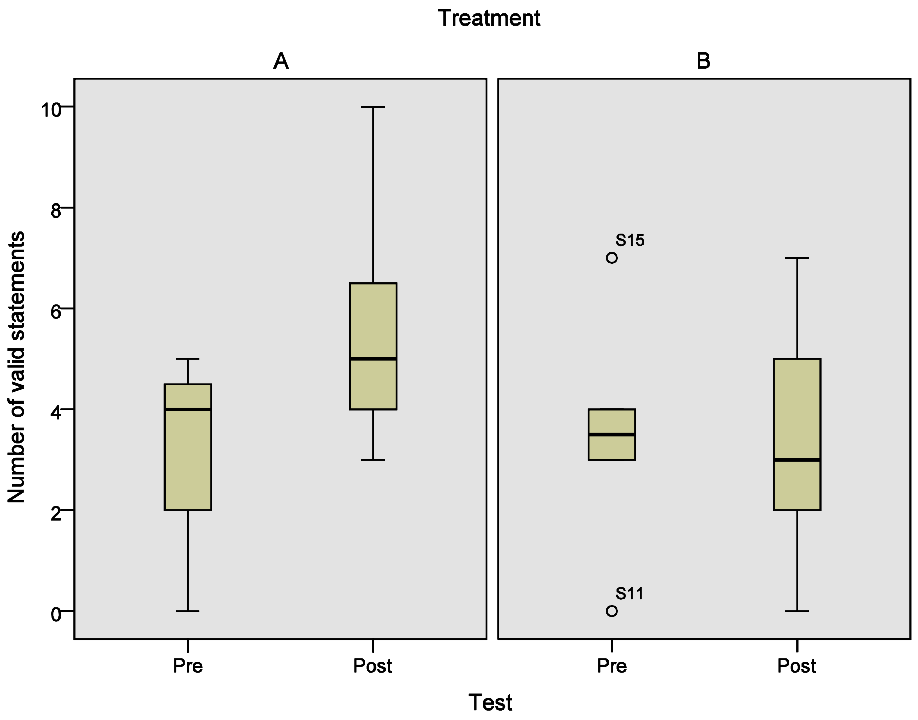


Figure 13 Box plot of the total number of valid statements mentioned by students in the pre and post written tests for the control group (A) and the students using the Teachable Agent mode (B).

Table 3 ANOVA results of comparison between treatment and control group pre and post modelling written answers to the GCSE exam question.

		Sum of Squares	df	Mean Square	F	Sig.
Number of concepts mentioned	Between Groups	20.575	3	6.858	1.229	0.323
	Within Groups	122.810	22	5.582		
	Total	143.385	25			
Number of valid statements	Between Groups	26.441	3	8.814	1.772	0.182
	Within Groups	109.405	22	4.973		
	Total	135.846	25			
Number of undefined causal statements	Between Groups	6.592	3	2.197	2.925	0.056
	Within Groups	16.524	22	0.751		
	Total	23.115	25			
Number of non-causal statements	Between Groups	0.881	3	0.294	0.157	0.924
	Within Groups	41.119	22	1.869		
	Total	42.000	25			

5.4. Discussion

Q1: Does the TA mode give students a better understanding of the system behaviour?

The analysis of the written texts did not reveal any significant differences between the treatment and control groups in terms of the total number of concepts they identified or in terms of the number of valid statements they made in their answers. Additionally, no significant effect of the treatment or the control was observed between the pre and post tests. However, it should be noted that this analysis was limited by the total number of students successfully completing and submitting all the task documentation. It is possible that any potential effect of the TA mode treatment (or indeed the control modelling activity) was masked by the activity being based on a topic that the students should have already known. Therefore, it is possible that their belief that they already knew and understood the topic precluded them from altering their response in any great detail between the pre and post test. However, it is apparent that this effect was the same for both groups and that they generally did not show much of a response to being confronted by models that contained more information that they had included in their pre-test responses. It is therefore possible that the length and design of the modelling activities were inadequate to stimulate the students in terms of the understanding of the full causal explanation for the problem and it's applicability to their written answers.

Q2: Does the "If, Then" question style of the TA contribute to the way students build causal arguments in the written text?

It was hypothesised that the structured "if, then" style of questioning and model behaviour exploration of the TA mode would give students clear scaffolding towards building logical causal arguments and expressions that they may not gain from open-ended simulation of complete models in LS2. It was anticipated that this might translate into the way they structured their written answers. However, no difference was observed between the pre and post tests for both the treatment with the structured exploration of behaviour and the control with open-ended model simulation. It is most likely that a single exposure to this learning mode is insufficient to stimulate such a response and transference of skills into their written answers. Therefore, it could be expected that over time and repeated modelling activities the effect could become apparent. The structured model exploration utilised by the TA mode is therefore worth pursuing and also transferring to other modelling activities and pedagogical approaches (use cases) for students exploring models.

Q3: Does the TA mode contribute more to students building better, more correct, models than just through pre-defining the ingredients that should be included in the model?

The models created here by both the control group and the group using the teachable agent mode do not show any indication as to whether students built better models in the TA mode. Students from both sets left out quantities from the pre-defined lists and therefore had incomplete models in terms of the causality they represented. Additionally both groups exhibited errors in their implementation of causal dependencies and if anything the group using the teachable agent tended to introduce slightly more errors that those working independently. Of the students who submitted models from before and after their question/answer interaction with the agent and the agent's interaction with the challenge none clearly showed an improvement in their model. Indeed, one student went on to introduce more errors into their causal model. However, much of this may relate to the students naivety in terms of what causal dependencies represent and how to implement them rather than a lack of domain knowledge or understanding. It is also possible that the students were unclear how to go about improving their models once their agent got answers incorrect. This is something that needs to be handled by the virtual characters during the teachable agent activities. It is possible that some other form of diagnosis, help or feedback could be introduced here to help students identify where their knowledge or implementation is lacking or incorrect.

6. General discussion

6.1. Conceptual understanding

The current study indicates that in the short term the modelling tasks used in the evaluations did not alter or improve students understanding of the topics, even when they were topics that the students should already be well grounded in (e.g. osmosis or photosynthesis). It is possible that either the students did not really understand these topics (current educational approaches are deficient at promoting systems and causal thinking) or that the students were distracted or even confused by attempting to comprehend the modelling approach. It could be anticipated that students learning by modelling will potentially move through three phases as they are introduced to the DynaLearn software. In these three phases students could be expected to:

- Fail to gain a greater conceptual understanding of a topic because they are focussed on understanding (or failing to understand) conceptual modelling (***Novice***).

- Gain a greater conceptual understanding of a topic by learning how to model directly using domain knowledge provided to them (***Apprentice***).

- Gain a greater conceptual understanding of a new topic by refining and consolidating information using a modelling/systems thinking approach (***Master***).

Therefore, it is probably unrealistic to expect a rapid response with students who are working at the novice or apprentice level. Qualitative modelling and systems thinking will be a completely new concept to them and as such it will take time for educational benefits to become apparent. Therefore, ideally the learning by conceptual modelling approach would be holistically grounded within the science curriculum at an early stage to enable students to move from novice to master alongside the development of understanding of domain knowledge, working from simple to more complex topics as their science education progresses. This requires that the learning by conceptual modelling pedagogy and curriculum identifies suitable structures and learning activities through which students can move from novice to mastery in application of qualitative modelling and systems thinking to learning. It is possible that the transition from novice to mastery can be enhanced by the application of learning activities utilising the different modelling modes of model exploration, evolution and creation such that they learn about the modelling approach starting with complete models, working through incomplete models to a point where they are able to model a new topic from first principles (Figure 14). This would enable the development of learning by modelling (qualitative reasoning and systems thinking skills) alongside the development of conceptual understanding, whilst at the same time the modelling approach, its ingredients and vocabulary, are fully grounded within a domain context. The development of these reasoning skills is discussed in the following section.

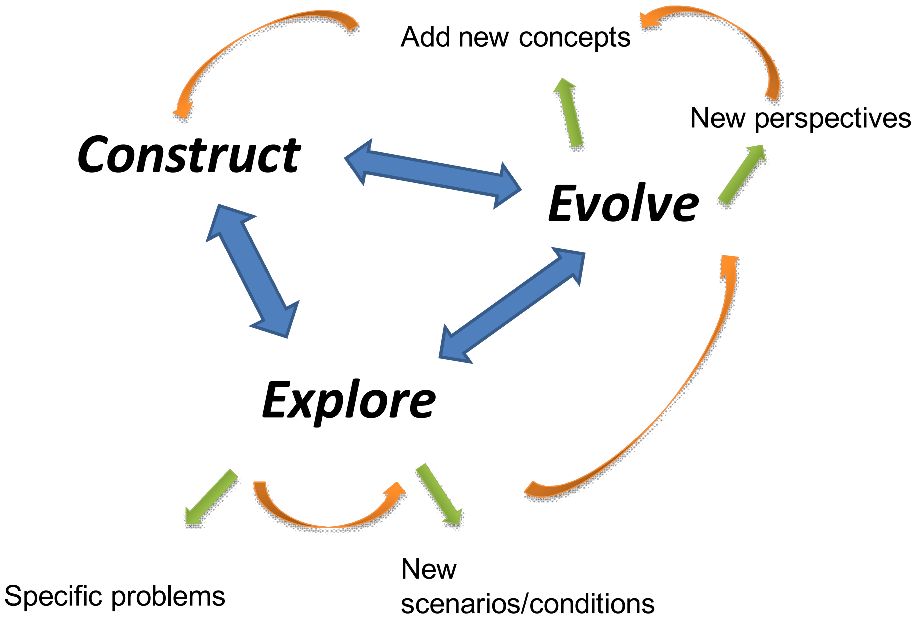


Figure 14 different modelling modes that could be utilised as learning activities within the DynaLearn software and curricula to facilitate the transition from novice to modeller in terms of modelling and learning conceptual information.

6.2. Reasoning skills and causality

The results of the studies here indicate that the development of better reasoning skills will take more time than was available in the current settings. As such, short scale evaluations are unlikely to show clear development of these skills. Therefore, it is probable that reasoning skills, in the context of application of systems thinking and qualitative process representations of causality to natural thought processes, spoken and written reasoning will only develop once the student has been able to fully understand the language and approach of qualitative modelling. Naive or novice modellers will always face a period of learning where they feel are primarily learning about modelling rather than the domain concepts they are exploring. It is apparent from the studies reported here that students initially struggle to implement and explain concepts using this approach, and that in most cases it was the implementation of the causal information that proved the most difficult. Therefore, the DynaLearn approach must provide some level of scaffolding to support learners through the development of these skills. This could be achieved using suitable modelling curricula, learning activities (specific pedagogy and use-cases) and the application of the value-added technologies in the virtual characters and the ontology-based model feedback. Indeed, specific learning activities and use cases should be fully developed and explored to optimise the use of these technologies within a structured learning activity.

The development of reasoning skills around causality and qualitative process theory for novices working in the DynaLearn software will come mostly from the simulation and exploration of model behaviour (exploration modelling mode). As such the development of the "diagnosis" mode and the feedback given by interactive virtual characters, such as the teachable agent, will have a great contribution to teaching the modeller about causal reasoning within the software. At the moment there are no easy avenues for a student to explore within the software to learn about the built in reasoning logic. This is something that the inbuilt help and diagnosis modes will need to address. Additionally, this is an area where well defined use-cases and appropriate modelling activities (advanced lesson plans and modelling curricula) need to be developed to optimise learning whilst using the software.

Furthermore, the structured use of the different learning spaces can be used to enhance qualitative systems thinking and notions of causality. The application of reasoning logic implemented in the different learning spaces could be explored to provide learning activities specifically geared around developing reasoning skills and learning about the reasoning logic used by qualitative conceptual modelling. In particular the DynaLearn environment encompasses three main types of causal information:

- Basic causality between quantities (LS2 and 3).
- Qualitative process theory (Direct influences, Proportionalities) (LS4 – 6).
- Conditional information:
    - Quantity spaces and thresholds (LS3 and above);
    - Correspondences and (in)equalities (LS3 and above);
    - Condition and consequence representations (LS5 and 6).

There is a major transition in the software between the notion of basic causality when students merely represent notions of positive and negative relations between quantities (although it should be noted that these relations are specifically proportionalities within the internal reasoning of the software) and causal differentiations where students implement notions of processes, direct influences and proportionalities between quantities. Therefore, the appropriate use of the different learning spaces and the activities used as transitions between these will have great importance for students developing reasoning skills based around qualitative process theory and systems thinking. Of particular importance will be the development of learning activities and use-cases for the implementation of models using basic causality in LS2 and LS3 and the transition from that to causal differentiation at LS4. The development of modelling activities using the basic notions of causality is important as the specific representation of the positive and negative relations as proportionalities means that only some causal ideas can actually be correctly represented and simulated in these learning spaces. This has implications for expert models created at these levels (WP6) and for the learning activities undertaken within them. This also has implications as naive modellers will have very general views of what constitutes a positive and negative relationship and what it means for their reasoning and yet within the software they have very specific meanings for the internal reasoning logic of the software. However, this does also provide opportunities to stimulate students to explore their notions of causality in more detail (including causal differentiation, ambiguity and inconsistencies) and as such learning activities and use-cases can be developed to exploit this.

6.3. Motivation

The studies undertaken here did not specifically focus on the motivation of students whilst using the software. The answers given by the students indicate that they found it an interesting and challenging activity and some of them indicated that they found modelling a motivating activity (even without any of the added value technology such as the virtual character interaction). From the verbal feedback given by the PGCE students it was clear that many of them struggled to see how the approach might be applicable to them in their current teaching practice. However, those that experienced the teachable agent mode and virtual character quickly identified how interactive modelling activities could be developed. However, many pointed out learning activities would need to be captivating, engaging and flexible so that students of different ability could be handled within the same activity. They highlighted the need to support the weaker students so they were not put off the approach whilst at the same time keeping the more advanced students active and motivated to work when they finished the initial tasks they had been set. This is something that needs to be addressed using the curricula, pedagogical strategies and the implementation of learning activities through interaction with the VCs.

6.4. Software issues

The evaluation studies reported here did not highlight many issues that related directly to the software usability or the design. Most students appeared to identify and work with the icon system and the functionality of the tools fairly easily. The two main features that were requested were the addition of an “undo” button and a change in the default setting of the “add quantity” and “add configuration” screen. Currently, the default setting when clicking on either of the two “add” functions opens the dialog boxes in a setup where the user is editing an existing ingredient rather than adding a new ingredient. This led to numerous instances of students overwriting existing ingredient definitions.

As the software is still a beta prototype with constant changes and updates of functionality it suffers from some level of instability. As such the evaluation exercise was also used to identify bugs and issues that caused the software to crash. Although using software that risks crashing during an evaluation exercise is undesirable in terms of collecting data and promoting the software to potential stakeholders (e.g. the group of trainee teachers) it is unavoidable in this development phase. What is of note is that some students seemed to suffer from more inexplicable crashes than others and it is likely that some naive users attempt to do things in the software that a developer would never try to do (mostly through knowing how something is meant to be done). Therefore, testing of this sort with naive users is probably essentially for testing software functionality and stability.

One final aspect of the software that was an issue during the evaluation was the successful installation of the different component software required to run the DynaLearn system. This required support from the Universities IT department to test and integrate the packages onto student image computers which have severely restricted local access rights. This also required settings to be adjusted in the university firewall settings. Whilst this is not necessarily something that is the responsibility of the DynaLearn project to address it is important that we make installation as simple as possible in the future and endeavour to fully integrate the implementation of the software.

6.5. Stakeholders

In addition to obtaining evaluation of the DynaLearn approach from students it is also important that evaluation and feedback is sought from other stakeholder such as teachers and education coordinators who would be responsible for teaching with and implementing the approach within their classrooms and overall curricula. The evaluations here using PGCE students and working with the University of Hull Centre for Educational Studies provided some insight into the requirements and thought of teachers. However, in most cases they initially struggled to identify the potential use of the approach in schools. However, the feedback given by the students who used the teachable agent mode indicated how this structured learning activity gave them a clearer indication of what could be done. This is an area of study that needs developing and further work with different stakeholders is required. The links developed with the Centre for Educational Studies will provide an avenue for working with stakeholders including local and national teaching organisations.

7. Conclusions and Future plans

7.1. Conclusions

The evaluations presented here highlight three main conclusions:

- Students initially have problems understanding and representing causal relations – the curricula, learning activities and help/feedback need to be structured in a way to scaffold students in learning the modelling approach alongside learning domain knowledge.
- The DynaLearn software comprises many different technological components and many gives rise to many different opportunities for learning activities – future work is required to identify a clear suite of learning activities, use cases and lesson plans that can exploit these as an integrated approach.
- As yet the evaluations activities have focussed on the conceptual modelling component of the software and its effects on conceptual understanding and reasoning skills. So far the use of the virtual characters and the semantic technology has received minimal attention (mostly due to the implementation schedule) – further evaluation of the software should focus on these aspects to provide insights into the effectiveness of integrated learning activities utilising these components.

7.2. Evaluation opportunities

The researchers at the University of Hull have identified four possible settings that could be used for future evaluation activities working towards D7.3.2:

1) Small groups of volunteer students from within the Department of Biological Sciences undertaking short evaluation studies outside of normal classes.

2) PGCE students from the Centre for Educational Studies may be available to undertake short evaluation activities with the potential that they could integrate it with some of their class-based activities within the university and within their teaching practise.

3) Researchers at UOH are exploring the opportunities to run short workshop sessions at the University with groups of local science teachers and A-Level students.

4) The staff at the Centre for Educational Studies are involved in delivering a number of summer schools and may be able to offer evaluation opportunities within these.

These activities would need to be implemented in the spring semester at the University prior to summer examinations (May/June) and the closure of the university for the summer (mid-June). Activities with summer schools may be available in August.

References

AQA (2008) General Certificate in Secondary Education, Additional Science Unit Biology B2 Exam Paper, Foundation Tier, January 2008.

AQA (2009) General Certificate in Secondary Education, Additional Science Unit Biology B2 Exam Paper, Higher Tier, June 2009.

Bredeweg, B. *et al.* (2008a) DynaLearn – Engaging and informed tools for learning conceptual system knowledge. DynaLearn, EC FP7, STREP Project no. 231526, Description of Work (DOW).

Bredeweg, B. and K. Forbus. (2003) Qualitative Modeling in Education. *AI Magazine*, **24**(4), 35-46.

Bredeweg, B., Linnebank, F., Bouwer, A. & Liem, J. (2009a) Garp3 - Workbench for Qualitative Modelling and Simulation. *Ecological Informatics* 263-281.

Bredeweg, B., André, E., Bee, N., Bühling, R., Gómez-Pérez, J.M., Häring, M., Liem, J., Linnebank, F., Thanh Tu Nguyen, B., Trna, M. and Wißner, M. (2009) Technical design and architecture, DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D2.1.

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. (2008b). Towards a structured approach to building qualitative reasoning models and simulations, *Ecological Informatics*, **3**(1): 1-12.

Forbus, K.D. (1984) Qualitative Process Theory. *Artificial Intelligence* **24** 85-168.

Gracia, J., Trna, M. Lozano, E., Tu Nguyen, T., Gómez-Pérez, A., Montaña, C. & Liem, J. (2010) Semantic repository and ontology mapping. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D4.1.

IPCC (1996) Climate Change 1995: The Science of Climate Change. Contribution of WGI to the Second Assessment Report of the Intergovernmental Panel on Climate Change. J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg and K. Maskell (eds.), Cambridge University Press, Cambridge, UK, 572 pp.

IPCC (1997) An Introduction to Simple Climate Models used in the IPCC Second Assessment Report. IPCC Technical Paper II. ISBN: 92-9169-101-1.

Liem, J., Beek, W. & Bredeweg, B. (2010) Multi use level workbench. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D3.1.

Lozano, E., Gracia, J., Gómez-Pérez, A., Liem, J., van Weelden, C. & Bredeweg, B. (2010) Ontology-based feedback on model quality. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D4.2.

Mioduser, D., Salles, P., Noble, R., Zitek, A., Benayahu, Y., Zurel, D., Lieba, M., Zuzovsky, R. & Nachmias, R. (2010) Lessons and assignment schemata. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D7.1.

Noble, R. (2010) University of Hull – Basic topics and models. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D6.2.2.

Nuttle, T., & Bouwer, A. (2009) Supporting education about environmental sustainability: Evaluation of a progressive learning route for qualitative models. *Ecological Informatics*, **4**, 396-404.

Salles, P., Assumpção Costa e Silva, P., Gontijo de Sá, I., Noble, R., Zitek, A., Uzunov, J & Mioduser D. (2009) DynaLearn environmental science curriculum requirements. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D6.1.

Wißner, M. Häring, M., Bühling, R., Beek, W., Linnebank, F., Liem, J., Bredeweg, B. & André, E. (2010) Basic help and teachable agent. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D5.3.

Wißner, M. Häring, M., Mehlmann, G., Bühling, R., Milosevic, U., Liem, J., Linnebank, F., Beek, W., Bredeweg, B. & André, E. (2010) Basic tutorial tactics for virtual agents. DynaLearn, EC FP7, STREP Project no. 231526, Deliverable D5.2.

Appendix A – Written question sheet used in the pilot study

Problem

Human induced emissions of greenhouse gases are contributing to global climate change, acting through altering the global atmospheric heat balance. These changes may have consequences for the survival and distribution of species sensitive to changes in thermal regime.

1) Describe the key concepts involved in the problem and phenomena (described above) and the relationships between them.

2) Identify and describe the main processes that control the phenomena and would influence the consequences of the stated problem.

3) Predict the consequences of the two following scenarios on a thermally sensitive species:

- **increasing emissions of greenhouse gases**
- **decreasing emissions of greenhouse gases**

and describe the causes/mechanisms for these consequences

4) For each of the scenarios given in question 3, and for each of the consequences you have predicted, describe under what conditions they may occur. Also identify any conditions under which each scenario may result in alternative consequences.

Appendix B - Norm concept map provided to students (Pilot Study)

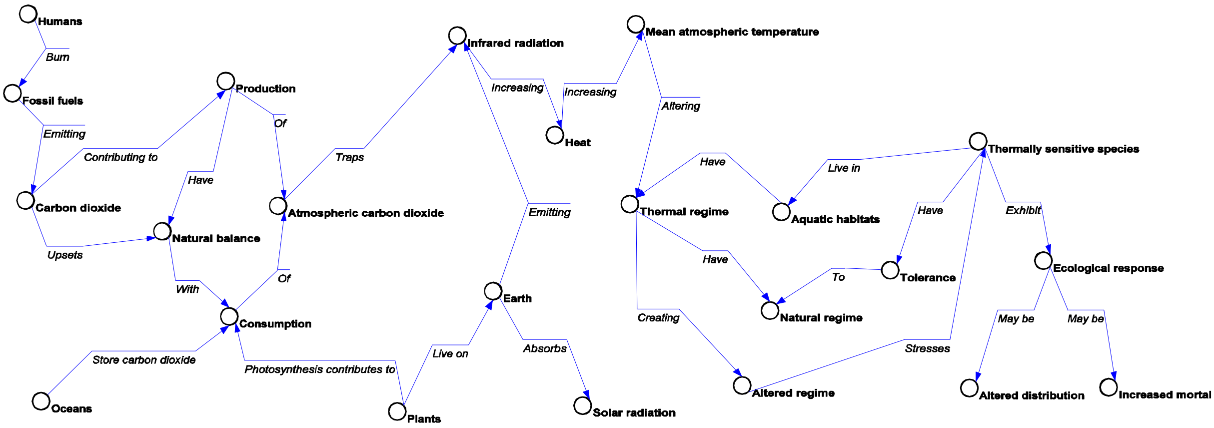


Figure A1 Norm concept map of the climate change problem used in the pilot study to direct the students modelling in the LS2 activity and to form the basis for defining the prior list of key concepts.

Appendix C - Norm LS2 model provided to students in pilot study

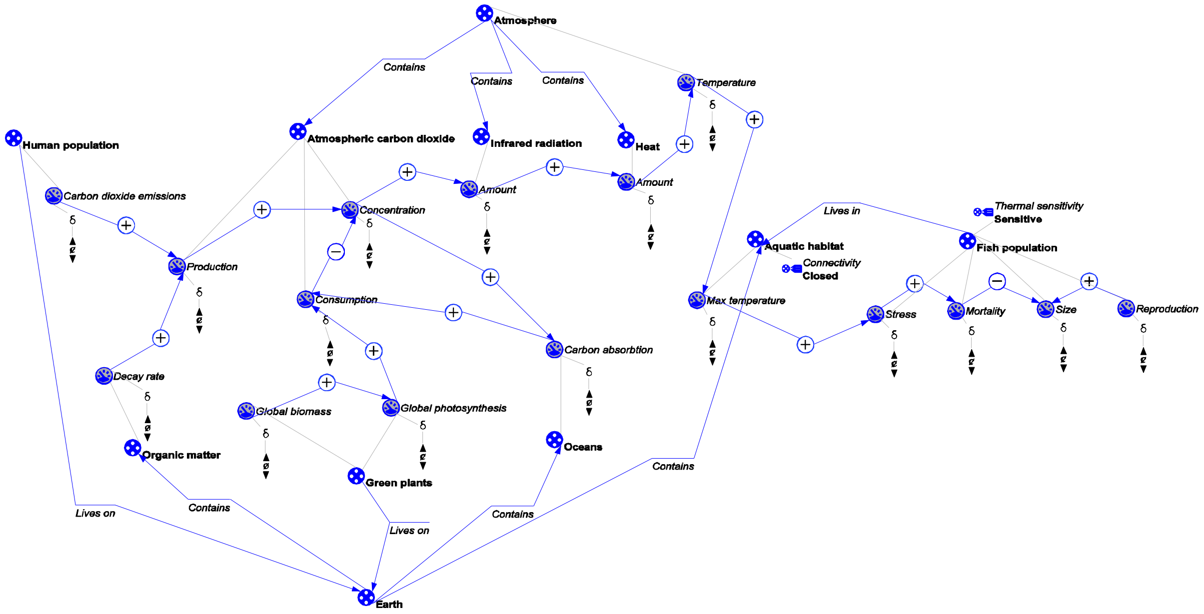


Figure A2 Norm basic causal model (LS2) of the climate change problem used in the pilot study to direct the students modelling in the LS4 activity.

Appendix D – Questionnaire used in the pilot study and student answers

Section 1

Question To what extent do you agree/disagree to the following statements:	Agree strongly	Agree	Neutral	Disagree	Disagree strongly
1) Conceptual modelling was a totally new approach to learning for me	**2**	**2**			
2) The modelling approach gave me new insight into the phenomena of climate change	**1**	**1**	**2**		
3) Building a model helped me focus on which concepts were important for addressing the stated problem	**1**	**2**		**1**	
4) Being able to simulate the models helped me to develop my understanding of potential system behaviour	**1**	**3**			
5) Modelling was a motivating task	**1**	**2**	**1**		
6) Building conceptual models made me think more clearly about the causes and effects in environmental systems	**1**	**1**	**2**		
7) The process of modelling motivated me to learn more about the phenomena		**1**	**2**	**1**	
8) I found the modelling approach interesting	**2**	**2**			
9) I clearly understood the goal of the modelling tasks	**2**	**2**			
10) Working on the modelling task helped me focus my written answers	**1**	**1**	**1**	**1**	

Section 2

Question How easy or difficult did you find the following activities?	Very easy	Easy	Easy/difficult in parts	Difficult	Very difficult
11) The overall modelling activity			**3**	**1**	
12) Identifying and extracting the relevant and key information from the text		**2**	**1**	**1**	
13) Drawing a concept map		**1**	**2**	**1**	
14) Identifying and describing entities and quantities in the system			**2**	**1**	**1**
15) Identifying and representing the basic causal relations in the model		**1**	**1**	**2**	
16) Simulations		**2**	**1**	**1**	
17) Differentiating between direct influences and proportionalities to be		**1**		**1**	**2**
18) Defining the quantity spaces		**1**		**3**	
19) Identifying and understanding the reasons for ambiguous behaviour in simulations was		**1**	**1**	**2**	
20) The functionality of the software		**1**	**3**		

Numbers indicate the number of students who gave that answer.

Appendix E – Stimulus questions for student logs in the pilot study

Please take some time to reflect on your experiences during each session, and record them in a diary to be submitted anonymously at the end of the study.

Please consider the following types of experiences:

- What was new to you?
- What was similar to things you had done before?
- Did the task make you think in a different way?
- What did you find easy?
- What did you find hard?
- Was there anything you hadn't grasped by the end of the session?
- How easy did you find the software to use?
- Are there any examples of where the modelling approach helped you understand the concepts?
- Are there any examples of additional questions that the modelling approach raised in your mind about the concepts covered?

Appendix F – Marking schemes used for pilot study instruments

List of key concepts

Key concepts	Scoring
Anthropogenic source of greenhouse gases Greenhouse gases Natural equilibrium Carbon capture - oceans Carbon capture - photosynthesis Organic decay as natural source of CO_2 Greenhouse gases concentration change Solar radiation as energy source* Infrared radiation capture by GH gases Increased OR altered radiation/heat balance Change in atmospheric temperatures Climate change Relationship between climate and habitat Species tolerance OR adaption* Ecological response due to stress Shift in species distribution Shift in species mortality / Extinction	***Concept maps***
	0 - Missing/incorrect/unclear 1 - Included partially 2 - Included clearly
	Basic causal models
	0 - Missing/incorrect/unclear 1 - Included partially 2 - Included causally

Concept maps

Component	Variables
Layout	Web Hierarchy
Concepts	n Total n Structural n Process n Change n Integrated
Relationships	n Total n Structural n Process n Change n condition/causal
Completeness	Complete Incomplete (checking for evidence of implementation mistakes)

Basic Causal Models (LS2)

Ingredient	Variables
Entities	n Correct n Incorrect - quantity n No assigned quantity n Total
Quantities	n Correct n State variable n Process n Incorrect - entity n Incorrect - assignment n Total
Configurations	n Correct n Incorrect - concept n Incorrect - implementation n Total
Causal dependencies	n Correct n Incorrect - concept n Incorrect - implementation n Total

Written texts

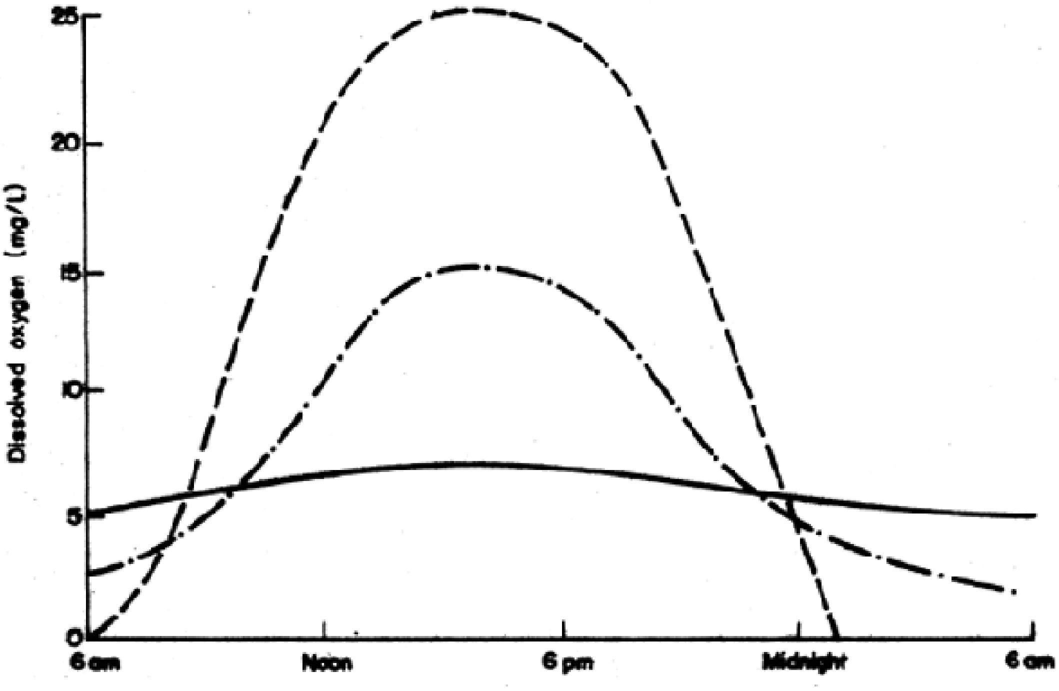
Question 1 – Key Concepts	*Question 2 – Identification of processes*
Number of key concepts identified (max 8) Total number of statements made	Number of processes named Number of non-process items
Question 3 – Causal hypothesis and explanation	*Question 4 – Conditional information*
Number of valid causal statements Number of non-causal statements Number of undefined or non-sequential causal statement* Number of superfluous statements	Number of valid conditions identified Number of conceptually incorrect conditions Number of non-conditional statements Number of valid alternative consequences identified Total number of statements

* In this context undefined causal statements include those where the “because” element was missing from an “if, then, because” format, or those where the statement was correct but bypassed an important conceptual step.

Appendix G – Written test used in Evaluation activity 2

This question was based on the style and content of a GCSE exam question about photosynthesis (AQA 2008).

A student measured the changes in oxygen concentration in three lakes over a 24 hour cycle of night and day. The results obtained are shown below:



Describe and explain the patterns observed on the graph.

Appendix H – LS2 model of photosynthesis used in Evaluation 2

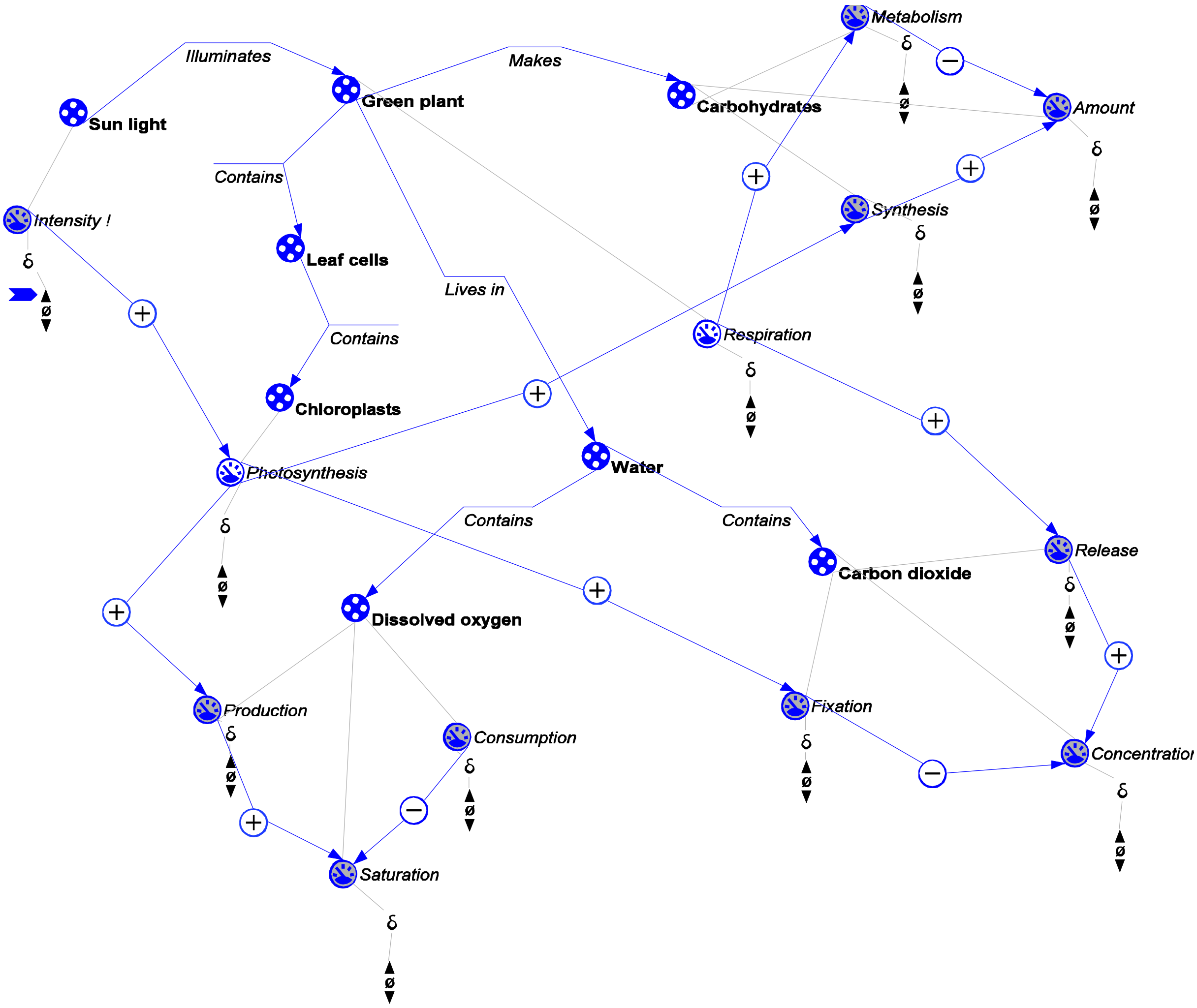


Figure A3 Norm basic causal model (LS2) of the photosynthesis problem used in the PGCE study focussing on grounding of terms.

Appendix I – Marking schemes used for Evaluation 3A instruments

Modelling Concepts

Key concepts	Scoring
Cells pump chloride ions across cell membrane into mucus Increased amount of chloride in the mucus Decreased amount of chloride ions in the cell Increased concentration of chloride in the mucus Decreased concentration of chloride ions in the cell Creation of concentration gradient across the membrane Creation of osmotic potential across the membrane Water moves down osmotic gradient Amount of water in the cell decreases Amount of water in the mucus increases Movement of water out of the cell increases the concentration in the cell* Movement of water into the mucus decreases the concentration of the mucus*	***Basic causal models*** 0 - Missing/incorrect/unclear 1 - Included partially 2 - Included causally

* Full causal implementation of these two concepts would introduce feedbacks and inconsistencies into a simulation in LS2. They were not implemented in the TA model used in 3B but are included here for the sake of completeness.

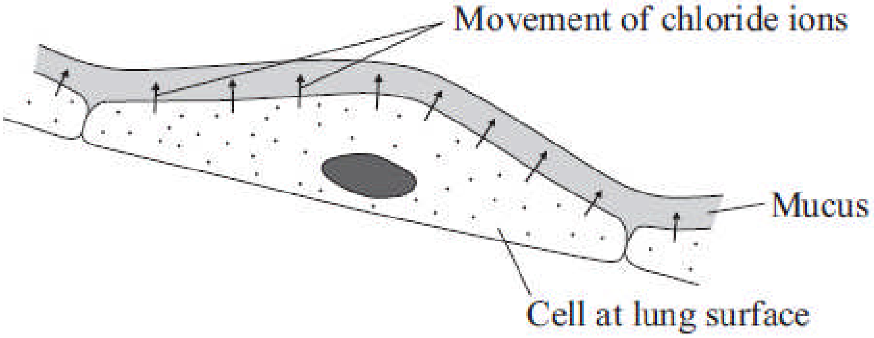
Modelling ingredients

Ingredient	Variables
Entities	n Correct n Incorrect - quantity n No assigned quantity n Total
Quantities	n Correct n State variable n Process n Incorrect - entity n Incorrect - assignment n Total
Configurations	n Correct n Incorrect - concept n Incorrect - implementation n Total
Causal dependencies	n Correct n Incorrect - concept n Incorrect - implementation n Total

Appendix J – Written pre/post test used in Evaluation activity 3

This is an example GSCE Additional Science Unit Biology B2 question from June 2009 (AQA 2009).

Question:



The diagram shows how, in a healthy person, cells at the lung surface move chloride ions into the mucus surrounding the air passages.

Explain why this movement of chloride ions causes water to pass out of the cells into the mucus.

Appendix K – Norm model used in the TA mode activity

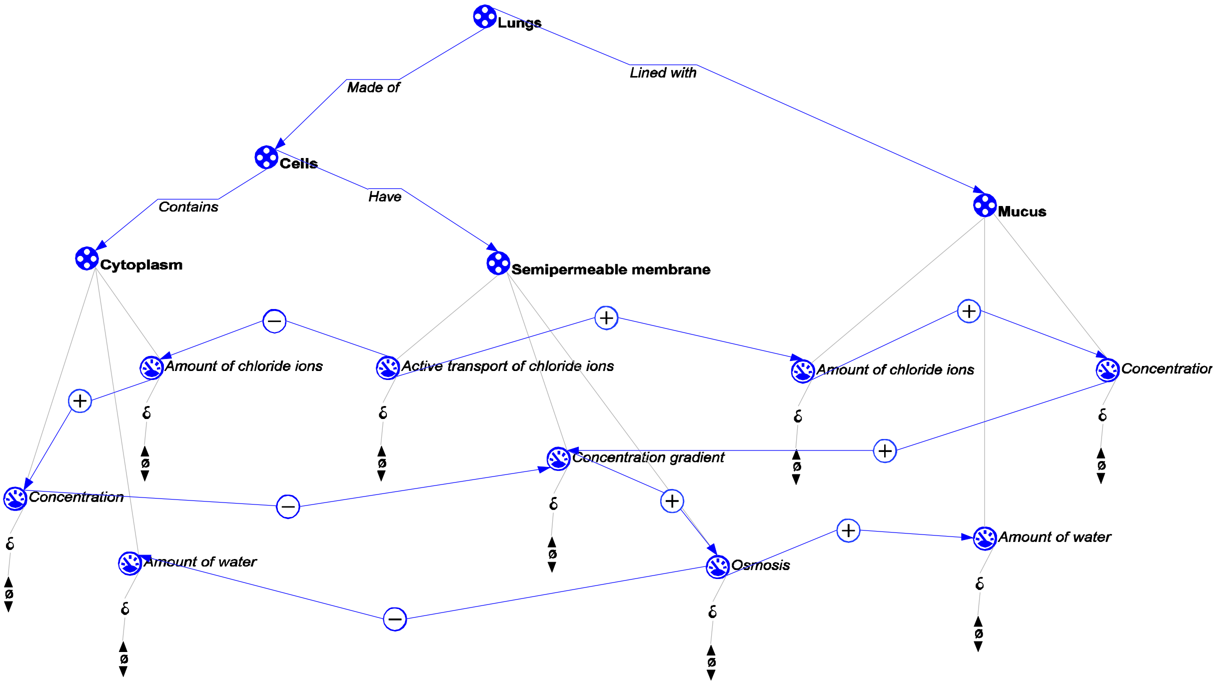


Figure A4 Norm basic causal model of the osmosis problem used in the PGCE study looking at the TA mode. The model was also used to form the basis for defining the prior list of key concepts used to evaluate the written answers.

Appendix L – Marking schemes used in Evaluation 3B instruments

Model status

Ingredient	Variables
Entities	All implemented?
Quantities	All used? n Incorrect - assignment
Configurations	All used? n Incorrect - implementation
Causal dependencies	n Incorrect - concept n Incorrect - implementation

e-mail: Info@DynaLearn.eu
website: www.DynaLearn.eu