Deliverable number:	D7.4
Deliverable title:	Final report on DynaLearn evaluation studies

Delivery date:	2011/12/31 - Extension deadline 2012/03/31
Submission date:	2012/05/20
Leading beneficiary:	Tel-Aviv University (TAU)
Status:	Final
Dissemination level:	PU (public)
Authors:	D. Mioduser, R. Zuzovsky, P. Salles, A. Souza, A. Portella, E. Braga, L. Wilhelms, M. Pereira Resende, G. Morison, P. Assumpção da Costa e Silva, I. Gontijo de Sá, L. Pires, A. Simões, R. Noble, I. Cowx, P. Borisova, J. Liem, Y. Uzunov, D. Zurel, M. Leiba, H. Benayahu, R. Nachmias, J. Ram, A. Zitek, M. Poppe, M. Stelzhammer

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 presents an integrative overview of the results of the evaluation activities conducted by WP7 partners in two rounds of evaluation of DynaLearn. It refers to 49 evaluation activities conducted with 736 participants. In the deliverable, the overall evaluation framework is presented, along with the questions addressed, the main insights gained and the main conclusions drawn. As well, a chapter on pedagogical issues deriving from the activities conducted is included.

Internal review

UVA: Bert Bredeweg

UHULL: Richard Noble

Acknowledgements

The authors would like to thank WP7 partners for the insightful discussions on the different evaluation issues and the development of evaluation plans and instruments. We would also like to thank the participants in our evaluation activities, students and teachers in all participating countries.

Document History

Version	Modification(s)	Date	Author(s)
01	Draft for TAU group internal review	8.5.12	Mioduser, Zuzovsky, WP7 partners
02	Draft for WP7 and internal reviewers	12.5.12	Mioduser, Zuzovsky, WP7 partners
04	Final version	20.5.12	Mioduser, Zuzovsky, WP7 partners

Contents

1. Introduction

This document reports the results of the evaluation activities of DynaLearn conducted by WP7 partners in two rounds of evaluation. Detailed description of the evaluation activities conducted and the data collected were presented in the set of deliverables reporting about phase one (D.7.2.1-5) and phase two (D.7.3.1-5) of the evaluation. The evaluation activities were designed with the aim to address the questions posed in the DOW, in correspondence with the functionalities and features afforded by the DynaLearn modelling environment. The main questions were operationalised and decomposed into specific questions and integrated in an evaluation framework guiding the development of the activities, methods, procedures and instruments.

In the following section the evaluation framework is briefly described, as well as an overview of the evaluation activities. A brief account of the activities conducted is presented in chapter 2. A summary of the results and the main insights gained appears in Chapter 3, followed by a discussion on pedagogical issues in Chapter 4. Chapter 5 presents the overall discussion and conclusions of the whole evaluation plan.

1.1. Evaluation framework

The evaluation framework served as inquiry space for the operational formulation of the evaluation questions and the design of the evaluation activities aiming to answer to these questions. In the project, the main goal of the evaluation was:

To assess the contribution of learning by conceptual modelling with DynaLearn on students' understanding of environmental systems.

In specific terms, the evaluation aimed to assess the effect of Dynalearn's key features and the process of modelling on students':

- Conceptual understanding (CU) - their learning of content knowledge related to the behaviour of complex ecosystems.
- Scientific reasoning, Qualitative reasoning (QR), and System Thinking (SQS) - their acquisition of scientific reasoning skills and ability to cope with complexity, through QR approach and language.
- Motivation and attitudes (M/A) - towards learning science and learning by modelling.
- Self-directed learning (SDL) - supporting the growth of Independent learning skills and practices
- Learning (L) - general aspects of learning enhanced by DynaLearn integrated functionalities.

The above, as function of learning with an environment encompassing:

- Conceptual Modelling (CM) - in terms of DynaLearn's specific modelling language, modelling process and 6 modelling levels - the Learning Spaces.
- Conversational virtual agents (VC) - as these act in various functions and roles while interacting with the learner.
- Semantic Technologies (ST) - supporting the individualisation of learning via DynaLearn tools for ontology mapping, diagnostic procedures, and the semantic repository.

The evaluation inquiry space is thus depicted in Figure 1.1. The cells in the Figure indicate the issues covered in both phases of evaluation activities. In the first phase most questions and data collected related to the effect of conceptual modelling with DynaLearn on students' conceptual understanding, acquisition and application of scientific skills and system thinking, and on motivational aspects (questions of types 1a, 1b and 1c). In the last stage of this evaluation phase, with the availability of features related to DynaLearn's semantic technologies, initial evaluations were also conducted about their effect on learning (Questions of types 3a and 3b).

The second phase of evaluation mainly addressed questions related to the effect of the features afforded by the VC and ST on students learning and growth of skills and modelling capabilities (questions of types 2a-d, and 3a-d).

In the DOW the overall set of questions to be addressed is presented, formulated in general terms (pp. 20-21). These questions, and the cells in the inquiry space into which they are inscribed, 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? And why or why not? **[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? **[1cd, 2cd, 3cd]**
6. Do learners actually learn better when using the full set of DynaLearn results? **[general - L]**
7. Are students more motivated to take on science curricula? **[1c, 2c, 3c]**

	1 CM	**2 VC**	**3 ST**
a CU	**1**	**2**	***1*+2**
b SQS	**1**	**2**	***1*+2**
c M/A	**1**	**2**	**2**
d SDL	***1*+2**	**2**	**2**
L	**1 + 2**		

1 = 1^{st} phase foci

2 = 2^{nd} phase foci

***1* = initially addressed in 1^{st} phase**

Figure 1.1: Evaluation inquiry space

For question six relating to the most general issues and aims of the DynaLearn project, the answers were gradually constructed upon the cumulative results for the various sub-questions along the project, and are addressed in the discussion and conclusions chapter.

Concerning question seven, the relevant target population are school-age students. In many educational systems these students face the need to choose learning tracks, and later on to make decisions about academic studies. This students (Junior-high and high-school students) were the participants in a considerable number of evaluation activities, e.g. by BOKU, FUB and TAU (see details in section 2.3.3).

As mentioned above, the questions in the DOW were operationalized and specific sub-questions were generated for planning and conducting the evaluation activities. The specific evaluation questions addressed, and variables and interpretative schemes for the results are summarized for each evaluation activity in the following chapters.

2. DynaLearn evaluation activities - general

2.1. Introduction

Phase one WP7 evaluation activities were conducted mostly during 2010. A total of 24 evaluation activities were conducted, varying in evaluation goals, sample size, research design, and duration (details are provided in the methods section).

Overall, the first phase of evaluation was of **exploratory** nature. The activities were built to examine the very first "educational encounter" of DynaLearn with learners in real learning situations and in real learning settings (e.g., school, college, and university classes). The activities were conducted within the possibilities and constraints dictated by the conditions in which they were implemented, such as:

- They were conducted while components of the software were gradually completed and released, affecting decisions about the features on which they were focused at each stage.
- Instruments and analyses schemes had to be developed mostly from scratch, due to the novelty of the learning environment and its features. Previous research on Learning by Modelling with other tools and approaches could only serve as starting points or reference. As a result, much effort had to be put in instruments' development.
- Environmental Science is not defined as independent subject in most countries. At the school level its topics and concepts are embedded in the curricula of different subjects (e.g., Earth or Life Sciences, Social SciencesAt the University level there are Academic programs on the subject (e.g., as in TAU's Porter School of Environmental Science) however, in most cases the individual courses are focused on specific disciplinary themes. As a consequence, to consolidate appropriate settings for the evaluation activities and reach reasonable samples demanded the solution of administrative (and even ethical) procedures. In some activities even affected the possibility to control continuous participation and even impede abandonment.

Phase two evaluation activities were conducted during 2011, with specific foci on the features and affordances of the semantic and virtual characters technologies integrated in DynaLearn. A total of 25 activities were conducted. As in the first phase, the activities were planned, and their main goals were defined, within the possibilities and constraints deriving from both technological and implementation factors:

- Also in phase two the activities were conducted while components of the software were gradually completed and released. This reality affected decisions about evaluation questions pursued at each and every stage, and changing foci as more modules and features were available by the partners for planning the activities. This "process evaluation" modality served as formative evaluation (as opposed to a summative evaluation which would only possible after the complete environment is in place), and has been challenging and rich in opportunities to learn "while on the move" about the learning value of the evolving features.
- In the second phase a more varied mix of participants, configurations and methods than in phase one was implemented. These included case-studies in which intense data-collection was carried about few participants, one-group-based designs for which repeated measures of aspects of the learning process were done, and group comparisons by various variable (e.g., a group using -for getting recommendations- a fixed reference model in contrast to a group using automatic-reference-model selection by the system; or experimental/control groups using/not-using a given feature).

- Many of the phase two instruments were built upon the ones developed and implemented in the first phase. Thus, the evaluations were built on previous experiences and on lessons learned for the development of new data-collection instruments, scoring methods and interpretation procedures.

2.2. Questions addressed

As already mentioned, the general evaluation goal addressed in the evaluation activities was:

> **To assess whether DynaLearn, by its different technologies, in its different learning spaces, and in various pedagogical modalities - contributed to students' conceptual understanding of complex systems, to their system thinking and scientific inquiry skills, and their motivation for learning by modelling and self-directed learning.**

This general question was operationalized in the studies conducted by all WP7 partners in both evaluation phases, in categories of questions and sub-questions dealing with the following issues:

- Conceptual understanding of Environmental Science concepts and knowledge
    - Learning of scientific contents - changes in knowledge state along the learning cycle (working in the different Learning Spaces) and/or at its end
    - Ability to apply the gained knowledge in new contexts and situations
- System thinking and scientific reasoning skills
    - Growth of system thinking approach and skills
    - Coping with complexity
    - Causality - understanding causal relationships and processes and change along the Learning-by-Modelling activities
    - Growth of Scientific Reasoning skills
    - Support for the formulation of scientific arguments
- Motivation - towards Learning by Modelling (LbM), learning Science by means of LbM and learning with DynaLearn
- Modelling capability and behaviour
    - Understanding and learning the QM approach and language
    - Modelling capability in different LS
    - From naïve to expert modelling
- Contribution of semantic technologies and VC to students' learning
    - Basic help
        - Effect on student modelling and learning
        - Support of students' independent modelling performance
    - Grounding
        - Effect on the quality of student models
        - Facilitation of self-directed learning
        - Functionality issues
    - Model-based recommendations

        - Effect on quality of student models
        - Students' understanding of recommendations
        - Extent of (correct) equivalence between terms in student and reference models
        - Effect on students' learning
    - Teachable Agent
        - Contribution to better understanding of the systems behaviour
        - Contribution to students' construction of better models
        - If/Then questions effect on causal arguments
- Social interaction during the LbM process
- Software usability
    - Observed usability aspects
    - Observed problems and difficulties
- Teachers' perceptions and attitudes towards LbM with DynaLearn
    - Teachers' perception of DynaLearn, and its role and potential contribution in teaching/learning processes
    - Teachers' motivation to work with qualitative models and to integrate DynaLearn in their teaching

In addition to these categories of questions, two additional pedagogical issues were addressed during the evaluation activities:

- Characterisation and definition of a repertoire of pedagogical modalities for integrating DynaLearn in teaching and learning processes.
- The development of assessment instruments - first used as measurement and data-collection instruments during the evaluation then refined as pedagogical instruments for assessing students' learning.

2.3. Methodological aspects

Most evaluation activities were primarily targeted at assessing the impact of LbM with DynaLearn on several target learning-layers: the acquisition of domain-content knowledge (conceptual understanding); understanding causal relations (causal understanding); understanding the complexity of ecological systems (complexity understanding); acquisition of modelling capabilities; growth of scientific reasoning skills; and motivation to learn science by modelling with DynaLearn.

For the implementation of these evaluation activities the design of non-conventional evaluation instruments was required. The development of these instruments demanded the integration between knowledge stemming from previous research, and new approaches and methodological solutions emerging while coping with **novel learning situations**, devised for the use of the **new learning environment**, for learning **new contents and concepts** in the DynaLearn curriculum.

In addition, the whole evaluation plan was **formative** in nature. Evaluation activities accompanied the development process of the learning environment and its different technologies. The direct implications were that not all features were available at given stages; that evaluation targets were incorporated as new features became available; and that evaluation data indicating faults or difficulties that was relevant at a particular stage was no longer relevant in later stages when new software versions and features became available.

2.3.1. Instruments and evaluation criteria

A wide range of instruments were used by WP7 partners along the two rounds of evaluations. A detailed description of these is included in the sets of deliverables written by the partners. Here we will present only a succinct overview of the main instruments.

A widely used source of data were the concept maps drawn by the learners themselves, or cognitive maps produced by the researchers on the basis of primary documents (textual) written by the learners. These instruments were used to obtain initial, pre-modelling data about students' "conceptual understanding" and "complexity understanding". These data were compared to data obtained from the analyses of the models constructed by the students using DynaLearn. For the analysis of concept maps and models a series of supporting tools had to be developed, e.g., sets of criteria, scoring guides, content analyses guides.

Two analytical tools were used for the analysis of video data and textual content analysis. For textual content analysis, Atlas TI software was used. This software enables identification and counting of students quotations. The quotations were then used to detect causal understanding, conceptual understanding (fit with expert answers) and for building cognitive maps. The other tool used was University of Wisconsin's Transana software that allows researchers to transcribe and analyze video data. Short video clips are organized into meaningful categories. These clips are transformed into wave files that are the basis for time sensitive coding aspects. Atlas TI and Transana were used by BOKU and enabled to identify differences between pre- and post-tests (Atlas TI) and the the mix of behaviours students exhibited while working in different learning spaces as assessed using video analysis (Transana).

For other analyses of qualitative data such as data obtained through the drawing of concept maps (Novak, Gowin & Johansen 1983; Ruiz-Primo & Shavelson, 1996) and the diagrams produced by DynaLearn we adopted cross-case display techniques (Miles & Huberman, 1994). The data gathered in this way was later on summarized for better description and interpretation (TAU comparison for concept maps and final models). Another way to assess models or concept maps was by comparing them to a reference (norm model) that was created (as was done by UHULL).

Another common instrument used in different evaluation studies were domain-specific tests usually containing open-ended questions, or mix of closed and open-ended questions addressing conceptual understanding and system thinking. These tests and their scoring guides were created by all partners.

Similarly, motivation/attitudes questionnaires have been developed and used by all partners. An initial set of questions was developed, and subsequently refined and expanded for its implementation in the different evaluation activities. As more software features were available and became targets for the evaluation, the motivation and attitudes questionnaires were expanded to cover these as well.

In the last stages of the evaluation focusing on the model-base recommendation features, students and system data stored in the system (e.g., student logs, repository records, recommendation files) served for the analyses of the learning process (as reported in Noble & Cowx, 2012).

2.3.2. Participants

The sample size in the evaluation activities were of two types. In many studies the sample size was small - from individual case studies to small groups up to 5 participants. Different reasons guided the decision to work with small groups or case studies, such as: the deliberate intention to conduct intense qualitative studies with few participants; logistic aspects such as the fact that in secondary schools conducting the evaluation interfered with the students' tight learning schedule, or in the university it interfered with other programs; or software related issues such as the fact that in given stages it was in development phase and getting permission to test it in regular classrooms as part of regular classes was problematic.

Other evaluation activities were conducted with groups comprising 15-30 participants. In these cases whole classes of students (junior-high, high-school or university students) participated in studies of varied designs (e.g., one group using repeated measures, one group with sub-group comparisons, or experimental/control groups comparisons).

A summary of the samples in each study conducted is shown in Table 2.1.

2.3.3. Design, data collection instruments and analyses procedures

The most common evaluation design used was the one Group Pre-test-Post-test Design. In cases where information was gathered during a sequence of lessons, the design was a "Single Group Interrupted Repeated Measure Design" – in these cases, the researcher recorded measures for the same group after each intervention.

In some cases when the effect of different software features were evaluated, the design of the studies was a quasi-experimental one with non-equivalent experimental and control groups. Randomly assigning treatment to groups was hard to obtain in the context of most evaluations studies.

When the unit of analysis was individual students, the small sample size required non-parametric statistical procedures. However, when the analysis was done on a large number of qualitative narratives or video segments regular statistical procedures were applied.

A summary of the designs implemented by each WP7 partner in each evaluation phase, and their reference code used in the remaining of this deliverable, is shown in Table 2.1.

Table 2.1: Summary of setting and methodological features of the evaluation activities

Phase one of evaluation - 24 activities				
Activity	**Sample**	**Duration**	**Design**	**Code**
BOKU1	2 (HS)	4 meetings / 24 hours	Case study / intensive data collection	CS
BOKU2	29 (GS)	4 hours	One group pre-post-measurement	1G/Pr-Po
BOKU3	2	Multiple sessions	Case Study w/repeated measures	CS
BOKU4	3 (GS)	4 sessions / 10 hours	Case study / intensive data collection	CS
FUB1	60 (HS)	8 meetings / 12 hours	Experimental/Control	E/C
FUB2	60 (HS)	8 meetings / 12 hours	Experimental/Control	E/C
FUB3	27 (Teachers)	12 meetings / 60 hours	One group w/repeated measures	1G/Rm
FUB4	4 (HS)	6 meetings / 8 hours	One group (small) post-measurement	1sG/Po
FUB5	10 (GS)	30 hours	One group post-measurement	1G/Po
FUB6	13 (Teachers)	4 meetings / 12 hours	One group pre-post-measurement	1G/Pr-Po
FUB7	21 (HS)	18 hours	Experimental/Control	E/C
FUB8	49 (HS)	18 hours	Experimental/Control	E/C
FUB9	35 (HS)	12 hours	Experimental/Control	E/C

FUB10	5 (HS)	3 meetings / 9 hours	One group (small) w/repeated measures	1G/Rm
IBER1	5 (GS)	6 meetings / 12 hours	1 group (small) pre-post-measurement	1G/Pr-Po
IBER2	10 (S)	6 meetings / 12 hours	One group pre-post-measurement	1G/Pr-Po
TAU1	10 (GS)	7 meetings / 14 hours	One group w/repeated measures	1G/Rm
TAU2	23 (HS)	6 meetings / 24 hours	Experimental/Control	E/C
TAU3	14 (GS)	1 meetings / 3 hours	One group pre-post-measurement	1G/Pr-Po
TAU4	30 (HS)	5 meetings / 15 hours	One group pre-post-measurement	1G/Pr-Po
UHULL1	4 (US)	4 meetings (12 hours)	Case Study w/repeated measures	CS
UHULL2	23 (US)	3 hours	E/C - incomplete due to technical failure	-
UHULL3	18 (GS)	3 hours	One group pre- post-measurement	1G/Pr-Po
UHULL4	18 (GS)	3 hours	Experimental/Control	E/C
Phase two of evaluation - 25 activities				
Activity	**Sample**	**Duration**	**Design**	**Code**
BOKU5	2 (HS)	4 sessions / 10 hours	Case study / intensive data collection	CS
BOKU6	31 (GS)	2 sessions / 8 hours	One group pre-post-measurement	1G/Rm
BOKU7	Stakeholders	2 sessions	Focus groups	FG
FUB11-22	42 (HS; ~3 per)	Short/long sessions	One group (small) w/repeated measures	1sG/Rm
FUB23	16 (HS)	Class sessions	One group w/repeated measures	1G/Rm
FUB24	27 (HS)	Class sessions	One group w/repeated measures	1G/Rm
FUB25	19 (HS)	Class sessions	One group w/repeated measures	1G/Rm
FUB26	27 (HS)	Class sessions	One group w/repeated measures	1G/Rm
FUB27	27 (HS)	Class sessions	One group w/repeated measures	1G/Rm
IBER3	23 (HS)	3 sessions / 9 hours	One group w/repeated measures	1G/Rm
TAU5	9 (HS)	1 meetings / 3 hours	One group pre-post-measurement	1G/Pr-Po
TAU6	15 (HS)	2 meetings / 16 hours	One group w/repeated measures	1G/Rm
TAU7	3 (HS)	1 meeting / 4 hours	Case study / intensive data collection	CS
UHULL5	8 (US)	2 sessions / 6 hours	One group w/repeated measures	1G/Rm
WP7	10 teachers	Questionnaire	One group post-measurement	FG
Activities **Participants**	**49** **736**			

HS: High School; **US**: Undergraduate students; **GS**: Graduate students; **S**: mixed-degrees University students **FG**: Focus group (stakeholders; teachers); Shaded text in this table and in the text refer to phase one.

3. Main insights from the evaluation of DynaLearn

A succinct account of the main insights from the evaluation is presented in the following sections by the main categories of questions detailed in section 2.2. Each section includes a set of main claims, followed by examples of observations from the whole set of evaluation activities. Examples and observations from each evaluation activity are preceded by indication of the study, e.g., BOKU2, and its design's reference code from Table 2.1, e.g., 1G/Pr-Po. The inclusion of the design reference code aims to clarify the methodological context within which the mentioned observation was collected, e.g., a small group in-depth case study or a large class multiple-sessions evaluation. As well, the shaded text indicates that the activity has been conducted in phase one of the evaluation, as in Table 2.1.

Obviously, due to the nature of this report as overall summary of the evaluation activities, the main insights are presented in the following without entering into specific details or relating in-depth to the data layer of the studies. Full detailed descriptions are presented in the deliverables prepared by all WP7 partners at the end of the two rounds of evaluations.

Although parts of this report were already included in the summary report of phase one of the evaluation (extra deliverable D7.2.6, Mioduser et al., 2011), there are two main reasons for their inclusion in this deliverable: (a) This is the final report about the whole evaluation process of DynaLearn, as stated in the DOW - in order to depict the complete picture, reporting about both evaluation phases is required; (b) The evaluation was planned and conducted vis-à-vis the software development process, and specific aspects of students' learning (e.g., conceptual understanding, acquisition of the modelling language and skills) were addressed mainly in phase one - leaving out the summary of findings from phase one actually means leaving out evaluation observations about a complete set of questions. However, we want to note that in this report we refer mainly to evaluation results from phase one which are relevant to the whole evaluation. Thus, observations related to e.g., software features that were afterwards significantly modified, or observations about learning processes that in phase one were preliminary and constrained by many factors, were obviously excluded from this report. In addition, evaluation results about a given issue from both evaluation phases appear in the same section, since these are conceptually part of the same topic.

3.1. Conceptual understanding of Environmental Science contents

An essential focus for the evaluations was the contribution of DynaLearn to students' conceptual understanding of ecosystems, and of concepts in Environmental Science. Evidence collected in the evaluation activities conducted by all WP7 partners, mainly in phase one of the evaluations, are clearly indicative of this contribution.

Overall, two main conclusions can be drawn from the observations in the evaluation activities. The first is that the main gain at the conceptual understanding level relates to students' acquisition of a systemic perspective in learning about ecological systems. The second is that previous content-knowledge plays a crucial role in the model construction activity. Alternatively, the acquisition of the new knowledge and concepts required for complex modelling demands an extended learning process. This was observed in activities comprising long-term interventions, in contrast with the limited effect of short-term interventions.

Data on the effectiveness of LbM showed a wide range of results, from moderate -and even no effect-to significant effect on students learning. Examples of observations collected are (presented in ascending extent of effect):

[FUB1,2,8,9 - E/C] Significant difference was observed in conceptual understanding gain in the experimental groups in [FUB1,2,8,9]. However, post-test differences between experimental and control groups were not significant in [FUB8,9].

- [BOKU2 - 1G/Pr-Po] The use of DynaLearn led to significant increase in content knowledge and conceptual understanding. Topics explored with DynaLearn were graded higher in the final exam that also contained topics not covered through DynaLearn intervention.
- [TAU2 - E/C] Students' explanations to a set of twenty key concepts in ecology were improved from pre- to post-testing (large effect sizes in both groups). However, both groups obtained low average scores.
- [IBER2 - 1G/Pr-Po] All students performed very well on knowledge questions related to topics explored with DynaLearn compared to topics explored without DynaLearn.
- [FUB25 - 1G/Rm] There was a significant difference between pre-test and post-test means, showing an increase in students' conceptual understanding.
- [FUB1, 2, E/C] Significant difference was observed in students' gain of conceptual understanding in the experimental group, and in comparison with the control group.

It is suggested [UHULL D7.2.2] that students learning by modelling will potentially move through three phases as they are introduced to the DynaLearn software from novice to apprentice and master phase. 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 level.

In addition, differences in the results could additionally be explained by motivational factors transmitted by the teacher, well known to have the potential to significantly influence learning success in general.

Observations about cognitive strategies applied that can be seen as indicative of conceptual understanding were obtained in many evaluation activities, such as:

- [BOKU3,4 - CS] Significant increase of the abstraction level of representing knowledge was observed.
- [TAU1,2 - 1G/Rm and E/C] Comparing student concept maps (previous to the modelling process) and final models yielded the following findings: Student final models contained fewer entities (only the relevant ones); quantities that were ignored in the concept maps were properly addressed in the models; differentiated representations of causal relationship were included in the models. The models were built around specific research questions and hypotheses that were tested through simulations and led in some cases to new questions and insights.
- [TAU2 - E/C] Students' capability to apply the gained knowledge in new contexts and situations was observed.

Generally, although there were mostly indications that students who used DynaLearn exhibited better conceptual understanding (in activities by TAU, BOKU, IBER), there were also cases that did not show this gain (UHULL, FUB).

Two explanations are given to the limited gains in conceptual understanding in several of the activities conducted: The first relates to the fact that in many activities DynaLearn intervention was a short term one, not sufficient enough to alter or improve students' understanding of the topic. The other explanation relates to the modelling activity itself that shifts students' investment of efforts from mastering the knowledge domain to mastering modelling capabilities.

3.2. System thinking and scientific reasoning skills

3.2.1. On system thinking and coping with complexity

Students' system thinking and ability to represent a system's structural and behavioural features were contributed by the work with DynaLearn. Along the learning processes, growth of skills and abilities was observed.

Overall, the students acquired rapid mastery of the skills and procedures required for constructing complex models with DynaLearn. As the modelling sessions advanced, their products reached high levels of complexity. At the end of the learning cycles, an increase in students' ability to represent a system's structural and behavioural features was observed. Analyses of models and explanations given by the students along the activities showed clear advances toward systemic view and understanding of the complexity and causal relationships (chain and loops) that stand behind the system behaviours.

Examples of observed performances indicative of growth in System Thinking are students' [TAU2 - E/C]:

- Progressive ability to define and refine the foci (the essential properties) of the model to be constructed
- Defining criteria for reducing the amount of model ingredients while preserving its meaningfulness
- Moving from a linear representation of a system's structure to a hierarchical ones, and then to a web-like configuration
- Evolving ability hypotheses formation and testing
- Generation of new questions and inquiry processes beyond the original information used for generating the model
- Perceiving the value of the models constructed not only in terms of the specific phenomena modelled, but as paradigmatic examples of complex systems in other areas of study

Also observed was growth in students' skills such as the referred in the literature as comprising system thinking (e.g., Draper, 1993; Assaraf & Orion, 2005), e.g., the ability to identify components of a system and processes within a system; the ability to organize the system's components within a framework of relationships; understanding the hidden dimensions of a system; or thinking temporally - retrospection and prediction.

- [TAU2 - E/C] Pre- post-data showed changes in the experimental group's perceptions of the system, as reflected in their representations and models: increases in web-type representations, in adopting ecological organizing principles, and in the complexity of relationships among entities were observed. In contrast none of these changes were observed in the control group.

- [IBER1,2 - 1G/Pr-Po] At the end of the course, the students seemed to have acquired the skills to correctly distinguish different concepts in a scientific paper as corresponding to a particular model ingredient type in DynaLearn. That is, which are part of the structure of the system, and which aspects are dynamic (the quantities). Furthermore, they seemed to be able to choose the correct causal relationship between the quantities. This is, they can successfully identify the processes that are important in the system. This suggests that the students acquire a new set of analytical skills with which they can analyze topics.

- [TAU4 - 1G/Pr-Po] Indication of system thinking was also obtained by analysing the changes in configuration of students' representations of the systems under study. Most students in all representational tasks (concept map, model 1, model 2) represented the system in hierarchical configuration. Only few used net-like configuration. However, concerning the type of relationships represented (e.g., structural or process/causal), at the end of the learning cycle the vast majority of these were process/causal relationships.

- [FUB case studies] observations about students perceptions of processes and their consequences indicate that:

    - After doing exercises using Is and Ps, students were able to understand better process-related phenomena and they showed increased capability to identify quantities and entities in the models.

    - After building models students´ scores were higher than in the first model, showing they were able to improve their qualitative reasoning and systems thinking skills.

    - The modelling activities promoted students' ability to recognize and implement model ingredients and build models of better quality.

The software definitely allows students to tackle complex problems. However, the development of system thinking implies an essential transition in students' learning: a transition in paradigmatic approach towards the inquiry of phenomena in the world. The vast majority of science teaching curricula and learning materials (particularly at the school level) aims to teach "classical science" and its methodology. Novel theoretical and methodological approaches in development in the sciences community for several decades, that view phenomena from the perspective of complexity and systems theories, are almost absent from existing curricula. This transition is not a trivial one, and demands the development of novel pedagogical approaches and essentially - it requires time, or long-term involvement in learning processes involving system thinking.

3.2.2. Understanding causality

Understanding causal relationships within a system and between it and its environment is critical for understanding processes at different levels of the system's behaviour and for predicting or hypothesizing about its behaviour under changing circumstances. DynaLearn allows representing causal relationships in increasing level of complexity along its different Learning Spaces. In the evaluation activities conducted observations were collected about students' modelling work in all LS's. The insights gained are presented below.

About student ways to perceive causality, distinction can be drawn between two main layers. The first relates to students' ways of expressing "explicit" causal relationships, namely, those that are attached to the components as represented in the model (e.g., direct relationships between quantities or causal chains). The other is the "hidden" layer, relates to their understanding of how the overall behaviour of the system results (emerges) from the causal configuration of its components.

- [TAU1,2 - 1G/Rm, E/C] For the first layer students represented causal relationships ranging from single/unidirectional relationships, through different configurations of one-to-many and many-to-one relationships, causal chains, to simple and complex feedback loops.

  For the second layer, our main observation was that along the modelling activities, students' perceptions (hence representations) of causal patterns and configurations became more complex. Even in the high school students group, most of them succeeded in expressing long causal chains in their models (in contrast with a non-modellers control group). Also in most cases students' showed high ability to predict causal chains and loops in alternative scenarios for the systems modelled.

  For the higher layer - understanding of the overall behaviour of the system- we found ample evidence in students' explanations. For example, undergraduate students' generation of inquiry questions beyond the questions in the scientific paper used by them as reference to build their models, is clear evidence not only of their understanding of the causal configurations among factors in the original experiment, but also of their ability to suggest and explore alternative causal configurations. One school student's comment is illustrative of the insights gained: "*The modelling activity taught me that some changes have long-term and far effects – If you touch one thing, everything can change*".

Evidence for students' understanding was noticeable in observations collected about their representations and verbal and written explanations, and about their models, as a result of the learning activities. These observations were collected in activities conducted with case studies as well as with small and large groups. Sample observations are:

- Significant increase in the use of causal relations was observed, especially graphical ones [BOKU2 - 1G/Pr-Po] and verbal expressions [BOKU1 - CS]. Data from two case studies extensively analyzed showed a significant increase in use of causal verbal expressions between pre- and post-tests, whereas wrong causal relations did not occur in post-tests.

- [TAU2 - E/C; TAU6 - 1G/Rm; IBER2 - 1G/Pr-Po] At the end of the learning cycles, an increase in students' perceptions and representations of multiple-variables causal relationships, causal chains and feedback loops was observed. This is indicative of students' evolving understanding of the complexity of a system and of the type of causal configurations provoking its behaviour.

- [BOKU5 - CS] - significant increase in causal understanding (by pre-/post-test measures) has been observed with two students who completed a series of activities with DynaLearn that included the use of grounding and feedback features.

- [BOKU6 - 1G/Rm] The comparison of students' LS2 models before and after feedback from the expert model showed a significant progress and improvement - the number of correct causal relationships increased significantly.

- [FUB11 - 1SG/Rm] Analyses of students' performance showed that the number of wrong causal relations in their models decreased as a result of the modelling tasks.

Difficulties in students' appropriate definition of causal relationships could be attributed to various aspects, some of them related to students' understandings (or missing- or miss-conceptions), and other to the affordances of DynaLearn in its different Learning Spaces.

- [UHULL3 - 1G/Pr-Po] In many cases students missed many of the key concepts required to build appropriate models. 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.

- [UHULL3 - 1G/Pr-Po] 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 LS2. Although the positive and negative relationships available can be viewed as general causal relations 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.

- A set of short-term evaluation activities at FUB included questions focusing on students understanding of causal relationships, as reflected in their explicit reference to causal Inferences in written texts, in their expression of trivial and non-trivial conclusions, and in their models [FUB1,2,7,8,9,11, 13]. Interpretation of the whole set of results suggest that while DynaLearn appears to foster the growth of significant scientific reasoning skills, short time interventions are not enough to support their development and consolidation.

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 are of great importance for students evolving understanding of causality.

3.3. Motivation towards LbM and towards learning Science by LbM

Data on motivational aspects and students and teachers attitudes towards LbM and towards learning Environmental Science with DynaLearn were collected by all partners in both evaluation rounds using questionnaires sharing common contents and structure. In phase two of the evaluation specific questions addressed the ST and VC features that were at the centre of this phase's evaluation activities.

In general students' feedback concerning DynaLearn and the contribution of LbM was rated positively, indicating the general acceptance of the whole DL approach by students. Overall, students' answers to motivation and attitudes questions in all questionnaires administered showed similar trends, thus a succinct overview of the observations by each partner is presented here.

- [IBER] The motivation questionnaire results indicate that students find DynaLearn software easy to learn, and they think that the software can be applied more widely in other curricula. All students indicated that they would use DynaLearn in other subjects. Students also indicate that they would like to have better learning materials so they can work more independently. These results suggest that students are motivated to take on more science curricula with the DynaLearn software.

- [UHULL] 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 characters 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.

- [TAU] Overall, the marks for motivational issues related to learning with DynaLearn were high. Looking at specific aspects, the higher scores were related to the students' perception of the software's motivational value for building ecological models (in the course and in the future as well), and to the contribution of the work with the software to their system thinking. At a lesser extent the modelling work was perceived by the students as contributing to their conceptual understanding and learning subject topics. Specific motivational responses were collected in activities focusing on the interaction with the VC. Students perceived positively this interaction. They indicated that it contributed to their learning and to their modelling work.

- [BOKU] The motivation questionnaires yielded in general only positive feedback to all questions asked. In detail, case-studies students very much liked the lesson and learning activities, indicating that modelling with the software led them to better understandings, and they highly agreed that modelling with the software could be also used in other learning topics. They less agreed that the software provides a very comfortable way of learning (modelling with DynaLearn was experienced as being challenging). Motivation data collected at [BOKU2] with more experienced students shows more heterogenic results. The highest agreement was documented for the applicability of the software to other learning topics. Furthermore they liked the lesson and learning activity supported by DynaLearn and they found it very interesting to work with DynaLearn. Also the importance of building models in different LS's was ranked as high. As the students were already well informed about the issue that was explored by DynaLearn, the activity did not much contribute to a new understanding of the system.

- [FUB] Students found it interesting and motivating to work with DynaLearn. In the case of deaf students they commented on how the use of qualitative models could help the deaf to learn concepts and to improve their writing skills. Students noted explicitly their perception of the contribution of DynaLearn to their learning. Teachers recognized the high potential of the modelling activities for the development of a number of competences and skills, including the ability to make inferences, analogies and deductions while analysing the behaviour of a system; formulate hypotheses and predict results; analyse and compare possible solutions to the same problem. One of the teachers said: “To me, qualitative models refine the scientific method, allowing the student to formulate hypothesis and predict results in a consciously way.” Difficulties were attributed mainly to modelling at the more complex levels, and to particular functional aspects of the software

One of the aims of DynaLearn as constructivist learning environment is to foster students' self-directed learning (SDL, Gibbons, 2002). It is clear that appropriate balance should be achieved between SDL and teacher-directed learning (TDL), aiming to create the necessary motivation and perception of self-confidence for the student to take control of her learning. Numerous observations of the gradual transition towards a more independent learning modality were collected by all partners. However, these situations should be formally and systematically devised as pedagogical process to be used in the future as models for the pedagogical implementation of DynaLearn.

Still in the dimension of the interaction between pedagogical approaches and students' motivation and attitudes towards learning, an important transition is required. For years the main teaching modality in which students are involved is the lecture-based or information-delivery modality. Learning about systems with DynaLearn demands a clearly different approach: learning by constructing models, and inquiry-based learning of phenomena represented in models. This transition implies once again a pedagogical model in which the "model constructor" responsible for her own learning stands at the centre of the scene (Papert, 1991). The constructionist idea stating that the students construct their inner world by constructing in the outer world demands a supporting pedagogy. Our observations of learning processes with DynaLearn serve as promising background for the development of these pedagogies.

3.4. Modelling capability and skills

Data on the gradual construction of modelling skills and capabilities were collected in many of the evaluation activities conducted. The observations are related to two main levels: (a) the conceptual level, focusing on students' gradual development of the QM approach and their ability to express phenomena in terms of qualitative models; and (b) the software level, related to students' work with the specific tools and features of DynaLearn, and the difficulties encountered. Naturally, the most detailed observations were obtained in the evaluation activities conducted as small-group case-studies, in which the learning process was intensively analyzed in detail (e.g., in BOKU or UHULL). However, substantial data were collected as well in activities in which repeated-measures were implemented, allowing comparison of stages in the modelling process.

At the conceptual level, it was evident that entering the realm of QM demanded a change in students' perceptions and approaches towards the inquiry of phenomena, contrasting with the perspective characterizing most science teaching and learning in educational systems. Evident as well is the fact that this transition demands time and involvement in recurring opportunities "to do the work" - to experience modelling tasks of varying types and complexity. Observations in the different activities unveiled different characteristics of this process, as in the following examples.

- Changes in perspective were reflected in the way students reformulated the phenomena under study for its representation in the modelling process. In [TAU2 - E/C] most students described the phenomenon to be modelled in the first session of the course in terms of a general question (e.g., "How do the wind and the waves affect the patella attachment to the rock?"). At the advanced modelling sessions they shifted to a language more focused on systemic and causal relationships (e.g., "The relationship between crabs, barnacles and the limpet"). The changes in the description of the aim of their modelling activity imply a change in perspective: from a focus on the local and specific aspects to be modelled to a more systemic view of the phenomena. The more generic descriptions imply also that the students were able to view the phenomena as particular instance of broader categories, in which multiple-variables causal relationships take place. Similar perceptions of modelling as novel approach were also observed in [FUB4 - 1G/Po].

- [BOKU1 - CS] While modelling, as students were not familiar with the ecological topic, they spent a lot of time in LS1 for picking up information (asking the teacher/other-student, looking in the internet, into additional materials). Topic-related questions decreased significantly with the progression to higher LS's, where the work focused on single processes. In addition, conversation on the modelling process and aspects of the activities (between students especially in LS2, and with the teacher especially in LS4) increased from LS1 to LS4.

- [BOKU1 - CS] In LS2 students worked independently on their models which implicated more time for thinking. LS2 also allowed students to easily translate their mental model into a dynamic model instead of having to invest too much effort in identifying relevant variables and relationships between them.

- The acquisition of the Qualitative Reasoning language and different expression tools showed uneven patterns. The basic ideas and procedures behind the diagrammatic approach were rapidly grasped by most "naïve modellers" [e.g., in UHULL3 - 1G/Pr-Po; FUB8 - E/C; TAU5 - 1G/Pr-Po]. In contrast, difficulties were observed in mastering more complex ideas in the modelling language, such as those related to "direct influences" and "proportionalities", causal dependencies that respectively represent processes, the initial cause of changes in the system and the modelling element used to propagate the effects of processes [e.g., UHULL1 - CS; TAU1 - 1G/Rm; FUB7 - E/C]. It is evident that there is a need to plan appropriate pedagogical interventions to overcome these difficulties.

- [UHULL1, 3 - CS, 1G/Pr-Po] Additional insights were gained on the characteristics of naïve modellers. Their models showed great levels of variability in terms of complexity, mainly relating to the numbers of configurations and causal relations used rather than the number of entities and quantities used. Growth in confidence and understanding about how the modelling environment works resulted in better disposition to enter more complexity in the models. As well, crucial variables that appear to affect the modelling ability and pace are previous knowledge and individual cognitive styles (e.g., for solving problems). The ways these variables affect the modelling process should be systematically examined in future research.

- In general, activities lasting at least several weeks allowed observation of the gradual acquisition of modelling skills and methods. At the beginning of the learning cycle students had difficulties in discerning the set of necessary ingredients to construct the model, thus including a wide scope of components; along the modelling process, students' modelling became more focused and clear-cut concerning the distinction between necessary and unnecessary components [e.g., IBER1,2 - 1G/PrPo; TAU1 - 1G/Rm; UHULL3 - 1GPr-Po]

- [TAU4, 5 - 1GPr-Po] Observations showed that students' modeling ability along the classes increased in mastery. Although the models were in most cases of the hierarchical type, these became gradually more complex as well as more conceptually focused. Models produced by students in the advanced sessions represented dynamic features (processes) rather than structural features of the systems. As well, the progression of models reflected a change in construction principle, from "telling the story of the system" (in the form of linear chaining of entities) to web-like representations of conceptual features of the system.

- Additionally, students adopted a range of strategies for coping with modelling tasks [TAU1,2 - 1G/Rm, E/C], as in the following examples: balancing between "trial-and-error" and "goal-oriented" modelling; "modelling-to-get-the expected-result" according to disciplinary knowledge - aiming to probe that meaningful (and known) results can be achieved with the software prior to using it as inquiry tool for new hypotheses; or differential attitudes towards "following-formal-instructions" while modelling [UHULL3 - 1G/PrPo].

An important set of observations relate to the contribution of the different Learning Spaces to students' learning.

- [BOKU1 - CS; BOKU5 - CS] A detailed mapping of students modelling behaviour in different LS was depicted. In brief: Meaningful questions about modelling with the software were raised by the students along the work in the LS's, reaching levels of complexity that corresponded with the complexity of the features at each LS. Additionally, the main mistakes (software related) made by the students in each LS were mapped. This mapping served as basis for devising appropriate pedagogical solutions for supporting students' work in activities conducted in phase two of the evaluation (reference to this issue also appears in section 4.2.2 on teachers' perceptions of the foci of the learning spaces)

- [FUB5 - 1G/Po; IBER2 - 1G/PrPo; UHULL5 - 1G/Rm] LS4 was considered by many students as the LS that most contributed to the understanding of the concepts represented by the models, probably because this is the first level at which causality differentiation can be applied and the consequences can be observed in action.

- [FUB22 - 1Sg/Rm] The compositional modelling and the hierarchic approach were seen as advantages of the LS6 by students, and they found very worthwhile to learn using LS6. The students were able to understand the definitions of process and feedback, and that these important concepts are not very well approached in their regular science classes.

- [TAU7 - CS] A detailed look at the work of an experienced modeller (case study conducted with a high-school student who had participated in previous evaluation activities) supplied evidence about how repeated involvement in modelling tasks supports the mastery of advanced modelling skills and strategies, and the construction of quite complex models. This supports once again the claim that appropriate development of modelling capabilities demands long-term learning processes and involvement in progressively complex modelling tasks.

Overall, the evaluation activities allowed the composition of a valuable picture of students' gradual mastery of modelling capabilities, of difficulties encountered, and of affordances of the software in each LS.

3.5. Contribution of semantic technologies to students' learning

Almost all evaluation activities in phase two of the evaluation carried by all partners included components related to the examination of the contribution of the semantic technologies to students' learning (few activities although were conducted in phase one, at the time these features were only at preliminary stages of development). The summary of the main insights related to these features will be presented around two main themes: Grounding and model-based recommendations..

3.5.1. Grounding

The "grounding" feature was not initially intended to be a pedagogical resource or modality, but a required stage of the data-completion process of the models to be incorporated in the repository. However, already in evaluation phase one its potential as pedagogically-rich learning mode was identified, and several evaluation activities were defined to assess its value for learning and modelling.

Different results were obtained by the activities in both rounds of evaluation, reflecting the gradual improvement and completion of the functionalities of the grounding feature over time. Evaluation results related to different aspects of the learning process affected by students' involvement in grounding the models, such as their understanding of terms and concepts or their ability to modify and improve their models. Sample results at different stages were:

- [TAU3 - 1G/PrPo] Grounding tasks supported graduate students' acquaintance with unknown concepts, and affected the quality of their models For the unknown concepts, students' inquiry of their meaning using the grounding feature allowed their appropriate integration into the models. Before the grounding task about 40% of the students created quality models of the highest scores. After the grounding, all but one model got the highest score. The main effect of the grounding activity was on the students' model construction and revision processes. Data collected on various parameters of students' work (i.e., identifying relevant entities and relationships; identifying their correct configuration in the model; number of elements in the causal chain; and their correct ordering) indicated an increase in students' capabilities following the grounding activity. Similar contribution of grounding activity was observed in [FUB15 - 1sG/Rm].

- [TAU4 - 1G/PrPo] In an activity conducted with High-school students, their performance showed a shallow approach. In most cases they did not go into deep analyses of the options offered, opting for the first definition or meaning in the candidates list. Moreover, performing the grounding, which means complementing the representation with a layer containing information and definitions on its ingredients, did not trigger students' reflection and revision of their original models which is at least one of the aims of providing students with accurate information. Although DynaLearn's grounding feature was not conceived as pedagogical resource, we found that this feature could be of value when we came to develop learning tasks. This feature prompts the students to find information which can help them in constructing their models. Given that the existence of (expert) grounded terms and models affects the results of a student's grounding inquiry, this can be seen as "implicit guidance" that bounds the student's search for information about her/his model.

- [UHULL5 - 1G/Rm] Observations about student-created anchor terms: All of the students had to create at least one anchor term; four (of seven in the group) had to create three or more and two of the students created anchor terms for the majority of their model ingredients. Most student models had a very low overlap with the reference model, with the closest overlaps being a maximum of four common groundings. The majority of anchor terms created were for compound quantity terms (e.g. Salt excretion) of for terms that were more descriptive in nature (e.g. steady-state system). By definition the creation of an anchor term by a student will result in that term not being used for model alignment and the more anchor terms created by a student the less likely that the repository will be able to find a suitable suite of reference models (without full ontology matching).
- It is not clear from the studies (e.g., UHULL, TAU] whether the students understood the grounding interface and the relevance of the information that is presented to them there.

Grounding activities served the students to think about the information layer which can be of substantial help for constructing their models. The existence of (expert) grounded terms and models affects the results of a student's grounding inquiry, and this can be seen as "implicit guidance" that bounds the student's construction process of her/his model. Still, at the stage at which the evaluations were conducted, several technical as well as conceptual questions regarding the grounding process remained open for further discussion.

3.5.2. Model-based recommendations

A major feature in DynaLearn aiming to support students' independent learning is the model-based recommendations. This feature and functionality were made available gradually for evaluation in the last stages of the project, and activities were conducted while their development continued.

The main questions addressed in the evaluation activities were:

- [UHULL5 - 1G/Rm]:
  - Does the semantic technology provide students suitable recommendations for model improvement?
  - Does the automatic algorithm for reference model selection provide suitable feedback in comparison to manually selected reference models?
  - Do the students understand the recommendations they receive from the semantic technology?
- [BOKU5 - 1G/Pr-Po]:
  - Do the semantic technologies and model-based recommendations affect students' conceptual understanding and knowledge gain?

As mentioned, the technology was assessed in the evaluation activities even though its development has not been completed, and both technical and conceptual problems interfered the running of the activities. In this sense, the evaluation results became important source of data for informing the developers about issues to be addressed in order to improve the offered features.

- [UHULL5 - 1G/Rm] About the recommendations generated and their effect on learning:
  - Of the feedback generated for each student on average only just less than half (49%) of items could be evaluated as being correct and relevant to the students model.
  - The largest source of incorrect suggestions was in the **terminology** category (Difference). Although terminology suggestions were evaluated as incorrect for a number of reasons, the main source of error was deemed to be terminology suggestions with no apparent logic (two seemingly unrelated items). **Extra term** suggestions made up around 17% of all suggestions generated, and on average 78.6% of the suggestions were correct and relevant to the student model. **Missing term** suggestions made up 28% of the suggestions generated and on average only 56.8% of the suggestions generated were correct. **Extra dependency** suggestions made up only 8% of suggestions, mostly due to the low numbers of dependencies implemented by the students in the first place. The majority of the extra dependency suggestions were correct. **Missing instances** made up 12% of all suggestions generated, however 100% of these suggestions were evaluated as dubious or incorrect.
  - Concerning students' perception of the recommendations, in general they scored the suggestions positively for understanding. However, in accordance with the uneven results summarized above, despite the relatively high level of understanding the students generally scored the suggestions, especially the difference suggestions, poorly for relevance in the context of their own model.
  - The models created by the students scored poorly both before and after recommendations with the students generally failing to implement any causal statements about the key concepts predefined (for a model on osmoregulation in *Artemia*). The students made few if any changes to their models. No student managed to improve their model.
- [BOKU4 - CS] The feedback significantly improved the models of the learners. The missing terms functionality worked well and was especially used to determine missing entities and quantities. The dependencies feedback was mostly completely arbitrary and wrong.
- [BOKU5 - 1G/Pr-Po] - Pre and post results collected in a modelling task before and after using the feedback feature:
  - The evaluation of the 'Feedback' showed that the quality of students models compared to an expert model increased in a very short time period (20 min.). The comparison of the eight LS2 models before and after feedback from the expert model showed a significant progress and improvement of the students' models. The number of entities and especially of quantities and correct causal relationships increased significantly.
  - Students liked the feedback functions, but indicated a lack of clarity in the structure of the feedback window.

The above results lead to several conclusions. In one hand they are indicative of the still uneven quality of the recommendations generated. In the other hand they unveil the potential of this feature for supplying students with valuable information to improve their modelling performance.

Important points needing further and thorough examination relate to the extent to which: (a) relevant recommendations help students in their modelling process; and (b) students master the appropriate skills and strategies required to evaluate the appropriateness of the recommendations and their potential contribution to their modelling.

The situation in which students make decisions about how to proceed on the basis of feedback they receive from the system, unveils substantial questions (to be further studied) about students' perception of the feedback. One important issue is the tension between students perception of the authoritative nature of the feedback (i.e., "the computer is always right"), and the need for their critical evaluation of the feedback to adopt only aspects that seem to advance understanding and modelling. As well, an equally important issue relates to the pedagogical side of the situation, about the tension between a teaching process oriented towards the right model (hence feedback should guide to replicate a given expert model in the repository) or towards mindful modelling supported by the critical evaluation of reference-model-based recommendations.

3.6. Contribution of the interaction with the VC to learning

Evaluation activities focusing on students interactions with the VC were conducted mainly during phase two of the evaluation, when these features became available. Two main foci of these activities were the "basic help", and the "teachable agent" (TA).

3.6.1. Basic help

Evaluation activities involving the "basic help" features were conducted mainly by UHULL and BOKU. The main goals of the evaluation were to assess whether these features:

- Enable students to independently learn how to build models
- Support the model building process itself

Overall, the evaluation results indicate that the "basic help" features need further development and refinement in pursue of the goal of supporting the independent modeller. Sample observations about the extent of use of the features and their contribution as perceived by the students are:

- [BOKU5 - CS] In all LS, help was needed to explain the modelling terms. As the basic help only explains "How to" and not "What is" at logical and conceptual levels, the required help was requested from the teacher.
- [UHULL5 - 1GRm] Students made little use of the "How to" support during their model building activity. This could be the result of students' lack of technical knowhow of using this support. The majority of these regarded how to add ingredients to their model. However, in most cases the students reported that the instructions were clearly understandable. The fact that the students made only a limited number of calls on the support technology indicates that the technical aspect of the model building process is fairly intuitive and that the contextual and step by step help is easily understood. However, the general comments provided by the students indicated that the technical know-how was not sufficient to support them building models and that given their lack of model building experience they needed much more support concerning what to do and how to go about building models from a **conceptual** viewpoint.
- [BOKU5 - CS] Students’ comments addressed the need for more detailed information addressing aspects at higher levels than the merely technical, e.g., "the hamster just said obvious things, but not say why something went wrong or what you can do instead", "hamster just gave little information, no detailed information or examples for better understanding".

3.6.2. Teachable agent

Evaluation activities focusing on the Teachable Agent (TA) were implemented in preliminary form in the last stages of the first phase, but more centrally in the second phase of evaluation. Results were collected in a varied range of activities, with different populations and samples. A summary of the main insights follows.

- [UHULL4 - E/C] The main goals in this activity were to examine the contribution of students' interactions with the TA to their understanding of the system's behaviour and to the quality of their models. In a more specific issue, the contribution of a particular question-format to students' causal reasoning was examined. Overall, no significant difference was observed between the experimental and control groups in the different aspects examined:
    - The analysis of the students' written tests did not reveal any significant difference between the treatment and control group. A probably reason for this result is the previous knowledge which both groups already had on the topic under study.
    - No difference in favour of being scaffolded by "if, then" type questions for building causal arguments was observed. It is most likely that a single exposure to this learning mode is not effective for supporting the development of the target skills.
    - The models created by both control and experimental did not show significant differences. Students from both groups left out quantities and had errors in building their model. TA scaffolding did not improve the quality of the models. It can be concluded that the short term intervention is insufficient to affect in significant manner the model construction process.
- [TAU5 - 1G/Pr-Po] The degree of match (or mismatch) between students' and expert models of a marine ecosystem, as reflected in the TA's answers to a quiz, indicated what aspects in the student model require revision and modification. Overall students obtained maximum scores for the inclusion of all required entities. To a lesser extent they succeeded in specifying quantity spaces and quantities (75%), and the groups mean for defining correct causal relationships was even lower (45%). Considering together these findings against the positive perception of the VCs expressed by the students, it seems that students subjectively perceived the value of the VC-based tasks higher than its actual effect on their work on the models (evident motivational effect).
- [FUB: 12 - 1sG/Rm; 17 - 1sG/Rm; 26 - 1G/Pr-Po] In a series of activities conducted by FUB with small groups and in one case with a bigger sample, several insights were obtained:
    - There were indications of increase in understanding and quality of models built (e.g., an increase in the percentage of questions correctly answered by TAs on the Quiz).
    - However in the bigger sample, no significant difference was observed between pre and post measurements following the TA-based tasks.

The above results are in line with previous work showing that while the work with VCs is perceived by students as pleasant, helpful and motivational, no significant improvement on learning took place (e.g., see Conati & Maske, 2009, who conducted a study comparing students learning with versions of an educational game with and without virtual agents). In our evaluation activities moderate effect on learning outcomes was observed in several cases; however the main effect of the interaction with the VCs can still be found at the motivation and attitudes levels.

3.7. Self-directed learning

One of the goals for the development of the interactive learning environment, DynaLearn, is to support learners in acquiring gradual mastery of modelling and system thinking skills and becoming independent learners. Creating the conditions for, and supporting the actual construction of, self-directed learning skills by the students is undoubtedly a serious educational challenge. Along the evaluation activities, we collected important insights about situations in which aspects of this goal have been achieved, and about other situations indicating what software and pedagogical development efforts are still required to fulfil the goals. Examples of both types of insights follow.

- [TAU6 - CS] The case-study allowed us to witness the course of ongoing reflection "in" and "about" action of a student (an experienced modeller) while building his models - reflection took place **while** constructing the models **about** aspects of the modelling process and the thinking involved From the 'think aloud' protocol it can be concluded that substantial learning and reasoning process took place. This was reflected in the student's modelling products as well as in reflective statements that dealt with the logic behind modelling. The students' comments related to the meaning of the actions taken. This denotes the students' ability to think along the modelling process about "what to do" as well as "why to do that" - or thinking-in-action (Schon, 1983). As well, the reflective statements show the student’s ability to examine critically their own work and discard actions or decisions taken considered to be unnecessary. At the end of the session the student decided to move to what they considered "the real task": using the model to explore the system's behaviour, and in particular the question that triggered the whole model construction process. The fact that the student is an experienced modeller reinforces the claim that long-term recurrent modelling experience is necessary for the gradual construction of independent capabilities.

The features expected to contribute most to the development of independent learning capabilities are obviously those related to the help and feedback supplied by the learning system. Due to the evolving nature of the software development process, the various features were only available gradually and in different stages, thus imposing constraints to the design of the evaluation activities. In addition logistic aspects related to schools' and academic calendars imposed constraints in the evaluation time-tables. As a consequence, some features were only assessed in a preliminary state or with the interference of unexpected technical difficulties. In these cases the main contributions of the evaluations were at the level of insights about directions to be taken (both technologically and pedagogically) in pursue of the key goal of supporting self-directed learning. Observations of this kind as collected in BOKU and UHULL activities were:

- [BOKU5 - CS] Basic help did not provide enough to support for independent model building. More information is required, e.g., about model building strategies or affordances at each LS. The feedback window was experienced as not insightful enough. The inclusion of useful feedback about the entity structure and causal relations is recommended.

- [UHULL5 - 1G/Rm] Due to technical problems students interaction with feedback features were limited. There is still work to be done to improve the technical performance of the recommendation process, the relevance of the suggestions generated and the stability and capacity of the repository. There is a need for the software to support naive modellers through improved filtering and ordering of suggestions, support and meta-feedback on suggestions from the virtual characters and the need for a greater range of structured model building activities/scaffolding.

The evaluation observations reinforce the need to assess systematically the learning value of diverse types of help and support in correspondence with students needs at different stages of the learning process, to inform the further development of DynaLearn (e.g., see Or-Bach & Bredeweg, in press, about students approach towards, and use of, various support types). Overall, the observations suggest that future lines of work on help and support features (either solely by the software or as pedagogical modes with the software) should focus on the logical and conceptual aspects of the modelling process, in a modality that supports students' self-reflection about, and self-elaboration on, their own modelling process. There is need for additional layers of support above the layers of functional help and structure-based recommendations.

3.8. Software issues

Issues related to students' interaction with the software, such as the learning process of its features and functionalities or difficulties encountered, were addressed explicitly in some evaluation activities, and in others observations were collected as additional data alongside the main data collection.

It should be noted that the evaluations were spread over a long period, and during this time successive versions of the software were released in which previously lacking functionalities were completed or bugs were solved. Thus, data about focal (and temporary) technical issues collected in the first phase of the evaluation are no longer relevant, and are not mentioned here.

Overall, students in all activities conducted reached rapidly mastery of the procedures and methods required to work with the software. General observations from the partners' reports follow.

- [FUB] Students are very enthusiastic about the software, even saying that it is easy to model concepts in a model and indicating that they think the software can be widely applied different scientific disciplines.
- [IBER 1,2 - 1G/PrPo] Some students indicated that the software is difficult to use initially, but becomes easier to use in time. Other students indicated that they have no trouble using the software at all. Students indicated that they will use the DynaLearn software for the rest of their education.
- [TAU2 - E/C] High school students reached immediate mastery of the software's features and functions, even in shorter time than the observed in the undergraduate course.
- [BOKU] Students commented on what they liked most - "learning-by-doing".
- [UHULL1 - CS] Overall most students indicated that once they had identified the appropriate icons the software was fairly easy to use.

In the evaluation activities conducted towards the end of phase two, using the most updated versions of the software, several usability aspects were addressed. At this stage, most observations were related to three main categories: ease and clarity of usage, conceptual issues stemming from software features, and suggestions for further development.

- [BOKU5 - 1G/Pr-Po] Generally, the use of the software does not seem to be as self-explaining as it should be. Interacting with the interface was not experienced as being easy, which might be also related to the lack of appropriate support by the VCs. Interestingly the students rated the identification of entities and quantities as easy, although they had significant problems in building up the appropriate entity structure of the model with their associated quantities. In contrast causal relations and state values seem to be quite self-explaining.
- Two main technical difficulties raised in several evaluation reports relate to: (a) the software installation process ("still challenging"); and (b) technical difficulties with ST-based features (e.g., server problems, communication failures, slow responses).

- Concerning the VCs, there was consensus through the vast majority of the evaluations about their motivational nature. However, students' and researchers' perceptions were that they are most suitable for young learners (junior-high and perhaps high-school students).

Examples of conceptual aspects related to software and interface features were already presented in previous sections of this report, , e.g., about the need for higher layers of help addressing more conceptual and logical issues, or modifications in the grounding or feedback windows to support better interactions of the students with the underlying functionalities.

It should be noted that many observations concerning software features as reported in the partners' deliverables were correct at the time the activities were conducted. By the very nature of an ongoing development project, features were completed, modified or improved in later stages, often as a result of feedback obtained from the evaluation activities (see section 3.8.1). Examples are the significant improvement in interface features, or in the installation process.

3.8.1. Contribution of evaluation feedback to the development of software features

An important result of the evaluation activities were feedback, recommendations and requests related to software features. Based on these many features were modified or introduced in different stages of the development process. The feedback and subsequent software modifications pertained to several main categories. In the following a few examples are presented:

- Feedback about functional difficulties unveiled while planning and conducting the evaluations (e.g. difficulties with LS5 in the early stages of the project, or with the grounding). These were immediately corrected and modified.
- Feedback about aspects of the grounding process - resulted in modifications and inclusion of features (e.g., compound wording).
- Feedback about language localization needs - resulted in the inclusion of multiple-languages capabilities.
- Feedback about difficulties in the installation process - resulted in the development of the installer and its significant improvement in the last releases.
- Feedback about the quality of the interactions with the VC, r the model-based recommendations feature - resulted in continuous improvement of interface and interaction features.
- Feedback about the pedagogical added value of software features - resulted in their development also as learning resources (e.g., grounding as learning task; pattern models as instrument for both feedback and instruction).
- Requirements based on research purposes - resulted in the devise of appropriate features (e.g., data export on semantic feedback or the quiz, for evaluation purposes).
- Requirements of support material - resulted in its development (e.g., sets of video-clips focusing on software features or aspects of the model construction process).

This is by no means and exhaustive list. Many feedback/treatment situations were generated while conducting the evaluations and were addressed in very short time. Most feedback recommendations about software features can be found in the set of deliverables produced by WP7 partners, and the way these were considered for development purposes in the deliverables and actual results (the software) of the technology partners' work.

4. Pedagogical observations

In addition to the evaluation results, two important outcomes were produced during the evaluations planning and implementation process. The first is a preliminary repertoire of pedagogical modalities and solutions for integrating DynaLearn in teaching and learning processes. The second is a collection of evaluation instruments which in turn might become meaningful assessment tools.

In this chapter we summarise pedagogical insights originated stemming from two sources. The first are the evaluation plans, evaluation instruments, actual activities implemented and conclusions drawn about pedagogical issues as presented by the partners in the deliverables of the two rounds of evaluations. The second source are the responses of ten teachers who taught with DynaLearn, collected using a questionnaire administered towards the end of the project (see Appendix --B).

4.1. Teaching with DynaLearn - lessons learned

This section summarises pedagogical aspects and conclusions as presented by WP7 partners in the deliverables. The vast majority of the evaluation activities were conducted in "real-life" learning settings (as opposed to "lab-like" settings), being these regular or specially developed courses for High School or University students. The immediate implication is that pedagogical procedures and processes had to be planned for running the activities. For this purpose, lesson plans including a range of pedagogical ingredients were devised, among others:

- Ways to introduce students to the approach and main concepts of qualitative reasoning and conceptual modelling.
- Support for the gradual activation and consolidation of system thinking skills.
- Pedagogical sequences supporting the gradual acquisition of modelling capabilities using DynaLearn.
- Strategies and methods for helping the students in their modelling processes.
- Repertoire of prototypical examples for introducing DynaLearn features at each LS.
- Alternative ways to trigger the modelling process (e.g., working on questions described in a scientific paper, or on a situated dilemma).
- Ad-hoc activities for working with specific features (e.g., grounding or TA).

In general terms many conclusions were drawn by the partners about the need to develop a comprehensive pedagogical approach and practical implementation guidelines for DynaLearn. Some are presented here:

- [UHULL] Students need support concerning what to do and how to go about building models from a conceptual viewpoint (i.e. they are unlikely to ask how to do something unless they know what they want to do or should be doing). As such naive students will need suggestions about what to do at an early stage of learning how to model. To achieve this, students will need structured activities/scaffolding to learn about the concept and process of modelling.

- [BOKU] There is need for motivational and challenging activities, learning units linked to the curriculum and to local demands. As well, training on good modelling practices and the use of modelling resources (e.g., generic modelling patterns) is required.

- [BOKU and others] For teachers - thorough teacher training plans as well as appropriate readymade pedagogical building-blocks are required.

- [TAU] A pedagogical approach based on students' construction of knowledge and skills by being engaged in modelling processes is required. The activities and lesson plans should afford, besides the active involvement in doing, reflection "on" and "about" action (i.e., **while** modelling **about** the modelling process itself).

A recent development highly relevant to the development of appropriate pedagogies for DynaLearn are the constructs developed in WP6 - "model patterns" (Salles et al., 2012). Developed within the context of the curricular work for DynaLearn, these constructs might become powerful cognitive schemas for approaching systems analysis and modelling. This approach obviously demands the development of appropriate pedagogical strategies and didactical means.

4.1.1. Sample pedagogical modes developed and implemented

A pedagogical mode used in many of the evaluation activities was based on a collaborative configuration, in which students worked in dyads. The modelling process provided collaboration artefacts, serving as anchors for discussing, justifying and explaining the models and simulations. In addition, the idea of implementing structured collaboration was explored in a study conducted in the University of Amsterdam, using the Pair Modelling technique (in this technique, the partners are assigned roles -i.e., "modeller" and "reviewer"- changing them alternatively during a session). The study (Or-Bach & Bredeweg, 2011) showed encouraging results based on assignments scores, observations and a questionnaire. Students' attitudes were neutral on average, but the average score of the group that employed Pair Modelling was significantly higher than the average score of the control group that employed unstructured pair collaboration.

Another mode used in the activities was devised as translation of information from scientific papers, presented in conventional form (e.g., research questions, methods, results, discussion) into models using the CM approach and language and resources in DynaLearn. This mode requires the students to analyse an original scientific text, reconstruct the phenomena under study in terms of the modelling environment, and reformulate the questions to allow their exploration using the constructed models and their simulations.

Interesting variations of model manipulation tasks were devised for different evaluation activities. In a study by BOKU students got feedback about their models from two sources: DynaLearn model-based recommendation, and a peer student. By this, students engaged in intense discussions about each other's models, in aspects such as terminology or missing ingredients. In several activities by TAU the grounding feature was implemented while completing a "blind model" - students were supplied with the "skeleton" of a model and a "resources box" of model ingredients and types of relationships. They were requested to complete the blind model before the grounding procedure, and revise it once again in light of the new information obtained during the grounding. In an activity by FUB students were requested to debug a faulty model. Model debugging was implemented to support better understanding of processes in the modelled phenomenon. These are but a few examples, and many more are presented in detail in WP7 deliverables.

4.1.2. Data-collection instruments and their value as assessment instruments

An important pedagogical outcome or by-product of the evaluation activities is the set of instruments used for measurement and data-collection during the evaluation. These can now be refined and adapted for use as pedagogical instruments for assessing students' learning. It should be noted that in many aspects the development of evaluation activities and instruments had to be done "from scratch", as previous research literature on many of the issues examined is neither abundant nor consistent. While the measurement of learning gains in terms of content knowledge using structured instruments is of common practice for decades, the measurement of the gradual construction of System Thinking, QM skills, and scientific skills required to cope with complexity still demands significant research effort.

During the evaluation activities we have developed a set of instruments and scoring guides for analysing students concepts maps, models of various levels of complexity, video-data, open texts and structured questionnaires that comprise a valuable methodological infrastructure. On this basis it is possible now to refine these instruments and develop new ones along similar lines for pedagogical use.

Examples of instruments specifically developed for DynaLearn evaluations are:

- Content-related questionnaires used for pre- and post-test measurement of students' knowledge.
- Motivation questionnaires - which evolved in stages in correspondence with DynaLearn features implemented at each stage.
- Evaluation tasks, designed to serve for both learning and assessment (e.g., the "blind model" task used for grounding [TAU4,5; FUB19]; "debugging a model" task [FUB13]; or assignment worksheets [BOKU5; UHULL5]).
- Scoring schemes and guides for assessing structural and dynamic features of students' concept maps and models - these instruments were refined over time to include focal criteria about issues such as structural configuration, processes, causal relationships, or guiding principle for constructing a model (e.g., systemic perspective built upon actual content, or formal/taxonomic perspective imposed to content). These scoring schemes can be easily adapted to serve as powerful assessment instruments of students gradual development of system thinking and modelling skills.
- Analysis procedures and scoring guides for video data using qualitative content-analysis software, as well as criteria for analysing student think aloud and reflective protocols, have been developed for the studies. Many of the criteria and categories defined for characterizing students performance, might be adapted for the design of assessment instruments.

4.2. Teaching with DynaLearn - expert-teachers questionnaire

A group of eleven teachers, who were involved in the different evaluation activities using DynaLearn, were asked to report about their experiences. The group consisted of six teachers who reported four or five years of experience with QR modelling, considered expert teachers; three teachers with only three years of experience - semi expert teachers; and two novice teachers with only one or two years of experience.

A set of questions included in a semi-structured questionnaire guided their reports. The questions posed to them were related to several pedagogical issues. The questionnaire (see Appendix B) comprised 4 main sections:

1. **Personal view** of the rationale for using DynaLearn for teaching Science and other subjects. Questions in this section relate to the teachers' vision of the pedagogical justification, expected added value and pursued goals entailed in the use of DynaLearn for teaching. Among these are the perception of Systems related knowledge and skills fostered by DynaLearn (Appendix B, questions 3-6).

2. **Pedagogical issues**. Five questions in this sections aim to unveil teachers' considerations about a series of pedagogical aspects: prerequisite knowledge and skills required to learn with DL; pedagogical foci at each Learning Space; pedagogical added value of DL various features and learning tools; characteristics of various teaching modes implemented with DL (Appendix B, questions 7-11).

3. **About the learners' learning**. Three main questions addressing: the extent to which learning goals related to system thinking skills and worldview have been achieved; evidence of students' independent and self-directed learning; characteristics of students' learning process with DL (Appendix B, questions 12-14).

4. **Difficulties and problems** faced during the teaching processes; either pedagogical or technological (Appendix B, questions 15, 16).

In the following sections we present a summary of the insights supplied by the teachers (a more detailed account is beyond the scope of this report, and is currently under preparation for inclusion in a journal publication).

4.2.1. Perception of the pedagogical rationale for using DynaLearn

Most teachers perceive DynaLearn as tool for supporting learning as well as the development of high-order skills. The learning environment is perceived as affording processes typical of constructivist/constructionist pedagogies. In teachers' words:

- "Learners engage in activities that require answers and solutions for problems - learning by doing and learning by problem solving"; "… to support the development of a causal explanation … to enable students to gain a better understanding of systems/problems"; "… to develop hypothesis forming skills with consistent evaluation of possible behaviours, and explanations for scientific observations"; "Dynalearn can help students organize the main concepts in a system and the relationship between them. In higher levels it enables to represent a phenomenon, hypothesise and test it".

Less experienced teachers focused on less comprehensive aspects, more attached to functional issues such as motivational (e.g., "enhance attention and involvement in science learning") or representational (e.g., "new means for representing systems or scientific modelling") aspects.

When asked about the pedagogical added value of LbM with DynaLearn, teachers emphasised aspects in the learning environment in clear affinity with their perception of DynaLearn's rationale (previous question). In most cases, the answers referred to features that support students' development of system thinking and inquiry skills, of their ability to understand and explain causal relationships, and their acquisition of a language for representing systems and exploring their behaviours. Among their answers are:

- "The main added value is related to developing systems/causal thinking enabling students to develop hypothesis forming skills with consistent evaluation of possible behaviour and explanations for scientific observations"; "Reasoning about quantities without numbers; a powerful language for conceptual modelling based on a relatively small set of elements; the possibility of building models at different levels of complexity and the progression from simple to more complex models".

Also for this question, less experienced teachers highlighted features at a more functional level:

- "Developing learning skills by recognising causes and effects in ecological processes". "A better insight in the relations between the variables that play a role in a process - DynaLearn is a sort of dynamic concept map".

Teachers were asked to rank the importance of four main goals within the rationale of LbM with DynaLearn. Their answers are summarised in Table 4.1.

Table 4.1: Teachers' ranking of main goals for LbM with DynaLearn (in a scale of 5)

Importance of goal	Experts (N=6)	Semi (N=3)	Novices (N=2)	Mean
Fostering a systems worldview	4.5	4	4	4.2
Acquiring system thinking skills	4.5	3.3	4	3.9
Acquiring modelling capabilities	4	2.7	2.5	3.1
Mastering content/subject-matter	4.2	4	3	3.7

Consistent with their previous answers, teachers ranked higher the goals related to fostering the acquisition of a systems worldview and system thinking skills. While expert teachers ranked evenly all goals, less expert teachers clearly valued system-related goals higher than these related to modelling capabilities or mastery of specific contents. Lowest ranked (particularly by less experienced teachers) was the goal supporting the acquisition of modelling capabilities.

The ranking reinforces teachers overall perception of the potential of DynaLearn to open new learning venues for coping with scientific phenomena at with a systems perspective. Their comments to the goals complementing the ranking are very explicit in this matter:

- [supporting] "The ability to understand the relationships between concepts and how each one effects the other and the whole system"; "The structured approach of QR enables students to have a structure for reasoning and a vocabulary for describing causality and for how systems work"; "Conceptual modelling makes a contribution to mastering content/subject matter but only when used in a holistic approach with other methods and if the models built/used are grounded/situated in a real context for the student"; [about the systems worldview] "Things are related and influence each other in different ways with different strengths; biological systems often have history, they might act differently to the same input, while physical systems are expected to perform similar all the time"; "fostering a way of viewing the elements of the world or reality interconnected. Is the way we understand how elements can influence and change other elements of the world".

4.2.2. Pedagogical issues

The next series of questions answered by the teachers related to various pedagogical issues, e.g., prerequisite knowledge, goals for using each Learning Space, or the contribution of DynaLearn features to learning.

Concerning **prerequisite knowledge** considered necessary for LbM with DynaLearn, teachers answers covered a wide range of aspects. Sample answers were:

- Expert teachers: Knowing the QR approach and its inbuilt reasoning logic; knowing the modelling language and standards; knowing the modelling tools.

- Semi-expert teachers: Knowing how things are interrelated (conceptual knowledge) and knowledge about mental models and concept maps; minimum conceptual knowledge about the phenomena modelled.

- Novice teachers: Software capacity for model building; knowledge of systems elements and how to use the software.

While expert teachers consider mastering the QR language as the most important prerequisite, semi expert teachers focused more on conceptual understanding, and novice teachers on software-related capabilities.

DynaLearn comprises six LS, each space affording a different way to address the modelling tasks in increasing level of complexity. The teachers' ideas about the main goals to be pursued in each LS are summarized in Table 4.2.

Table 4.2: Affordances and goals in each LS as perceived by the teachers

LS	Teachers answers
1	• **Experts**: Starting point; to extract key knowledge from stimuli texts/learning materials and organise thoughts and ideas outside of a formal modelling language or approach; Organizing ideas; freedom to represent concepts in a simple way; helps to organize thinking • **Others:** To get used to concept maps; understand relations between entities; learning basic and important concepts
2	• **Experts**: To work on model structure, entities, quantities and basic causal relationships; to capture tendencies of change; to allow understanding of simple dynamics • **Others**: General reference to working with entities and quantities
3	• **Experts**: To explore key system states and threshold values; to introduce the concept of magnitude of a quantity; to understand changes in derivatives and magnitudes; to support conceptual understanding, causality, system thinking • **Less-experts:** General reference to links between quantities; coping with complexity with useful tools
4	• **Experts**: The most compact LS for developing basic models with notions of causality and capacity to consistently simulate; to differentiate processes and propagation; to represent feedback loops; • **Others:** Allow coping with increasing complexity; feedback loops
5	• **Experts**: Working with concepts requiring conditional knowledge; allow making explicit conditions for things to happen; allow building more realistic models; • **Others:** supports the idea of conditions; in general, affords the development of complex modelling skills and capability
6	• **Experts**: Most suitable for research projects; affords reusability of model fragments; reusability of knowledge and representation of hierarchy and inheritance; possibility to run different scenarios; most powerful/complex LS - full strength of DL • **Others:** General answers about LS affordance to coping with complexity

In general teachers were able to address the main affordances and strengths of each LS, and their potential contribution to the gradual development of understandings and skills. Understandably, while expert teachers supplied more detailed and knowledgeable answers (in relation to specific features of the LS), less expert teachers' answers were formulated in more general terms.

Asked about constraints or limitations in each LS, these were mostly addressed by expert teachers, and generally refer to two main issues: conceptual constraints (what is/is-not afforded in terms of modelling possibilities at each LS), and interface features perhaps to be addressed in further development stages (actually many aspects were already modified in the latest releases of the software).

Teachers were also asked to elaborate on the added pedagogical value of different **features** of DynaLearn. Answers were provided mainly by expert teachers. Teachers with less experience supplied few answers, and in many cases indicated only that a given feature was not implemented in their teaching. A summary of teachers answers follow:

- **Conceptual modelling**, affords: Conceptual understanding; thinking deeply and causally about systems; understanding a system's behaviour; critically evaluate consistencies in their ideas about systems; working on smaller pieces then joining them together.

- **Teachable Agent**, affords: Motivation; self-directed learning; opportunity to correct erroneous representations; reinforce modelling. Characters and challenges fitting best lower level and young students.

- **Quiz**, affords: Thinking causally about systems; motivation; self-regulated learning. Questions were raised as to whether questions should focus on meta-aspects of the system and its behaviour rather than only on the model and its design.

- **Diagnosis**, affords: Confronting students with the logic of their expectations; self-directed learning; motivation; valuable support for naïve modellers.

- **Feedback**, affords: Identification and correction of modelling errors; expanding ideas about the modelled system; motivation; reflection; self-directed learning. Concern about current interface and process and difficulties for naïve modellers.

- **Basic help**, affords: help in key steps of the modelling; motivation; self-directed learning. Remarks: "What-is" function considered not too relevant for the students; "Why?" function potentially most helpful.

In general, the added value of conceptual modelling is regarded in most cases as contributing to conceptual understanding and improved understanding of the systems behaviours. The main added values of the other features (Teachable Agent, Quiz, Diagnosis, Feedback and Basic Help) are at the levels of motivation, self-regulated learning, and assistance in correcting or improving the models.

In the final question in the pedagogical issues section teachers were requested to rank different teaching configurations or modes, and add their comments about each mode. The ranking is presented in Table 4.3.

Table 4.3: Teachers ranking of different teaching modes

Teaching modes	**Experts** (N=5)	**Semi** (N=3)	**Novices** (N=2)	**Mean**
Independent modelling task	3.2	2	2.5	2.6
One modelling task across all LS	4	4	4	4
Framing reference model	4.2	4.3	4.5	4.3
Not framed by reference model	3.5	5	4.5	4.3
Embedded in curricular activities	4.2	4.7	3	4
Not related to regular curriculum	2.7	2	3	2.6

In general, expert teachers prefer evolving modelling tasks, modelling processes compared to a reference model, and modelling activities embedded in the school curriculum. Less expert and novice teachers also prefer an evolving modelling tasks, and equally prefer modelling processes either framed or not by reference models. Overall, independent (one-time) tasks, and tasks detached from curricular frameworks, are the less preferred teaching modes.

In the context of the learning setting (individual, small group, large group), the majority of the respondents ranked highly working in small groups and in dyads, stressing the importance of: (a) the need to support closely all (individual) students' learning, and (b) the value of the interaction with peers for the reflection and modelling process.

4.2.3. The learners' learning

Teachers were requested to rank their perception of students learning and performance in relation to the main goals ranked in the first part of the questionnaire. Novice teachers did not answer to this question. The ranks appear in Table 4.4.

Teachers ranked their perception of students' attainment in similar pattern than their ranking of the importance of the goals (see Table 4.1). While expert teachers ranked attainments slightly lower than importance of goals, less experienced teachers ranked attainments slightly higher (except for mastery of subject matter). In general teachers perceived the best achievements to be those related to a systems worldview and skills, and the lowest achievements to be those related to modelling capabilities.

Table 4.4: Teachers ranking of students attainments

Students learning	**Experts** (N=6)	**Semi** (N=3)	**Novices** (N=2)	**Mean**
Fostering a systems worldview	4.2	4	--	4.1
Acquiring system thinking skills	3.7	3.5	--	3.6
Acquiring modelling capabilities	3.3	3	--	3.1
Mastering content/subject-matter	4	3.5	--	3.7

Concerning the idea of supporting self-directed learning (SDL), only part of the teachers referred explicitly to it. In general, the features of diagnosis and feedback were regarded as important factors affecting the pursuit of this goal. Typical descriptions of evidence for SDL in students' performance were:

- "Students looking for literature support to develop their models; proposing and planning new models or extensions on models they were building; correcting errors in their models by themselves; exchanging of modelling experiences with peers".
- Students started to integrate personal viewpoints in LS2 and LS4, and also were stimulated to look for different issues in the internet stimulated by the modelling activity".
- "Building concept maps (LS1) before starting the modelling activity; producing sketches in LS2 and LS4, before moving to LS5 and LS6; activating and de-activating model fragments in LS6 to test alternative representations; advanced modellers using assumptions to activate and de-activate groups of model fragments and test options on how to improve their models".
- "The feedback and diagnosis is the most important feature that gave students support for independent, self-directed learning. In my experience, when I compared the level of acquisition of knowledge between students with modelling activities and regular classes we saw clearly their SDL capability".

The last item in this section, focused on perceived difficulties faced by the students during the modelling processes. These were described by teachers as pertaining mainly to three categories:

- At a general level the challenges encountered related to the acquaintance with the systems approach, the qualitative approach, the modelling language, and the conceptual modelling process.

- At a more specific level difficulties were raised concerning the modelling of causality; implementing conditional knowledge, assumptions, and attributes; working with model fragments - specifically mentioned in several cases was the understanding of Is and Ps.
- A third category of difficulties related to features of the software, mainly at the interface level, as well as the transitions between Learning Spaces.

Most teachers expressed their conviction that the challenges of the first two categories demand further elaboration at the pedagogical level, aiming to create models, learning sequences, progression of tasks and support materials to help students in assimilating the approaches and language fostered by DynaLearn.

4.2.4. Problems/difficulties faced

In an additional question teachers were asked to refer to the pedagogical and technological difficulties faced during their teaching experience. Their answers were mostly around three main themes:

- **As Teachers** - they related to teaching and pedagogical challenges: lack of time available to support all students; the demands of tutoring heterogeneous groups; their incomplete mastery of the conceptual approaches (QR, conceptual modelling); the demands posed by the need to master knowledge in many relevant aspects (e.g., content, system approach, modelling); the fact that LbM is a time consuming processes in comparison with other teaching modes.
- **Students** - in addition to the difficulties mentioned in the previous question, specific learning challenges mentioned were: students difficulties in understanding the tasks; in translating scientific texts information into models; in knowing what to do with the feedback received and use it to improve their models; difficulties in managing the complexity and scope of tasks and processes.
- **Technological difficulties**: Interface features; slow response from repository; (lack of) transference of models between Learning Spaces; crashes while working; permissions issues; faulty functionalities; installation difficulties.

Completing this section of the questionnaires teachers suggested recommendations for coping with the technological difficulties.

5. Concluding remarks

This document summarizes WP7 evaluation results obtained in activities conducted in two phases of evaluation during the years 2010-2012. All along the process, the main goals of the evaluation activities focused on: (a) assessing learning gains resulting from students interaction with DynaLearn, at three main levels: their conceptual understanding of ecosystemic phenomena, their growth of a systems worldview and system thinking skills, and their motivation to learn; and (b) informing the technology development partners about students' use of the software, and about requirements stemming from learning-related needs.

Due to the ongoing-development nature of the project, the evaluation has been mainly formative accompanying the different stages of the development of the software. Thus, the first phase focused mainly on the conceptual modelling workbench in its different LS, and the second phase on the ST and VC components of the learning environment.

DynaLearn is a complex project in which learning/pedagogical and technological aspects are intricately intertwined. Many of the initial questions, originally formulated in general terms, evolved into issues comprising multiple dimensions as we went into the planning and implementation of the evaluation activities. Examples of key learning and cognitive aspects that became emphasized were the growth of system thinking skills, of the capability to represent structural and dynamic aspects of complex systems in models, or the ability to explore and predict systems' behaviours. Examples of key pedagogical issues that became highlighted were the design of appropriate learning tasks and opportunities for a range of specific aspects of the CM language, or of DynaLearn's LS. The foci of the evaluation activities reflect the complexity and the scope of the project's manifold aspects.

In this concluding chapter we abandon the details and specifics of the evaluations as presented in the deliverables and summarized in the previous chapters, to present the set of main insights obtained along the project. For this purpose we leave the level of the operational questions used in the evaluations, one level up to the project's general questions as formulated in the DOW (Bredeweg et al., 2008):

1. **Does the diagrammatic approach (as organised in the DynaLearn setting) actually allow learners to address more complex problems?**

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?**

About these questions, the most evidences -and the most conclusive- have been collected in the evaluation studies. DynaLearn has proven to be of great potential for supporting students learning of systems and complex phenomena. Along the evaluations we have observed DynaLearn support of students: growth of causal System Thinking; acquisition of scientific reasoning skills; ability to learn about complex ecosystems; gradual construction of content knowledge; gradual development of CM approach and skills.

Students of different age levels and backgrounds (e.g., junior high and high-school, undergraduate and graduate) were able to construct models, to represent structural features as well as processes relevant to the phenomena modelled, to run simulations and explore the represented system's behaviours, and to make predictions and hypothesize about its behaviours. Data collected in many case studies and in small-group activities in which thorough analyses of students learning were conducted, indicated that the CM approach and language have been assimilated by the students as powerful intellectual tools for approaching systemic phenomena.

At the same time it became evident that changes in perceptions and approaches towards systemic phenomena and achievement of significant learning gains demand time: short term interventions were of modest impact on students' conceptual understanding, conceptual change and skills acquisition. Equally important, the acquisition of CM skills and mastery of DynaLearn's language and features is also time demanding.

Concerning the learning curve of CM capabilities, as the learning activities required working in higher LS, students faced more difficulties and experienced more cognitive demand. Although students grasped in very short time the essentials of the work with the software and its basic features, successful coping with more complex modelling tasks and mastery of the higher Learning Spaces' features seem to require a continuous and long-term learning process and appropriate pedagogical support.

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? And why or why not?**

By the time the evaluation activities were conducted, they focused on students interactions with the VC mainly in the "teachable agent" (TA) and "basic help" modes (it should be noted that current releases advanced significantly the features and modes offered).

In general, students perceived positively the interaction with the VC. These were regarded mainly as motivational and pleasant. In several evaluation activities students indicated that the interaction with the VC contributed to their learning and to their modelling work.

However concerning actual conceptual learning, and the improvement of the models, contrasting observations were obtained. For the "basic help" mode, moderate or no effect on model construction or improvement has been observed. The main interpretation of the findings is that for supporting the model construction process, the offered help being closely tied to the features of the software or to the evident ingredients of the model is not enough, and help at a more conceptual and logical level is required.

For the TA mode, a greater mix of results have been obtained, and clearly its motivational value and the involving situations it affords are the strengths of this mode.

Considering together these uneven findings, against the positive perception of the VC as expressed by the students and in particular by school-age students, we can conclude that they perceived the value of the VC-based tasks higher than its actual effect on their work. This supports once again the view of the VC as rewarding motivational component.

Overall, it is suggested that interaction with the VC fit best the school-age population rather than the University students population.

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

Concerning the ST, evaluation activities in phase two focused on grounding tasks and the model-based-recommendations feature.

Through the evaluations, the added pedagogical value of the grounding process emerged. Grounding, a feature originally conceived as a stage in the repository models' indexing process, was incorporated into evaluation tasks to assess (a) students extent of mastery of relevant terms and concepts, and (b) if and how the involvement in defining the information layer for the model ingredients does affect conceptual understanding and the modelling process itself. This involvement proved to be a valuable learning experience, and its evaluation provided rich results about its contribution to learning, and about further development requirements.

The model-based recommendations feature was available for evaluation only up to a given stage of its development, and additional features were completed in later stages. This feature was conceived as substantial resource for scaffolding learning. Indeed, the evaluations results are indicative of the promising value of it for supporting learners in their modelling process.

The difficulties encountered (besides technical issues meanwhile addressed in further releases) were related mainly to lack of clarity and relevance of the recommendations, and to insufficient support for important aspects of the "learning-to-model" process. In many cases, the automated procedure for reference model selection did align the student model to appropriate reference models, although the accuracy and relevance of the matching varied greatly (e.g., because of limited common grounding).

Concerning the modelling process, it was concluded that currently the inbuilt help and the semantic technology focus on the implementation of the model and the evaluation of the model post implementation, rather than in specifically supporting learning about the modelling approach or the process of formalising domain knowledge into the conceptual modelling framework. Such support and scaffolding would need to be formalised in future work through dedicated use cases and structured learning plans supported by features in the software.

Based on the potential of the ST features to support learning as unveiled in many evaluation observations, a series of recommendations for the further development of recommendation layers focusing on the conceptual and logical aspects of the modelling process were formulated (see next section about future work).

5. Does the personal autonomy cause learners to be more motivated?

Numerous observations of the gradual transition towards a more independent learning modality were collected. The implications of the independent modality at the motivational level, besides the positive perception of the learning process with DynaLearn, is the development of the sense of gradual mastery of modelling capabilities (as observed mainly in the long-term activities), and readiness to transcend the boundaries of the given task towards broader conceptual realms (among the observations, e.g., working back-and-forth between the modelling environment and scientific texts; integrating into the modelling task pieces of knowledge acquired separately in previous disciplinary courses; identifying a new question emerging from a recently modelled phenomenon, and building a new model for it).

However, the transitions towards self-directed-learning is a long process, and our observations, although promising, are only the beginnings of it. This transition implies the design of a pedagogical model in which the "model constructor" responsible for her own learning stands at the centre of the scene. The constructionist idea stating that the students construct their inner world by constructing in the outer world demands a supporting pedagogy. Our numerous observations of learning processes with DynaLearn serve as promising background for the development of these pedagogies.

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

No real opportunity to assess the full integrated software was given as more features were added when the evaluations were over. The idea of better learning is composed from all observations, each time a different aspect.

How is "better" defined based in our evaluation results? Among the salient aspects are: Better in terms of fostering analytic and synthetic capabilities; in acquiring tools for understanding the systemic character of phenomena; in approaching scientific topics as innovative science inquiry does; in acquiring tools for representing the phenomena under study - for objectivise knowledge and understanding into a representation that in turn becomes object for reflection and critical analysis (debugging ones own thinking); comparing ones own knowledge to expert knowledge - that is the need to develop critical evaluation skills to improve ones own learning; internalization of DynaLearn's meta-vocabulary in the progressive work in each LS - constructing the inner aggregate of intellectual tools vis-à-vis the transition from one LS to the next.

This sample list is in no way exhaustive, and can be expanded with numerous specific observations included in the whole set of WP7 deliverables. We assume that future studies to be conducted with the full version of the learning environment, in comparison with the way learning takes place in traditional teaching, will supply more comprehensive answers to the question.

7. Are students more motivated to take on science curricula?

This question is manly relevant for school-age populations. Higher education students have made their choices, and we meet them when they are already studying in specific tracks or fields.

Concerning junior-high and high-school students, this is a challenging question considering the short term experiences they have had with DynaLearn in our evaluation activities. Long-term interventions as well as long-term follow-ups are required to obtain reliable data on students motivation and choices related to science learning in the high-school and beyond.

However, we do have information (from many motivation questionnaires administered) about their current perceptions about learning science and LbM with DynaLearn. In general, most students' perceptions and motivational stance towards the work with DynaLearn on science topics were positive. Across most evaluation activities, students perceived as main learning gain the change in approach towards complex phenomena, and the acquisition of intellectual tools (CM skills and methods) for addressing these phenomena from a systems perspective.

5.1. A look into further work

In the last section of this report we want to refer briefly to recommendations and suggestions for future work stemming from the results of the activities conducted. In general, these relate to three main areas: the software, pedagogical issues, and future research.

Concerning the software, the evaluators recommendations address two aspects. The first relates to the upgrading and completion of existing features used in the evaluations. As specified in WP7 deliverables, the recommended modifications relate to several functionalities, stability of the software, of the repository servers as well, interface issues and ease of installation. At this point, many of the issues raised were already fixed, modified or completed in current releases.

The second aspect and more relevant for future work focuses on features required to support learning and modelling processes at the conceptual and high order thinking levels. It is suggested that help and feedback functionalities should address conceptual and logical aspects of the model construction process - such as the identification and definition of relevant ingredients, the definition of appropriate hierarchies and causal configurations, or the interpretation of simulation results. The combination of feedback generated ad-hoc about actual performance, with traces of a student's modelling history, should support the generation of appropriate scaffolding (including, e.g., differential "density" of support offered, fading it out when mastery is assessed). Other recommendations stress the need for support appropriate modelling practices and strategies.

A second area of recommendations for further work is centred in pedagogical issues. Being obvious that the attainment of the learning and cognitive goals fostered demands long term learning processes in which students engage in tasks of increasing complexity, there is evident need to design appropriate pedagogical solutions of varied types, e.g.,: learning tasks and sequences; templates for specific pedagogical modes; support resources (e.g., media pieces along the lines of the video-clips already developed); instruction-based-assessment and summative-assessment instruments; guidelines and tools for adapting the instruction to the individual needs of students in heterogeneous populations; pedagogical modes integrating the work with DynaLearn with laboratory, field and digital-library activities. A powerful construct developed by WP6 are the model patterns (Salles et al. 2012). The design of patterns-based pedagogical sequences is a challenge worth to pursue in future work.

Finally, among the results of a complex project such as DynaLearn unavoidably are new challenging research questions to be addressed in further work. The questions naturally address -individually or in any combination- the three main components of a DynaLearn-based learning situation, i.e., software, learners, pedagogy. Examples of issues to be addressed in future research are:

Research and development of the future layers of DynaLearn supporting learners' modelling at conceptual and high-order levels of the modelling practice.

Long-term study of learners' appropriation of system thinking skills as cognitive tools for scientific inquiry.

Study of students' internalization of model patterns as generic intellectual tools for studying systems.

Long term study of alternative pedagogical modes based on LbM with DynaLearn in "real life" classes and settings.

Research and development of reliable assessment instruments focusing on the acquisition of system thinking skills, modelling capabilities and conceptual understanding, for implementation in regular classes.

We believe that future research should be conducted at both the theoretical and the practical implementation levels. In one hand we would like to know more about theoretical issues related to DynaLearn, in the different disciplines involved (e.g., learning and cognition, AI in Ed., systems thinking, ontological studies, innovative pedagogies). In the other hand, we obtained enough evidence of the learning potential of DynaLearn and its potential to advance innovative ways of learning science as well. Thus, we believe that implementation studies aiming to incorporate DynaLearn into regular educational settings will be of great contribution to the teaching and learning of science.

References

Assaraf, O., & Orion, N. (2005). Development of System Thinking Skills in the Context of Earth System Education. *Journal of Research in Science Teaching, 42*(5), 518-560.

Borisova, P., & Uzonov, Y. (2010). *IBER evaluation of DynaLearn prototype*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.2.3.

Borisova, P., Liem, J. & Uzonov, Y. (2012). *IBER evaluation of DynaLearn final software*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.3.3.

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

Draper, F. (1993). A proposed sequence for developing systems thinking in a grades 4—12 curriculum. *System Dynamics Review, 9*(2), 207-214.

Gibbons, M. (2002). *The self-directed learning handbook*. San Francisco, CA: Jossey-Bass

Miles, M., & Huberman, A. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Mioduser, D. et al. (2010). *TAU evaluation of DynaLearn prototype*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.2.4.

Mioduser, D. et al. (2012). *TAU evaluation of DynaLearn final software*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.3.4.

Noble, R. (2010). *UH evaluation of DynaLearn prototype*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.2.2.

Noble, R., & Cox, I. (2012). *UH evaluation of DynaLearn final software*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.3.2.

Novak, J.D., Gowin, D.B., & Johansen, G.T. (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education,* 67 (5), 625-645.

Or-Bach, R., & Bredeweg, B. (2011). Pair Modeling with DynaLearn – Students' Attitudes and Actual Effects. *Interdisciplinary Journal of Information, Knowledge, and Management, 6*, pp. 119-135.

Or-Bach, R., & Bredeweg, B. (in press). Support options provided and required for modeling with DynaLearn. *Education and Information Technologies*.

Papert, S. (1991). Situating constructionism. In I. Harel & S. Papert. (Ed.) Constructionism. Norwood. NJ: Ablex.

Ruiz-Primo, M.A., & Shavelson, R.J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33* (6). 569-600.

Salles, P. et al. (2010). *FUB evaluation of DynaLearn prototype*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.2.1.

Salles, P. et al. (2012). *FUB evaluation of DynaLearn final software*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.3.1.

Salles, P. et al. (2012). *DynaLearn curriculum for environmental science*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.6.5.

Schön, D.1(983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Zitek, A., Poppe, M., & Stelzhammer, M. (2010). *BOKU evaluation of DynaLearn prototype*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.2.5.

Zitek, A., Poppe, M., Stelzhammer, M., & Schuller, V. (2012). *BOKU evaluation of DynaLearn final software*. DynaLearn, EC FP7 STREP project 231526, Deliverable D7.3.5.

Zuzovsky, R. & Mioduser, D. (in press). Learning about ecological systems by constructing qualitative models. In Z. Liebman, *Constructivist Learning*, Tel-Aviv: Am Oved.

Appendix A:

Summary of each evaluation activity by each WP7 partner

BOKU Evaluation Activities

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Design & Instruments
BOKU1	LS1 – LS2 LS4	Wind energy production and its relation to fish in rivers	Upper secondary students in technical school	2		

Evaluation Spaces:

	1 CM	2 VC	3 ST
a CU	2		
b SR/M	1		
c M/A	3		
D SRL			

Evaluation Design & Instruments:

O_1	X_1	O_2	X_2	O_3	X_3	O_4
Pre-test	Introduction lecture	Concept-map	Lecture hands-on model-ing (LS2	Student's LS2 Model	Demon-stration of modelling (LS4 model)	Stu-dent LS4 mo-del post-test

Evaluation Questions	Results & Conclusions
1. Modelling behaviour and social interaction (usability) How was the behaviour and social interaction during the modelling work?	Student behaviour differed significantly between the different LS - more picking information, more thinking time in LS_2 - increased conversation time from LS_1 - LS_2 - LS_4 especially with the teacher - increased number of questions related to modelling from LS_2 – LS_4
2. Causal understanding and content knowledge. How did the causal relations change during modelling (graphical and verbal expressions)?	The use of causal relations increased by 91% from pre- to post-test Wrong causal relation did not occur in the post-test Increased match between students' models and expert models and indicating acknowledging the advantage of LBM for representing causal relationship and indicating growth in causal understanding.
3. Motivation	Positive feedback to all questions asked. Better conceptual understanding; perceiving modelling with the software applicable for learning other topics; challenging; modelling LS4 models contributed mostly to their conceptual understanding

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>BOKU2</td><td>LS2
LS4</td><td>Aquatic ecology and river manage-ment</td><td>Post graduate Students at BOKU University</td><td>21</td><td><table>
<tr><td></td><td>1
CM</td><td>2
VC</td><td>3
ST</td></tr>
<tr><td>A
CU</td><td>1</td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>2</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>3</td><td></td><td></td></tr>
<tr><td>D
SRL</td><td></td><td></td><td></td></tr>
</table></td><td><table>
<tr><td>O_1</td><td>X_1</td><td>O_2</td></tr>
<tr><td>Pretesting</td><td>Modelling activity</td><td>Post testing final content exam</td></tr>
</table></td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Feedback on usability
2. Causal understanding
3. Motivation</td><td colspan="2">Decreased wrong causal relations
Increased graphical causal relations
Slight decrease in the use of verbal causal relations
The topic explored with DL was amongst the bet graded question in the final exam.
Highest agreement for applicability of the software for other learning topics as well.
Liking, interesting, the modelling activity needs to build models in different LS.
The activity did not contribute to better conceptual understanding, but students developed a better focus and used more causal relations.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>BOKU3</td><td>LS1-LS4</td><td>River continuum and river catchment</td><td>Master Students</td><td>2</td><td><table>
<tr><th></th><th>1 CM</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td>1,2</td><td></td><td></td></tr>
<tr><td>b SR/ M</td><td></td><td></td><td></td></tr>
<tr><td>c M/A</td><td>3</td><td></td><td></td></tr>
<tr><td>d SRL</td><td></td><td></td><td></td></tr>
</table></td><td>Continuous documentation of modelling behaviour problems, attitudes and motivation.</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Usability issues
2. What problems learners encounter using the software
3. Motivational issues</td><td colspan="2">Comments on what students liked – learning by doing
Did not like – Lack of guiding manual for different LS
Comments on LS</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>BOKU4</td><td>LS6</td><td>-</td><td>Master students (2)
Post-Doc (1)</td><td>3</td><td><table>
<tr><th></th><th>1 CM</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td>1,2,3</td><td></td><td></td></tr>
<tr><td>b SR/M</td><td>4</td><td></td><td></td></tr>
<tr><td>c M/A</td><td>5</td><td></td><td></td></tr>
<tr><td>d SRL</td><td></td><td></td><td></td></tr>
</table></td><td></td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Usability of OBF
Correctness (equivalence) between terms in student model and the reference model
2. Quality of the models supported by OBF</td><td colspan="2">Difficulties in using OBF window (list of problems).
Alternative tools (Text, Metadata on models) can compete with the OBF functionalities
Lack of well developed guidance for the whole process of model comparison.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>BOKU5</td><td>LS2-LS4-LS5</td><td>Nuclear radiation and its effect on the environment and humans</td><td>High school students with previous experience in modelling with DynaLearn</td><td>2 Students (one male, one female)</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td>1c</td></tr>
<tr><td>b
SR/M</td><td></td><td></td><td>1b</td></tr>
<tr><td>c
M/A</td><td></td><td></td><td>1d</td></tr>
<tr><td>d
SRL</td><td></td><td></td><td>1a</td></tr>
</table>
</td><td rowspan="2">Assignment

1. Identification of adequate learning space for investigating a phenomenon.
2. Getting feedback from expert model and ground terms for dbpedia.
3. Interpret simulation.
4. Debug incomplete LS5 model.
5. Run simulation on the model they debugged.

Instruments

• Observations
• Video recording and analysis using "Transana" software) related to social interaction questions and answers.
• Pre- and post-test using "Atlas" software.
• Motivation questionnaire.</td></tr>
<tr><td colspan="4">1. The effect of basic help, feedback and grounding.
a. On student social interaction during modelling.
b. On student self-regulated learning.
c. On student increase in content knowledge.
d. On student motivation toward LbM.

Evaluation Questions
1. a. How did student modelling and social behaviour change in different LS?
b. Which questions were asked during the modelling and when was help needed?
c. How often did students activate the VC and were they using the VC answers?
d. How did students' causal understanding change?
e. ?</td><td colspan="2">Results & Conclusions
1. a. In LS1 students spent a lot of time picking out information (from teachers, internet, other materials. In LS2 students worked independently. During the grounding session, they again needed help. In LS4 students worked more that 60% of the total session time independently.
b. In all LS, help was needed to explain the modelling terms. As the basic help only explains "How to" and not "What is".
c. Students regard VC help as not useful.
d. Comparing pre- and post-tests for degree of causal understanding using "Atlas" software, show an increase of 66.7%, and for degree of abstraction, an increase of 50%.
e. Students liked LbM. The modelling enabled them to better understand the topic and motivated them to learn more. The grounding functionality and the VC were linked. However, they were not rated as useful for modelling.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>BOKU6</td><td>LS1,O LS2</td><td></td><td>University students studying toward master degree with previous content knowledge</td><td>31 master degree students from the University of Natural Resources in Vienna</td><td>
<table>
<tr><th></th><th>1
CH</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td>1b</td></tr>
<tr><td>b
SR/M</td><td></td><td></td><td></td></tr>
<tr><td>c
M/A</td><td></td><td></td><td>1c</td></tr>
<tr><td>d
SRL</td><td></td><td></td><td>1a</td></tr>
</table>
</td><td>Evaluation Assignment and Instrument
Five Assignments:
1. Drawing concept map of the river canalization phenomenon
2. 2. Creating a causal model at LS2
3. Using the feedback tool for modelling LS2 models.
4. Identifying model bugs in LS2 model.
5. Ground several terms of readymade LS2 model.
Instruments;
1. Video analyses – using "Transana" software.
2. Observations.
3. Model comparison – comparing models made by students before and after, using the feedback feature. Evaluation criteria: number of corresponding entities, number of causal relationships, verbally or graphically represented, match with expert model.
4. Pre- and post-test and final exams – using "Atlas" software. Marking quotations, linking them with codes, categorization of causal relations, determining degree of abstraction.
5. Motivation questionnaires – toward the software feature. 33 Likert-type items and 6 open items.
<table>
<tr><td>O_1
Model Analysis

Pre-tests</td><td>X_1
Modelling and using all features of the software</td><td>O_2
Model Analysis
Post-tests
Motivation questionnaires</td></tr>
</table>
</td></tr>
</table>

Evaluation Questions	Results and Conclusion
1. The effect of using Basic help, Diagnostic feedback. S.T. model building recommendation, bug repair feature on: a. Self regulated learning. b. Conceptual understanding and knowledge gains. c. Motivation for LbM using DynaLearn features.	1b) • Direct-verbal and graphic causal relations increased in the post-tests. • Students also use higher number of concepts in the post-test. • Degree of abstraction in student statements increased by nearly 10%. 1c) • Positive reaction to the grounding and feedback functions • Students did not find the VC useful, although they liked being helped. • The interaction with the software is not experienced as being easy. • Running simulations contributes to better understanding of the system behaviour. • Learning by modelling was positively appreciated.

IBER Evaluation Activities

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>IBER1

Pilot Study</td><td>LS1

LS6</td><td>Water cycle
Nutrient cycle</td><td>First degree students</td><td>5 students</td><td><table><tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr><tr><td>a
CU</td><td>2</td><td></td><td></td></tr><tr><td>b
SR/M</td><td>1</td><td></td><td></td></tr><tr><td>c
M/A</td><td>3</td><td></td><td></td></tr><tr><td>d
SRL</td><td></td><td></td><td></td></tr></table></td><td><table><tr><th>O_1
Concept map
(LS1)</th><th>X
Course activities</th><th>O_2
Final model
(LS6)</th></tr></table></td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Does the diagrammatic approach (as organized in the DynaLearn setting) actually allow learners to address more complex problems? (1b)</td><td colspan="2">At the end of the course, the students seemed to have acquired the skills to correctly distinguish different concepts in a scientific paper as corresponding to a particular model ingredient type in DynaLearn. That is which are part of the structure of the system, and which aspects are dynamic (the quantities). Furthermore, they seemed to be able to choose the correct causal relationship between the quantities. This is, they can successfully identify the processes that are important in the system.</td></tr>
<tr><td colspan="5">2. Does the meta-vocabulary from which a conceptual interpretation is built, enable them to construct more fine-grained analysis of how students work?(1a)</td><td colspan="2">-</td></tr>
<tr><td colspan="5">3. Are students more motivated to take on science curricula? (1c)</td><td colspan="2">Students are very enthusiastic about the software, even saying that it is easy to model concepts in a model and indicating that they think the software can be widely applied different scientific disciplines (and the education thereof).
Some students indicated that the software is difficult to use initially, but becomes easier to use in time. Other students indicated that they have no trouble using the software at al. Students indicated that they will use the DynaLearn software for the rest of their education.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>IBER2</td><td>LS1, LS4</td><td>Intensive agriculture</td><td>First, second and third degree students</td><td>10</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td>2</td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>1</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>3</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><th>O_1</th><th>X</th><th>O_2</th></tr>
<tr><td>Concept map</td><td>Course activities</td><td>Final model (LS2, LS4)

Motivation measures

Grades on final exam</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Evaluation Questions</th><th>Results & Conclusions</th></tr>
<tr><td>1. Does the diagrammatic approach (as organized in the DynaLearn setting) actually allow learners to address more complex problems? (1b)</td><td>All students performed very well on questions related to the topic explored with DynaLearn compared to other topics explored without DynaLearn.</td></tr>
<tr><td>2. Does the meta-vocabulary from which a conceptual interpretation is built, enable them to construct more fine-grained analysis of how students work? (1a)</td><td>At the beginning of the modelling activity students tend to use all entities and relationships they identified. At the end of the activity, they committed themselves to only those essential for the purpose of modelling.
Seventy percent of the students considered modelling at Learning space 4 to be very informative, but risky, since it is required to differentiate between direct and indirect causal relationships. Fifty percent of the students preferred to work at LS3.</td></tr>
<tr><td>3. Are students more motivated to take on science curricula? (1c)</td><td>Students responded positively to the motivation questionnaire. Found the software interesting and learning by models useful. The software can easily be implemented for learning other topics.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>IBER3</td><td>LS2, LS4</td><td>Biodiversity
Urbanization
Invasive and nature</td><td>High school students</td><td>16 Students aged 15-17
Effective sample 12</td><td><table><tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr><tr><td>a
CU</td><td>1</td><td></td><td></td></tr><tr><td>b
SR/M</td><td>2</td><td></td><td></td></tr><tr><td>c
M/A</td><td>4</td><td></td><td></td></tr><tr><td>d
SRL</td><td>3</td><td></td><td></td></tr></table></td><td rowspan="3">$O_1 \longrightarrow X_1 \longrightarrow O_2$
$O_3 \longrightarrow X_2 \longrightarrow O_4$

O_1 – Pre-test 1 on biodiversity
– Written statements and motivation questionnaire on LbM.
X_1 Modelling activity on "Loss of biodiversity" due to "loss of habitat" – LS2.
O_2 Post-test 1 on biodiversity.
O_3 Pre-test 2 – urbanization.
X_2 Modelling urbanization (unknown topic) LS4
Modelling invasive native species LS4
O_4 Pos-test – urbanization written statements and
Motivation questionnaire on LbM.</td></tr>
<tr><th colspan="4">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="4">1. Does modelling improve conceptual understanding of a known topic?
2. Does modelling of a known topic teach analytical skills? (i.e., identify key concept, structural elements-entities and configuration, behavioural aspects, processes and state variables.
3. Does modelling a known topic improve analytical skills and conceptual understanding of an unknown topic.
4. Does modelling improve motivation toward LbM</td><td colspan="2">1-2 In most cases (8/12) no significant improvement occurred between pre- and post-testing. In 4/2 cases a significant decrease between pre- and post-test occurred.
4. 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.
Learning by Modelling:
Most of the students think that working with DynaLearn is an interesting way to learning.
Simulation:
For most of the students the simulation results were interesting and also provided them better understanding of the behaviour of the system.
Defining the correct relations (Is or Ps) between the quantities was not an easy task.
Motivation:
For all questions in the motivation part of the questionnaire, the students gave very positive answers.
Our overall conclusions on the experiments are that interaction with DynaLearn strongly motivates the students. Moreover, students reported that they have gained a better understanding of the topic that they have modelled. However, in our experiments we have not observed this better understanding in the results of the students. In 3 of the 4 significant results, the students actually performed poorer. We expected that the short interaction time was the main cause of this confusion. The fact that the last post-test showed an improved understanding seems to support this expectation. However, for more conclusive evidence, more research is needed with longer interaction times.</td></tr>
</table>

UHULL Evaluation Activities

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
UHULL1 Pilot Study	LS! - Concept Maps LS2 – basic causal model LS4 – causal differentiation model	Global Climate Change	University Biological sciences Students	4 - 7

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU	1		
b SR/M	3,2		
c M/A	4		
d SRL			

Evaluation Design & Instruments

X_1	O_1	X_2	O_2	X_3	O_3
Stimu-lus material	Con-cept map written assign-ment	Lectur demo hands on modelli ng norm CM	LS2 Mod el writ-ten assig nme nt	Demo hands on modelli ng normal basic model	LS4 Model written assign ment

Evaluation Questions	Results & Conclusions
1. Do students easily understand the DynaLearn modelling language and approach? (1b) 2. Do students build better models when working in the formalized dynamic learning spaces? (1b) 3. Does working with dynamic causal models enable students to write clearer scientifically correct causal arguments? (1b) 4. Do students find it easy to work with the DynaLearn software?(1c)	1. Difficulties in mastering the modelling language specifically regarding "direct influences" and "proportionalities" [needed to plan intervention to overcome this difficulty]. 2. Difficulties in producing good representation using less formal concept maps. No clear result that supports the advantage of formalized models [modelling is viewed as a promising practice]. 3. No indication of improvement in student writing of causal explanation [based on written assignment]. 4. After acquaintance, the software was fairly easy to use [needed improvement to ease handling the software].

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
UHULL2 Grounding of terms in models	LS2	Photosyn thesis	Post graduate Certificate in Education Students	23 - 37

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU			5
b SR/M			1,2, 3,4
c M/A			
d SRL			

Evaluation Design & Instruments

PowerPoint Presentation	
Hands-on modelling using a normal model	
X_1 Using grounding X_2 Using free text comment boxes	O_1 models written assignments O_2 models written assignments

Evaluation Questions	Results & Conclusions
1. Does the current repository, DBpedia and grounding facility enable students to ground all terms? (3b) 2. How many terms would currently require the creation of anchor terms in the repository? (3b) 3. What type of terms require the generation of anchor terms in the repository? (3b) 4. How variable and correct are definitions provided by a range of students? (3a) 5. Does grounding (choosing from a predefined list) give students a better conceptual understanding of terms than requiring them to provide their own definition? (3a)	NO RESULTS – FAILURE OF THE GROUNDING FACILITIES

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>UHULL3

Open-ended modelling in learning space 2</td><td>LS2</td><td>Osmosis</td><td>Post-graduate Certificate In education Students</td><td>18</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>1,2,
3,4</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td></td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><td>0_1
Pretest concepts Causal argumentation</td><td>X
Powerpoint diagrams, lecture, hands-on modelling</td><td>O_2
Posttest Models Evaluated according to criteria</td></tr>
</table>
</td></tr>
<tr><th colspan="4">Evaluation Questions</th><th colspan="3">Results & Conclusions</th></tr>
<tr><td colspan="4">1. How often do naïve modellers make errors in implementation of their model ingredients? (1b)
2. How easily do naïve modellers identify the important components that need to be included in a causal model to represent a scientifically accurate causal argument?(1b)
3. How variable is model complexity between students in terms of number of different ingredients?(1b)
4. Do naïve modellers work at appreciably different rates during a modelling session?(1b)</td><td colspan="3">1. The vast majority pick up the structural aspects of a model. Errors occurred due to including processes in the configuration.
2. Students did not identify or were not able to implement all key domain concepts [personified ontology based feedback is needed to direct students about relevant concepts].
3. The models developed showed high levels of complexity in terms of causal relations [the task is dependent on previous understanding of the domain].
4. Large variability among the students dependent on some cognitive problem solving styles [needed help and feedback support].</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr>
<td>UHULL4

Influence of teachable agenda mode</td>
<td>LS2</td>
<td>Osmosis diffusion</td>
<td>Post-graduate Certificate in education Students</td>
<td>18 - 37</td>
<td>
<table>
<tr><th></th><th>1 CH</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td>2c 1c</td><td>1c</td><td></td></tr>
<tr><td>b SR/M</td><td></td><td>3c 3c</td><td></td></tr>
<tr><td>c M/A</td><td></td><td></td><td></td></tr>
<tr><td>d SRL</td><td></td><td></td><td></td></tr>
</table>
</td>
<td>
<table>
<tr><th colspan="5">Experimental</th></tr>
<tr><td>O_1 Pre test</td><td>X_1 Building a model using pre-defined ingre-dients inter-acting with VC</td><td>O_1 Model B1 criteria for analy-sing the model for causal depen-dence</td><td>X_2 Review the model following the student charac-ter ques-tions and their perfor-mance on a quiz</td><td>O_2 Model B2 criteria for ana-lysing the model for causal depen-dence</td></tr>
<tr><th colspan="5">Control</th></tr>
<tr><td>O_0 Pre-test</td><td>X_1 Building Model using a list of modell ing ingre dients</td><td>O_1 Model A_1</td><td>X_2 Running simu-lations</td><td>O_2 Model A_2 post-test</td></tr>
</table>
</td>
</tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2"></th></tr>
<tr>
<td colspan="5">
1. Does the TA mode give students a better understanding of the system behaviour? (2a)
2. Does the "If, Then" Question style of the TA contribute to the way students build causal arguments in the written text?(2b)
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?(2)
</td>
<td colspan="2">
1. The analysis of the written test did not reveal any significant differences between the treatment and control group [maybe because initial knowledge both groups had on the topic.
2. No difference in favour of being scaffolded by "if, then" style questions used by the VC toward building logical causal arguments [single exposure to this learning mode is insufficient.]
3. The models created by both control groups (simulation) and experiment group (VC) do not show any differences students from both groups left out quantities and had errors I building the model. VC scaffolding did not improve the quality of the models.
</td>
</tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>UHULL5</td><td>LS1

LS4</td><td>Homeostatis mechanisms Osmo-regulation

Homeostatis and osmo-Regulation in the brine Shrimp Artemia</td><td>Volunteer students Department of Biological Sciences at Hull University</td><td></td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td></td><td></td><td>1,2,
3,4</td></tr>
<tr><td>c
M/A</td><td></td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td rowspan="3">Drawing paper-based concept map based on text.

Build model, ground their model terms, request feedback and recommendation for building the model.

Using basic help – "How to" feature.

Two experimental groups – one getting feedback from one fixed reference model. The second getting feedback from automatic selection from repository.

Analysing models: A – before grounding; B – after grounding; C – after recommendation of the S.T.

Criteria for assessing models produced were the following:
• Number of elements (entities, quantities, configuration, causal dependencies scientific correctness), correctly introduced.
• Implementing 7 key concepts that represent causal ideas regarding osmoregulation on a scale of 1-3, maximum score 21.
Criteria for assessing the use of the grounding mechanism:
• Number of anchor terms and why type students introduced to their models, high numbers indicated poor use of the domain terms in the DBpedia.

Analysing student logs related to their use of the "How to" basic help feature.

Student scores for understanding and using the information provided by the "How to" feature.

Criteria to assess the use of S.T. model-based recommendations and the "How to" support mechanisms:
• Number of times individual students used them.
• Level of understanding the information the features provided on a 1-3 scale scores for using the recommendations and level of usefulness on a 1-5 scale.

Observation of students' handling the software</td></tr>
<tr><th colspan="4">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="4">1. Does the "semantic technology?" feature provide students with suitable recommendations for model improvement?

2. Do the students understand the recommendations they receive from the semantic technology?

3. Does the semantic technology that provides model-based recommendation improve students' own models and improve their understanding of the target system?

4. Does the "How to?" feature support students' understanding and enable them to independently learn how to build models?</td><td colspan="2">1. Many recommendations were found to be either with errors, or not relevant. Most students received suggestions only concerning differences between their model and the reference model. Low number of common ground between the student model and the reference model naming term, and grounding term is key step and should be an initial step before the students build a complicated model to avoid many nomenclature errors.
2. Students, on the most part, understood the recommendation they received from the S.T. However, they considered most of the recommendations to be not valid or not relevant.
3. While the recommendation provided by the S.T. did not improve students' own models, the modelling activity itself shifted students' focus from general concepts of homeostasis to those related the target system.
4. Students made little use of the inbuilt "How to" support during their model building activity. This could be the result of students' lack of technical knowhow of using this support.</td></tr>
</table>

TAU Evaluation Activities

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>TAU1</td><td>LS6</td><td>Marine eco-systems</td><td>Under-graduate students of biology</td><td>10 students</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td></td></tr>
<tr><td>b
SR/
M</td><td>1.2</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>3</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><td>O_1
Concept map</td><td>X
Progression of modelling tasks</td><td>O_2
Student final models interview motivation questionnaire</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">Does student involvement in Qualitative Modelling (QM) of ecological systems improve:
1. Understanding of ecological systems?(1b)
2. Modelling capability and scientific reasoning skills? (1b)
3. Motivation to learn science by QM? (1c)

Additional evaluation objectives for this activity were:
1. To provide descriptive information on the integration of QM in regular course.
2. To serve as pilot study for the definition of the evaluation framework and the development of evaluation instruments.</td><td colspan="2">1. & 2. Comparing student concept maps and final models yielded the following findings. Student final models contained less entities (only the relevant ones), quantities that were ignored in the concept maps were properly addressed, differentiated representation of causal relationship. The models were built around specific research questions and hypotheses that were tested through simulations and led in some cases to new questions and insight.
3. Lower scores for increasing conceptual understanding overall. The students showed interest and motivation to learn with the software, also in the future. The main gain perceived was about QM contribution to system thinking and understanding the complexity of the ecosystem.
Less strong, was considered the contribution of QM to learning subject topics.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>TAU2</td><td>LS3, LS4</td><td>Marine eco-system</td><td>Junior High students</td><td>21</td>
<td>
<table>
<tr><th></th><th>1 CM</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td>1</td><td></td><td></td></tr>
<tr><td>b SR/M</td><td>2.3</td><td></td><td></td></tr>
<tr><td>c M/A</td><td>5</td><td></td><td></td></tr>
<tr><td>d SRL</td><td>4</td><td></td><td></td></tr>
</table>
</td>
<td>
<table>
<tr><th>O_1</th><th>X_1</th><th>O_2</th><th>X_3,X_4</th><th>$X_{3, 4, 5}$</th></tr>
<tr><td>Pre-test exp.</td><td>Trip</td><td>Concept map</td><td>Modelling</td><td>Models and model docu-mentation</td></tr>
<tr><td>Pre-test control group</td><td>Trip</td><td>Concept map</td><td>Web-inquiry</td><td>Post-test Challenging question-naire and motivation question-naire</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">Does LbM with DynaLearn contribute to Junior High School students:
1. Conceptual understanding of a set of key concepts that represent the relevant content-domain (ecological systems)? (1A)
2. Ability to model a complex system and represent it at different levels of complexity using the qualitative reasoning approach embedded in DynaLearn. (1B)
3. Gradual acquisition of scientific and reasoning skills due to the modelling activity? (1B)
4. Capability to apply the knowledge and skills gained for approaching new ecological phenomena?
5. Motivation and attitudes toward learning by modelling?</td>
<td colspan="2">Students explanations to a set of 20 key concepts in ecology were improved from pre-testing to post-testing (large effect sizes in both groups). However both groups obtained low average scores (to short intervention).
The following pre-post changes occurred in the experimental group. Increase in net type representation and increase in adopting an ecological organizing principal In comparison none of the representation in the control group was net-like and less used the ecological organizing principal in these maps.
Analyses of models and explanations given to them along the course showed clear advances toward systemic view and understanding of the complexity and causal relationships (chain and loops) that stand behind the system behaviours.
Responses to the challenging questionnaire showed a tendency to adopt structure – behaviour – function type of explanations delineation of long chains of causal events. These characteristics were more profound in the experimental group.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>TAU3</td><td>LS2</td><td>Sea pollution</td><td>Undergraduate biology students</td><td>14</td><td>
<table>
<tr><th></th><th>1 CM</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td></td><td></td><td></td></tr>
<tr><td>b SR/M</td><td></td><td></td><td>1</td></tr>
<tr><td>c M/A</td><td></td><td></td><td></td></tr>
<tr><td>d SRL</td><td></td><td></td><td>2</td></tr>
</table>
</td><td>
<table>
<tr><th>O_1 Blind model exercise and explanations</th><th>X_1 Grounding task</th><th>O_2 Find model and explanations</th></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">Does the grounding process contribute to:
1. The quality of the models constructed by the students (relevant entities and relationships, correct configuration causal chains).
2, Facilitate self-directed learning of concepts and causal relationships in the modelled system.
(Select appropriate definitions dealing with unknown terms construct knowledge related to an ecological phenomenon)</td><td colspan="2">Before the grounding task about 40% of the students created quality models of the highest scores. After the grounding all but one student's model got the highest score observational data indicated independent successful work in completing the grounding task.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>TAU4</td><td>LS1, LS3</td><td>Effect of pollution and flood on population</td><td>30 Eighth grade high school students</td><td>15 Eighth graders</td><td><table><tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr><tr><td>a
CU</td><td>1a</td><td></td><td></td></tr><tr><td>b
SR/M</td><td>1b</td><td></td><td>2</td></tr><tr><td>c
M/A</td><td></td><td></td><td></td></tr><tr><td>d
SRL</td><td></td><td></td><td></td></tr></table></td><td rowspan="3">X_1 - O_1 - X_2 - O_2

X_1 Reading text describing the effect of pollution and a flood on the population of soft turtles in a nearby river.
O_1 Answering a set of related questions, Drawing concept maps.
Modelling the same phenomenon – LS2.
Modelling a new phenomenon – LS3.
Analyzing CM and models produced using a similar set of criteria (structure, relationship, quantities, organizing principle).
X_2 Filling a 'blind model' with ingredients from a given list, answer questions related to the model, ground the term in the model, modify the model if needed.
Students were given a "blind model" and a list of 7 entities (some unfamiliar) and 4 quantities to fill in the blind model. They were requested to describe the relationship between entities, look for the unknown term in the repository, revise the model if needed, and change their interpretation of the behaviour of the system described.
The evaluation criteria of the activity before grounding were proportions of correct fit of entities, quantities, relationship and proper configuration of the model. The grounding criteria were adequacy of term chosen, the extent students were mining the repository for appropriate term and whether they revised their model.
O_2 Analyzing students' models before grounding for correct number of entities, quantities, relationships and configuration of the model.
Assess the models after grounding according to adequacy of information chosen, location of the information in the database.</td></tr>
<tr><th colspan="4">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="4">1. Does LbM with DynaLearn contribute to:
a. Conceptual understanding of ecological system.
b. Improved ability to model the system in growing levels of complexity from LS1 to LS3.
2. Does the use of the grounding mechanism improve students' ability to model a complex system more precisely?</td><td colspan="2">1. a. Responses of only 15 students to seven questions that followed the text revealed a moderate understanding. The group mean reached only 64% of maximum score points with only three students scoring above 80% of maximum scores. The test was given at the beginning of the activity and meant to assess students' prior knowledge before modelling.
b. Ability to model the system in different levels of complexity was assessed by several criteria:
• Student progress toward hierarchical and web-like configuration of a system;
• Similar representation of both structural and process relationship in the model;
• Representing relevant and accurate quantities and define correctly the direction of relationship between them, and
• The organizing principle in building a model follows a conceptual skeleton.
Concerning the type of configuration, students mostly used linear and hierarchical representation. They focused on representing processes and causal relationships. In most cases (70%), students specified quantities, half of which indicated the direction of the relationship between the entities. In the last modelling activity, all students organized the models around concepts.</td></tr>
</table>

	2. Results of this evaluation activity showed that almost all students incorporated all given entities correctly, a third of the students failed to provide correct quantities and correct configuration. The descriptions of the term chosen from the repository were the correct ones. In most cases, it was the first in the repository. Those students who participated in the activity and did the grounding (6/9), did not change their models. Student performance in the grounding activity showed a shallow approach, performing the grounding did not trigger students to revise their models.	

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
TAU5	LS2, LS3	Invasion species	High school students	15 Volunteers aged 14016

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU			1a
b SR/M			1b
c M/A	1c		
d SRL			1d

Evaluation Design & Instruments

O_1	X_1	O_2	X_2	O_3
Text related pretest	LS2 Model and TA task	LbM & TA motivation question-naire	LS3 model	a. Model analysis b. Text related post-test

O_1/O_3b Students' conceptual understanding of an 11 items' test that followed a text about the jellyfish invasion to the Mediterranean Sea – maximum possible score – 27 points. The test was administered twice after introducing the text on the first day and again after several modelling experiences at LS2 and LS3.

O_2 The effect of using the TA feature was assessed after the students created a model using the text on the jellyfish invasion and given set of entities and relationships that represent this phenomena. On the basis of their models, the TA responded to a quiz. The success of the TA to correctly answer the quiz indicated a match between student model and an expert model. Mismatch caused students to revise their model. A set of statements deduced from the expert model in relation to the structure, quantities used and causal relationship served as a standard against which statements that could be deduced from students' models were compared.

O_3a Motivational aspects were assessed only in relation to the activity where the virtual characters were involved. In this activity, both the motivation toward learning by modelling and toward the interaction with the virtual characters was assessed.

Evaluation Questions	Results & Conclusions
1. Contribution of student interaction with a TA virtual character to: a. Conceptual understanding b. Modelling capabilities c. Motivation for LbM using DynaLearn d. Motivation for using the V.C. feature	1a. The results show that students' conceptual understanding in the pre-test was already high – 78%. This score decreased slightly to 75% of the total maximum score. This decrease could be attributed to the very short time interval between the two tests administered and also probably to the students' lack of motivation to respond again to the same questionnaire that they had already completed. 1b. All students who participated in this activity obtained maximum point scores for including the correct entities. At a lesser extent, they succeeded in specifying quantities or quantity spaces (average score 75%) and the group mean for inferring correct causal relationship was only 49%. 1c. Students' appreciation of LbM increased during the activities with DynaLearn. The software was considered as contributing to their understanding, especially due to their own building of models and when comparing their models to expert models. Scoring on 1-5 scale ranged from 3.2 - 3.3. 1d. Students perceived positively their interaction with the virtual characters. They found them contributing to learning (3 on a 5-point scale) and helping in building a model (4 on a 5-point scale).

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>TAU6</td><td>LS5, LS6</td><td>General</td><td>Student with previous experience in DynaLearn</td><td>1 Student 16 years old</td><td><table><tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr><tr><td>a
CU</td><td></td><td></td><td></td></tr><tr><td>b
SR/M</td><td></td><td></td><td></td></tr><tr><td>c
M/A</td><td></td><td></td><td></td></tr><tr><td>d
SRL</td><td>1</td><td></td><td></td></tr></table></td><td rowspan="3">Transcription of students' "think aloud" and the models produced were analyzed and interpreted.</td></tr>
<tr><th colspan="4">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="4">1, To what extent does DynaLearn contribute to self-regulated learning through reflection "on" and "in" action, while building models.</td><td colspan="2">From the 'think aloud' protocol it can be concluded that substantial learning and reasoning process took place. This was reflected in the students' modelling products as well as in reflective statements that dealt with the logic behind modelling. The students' comments related to the meaning of the actions taken. This denotes the students' ability to think along the modelling process about "what to do" as well as "why to do that".</td></tr>
</table>

FUB Evaluation Activities

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
FUB1 Tree and shade Algal bloom	LS6	Global warming Algal bloom	Deaf students 15-29 years old	30

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU	1		
b SR/ M	2,3		
c M/A	4,5		
d SRL			

Evaluation Design & Instruments

O_1 Pre-test written essay	Experimental X_1-X_8 : Lecture Private Modelling	O_2 Post-test questionnaires
O_1 Pre-test Algal bloom written essay questionnaire Motivation and usability	Control X_1-X_2 Lecture ppt. Presentation	O_2 Post-test Algal bloom Written essay questionnaire

Evaluation Questions

1. Do learners present significantly better scores in tests involving concepts on environmental science when using Garp3 conceptual modelling workbench and exploring qualitative models?
2. Does the meta-vocabulary used in qualitative models, from which a conceptual interpretation is built; provide learners a domain independent analytic instrument that enables them to understand more fine grained and thorough analyses of how systems work?
3. Do the students significantly improve their capacity of making inferences (causal reasoning) after using qualitative models as learning tools? Non-trivial conclusions?
4. What is the students' perception of the software, the modelling activities and the use of qualitative models in their learning process?
5. Are the students motivated to work with qualitative models and use Garp3 software?

Results & Conclusions

1. The experimental group presented significant improvement in conceptual understand on the results of post-test after the use of qualitative models;
2. Although a significant difference between the pre and post-test results of the control group indicate a learning effect in the expositive lecture, comparison between the post-tests results in the two groups support the conclusion that the use of qualitative models has produced better results;
3. Significant differences were found between the experimental group's pre and post-tests, in total number of inferences and number of non-trivial conclusions, but no significant differences were found on the number of trivial conclusions;
4. Significant differences were found between post-tests of experimental and control groups, in number of inferences and number of non-trivial conclusions, but no significant differences were found on number of trivial conclusions.
5. The modelling language was accessible to the students and they considered excellent or good the clarity of the representation. About the use of QR language to describe the topics addressed by the models, 20% of the deaf students considered it only regular, probably by the level of complexity of the language. The understanding of the concepts worked out with the support of the models by 80% of the respondents. The students have a positive view on the contribution of qualitative models to their learning process, and would be keen in using these models in the classroom.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB2</td><td>LS6</td><td>Trees and shade, Climate change, Economy and global warming</td><td>Public school hearing students age 15-18</td><td>60
(30 x 2)</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>2,3</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>4,5</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><td>O_1
Concept map</td><td>X
Experimental
X_1 –X_8

Course activities</td><td>O_2
Final model (LS2, LS4)
Motivation measures
Grades on final exam</td></tr>
<tr><td>O_1</td><td>Control
X_1 -X_1</td><td>O_2
Final model (LS2, LS4)
Motivation measures
Grades on final exam</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Do learners present significantly better scores in tests involving concepts on environmental science when using Garp3 conceptual modelling workbench and exploring qualitative models?
2. Does the meta-vocabulary used in qualitative models, from which a conceptual interpretation is built, provide learners a domain independent analytic instrument that enables them to understand more fine grained and thorough analyses of how systems work?
3. Do the students significantly improve their capacity of making inferences (causal reasoning) after using qualitative models as learning tools?
4. What is the students' perception of the software, the modelling activities and the use of qualitative models in their learning process?
5. Are the students motivated to work with qualitative models and use Garp3 software?</td><td colspan="2">1. The experimental group presented significant improvement in conceptual understand on the results of post-test after the use of qualitative models;
2. A significant difference between the post-tests results in the two groups support the conclusion that the use of qualitative models has produced better results than the expositive lecture.
3. In the experimental group, the number of total inferences and of non-trivial conclusions in the post test is highly significantly bigger than in the pre-test, and the number of trivial conclusions was significantly smaller in the post test;
4. When the two groups were compared, the difference in the total number of inferences was not significant; the number of non-trivial conclusions was not significant, but it was exactly on the level of significance (5%); and the number of trivial conclusions was significantly higher in the control group.
5. The analysis of the questionnaires has shown that the students were very satisfied with the use of the didactic material to explain scientific concepts and approved the use of qualitative models and Garp3; the modelling language was accessible to them and they considered good the clarity of the representation. About the use of QR language to describe the topics addressed by the models and to express causality, the students considered it only regular, probably by the level of complexity of the language. The understanding of the concepts worked out with the support of the models for 80% of the respondents. The students have a positive view on the contribution of qualitative models to their learning process, and all the students would be keen in using these models in the classroom.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB3

Teaching with models</td><td>LS1-LS4</td><td>Q.R. and Qualitative modelling</td><td>Public school teachers Integrated science at elementary level and specific domains at secondary level All teachers teach hearing and deaf students</td><td>27</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td>1</td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>1</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>1</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><td>X
O_1</td><td>X
O_2</td></tr>
<tr><td colspan="2">Teachers written report collected during the course and in schools; Questionnaires</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Teachers' perception of the contribution of Q.M. to:
• Development of written scientific text
• Vocabulary
• Solving problems
• Producing explanations, formulate hypotheses predict results and produce arguments
• Teachers opinion Q.M.</td><td colspan="2">1. The teachers were able to explore the models presented and recognized the high potential for the development of a number of competences and skills, including the ability to make inferences, analogies and deductions while analysing the behaviour of a system; formulate hypotheses and predict results; analyse and compare possible solutions to the same problem. The teachers also reported substantial improvement on the students’ writing skills. We can conclude the didactic material compiled in a DVD was well accepted for the deaf students and considered a bit slow for the hearing students. The teachers considered it a valuable tool for science teaching contributing for learning scientific concepts and the development of reasoning skills.</td></tr>
</table>

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Design & Instruments
FUB4	LS2	Trees and shade Algae bloom, Erosion, Air pollution, Dengue Fever	Q.R. Experienced female deaf students	4	(see Evaluation Spaces table below)	(see Evaluation Design table below)

Evaluation Spaces:

	1 CM	2 VC	3 ST
a CU			
b SR/M			
c M/A	1,2		
d SRL			

Evaluation Design & Instruments:

X_1	O_2
Lecture text demonstrations modelling	Questionnaires

Evaluation Questions	Results & Conclusions
This evaluation study aimed at answering the following questions: 1. What is the students' perception of the DynaLearn software, the modelling activities and the use of qualitative models in their learning process? 2. Are the students motivated to work with qualitative models and use DynaLearn software?	• Most students said the conceptual modelling was a new approach, and it made them think about systems in a different way, giving new insight into phenomena and processes on ecology; • All the students ticked "fully agree" when was said the process of modelling motivated them to learn more about the phenomena, recognising how the conceptual modelling could help them to learn about other topics; • Most students understand the modelling approach and the goal of the modelling tasks; • The students rated as "easy" and "hard in part" the conceptual modelling, find the conceptual mapping task and identification and describing of entities and quantities in system; • Describing system structure in a basic casual model and work with simulations was easy to most of them. Having had the experience with Garp3, the students quickly got the main features of the DynaLearn interface. Manipulating models was relatively easy for them, and their impression of the software was very positive. All the four students considered DynaLearn more accessible than Garp3, and comment on how the use of qualitative models could help the deaf to learn concepts and to improve their writing skills. One of the students suggested having all the words in the software written in Portuguese, in order to make it easier for understanding the models.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB5</td><td>Introduction to LS1-LS4; most activities in LS6</td><td>The modelling process</td><td>Graduate ecology students at the university level</td><td>10
(6 PhD,
3 MA,
1 Re-searcher PhD)</td><td><table><tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr><tr><td>a
CU</td><td>1</td><td></td><td></td></tr><tr><td>b
SR/M</td><td>2</td><td></td><td></td></tr><tr><td>c
M/A</td><td>3</td><td></td><td></td></tr><tr><td>d
SRL</td><td></td><td></td><td></td></tr></table></td><td><table><tr><th>X_1</th><th>O_1</th></tr><tr><td>Free modelling activity</td><td>Final models
Motivation and software usability questionnaires</td></tr></table></td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. What is the students' perception of the software, the modelling activities and the use of qualitative models in their learning process?
2. Do they foresee the use of qualitative models and DynaLearn as part of their research activities?

3. Are the students motivated to work with qualitative models and use DynaLearn software?</td><td colspan="2">1. Some students said the conceptual modelling was a new approach, and the most said it made them think about systems in a different way, giving new insight into phenomena and processes on ecology;
2. All the students ticked "strongly agree" when was said the process of modelling motivated them to learn more about the phenomena, recognising how the conceptual modelling could help them to learn about other topics;
3. Most students understand the modelling approach and the goal of the modelling tasks;
4. The students rated as "easy" and "difficult in part" the conceptual modelling, find the conceptual mapping task and identification and describing of entities and quantities in system;
5. Describing system structure in a basic casual model was easy and work with simulations was difficult in part to most of them.
6. The students rated as easy the qualitative modelling used in classes, and evaluated the experience of working with the DynaLearn as very interesting;
7. They said to build models in different specific Learning Space of DynaLearn is very important for understanding, and the LS4 was the LS who did contribute most with understanding of the concepts represented;
8. The students said completely agreed the modelling enabled them to better understand the complexity of the ecological and environment science, and the software could also be used in other learning topics.
Despite considering the conceptual modelling approach difficult in part, in general, most graduate students that attended the course considered that this approach improved their way of thinking about the environmental systems and their behaviour, as well as about their phenomena. Among them, 66% considered the conceptual modelling as a completely new approach, while 44% already knew this approach. However, they all felt that this technique made them think in a different way about the systems.
All students found it interesting, in different levels, work with the DynaLearn software, and</td></tr>
</table>

	most of them considers it easy or very easy to operate it, although several students found it difficult. The LS4 was considered the learning space that most contributed to the understanding of the concepts represented by the models, probably because this is the first level at which causality differentiation can be applied and the consequences can be observed in action. The main negative points raised about the use of the DynaLearn software treated on the lack of commands that are found in other software commonly used in Brazil and which allow, in general, the actions of "undo" and "redo" things and "copy" fragments of the models built.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB6

Teaching with models</td><td>LS1-LS4</td><td>Pollution, deforestation, bio-magnification, population dynamics</td><td>Secondary school teachers from different disciplines
Some with modelling experience</td><td>23</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td></td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>2,3</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>4,5</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><th>O_1</th><th>X</th><th>O_2</th></tr>
<tr><td>Pre-activity questionnaire

Written pre-test (causality)</td><td></td><td>Questionnaire on motivation and usability
Written post-test (causality)</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">Software and motivation
1. What are the teachers’ perceptions of the software, the modelling activities and the use of qualitative models in their learning process?
2. Are the teachers motivated to work with qualitative models and use DynaLearn software?</td><td colspan="2">1. Most teachers rated as “hard” the modelling used to develop the educational activity, but evaluated the experience of working with DynaLearn software as “interesting”, helping them to understand the ecological problem after explore the topic in DynaLearn;
2. They said to build models in different specific Learning Space of DynaLearn is important for understanding, and the LS2 was the LS who did contribute most with understanding of the concepts represented;
3. The teachers agreed the modelling enabled them to better understand the complexity of the biodiversity loss process, and the software could also be used in other learning topics.
The test about causality detected an increase in the average score in post-test greater than in pre-test, which means that secondary school teachers can recognize more causal relationships after modelling activities. The result supports the conclusion that the use of the software can improve the ability of make and identify causal inferences. Despite they thought hard to use modelling in education activity, the use of DynaLearn was interesting manly to understand environmental issues. This is probably related to diagrammatic approach, the capacity to make predictions and to observe systems behaviour as a whole.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB7

Conservation biology</td><td>LS1-L4</td><td>Conservation issues
Hydrological erosion</td><td>Secondary school students
15-18 year olds</td><td>21</td><td><table><tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr><tr><td>a
CU</td><td>1,2</td><td></td><td></td></tr><tr><td>b
SR/M</td><td>3,4,
5</td><td></td><td></td></tr><tr><td>c
M/A</td><td>4,5</td><td></td><td></td></tr><tr><td>d
SRL</td><td>6,7</td><td></td><td></td></tr></table></td><td><table><tr><th>O_1</th><th>X_1-X_4</th><th>O_2</th></tr><tr><td>Pre-test causality and Conservation concept
Pre-motivation questionnaire</td><td></td><td>Post-test causality and Conservation concept
Post-motivation questionnaire</td></tr></table></td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Do learners present significantly better scores in concept tests when using DynaLearn’s conceptual modelling workbench?
2. Do the students present significantly better scores in concept tests while moving from LS1 – LS2? And from LS2 – LS4?
3. Does the diagrammatic approach (as organized in the DynaLearn settings) actually allow learners to address more complex problems?
4. 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?
5. Do the students present significantly improve their capacity of making inferences (causal reasoning, analogies)?
6. Are the students motivated to work with qualitative models and use DynaLearn software?
7. What is the students’ perception of the software, the modelling activities and the use of qualitative models in their learning process?</td><td colspan="2">1. Using the result obtained in the paired t test with bootstrapping we observe a significant difference between pre and post-test in questions on Conservation Biology, and the mean scores increase from pre to post-test. This means that the approach used can collaborate with concept understand of dynamic systems.
2. They said to build models in different specific Learning Space of DynaLearn is very important for understanding, and the LS4 was the LS who did contribute most with understanding of the concepts represented.
3. The students agreed the modelling enabled them to better understand the complexity of the biodiversity loss process, agreed that the software provides a comfortable way of learning and it could also be used in other learning topics.
4. Some students said the conceptual modelling was a new approach, and it made them think about systems in a different way, giving new insight into phenomena of environmental science. Most students agreed that the process of modelling motivated them to learn more about the phenomena, recognizing how the conceptual modelling could help them to learn about other topics.
5. Some students rated as “difficult” and most as “neutral” the qualitative modelling used in classes, and evaluated the experience of working with the DynaLearn as very interesting. Most students understood the modelling approach and the goal of the modelling tasks.
6. The students rated as “easy” and “hard in part” the conceptual modelling, find the conceptual mapping task and identification and describing of entities and quantities in system.
7. Describing system structure in a basic casual model was hard in part and work with simulations was hard in part to most of them.</td></tr>
</table>

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr>
<td>FUB8

Conservation biology; metapopulation</td>
<td>LS1-LS4</td>
<td>Conservation issues
Ecosystems
Human Activity
Communities
Population</td>
<td>Secondary public school Students' aged 15-18</td>
<td>49</td>
<td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td>1,2</td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>3</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>4,5</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td>
<td>
<table>
<tr><td>O_1

Pre-tests
Written essays
Diagram-matic products

O^1
Pre-tests
Written essays
Diagram-matic products</td>
<td>X_1

Experimental groups (modelling)
1. Trained – group A
2. Not trained – group B

X_2
Control group C traditional</td>
<td>O_2

Post-test
Written essay
Modelling products

O_2
Post-test
Written essay
Modelling products</td></tr>
</table>
</td>
</tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr>
<td colspan="5">
1. Do learners present significantly better scores in tests involving concepts on environmental science when using DynaLearn’s conceptual modelling workbench to build and explore qualitative models?
2. Do the students present significantly better scores in concept tests while moving from LS1 – LS2? And from LS2 – LS4?
3. Do the students present significantly improve their capacity of making inferences (causal reasoning, analogies)?
4. Are the students motivated to work with qualitative models and use DynaLearn software?
5. What is the student's perception of the software, the modelling activities and the use of qualitative models in their learning process?
</td>
<td colspan="2">
• The experimental group B presented significant improvement in conceptual understand on the results of post-test after the use of qualitative models;
• The experimental groups put together A+B presented significant improvement in conceptual understand on the results of post-test after the use of qualitative models;
• The experimental group A considering only students that had trained in DynaLearn presented significant improvement in conceptual understand on the results of post-test after the use of qualitative models and the time to familiarize with software.
• The difference on posterior knowledge (post x post) on the concepts addressed during the course, the three groups, experimental A and B and control, was not significant, the significant difference is an important condition to assess the experimental treatment effects on the students’ behaviour.
• All groups presented significant improvement in conceptual understand of meta-populations considering the questions 14 and 15 in population biology test on the results of post-test after the use of qualitative models;
• The difference on previous (pre x pre) and posterior (post x post) knowledge on the concepts addressed during the course, the three groups, experimental A and B and control, considering the questions 14 and 15 in population biology test, was not significant, an important condition to assess the experimental treatment effects on the students’ behaviour;
• All groups presented not significant improvement in conceptual understand of meta-
</td>
</tr>
</table>

	populations on the results of post-test after the use of qualitative models; • The difference on posterior knowledge (post x post) on the concepts addressed during the course, the three groups, experimental A and B and control, was not significant, the significant difference is an important condition to assess the experimental treatment effects on the students' behaviour. • They said to build models in different specific Learning Space of DynaLearn is important for understanding, and the LS2 (group A) and LS1 (group B) was the LS who did contribute most with understanding of the concepts represented; • Most students rated as "easy" the qualitative modelling used in classes, and evaluated the experience of working with the DynaLearn as interesting; • The students agreed the modelling enabled them to better understand the complexity of the biodiversity loss process, agreed the use of the software provides a very comfortable way of learning and the software could also be used in other learning topics; • Some students said the conceptual modelling was a new approach, and the most said it made them think about systems in a different way, giving new insight into phenomena of environmental science; • Most students ticked "agree" when was said the process of modelling motivated them to learn more about the phenomena, recognising how the conceptual modelling could help them to learn about other topics; • Most students understood the modelling approach and the goal of the modelling tasks; • The students rated as "easy/hard in part" the conceptual modelling, as well the conceptual mapping task and identification and describing of entities and quantities in system; • Describing system structure in a basic casual model and work with simulations was easy to most of them.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB9</td><td>LS1-LS4</td><td>Conservation biodiversity
Deforestation
Pollution
Meta-populations</td><td>Secondary public School students</td><td>15-control
5-experi-mental
15-18 year olds</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td>1,2
4,5</td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>3</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>6,7</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><td>O₁

Pre-tests
Written essays
Diagram-matic products</td><td>X₁

Experimental group (modelling)
1. Trained – group A
2. Not trained – Group B</td><td>O₂

Post-test
Written essay
Modelling products</td></tr>
<tr><td>O¹
Pre-tests
Written essays
Diagram-matic products</td><td>X₂
Control group C traditional</td><td>O₂
Post-test
Written essay
Modelling products</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. Do learners present significantly better scores in concept tests when using DynaLearn’s conceptual modelling workbench?
2. Do the students present significantly better scores in concept tests while moving from LS1 – LS2? And from LS2 – LS4?
3. Do the students present significantly improve their capacity of making inferences (causal reasoning, analogies)?
4. Do learners present significantly better scores in tests involving concepts on environmental science when using DynaLearn’s conceptual modelling workbench and building and exploring qualitative models?
5. Do the students present significantly better scores in concept tests while moving from LS2 to LS4?
6. Are the students motivated to work with qualitative models and use DynaLearn software?
7. What is the students’ perception of the software, the modelling activities and the use of qualitative models in their learning?</td><td colspan="2"><u>In regard to the concept of population</u>:
• The experimental (group A) and control groups presented significant improvement in conceptual understand on the results of post-test;
• The experimental and control groups showed no significant difference between pre-tests which means that the groups were homogeneous for the subject of the workshop;
• The difference in the results of post-tests between groups was not significant and the modelling approach didn’t produce better result in that conditions.
<u>In regard to the concept of meta-population</u>:
• Only experimental group (A) presented significant improvement in conceptual understand on the results of post-test;
• The experimental and control groups showed no significant difference between pre-tests which means that the groups were homogeneous for the subject of the workshop;
• The difference in the results of post-tests between groups was not significant and the modelling approach didn’t produce better result considering the design of the experiment.
• Most students rated as “easy” the qualitative modelling used in classes, and evaluated the experience of working with the DynaLearn as very interesting;
• They said to build models in different specific Learning Space of DynaLearn is very</td></tr>
</table>

	important for understanding, and both LS3 and LS4 was the LS who did contribute most with understanding of the concepts represented: • Most students fully agreed the modelling enabled them to better understand the complexity of the biodiversity loss process, and the use of it provides a comfortable way of learning and could also be used in other learning topics; • The students said the conceptual modelling was a new approach, and the most said it made them think about systems in a different way, giving new insight into phenomena of environmental science; • Most students ticked “totally agree” when was said the process of modelling motivated them to learn more about the phenomena, recognising how the conceptual modelling could help them to learn about other topics; • Most students understood the modelling approach and the goal of the modelling tasks; • The students rated as “very easy” the conceptual modelling, find the conceptual mapping task and identification and describing of entities and quantities in system; • Describing system structure in a basic casual model was very easy or easy/hard in part and work with simulations was very easy to most of them. • The results obtained in this evaluation were similar to the previous and in those conditions, the learning by modelling approach had similar effect as the expository classes. There was an improvement between pre and post-test in both groups, but the improvement in experimental group wasn’t greater than the improvement in control group as we expect. • Considering only the metapopulation subject was obtained a little different result: the experimental group had a better result in post-test comparing with pre-test, but the same wasn’t observed for control group. Nevertheless, no difference was detected in post-tests between experimental and control group.. • In general, the students had a very good opinion about the course and the modelling activities developed as a motivating task and as stimulate extra class activity. Most of the students looking for new ways, new tools to learning more and better and the activity were a very good experience in this point.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Design & Instruments</th></tr>
<tr><td>FUB10</td><td>LS1-LS4</td><td>Global warming</td><td>Secondary school students experienced in DynaLearn</td><td>5-8 students</td><td>
<table>
<tr><th></th><th>1
CM</th><th>2
VC</th><th>3
ST</th></tr>
<tr><td>a
CU</td><td>1</td><td></td><td></td></tr>
<tr><td>b
SR/M</td><td>2</td><td></td><td></td></tr>
<tr><td>c
M/A</td><td>3,4</td><td></td><td></td></tr>
<tr><td>d
SRL</td><td></td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><th>X_1 - X_3</th><th>O_1</th></tr>
<tr><td>Modelling activities</td><td>Student modelling products</td></tr>
</table>
</td></tr>
<tr><th colspan="5">Evaluation Questions</th><th colspan="2">Results & Conclusions</th></tr>
<tr><td colspan="5">1. 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?
2. Do the students present significantly improve their capacity of making inferences (causal reasoning, analogies)?
3. Are the students motivated to work with qualitative models and use DynaLearn software?
4. What is the students’ perception of the software, the modelling activities and the use of qualitative models in their learning process?</td><td colspan="2">Not yet available</td></tr>
</table>

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces				Evaluation Designs & Instruments
FUB11 Working with Is and Ps	LS01-LS02 E LS04	Deforestation; Habitation; and Energy	Secondary school students	04		1 CM	2 VC	3 ST	
					a CU	1, 2			
					b SR/M	3,4			
					c M/A				
					d SDL				

O_1	X_1	O_2	X_2	O_3	X_3	O_4	X_4	O_5	X_5	O_6
Pre-test	Introduction lecture	Conc Ept Map P Cycle	Evolving Deforestation model	Post Model 01 Complete Is and Ps	Evolving Habitation model	Post Model 02 Complete Is and Ps	Evolving Energy model	Post Model 03 Complete Is and Ps	Concept Mao S Cycle	Post test

Evaluation Questions	Results and Conclusions
1. Are the learners able to identify the key elements of the system (entities, quantities and processes)? 2. Are the learners able to identity processes, rates and to understand how concurrent processes shape system behaviour? 3. Are the learners able to differentiate processes (direct influences) and propagation the processes effects (proportionalities)? 4. What are the criteria for the learners to select the elements of the system to be included in the model (contour conditions)?	Students perceptions of processes and their consequences: - After doing exercises using Is and Ps, students were able to understand better process –related phenomena and they showed increased capability to identify quantities and entities in the models. - After building models students´ scores were higher than in the first model, showing they were able to improve their qualitative reasoning and systems´ thinking skills. - Improvement in their ability to recognize and implement model ingredients and so building models of better quality. - Better conceptual understanding of the domain studied after modelling activities as the motivation to talk more and with more confidence about issues related to environmental sciences. - Wrong causal relations decreased after modelling activities.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB12 Teachable Agents	LS02 - LS04	Deforestation; Photosynthesis; and Respiration	Secondary school students	05	see Evaluation Spaces table below	see Evaluation Designs & Instruments table below

Evaluation Spaces:

	1 CM	2 VC	3 ST
a CU		3	
b SR/M			
c M/A		1	
d SDL		2	

Evaluation Designs & Instruments:

O_1	X_1	O_2	X_2	O_3	O_4
Lecture 'Lake ecosystem'	Model building LS2	Support text 'Photosyntheses and respiration'	Model Building LS2	Quiz master and model version	Interview and motivation questionnaires

Evaluation Questions	Results and Conclusions
Do Virtual Characters using the Teachable Agent mode contribute to… 1. Increase learners' motivation to learn environmental science topics? 2. Improve the interaction between the learners and the software while revising and improving their initial model? 3. Produce better understanding of the concepts approached in classroom?	Students results: - Percentage of questions correctly answered by students pet in each quiz increased during the evaluation - Students' impressions about the use of VCs were: better option to motivate modelling than reading texts; observed knowledge improvement by model building; a better understanding of the model subject; a better conceptual understanding and the motivational aspects of better learning and model building using TAs. - Students find difficult to put the relations between the entities and between quantities, rather than other modelling activities. - Some students showed preference for using texts before the modelling activity. - All of them mentioned that Dynalearn is an efficient method for fixating and improve knowledge; it helps to meet different possibilities and to analyze consequences through simulations.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB13 Model Debugging	LS04	Phytoremediation	Secondary school students	03	(see evaluation spaces table below)	(see evaluation design table below)

	1 CM	2 VC	3 ST
a CU	2		
b SR/M	2		
c M/A	1		
d SDL			

O_1	X_1	O_2	O_3	O_4	O_5
Pre-test	Text about 'Problem of lake pollution'	Model debugging 01 'Environmental restoration in urban areas'	Model debugging 02 'Environmental restoration in urban areas'	Modelling report answering	Interview

Evaluation Questions	Results and Conclusions
Does conceptual modelling using a model debugging approach contribute to… 1. Increase learners' motivation to learn environmental science topics? 2. A better understanding of the concepts approached in classroom?	- Students understood better how to apply different causal relations and had the opportunity to learn concepts related to restoration of areas with environmental damage and eutrophication. - Students' results showed that they handle better the model after debugging it. - The students found more interesting debug models using the pets.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB14 Conceptual understading of Metapopulations	LS04	Metapopulations and Population biology	Secondary school students	02	(see evaluation spaces table below)	(see evaluation design table below)

	1 CM	2 VC	3 ST
a CU	1		
b SR/M	1		
c M/A			
d SD	1		

O_1	O_2	O_3	X_3
Pre-test	Modelling activity Is and Ps	Post- Test	Interview

					L		

Evaluation Questions	Results and Conclusions
5. Are the students able to explain the metapopulation concept and the behaviour of metapopulations only using the elements that they found in the model?	Students tests pointed that: - They really can understand the concept of metapopulation only by handling the model, and they also scored better the conceptual test after doing the I´s and P´s exercise in an expert 'Metapopulation' model. - They find that the QR language is a good approach to learn certain issues, mainly those related to ecology because of the better understanding of the relation between entities. - They also said that the LbM is better to help them for fixing and easily understand new concepts.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB15 Conceptual understading of Metapopulations	LS04	Metapopulations and Population biology	Secondary school students	02	(see table below)	(see table below)

Evaluation Spaces:

	1 CM	2 VC	3 ST
a CU	1		
b SR/M	1		
c M/A			
d SDL	1		

Evaluation Designs & Instruments:

O_1	O_2	O_3	X_3
Pre-test	Modelling activity Is and Ps	Post- Test	Interview

Evaluation Questions	Results and Conclusions
6. Are the students able to explain the metapopulation concept and the behaviour of metapopulations only using the elements that they found in the model?	Students tests pointed that: - They really can understand the concept of metapopulation only by handling the model, and they also scored better the conceptual test after doing the I´s and P´s exercise in an expert 'Metapopulation' model. - They find that the QR language is a good approach to learn certain issues, mainly those related to ecology because of the better understanding of the relation between entities. - They also said that the LbM is better to help them for fixing and easily understand new concepts.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
FUB16 Osmosis and diffusion	LS04	Osmosis and diffusion	Secondary school students	03

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU	1,2		
b SR/M	2		
c M/A	3		
d SDL			

Evaluation Designs & Instruments

O_1	X_1, O_2	O_3	O_4
Pre-test	Open partially developed models, complete them and run simulations	Post-Test	Motivation / attitude questionnaires

Evaluation Questions	Results and Conclusions
Does the involvement in a modelling activity improve students'… 1. Understanding of an ecological system? 7. Representation of the system - specification of entities, quantities and relationships in the system? 8. Ability to explain and predict the behaviour of the system?	There was an increase in the percentage of correctly answered exercises; The scores obtained in post-test were higher than those obtained in pre-test; Motivation questionnaires under analysis.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
FUB17 Teachable Agent – Food Chain	LS02	Food Chain	Fundamental school students in "acceleration class"	16

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU		1, 2	
b SR/M		1	
c M/A		1, 2	
d SDL		2	

Evaluation Designs & Instruments

O_1	X_1	X_2	X_3	O_2	O_3
Pre-test	Short presentation of DynaLearn modeling environment and model ingredients	Modelling activity Conceptual modelling only – lacustrine ecosystem	Modelling activity Teachable Agent – Food chain	Post-Test	Motivation and attitude Questionnaire

Evaluation Questions	Results and Conclusions
Does students' involvement in teaching virtual characters 1. Contribute for student's improvement of their models quality? 2. Improve their understanding of concepts and causal relationships within the modelled	There was no significant difference between pre and post-test. There was an increase in the percentage of questions correctly answered by Students Pets on the Quiz.

system?	In general students liked the interaction with the hamsters , they found themselves motivated by the VC.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB18 Use of Is and Ps.	LS04	Population, rainfall	Secondary school students	02	<table><tr><td></td><td>1 CM</td><td>2 VC</td><td>3 ST</td></tr><tr><td>a CU</td><td>1</td><td></td><td></td></tr><tr><td>b SR/M</td><td>2</td><td></td><td></td></tr><tr><td>c M/A</td><td></td><td></td><td></td></tr><tr><td>d SDL</td><td></td><td></td><td></td></tr></table>	<table><tr><td>O_1</td><td>X_1</td><td>O_2</td></tr><tr><td>Pre-test</td><td>Create and develop models, run simulations</td><td>Post-test</td></tr></table>

Evaluation Questions	Results and Conclusions
Does students' involvement in modelling activities contribute to.. 1. develop a better understanding of the concepts approached in classroom? 2. develop a number of specific reasoning skills?	There was a slightly change to increase the number of correctly answered questions for one student between pre and post-test. Other results are in analysis.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB19 Grounding Deforestation and Ecosystem Services	LS04	Deforestation and ecosystem services	Secondary school students	03	<table><tr><td></td><td>1 CM</td><td>2 VC</td><td>3 ST</td></tr><tr><td>a CU</td><td></td><td></td><td>1</td></tr><tr><td>b SR/M</td><td></td><td></td><td>2, 3</td></tr><tr><td>c M/A</td><td></td><td></td><td></td></tr><tr><td>d SDL</td><td></td><td></td><td></td></tr></table>	<table><tr><td>O_1, X_1</td><td>X_2</td><td>X_3</td></tr><tr><td>Open blind model, Replace blind entities and quantities by real ones</td><td>Grounding all terms of the final model</td><td>Interview</td></tr></table>

Evaluation Questions	Results and Conclusions
Does the involvement in a grounding activity improve students'… 1. Understanding of an ecological system? 2. Representation of the system - specification of entities, quantities and relationships in the	Results are in analysis

system? 3. Ability to explain and predict the system behaviour?	

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB20 Water erosion LS4	LS04	Water erosion	Secondary school students	03		

	1 CM	2 VC	3 ST
a CU	1		
b SR/M	2		
c M/A			
d SDL			

O_1	X_1	X_2	O_3	X_3
Pre-test	Creating single models using basic patterns	Answering the exercises after simulate the models	Post-test	Interview

Evaluation Questions	Results and Conclusions
Does students' involvement in modelling activities contribute to… 9. A better understanding of the concepts approached in classroom? 10. Develop several reasoning skills?	Students perceptions of processes and its consequences: - The students found that creating the causal dependencies the most difficult part of the activity - In general the students were able to define what is water erosion, its causes and consequences after the activity Other results are in analysis

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB21 Water erosion on LS5	LS05	Water erosion	Secondary school students	03		

	1 CM	2 VC	3 ST
a CU	1, 3		
b SR/M	2		
c M/A			
d SDL			

O_1, X_1	O_2, X_2	O_3, X_3	X_4
To reproduce a model in LS5	To complete a model in LS5	To answer the exercises	Interview

Evaluation Questions	Results and Conclusions
Does students' involvement in modelling activities contribute to… 11. A better understanding of the concepts approached in classroom?	Interview results: - In general, despite initial difficulties, students were able to build the LS5 models.

12. Develop several reasoning skills? 13. Understand and recognize conditional knowledge?	- They found LS5 more complex and difficult than the LS4. - In general, students were able to recognize conditional knowledge.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th colspan="2">Evaluation Designs & Instruments</th></tr>
<tr>
<td>FUB22

Algal bloom LS6</td>
<td>LS05
LS06</td>
<td>Algal Bloom</td>
<td>Secondary school students</td>
<td>03</td>
<td>
<table>
<tr><th></th><th>1 CM</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td>1</td><td></td><td></td></tr>
<tr><td>b SR/M</td><td>2, 3, 4</td><td></td><td></td></tr>
<tr><td>c M/A</td><td>1</td><td></td><td></td></tr>
<tr><td>d SDL</td><td>1</td><td></td><td></td></tr>
</table>
</td>
<td colspan="2">
<table>
<tr><th>X_1</th><th>X_2</th><th>O_1, X_3</th><th>O_2, X_4</th><th>O_3</th><th>O_4</th><th>O_5</th></tr>
<tr><td>To reproduce a model in LS5</td><td>To reproduce a model in LS6</td><td>To create a model in LS6 based on a model in LS5</td><td>To create a model in LS6 by themselves using hierarchy knowledge</td><td>Answer the exercises</td><td>Motivation and attitude Questionnaire</td><td>Interview</td></tr>
</table>
</td>
</tr>
<tr><th colspan="6">Evaluation Questions</th><th colspan="2">Results and Conclusions</th></tr>
<tr>
<td colspan="6">Does students' involvement in modelling activities contribute to…

1. A better understanding of the concepts approached in classroom?
2. Develop several reasoning skills?
3. Understand and recognize conditional knowledge?
4. Understand, recognize and create hierarchic patterns?</td>
<td colspan="2">Interview results:
The students found the activities difficult, but they also think that with practice they were getting better.
The compositional modelling and the hierarchic approach were seen as advantages of the LS6 by students.
They found very worthwhile to learn using LS6.
The students were able to understand what a process and a feedback are, that these important concepts are not very well approached in their regular classes.</td>
</tr>
</table>

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces				Evaluation Designs & Instruments					
FUB23 Water cycle with deaf students (Systems Thinking)	LS04	Water cycle	Secondary school deaf students	27		1 CM	2 VC	3 ST	O_1	O_2, X_1	O_3, X_2	O_4, X_3	O_5	O_6
					a CU	1			Pre-test	Water in soil exercise	Water in lake exercise	Water in subsoil exercise	Post-test	Motivation and attitude Questionnaire
					b SR/M	2								
					c M/A	1								
					d SDL	1								

Evaluation Questions	Results and Conclusions
Does students' involvement in modelling activities contribute to.. 1. A better understanding of the concepts approached in classroom? 2. Develop several reasoning skills?	Pre and post-test results: There was a significant difference (t = -3.699; df = 22; p = 0.001) between pre-test (mean=3,9±1,76) and post-test (mean=5,7 ±1,9). Showing an increase in their conceptual understanding. Questionnaire results: Most students found good or very good the activities they had using DynaLearn and it has opened new ways of thinking about natural systems. In general the students found easier to identify entities and quantities of systems. Also most students agreed that being able to simulate the models helped them to develop their understanding of the potential behaviour of systems.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size	Evaluation Spaces	Evaluation Designs & Instruments
FUB24 Water cycle with deaf students (Systems Thinking)	LS04	Water cycle	Secondary school hearing students	19	see below	see below

Evaluation Spaces:

	1 CM	2 VC	3 ST
a CU	1		
b SR/M	2		
c M/A	1		
d SDL	1		

Evaluation Designs & Instruments:

O_1	X_1	O_2, X_2	O_3, X_3	O_4, X_4	O_5	O_6
Pre-test	Water in atmosphere exercise	Water in soil exercise	Water in lake exercise	Water in subsoil exercise	Post-test	Motivation and attitude Questionnaire

Evaluation Questions	Results and Conclusions
Does students' involvement in modelling activities contribute to… 1. A better understanding of the concepts approached in classroom? 2. Develop several reasoning skills?	Pre and post-test results: There wasn't significant difference (t = 1.023; df = 18; P = 0.320) between pre-test (mean = 11.21; SD = 2.55) and post-test (mean = 10.57; SD = 2.14). Exercise results: It was ran a Repeated Measures Analysis of Variance on Ranks and there wasn't significant difference (X^2 = 1.200; df = 2; P = 0.549) between mean scores obtained by students in the exercises Questionnaire results: Most students found very good or good the activities they had using DynaLearn and they agreed or totally agreed that it has opened new ways of thinking about natural systems. In general the students found easier to identify entities and quantities of systems. Also most students agreed that being able to simulate the models helped them to develop their understanding of the potential behaviour of systems.

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
FUB25 Teachable agent Environmentalists vs woodcutters (Deaf students)	LS02	Environmentalists and woodcutters	Secondary school deaf students	5

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU		3	
b SR/M			
c M/A		1	
d SDL		2	

Evaluation Designs & Instruments

O_1	O_2, X_1	O_3, X_2	O_4, X_3	O_5, X_4	O_6	O_7
Pre-test	Model building version 1 and quiz	Model building version 2 and quiz	Model building version 3 and quiz	Model building version 4 and quiz	Post-test	Motivation and attitude Questionnaire

Evaluation Questions	Results and Conclusions
The Virtual Characters using Teachable Agent mode contribute to… 1. Increase learners' motivation to learn environmental science topics? 2. Improve the interaction between the learners and the software while revising and improving their initial model? 3. A better understanding of the concepts approached in classroom?	<u>Pre and post-test results</u>: There was no significant difference between pre and post-test. <u>Motivation and Attitude Questionnaire results</u>: Most students agreed that the quiz helped them to understand better the model and it also helped them to correct their initial mistakes. In general the students totally agreed or agreed that teach the Hamster motivated them to understand correctly the concepts.

<table>
<tr><th>Activity Title</th><th>Learning Spaces</th><th>Content Domain</th><th>Target Population</th><th>Sample Size</th><th>Evaluation Spaces</th><th>Evaluation Designs & Instruments</th></tr>
<tr><td>FUB26

Environmentalists vs woodcutters (Hearing students)</td><td>LS02</td><td>Environmentalists and woodcutters</td><td>Secondary school hearing students</td><td>27</td><td>
<table>
<tr><th></th><th>1 CM</th><th>2 VC</th><th>3 ST</th></tr>
<tr><td>a CU</td><td>3</td><td></td><td></td></tr>
<tr><td>b SR/M</td><td></td><td></td><td></td></tr>
<tr><td>c M/A</td><td>1</td><td></td><td></td></tr>
<tr><td>d SDL</td><td>2</td><td></td><td></td></tr>
</table>
</td><td>
<table>
<tr><th>O_1</th><th>O_2, X_1</th><th>O_3</th><th>O_4</th></tr>
<tr><td>Pre-test</td><td>Model building</td><td>Post-test</td><td>Motivation and attitude Questionnaire</td></tr>
</table>
</td></tr>
<tr><th colspan="6">Evaluation Questions</th><th>Results and Conclusions</th></tr>
<tr><td colspan="6">Does students’ involvement in modelling activities contribute to …

1. Increase learners’ motivation to learn environmental science topics?
2. Improve the interaction between the learners and the software while revising and improving their initial model?
3. A better understanding of the concepts approached in classroom?</td><td><u>Pre and post-test results</u>:
There was significant difference between pre and post-test (t = -2.327; df = 26; P = 0.028).
<u>Motivation and Attitude Questionnaire results</u>:
Most students found very good or good the activities they had using DynaLearn and they agreed or totally agreed that it has opened new ways of thinking about natural systems.
In general the students found easier to identify entities and quantities of systems.
Also most students agreed that being able to simulate the models helped them to develop their understanding of the potential behaviour of systems.</td></tr>
</table>

Activity Title	Learning Spaces	Content Domain	Target Population	Sample Size
FUB27 Teachable Agent Environmentalists vs woodcutters (Hearing students)	LS02	Environmentalists and woodcutters	Secondary school hearing students (Kilma's students)	27

Evaluation Spaces

	1 CM	2 VC	3 ST
a CU		3	
b SR/M			
c M/A		1	
d SDL		2	

Evaluation Designs & Instruments

O_1	O_2, X_1	O_3, X_2	O_4, X_3	O_5, X_4	O_6	O_7
Pre-test	Model building version 1 and quiz	Model building version 2 and quiz	Model building version 3 and quiz	Model building version 4 and quiz	Post-test	Motivation and attitude Questionnaire

Evaluation Questions	Results and Conclusions
Did the Virtual Characters using Teachable Agent mode contribute to… 1. Increase learners' motivation to learn environmental science topics? 2. Improve the interaction between the learners and the software while revising and improving their initial model? 3. A better understanding of the concepts approached in classroom?	<u>Pre and post-test results</u>: There was no significant difference between pre and post-test (v = 1.720; df = 26; *P* = 0.088). *Results under revision.* <u>Motivation and Attitude Questionnaire results</u>: Most students agreed that they really felt part of the Quiz and wanted to do well in the quiz. To answer a question correctly motivated the students. In general the students agreed or totally agreed that they also learned while they were teaching their Pets. Most students agreed that the quiz helped them to understand better the model and it also helped them to correct their initial mistakes. In general the students totally agreed or agreed that teach the Hamster motivated them to understand correctly the concepts.

Appendix B: Teachers Questionnaire

Expert-teachers questionnaire

[Please expand the spaces for the answers or add additional pages if needed]

A. Background

Name: ______________________________

1. Assess your own expertise with DynaLearn: Novice 1 2 3 4 5 Expert

2. What is your disciplinary background?

B. Personal view of the rationale for using DynaLearn

3. What is your pedagogical rationale for using DynaLearn in Science (or other subjects) teaching?

4. What do you see as the main added value of Learning by Modelling with DynaLearn approach?

5. Rank the contribution of LbQM to the following goals on a scale of “minor 1 to 5 high”?

	rank
fostering a systems worldview	
acquiring system thinking skills	
acquiring modelling capabilities	
mastering content/subject-matter	
Other	

6. Explain your understanding of each goal

fostering a systems worldview	
acquiring system thinking skills	
acquiring modelling capabilities	
mastering content/subject-matter	
Other	

C. Pedagogical aspects

7. What do you consider prerequisite knowledge and skills needed for Learning by Modelling using DynaLearn?

---

8. What characterized your teaching at the different Learning Spaces (approach, foci, resources besides DL, introduction of features, tasks administered)?

---

9. What do you think should be the preferred learning goal (Knowledge, skills, modelling capability) in each Learning Space? What are the strengths and weaknesses of each Learning Space?

LS1	
LS2	
LS3	
LS4	
LS5	
LS6	

10. What do you think is the pedagogical added value afforded by the different DynaLearn features?

Conceptual modelling	
Teachable Agent	
Quiz	
Diagnosis	
Feedback	
Basic help	
Other	

11. Rank your preference of the following teaching modes on a scale of “minor 1 to 5 high?

	rank	comments
Complete an independent modelling task for each LS		
One modelling task (phenomenon) across all LS - "evolving models"		
Individual / small group / whole class work		
Lesson plan framed by a reference (expert) model		
Open model-construction tasks - not oriented towards an expert-model		
Modelling embedded in normal curricular activities		
Modelling activity not related to the curriculum		

D. On the learners' learning

12. How do you perceive the extent to which the following goals have been achieved through your teaching with the LbM approach in DynaLearn, on a scale of “not-at-all 1 to 5 to a large-extent"?

	rank	comments
fostering a systems' worldview		
acquiring system thinking skills		
acquiring modelling capabilities		
mastering content/subject-matter		
Other		

13. Through your experience in teaching with DynaLearn - have you collected observations or evidence on students' independent/self-directed learning (SDL)?
    a. Please describe a few examples
    b. What DL features do you think supported SDL

---

14. Can you describe typical processes students go through while learning by modelling. What are their barriers? How do they progress?

E. On problems/difficulties faced

15. Please specify problems/difficulties encountered during the activities

	Teacher	Students
Learning/pedagogical		
Technological		

16. Do you have suggestions for the completion/improvement of the software?

F. Representative examples

17. Please supply descriptions of peak experiences in Learning by Modelling using DynaLearn

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