Deliverable number:	D7.3.4
Deliverable title:	TAU evaluation of DynaLearn final software

Delivery date:	2011/10/31
Submission date:	2012/02/01
Leading beneficiary:	Tel-Aviv University (TAU)
Status:	Version: final
Dissemination level:	PU (public)
Authors:	Mioduser, Zuzovsky, Zurel, Leiba, Benayahu, Nachmias, Ram

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

Abstract

This deliverable reports about the results of the second phase of evaluation activities of DynaLearn by Tel Aviv University (TAU). The evaluation activities were designed with the aim to address the evaluation questions posed in the Description of Work, in correspondence with the functionalities and features afforded by the modelling environment in this phase. We report the results of six evaluation activities: One with Undergraduate students at Tel Aviv University; two with Junior-High School students in regular Science classes; one with Junior-High School students attending a summer course in Marine Biology; one case study of an advanced student-modeller; one planned with expert-teachers. Results are discussed for each activity, and summarized into main themes including conclusions in the general discussion chapter.

Internal review

- UH
- UVA

Acknowledgements

The authors would like to thank our partners in WP7 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 in Tel Aviv University's Marine Biology courses in years 2010 and 2011; Junior High students in the Maabarot School's spring course on Science; and Junior High students in the summer course on Environmental Science. As well our thanks go to the University staff who assisted us with the technological facilities and labs during the activities. Thanks also to our reviewers for their contribution to improve this deliverable.

Document History

Version	Modification(s)	Date	Author(s)
01	Draft for TAU group internal review	2011-20-12	Mioduser, Zuzovsky, Zurel, Leiba, Benayahu, Nachmias, Ram.
02	Draft for internal review	2012-1-3	Mioduser, Zuzovsky, Zurel, Leiba, Benayahu, Nachmias, Ram.
03	Revised version	2012-1-20	Mioduser, Zuzovsky, Zurel, Leiba, Benayahu, Nachmias, Ram.
04	Final version	2012-2-28	Mioduser, Zuzovsky, Zurel, Leiba, Benayahu, Nachmias, Ram.

Contents

1. Introduction

This deliverable reports about the results of the second phase of evaluation activities of DynaLearn by Tel Aviv University (TAU). The evaluation activities were designed with the aim to address the evaluation questions posed in the DOW, in correspondence with the functionalities and features afforded by the modelling environment in this phase. The main questions were operationalized 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 framework will be briefly described, as well as an overview of the current phase evaluation activities. Report on the activities is presented in chapter 2, 3 and 4 of this report, and a summary discussion of the results of this phase in chapter 6.

1.1. DynaLearn pedagogical and technological components

The DynaLearn project's main goal, as stated in the Description of Work (DOW) is "*to develop an interactive learning environment that allows learners to construct their conceptual system knowledge, either individually or in a collaborative setting*". The learning environment under development integrates three well established, but as yet independent, technologies to create an individualised and engaging cognitive tool for acquiring conceptual knowledge in environmental science. The software integrates a diagrammatic approach to constructing qualitative conceptual models, ontology mapping and semantic technology to ground model building terms and compare models, and virtual character technology to provide individualised feedback and enhance motivation of learners (Bredeweg et al., 2010).

The DynaLearn approach and paradigm are grounded in the "Learning-by-Modelling" (LbM) pedagogical approach. In this approach, students are involved in interpreting, manipulating and constructing models of systems, fostering both conceptual understanding (of subject-related content knowledge) and the gradual development of robust scientific reasoning skills.

Integrating all above components, namely, the different technologies embedded in the learning environment and the target pedagogical and learning objectives with Dynalearn, we defined the framework for the design of the evaluation activities.

1.2. Evaluation framework

The evaluation framework is built as the inquiry space for the formulation of the evaluation questions and the design of the evaluation activities aiming to answer these questions. Our main goal for the evaluation is:

> To assess the contribution of learning by qualitative modelling with DynaLearn on students' understanding of ecological systems.

In specific terms, we aim 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 structure and behaviour of complex ecosystems*.

- Scientific reasoning, Qualitative reasoning, 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*.

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 agents (VC) - as these act in various functions and roles while interacting with the learner.
- Semantic Technologies (ST) - individualization 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. In the Figure also indicated as shaded cells the issues addressed in our first phase evaluation activities, and the issues addressed in phase 2. Most questions and data collected in phase 1 (see Mioduser et al. 2011) 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 the previous evaluation phase, with the availability of features related to DynaLearn's semantic technologies, we also conducted initial activities about their effect on learning (Questions of types 3a and 3b). In the second phase of evaluation activities reported here we continued our inquiry on aspects related to CM, but our main target was to address the effect of the additional components, namely VC and ST, on students' learning and acquisition of system thinking skills and modelling capabilities.

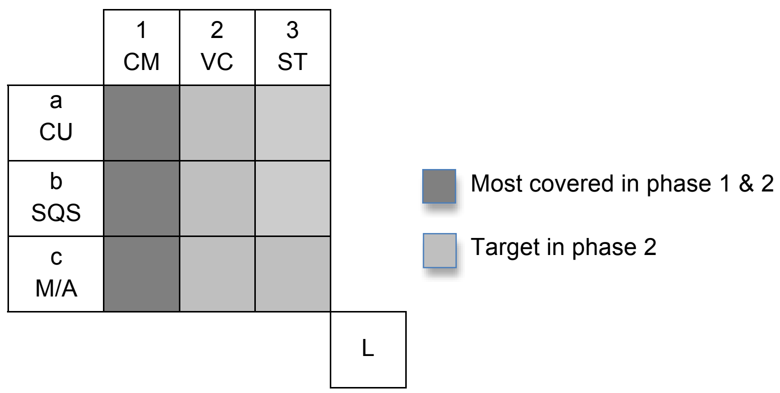


Figure 1.1: evaluation inquiry space

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? [general - L]
6. Do learners actually learn better when using the full set of DynaLearn features? [general - L]
7. And are students more motivated to take on science curricula? [1c, 2c, 3c]

Thus, according to the goals set for the second phase of the evaluation, the main questions directly addressed were 3 and 4.

For questions 5 and 6 relating to the most general issues and aims of the DynaLearn project, the answers: (a) will be gradually constructed upon the cumulative results for the various sub-questions along the project, and (b) will be addressed in the summary report under preparation (D7.4) while integrating data and evaluation results from all partners on the whole spectrum of questions examined.

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, variables and interpretative schemes for the result are detailed for each evaluation activity in the following chapters.

1.3. Evaluation plan for phase 2

Table 1.1: evaluation activities in phase 2

	Evaluation activity	Focus	Duration
1	Marine Biology Undergraduate course 14 students (whole semester course)	Effect of grounding task on modelling	2 classes x 2 hours each class
2	Maabarot High-School 30 students	Evolving conceptual understanding through successive modelling tasks	5 classes x 3 hours each class
3	Maabarot High-School 9 students	Effect of grounding task on modelling	1 class x 3 hours
4	High-School students focus group 15 students	Effect of interaction with VC’s (TA) on students’ modelling of a system	2 days x 8 x 90 minutes sessions
5	Case study	Advanced modelling by an “expert-modeller” High School student	1 day x 4 x 90 minutes sessions
6	Expert-teachers questionnaire – WP7 partners (and associates)	Perceptions and pedagogical insights concerning the integration of DynaLearn in enviromental science education	Planned activity

The evaluation plan for phase 2 is presented in Table 1.1. Six activities were planned with undergraduate students, Junior-High School students, and expert-teachers. In this deliverable we report about results obtained in activities 2 to 5. Evaluation activity 1 was conducted in the fall/winter semester 2010/2011 and a detailed report has been already included in Mioduser et al., 2011. Thus, we will present in this deliverable only a brief summary of the findings and integrate these in the general discussion about the issues examined in phase 2 (see section 5). Activity 6,an expert-teachers questionnaire, is ongoing and is aimed to become a chapter about a wide scope of pedagogical issues in the final deliverable D7.4. In section 6 the goals and instruments of this evaluation activity are presented.

2. Evaluation activity No. 2 and 3: Junior High students LbM

2.1. Introduction

During May 2011, an intervention of five two-hour meetings took place in a comprehensive high school at a kibbutz in the central part of Israel (Ma'abarot). The intervention was part of a biology class. Students were studying ecological systems and were introduced to the main concepts related to the topic before the intervention started. In the five meetings that followed, they were introduced to the idea of learning by modelling, and exercised the use of the DynaLearn software for actual modelling of ecological systems in increasing complexity of learning spaces. They were also introduced to the idea of grounding their models, as a means to better understand the meaning of the terms that appear in their models. The actual task conducted made use of the grounding feature embedded in DL, enabling connecting the terms in the student models, via semantic mechanisms, to information sources (i.e., Wikipedia, see Lozano et al., 2010). One of the meetings was aimed to evaluate the contribution of the grounding mechanism, this time, only toward the improvement of their conceptual modelling of a given ecological system through a more adequate choice of entities, quantities and relationships.

The evaluation questions for the whole intervention were as follows:

1. Does LbM with DynaLearn contribute to junior high students'
    a. Conceptual understanding of an ecological system?
    b. Ability to model the system in different levels of complexity from learning space 1 (concept map) to learning space 3 (causal models)?
2. Does the use of the grounding mechanism improve students' ability to model a complex system more precisely?

In the following sections the evaluation activity method, instruments and results are presented.

2.2. Method

2.2.1. Participants

Participants were 30 8th grade students in a school in central Israel. In the specific activity focusing on grounding the models, data collection and analyses were conducted with 15 students, arbitrarily selected from the whole group.

2.2.2. Variables

The variables for which the data have been collected and analyzed were defined in correspondence with the evaluation questions:

- Students' conceptual understanding of a text that describes the effects of flood and pollution on a population of soft turtles in a nearby river.
- Students' modelling capabilities - appropriate representation of entities, quantities, relationships and model configuration.

- Students' choice of accurate information from a semantic repository regarding the terms that appear in the students' models.

Detailed description of the data collected and scoring schemes for these variables appear in section 2.2.4.

2.2.3. Implementation instruments: The intervention and evaluation plan

The intervention lasted five meetings conducted once a week for two to three hours each time (total about 15 hours). The practical modelling work was conducted using DynaLearn, specifficaly in Learning Spaces 1/2/3. In Table 2.1 the intervention and evaluation activities are presented.

Table 2.1: Intervention and evaluation activities

Meeting	Content	Evaluation Activity
1.	Introduction to LbM. Reading a text describing the effect of pollution and a flood on the population of soft turtles in a nearby river (see Appendix A)	– Answering a set of questions related to main ecological concepts embedded in the text. (see questions – Appendix A).
2.	Following the text, students were asked to draw a concept map using DynaLearn (learning space 1) (CM).	Analyzing the concept map to compare it to the other models created. (See further criteria for assessing modelling capabilities).
3.	Modelling the same phenomenon at learning space 2 (M1).	Analysis of the models using similar criteria (see
4.	Modelling a different ecological phenomenon at learning space 2 and 3 (M2).	Analysis of the models using the same criteria.
5.	Grounding activity: Phase 1 – filling in given terms into a "blind model" and revising the model if needed. Phase 2 – looking for an unknown term of the model using the repository, revising the model if needed. Students were requested to describe and predict the behavior of the system under certain conditions.	Students were given a list of 7 entities, (some unfamiliar) and 4 quantities to fill in the blind model, and were requested to describe the relationship between the entities. (The blind model and the given components appear in Appendix B, C).

In the meetings the students were introduced to: (1) the qualitative modelling approach - its logic and terminology as manifested in the software; (2) the use of concept maps to represent the entities of a given system and types of relationships among them; (3) ways to express quantities in a qualitative model: assign values to entities and processes and define quantity spaces; (4) way to address processes occurring within the system (such as predation or competition) or caused by external agents (such as over-fishing or pollution); (5) running the models. For the specific activity focusing on grounding, the students were introduced to the procedures required for accomplising the task.

The modelling tasks were based on a text describing scientific phenomena. The text dealt mainly with issues related to pollution and its effect on ecosystems. The students were asked to read the text, understand the phenomenon described, analyze the components, relationship and process that were described in it, represent them in a concept map and gradually build qualitative models that represent the dynamics of the relevant ecological system. In addition, they performed the "blind model" task.

2.2.4. Study's design and Data collection instruments

The evaluation activity followed a "One group repeated-measures" design (Creswell, 2003), as shown in Figure 2.1. In it 'X' represents the exposure to a sequence of modelling tasks and 'O' represents an observation or measurement obtained by different instruments.

O	→	X	→	O	→	X	→	O
CU test		Progression of ModellingTasks CM, M1, M2		Process data: student models CM, M1, M2		"Blind model" and grounding task		Process data: “Blind model” and grounding task

Figure 2.1: the study's design

Data analysis instruments included the structured evaluation of students' concept maps, a widely used method for assessing conceptual understanding (Novak & Cañas, 2008; Ruiz-Primo & Shavelson, 1996; Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005). This task was administered before students were exposed to the qualitative modelling tasks. The progression in modelling capabilities was assessed through analyses of the intermediate products in students' modelling tasks.

2.2.5. Data analysis

In this evaluation activity different criteria and scoring guides were employed for each evaluation question. In the following section the criteria and scoring guides for each question are presented.

Assessing students' prior understanding of the phenomenon to be modelled

A detailed scoring guide enabled the scoring of student responses to seven questions that followed the text about the soft turtles. The maximum score that could be obtained was 25 score points.The score points that each student obtained in % from the maximum score point comprised his/her score.

Assessing students' modelling capabilities

The criteria used to assess students' modelling capabilities were as follows:

a. Students' ways of representing the configuration of the system (see Figure 2.2)

- "Linear chain" of entities and processes affecting each other (L)
- A "Sun representation" of element in a system all related to a central one (S)
- "Hierarchical representation" that take into account different levels in an ecosystem (H)
- “Hierarchical representation with within-levels links”

- "Weblike" (net like) representation where links of relationship within and between levels are considered (W)

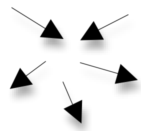
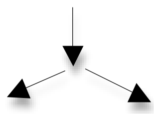
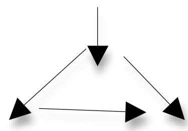
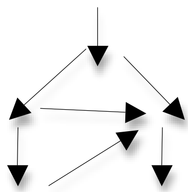
Linear or "chain" or type	
"Sun" type	
"Hierarchical" type with the number of levels in brackets	
"Hierarchical with within level links" type with the number of levels in brackets	
"Web" type – within and cross level links	

Figure 2.2: patterns of relationship among the system's components

b. Indication of structural vs. process relationships among the elements of the system. For this criterion, the number of links between entities that describe hierarchical or structural relationships vs. those that refer to processes have been registered.

c. Indication of quantities. The models were examined as to whether the students ignore quantities of in the system's model (0); refer to quantities (Q); refer to quantities and define the direction of the relationship between them (Q±).

d. The organizing principle used in the construction of the representation both in the concept maps and in the models. We examined whether the student follows a story and gives a description of an event (E) or s/he organizes the representation around meaningful concepts (C).

Assessing students modelling capability as a result of the grounding activity

Two sets of evaluation criteria were used for the grounding-based activity, for assessing the students' work before and following the grounding of the models.

The evaluation criteria employed for phase 1, modelling before the grounding activity, appear in Table 2.2. The criteria for phase 2, after grounding, in Table 2.3.

A mixed method approach, mainly qualitative in nature, was adopted in this study. Comparing between the initial concept of maps that represented the relevant ecological system studied with the intermediate and final products of the modelling process, enabled to trace growth in conceptual understanding of ecological concepts, the understanding of the ecological complexity underlying the ecosystems, and the acquisition of modelling capabilities and scientific reasoning skills.

Table 2.2: Evaluation criteria for student models before grounding

Criteria	Scale
Correct fitting of the 6 entities	Proportion of correct fit out of 6
Correct fitting of the 5 quantities	Proportion of correct fit out of 5
Correct fitting of the 4 types of relationships	Proportion of correct fit out of 4
Proper configuration of the model	1 – wrong or missing hierarchy; 2 - partially correct; 3 - completely correct

Table 2.3: Evaluation criteria for student models after grounding

Criteria	Scale
Adequacy of information chosen for each entity	1 – not adequate; 2 – partially adequate; 3 – adequate
Location of the information for each entity in the database	I - first in the list; II - second; III - third…
Revision of the model	0 - No revision 1 - Did revised the model

2.3. Results

2.3.1. Students' prior understanding of the phenomenon

Data collected with the prior-understanding questionnaire is presented in Table 2.4. Only 15 students of the 30 participated in the test. Their identification number is marked in Table 2.4. Students' prior conceptual understanding of key concepts related to the phenomenon described was moderate (mean - 64%), with only three students scoring above 80% of maximum score.

Table 2.4: Students prior understanding*

2	6	7	9	10	12	15	16	17	19	20	21	23	25	26
16/25 64%	15/25 60%	18/25 72%	6/25 24%	19/25 76%	21/25 84%	12/25 48%	16/25 64%	21/25 84%	17/25 68%	13/25 52%	10/25 40%	17/25 68%	16/25 64%	22/25 88%

(*) In the Table: student ID, number of correct answers out of 25, and score in %

2.3.2. Students' modelling of the system

Student models were analyzed using the criteria and scoring guides presented in section 2.2.5. The results are presented in Table 2.5. It should be noted that M2 required to model a new phenomenon, thus requiring transfer of the gained knowledge into a new situation.

Table 2.5: Students modelling performance*

ID	Configuration[1]			Proportion of structural-to-process links			Quantities[2]		Organizing Principles[3]		
	CM	M1	M2	CM	M1	M2	M1	M2	CM	M1	M2
2	L (8)	L 8	-	0/5	0/5	–	(0)	(0)	E	E	–
6	S8	S8	-	0/7	1/6	—	(0)	(0)	E	E	–
7	H(3)	H(2)	L3 (loop)	3/2	2/2	0/3	(0)	(Q)	C	C	C
9	H(1)	W	W	–	2/6	0/8	(Q)	(Q)	C	C	C
10	L(6)	H(2)	W	0/5	0/7	0/11	(0)	(Q)	E	E	C
12	L(6)	H(2)	L(3)	0/5	0/7	0/2	(0)	(Q±)	E	E	C
15	L(3)	H(2)	L(3) Loop	0/2	2/2	0/3	–	(Q)	E	–	C
16	H(2)	H(2)	L(3)	2/2	1/3	0/2	(0)	(0)	C	C	C
17	H(5)	H(3)	L(4)	3/6	2/5	0/3	(0)	(Q±)	C	C	C
19	S (6)	S(6)	–	0/6	0/6	–	(0)	–	C	C	–
20	L(9)	L(9)	H(1)	0/8	0/8	1/2	(0)	(0)	E	E	E
21	W	S(11)	S(9)	6/4	5/5	0/9	(Q)	(Q)	C	C	C
23	H(3)	W	H2	0/9	1/9	0/4	(Q)	(Q±)	E	E	C
25	W	H(3)	L(5)	8/11	0/7	0/10	(Q)	(Q±)	C	C	C
26	H(3)	H(4)	H(2)	0/5	2/9	0/4	(Q)	(Q±)	C	E	C

(*) Same criteria were applied for analysing student Concept Maps (CMap), Models 1 and 2 (M1, M2).

([1]) The symbols used referred to the type of configuration, in brackets the number of elements included.

([2]) The symbols refer to "0" without quantities, "Q" referring to quantities, and "Q±" quantities + causal directions.

([3]) Symbols refer to "E" following events, "C" following conceptual representation.

The analysis of students' modelling capability criteria revealed the following:

a. Concerning the overall configuration of the models, students' represented the phenomena using mainly linear and hierarchical representations.

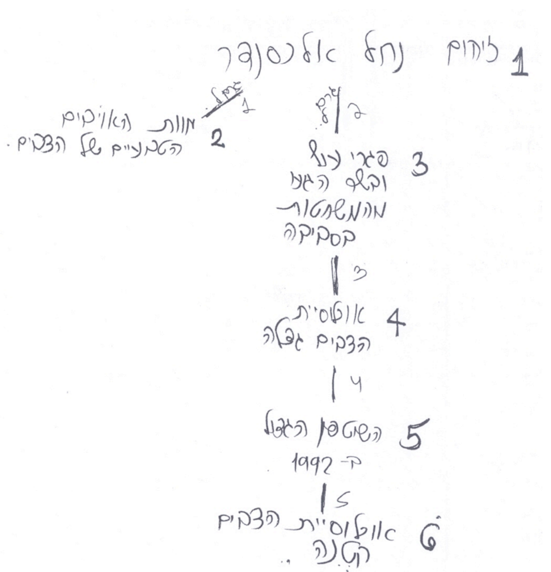
b. Only a few of the relationships defined were structural relationships, most of them described process or causal relationships reflecting the dynamic aspects of the system. Proportions for the CM were 30% structural and 70% process relationships, in M1 20% and 80% correspondingly, and in M2 almost 100% of the indicated relationships were of the process type. These results indicate that as students gained more experience in modelling (i.e., in analysing phenomena and defining the way to represent their essential features in a model); they were able to focus more on dynamics aspects and on the processes characterizing the interrelationships among the systems' components.

c. In all concept maps and in most of the constructed models that followed the concept map of the river habitat, most students ignored the definition of quantities. However, in their second modelling activity addressing a new ecological system, ten out of fourteen models included the representation of quantities, including, in half of these, indication of the direction of the relationship between the quantities (marked by '+' or '-'). Repeated experience in modelling seems to support deeper analyses of a phenomenon and refinement in the way of representing all its essential traits (e.g., quantities and relationships).

d. Models were also analyzed in terms of the organizing principle leading the construction of the model, whether it follows a "story" in correspondence with the text description, or is being built around key concepts distilled from the text description. In the first and second modelling activity (CM and M1), the use of both approaches was evenly distributed. In M2 however, all but one of the models were constructed around concepts. This implies an important change in perspective indicating the development of a deeper and generic systemic approach towards modelled phenomena.

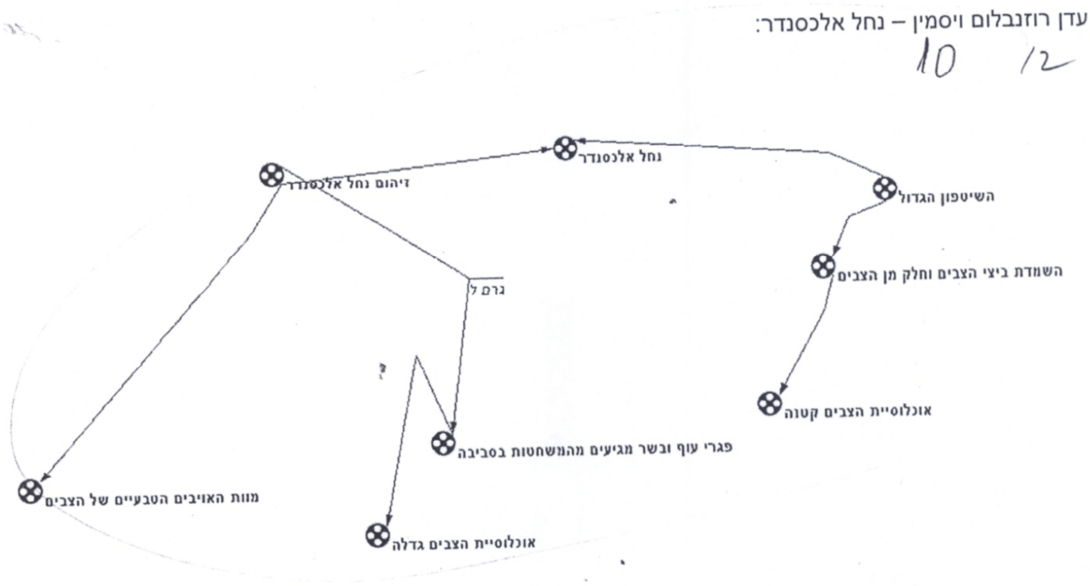
e. Concerning the models' representational structure, students' progress from a linear representation to hierarchical and web-like representation was rare. An example of such progress is represented in a set of models created by one of the participants, ID10. The following set of figures illustrates this progress:

- Figure 2.3 shows a linear representation, mostly of a "chain type" (1), of the soft turtles population.
- Figure 2.4 shows a two-level hierarchical representation of the same phenomenon.
- Figure 2.5 shows a web-like model with cross-level representation with specification of quantities, this time representing a new phenomenon of prey-predator relationship.



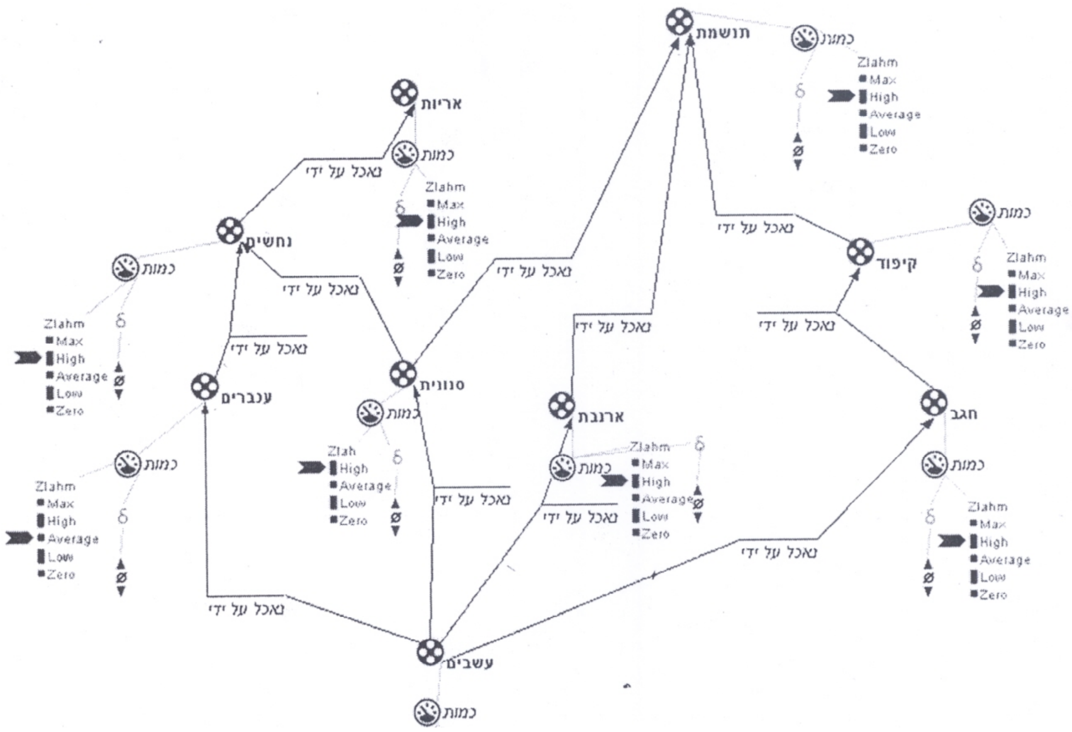
Concept map representing pollution in a river system. The map is built as linear hierarchy, "telling the story" of the river: organic waste from surrounding industries is deposited in the river >> the fish population increases >> big flood in 1922 >> turtle population decreases …

Figure 2.3: Linear representation in a Concept Map



The model in Figure 2.3 includes two branches, the left about the effect of pollution (cousing both a decrease in turtle predators and an increase in turtles poplation that feed on the pollution) and the right about the effects of the flood (causing eggs destruction and injuries to the turtles and thus a decrease in population).

Figure 2.4: Two-level hierarchical representation in a model



Representation of the prey/predator web indicating the different populations (e.g., reptiles, mammals, insects, grass), quantity spaces for the populations' size, and the interrelationaships among these.

Figure 2.5: Web-like model with cross-level representation

The linear representation in Figure 2.3 "tells a story" as an ordered sequence of events. Rather than a representation of the phenomenon or the system, it enumerates the chain of events and consequences in the river environment. Figure 2.4 shows ID10's M1 about the same phenomenon. This time the model consists of a three branches hierarchical representation, being the upper node the river itself, one branch indicating the negative effect of a serious flood on the size of the turtles population, and another indicating differentially the effect of pollution on the turtles as well as their predators populations. Figure 2.5 shows a significant change in modelling approach and mastery of the modelling (conceptual as well as technical) language. Moreover, the model is about a new phenomenon, indicating the student's ability to use the acquired intellectual instruments and language to address new and different systemic phenomena.

2.3.3. The effect of the 'grounding' activity on students' modelling

Following are the data collected in the evaluation activity which focused on the grounding task. Students were requested to: (a) complete a "blind model" integrating in it ingredients from a given list; (b) answer a series of questions about the modelled phenomenon; (c) ground the model; and (d) modify the model if they find it necessary. Their model and the grounding process were analyzed using the criteria presented in 2.2.5. The results of the analysis appear in Table 2.6.

Table 2.6: The Effect of the Grounding Activity on Modelling Capabilities and Choice of Adequate Information

Model Representation					Adequate Information and Position in the Database						
ID	Entities	Quanti-ties	Rela-tionship	Configu-ration	Mediter-ranean	Ships	Anti-fouling Paint	Sedi-ment	TBT	Gastro-pod	Anti-fouling Paint
1	6/6 100%	4/5 80%	3/4 75%	2	3 (I)	3 (I)	3 (I)	1 (IX)	3 (I)	3 (I)	0
2	6/6 100%	5/5 100%	4/4 100%	3	3 (I)	3 (I)	3 (I)	3 (I)	3 (I)	3 (I)	0
3	6/6 100%	0/5 0%	4/4 100%	1	3 (I)	3 (I)	3 (I)	- (-)	3 (I)	3 (I)	0
4	6/6 100%	0/5 0%	4/4% 100%	1	3 (I)	3 (I)	3 (I)	- (-)	3 (I)	3 (I)	0
5	6/6 100%	2/5 40%	4/4 100%	2	3 (I)	3 (I)	- (-)	3 (I)	2 (III)	3 (I)	0
6	6/6 100%	5/5 100%	4/4 100%	2	3 (I)	3 (I)	3 (I)	1 (IX)	3 (I)	3 (I)	0
7	6/6 100%	2/5 40%	4/4 100%	3	Did not "ground"						
8	6/6 100%	0/5 0%	0/4 0%	1 (linear)	Did not "ground"						
9	3/6 50%	3/5 60%	4/4 100%	2	Did not "ground"						
mean %	5.7/6 94%	2.3/5 47%	3.4/4 86%	1.8/3 63%							

Almost all students incorporated all given entities, however, the position of these entities in a meaningful representation was not always correct. Only 3 out of the 9 students represented a correct configuration, 3 partially correct, and 3 use wrong configurations. The three students who failed to provide correct configurations also did not provide correct quantities.

In almost all cases of grounding, the students chose the first information that appeared in the database, indicating that this tool is used in a very superficial way. However most of the information chosen was relevant, but no further effort to evaluate other candidate options (perhaps a better fit) was carried. It should be noted that we used a model that was not grounded before, and all candidate definitions were offered for evaluation by the students.

All students who participated in the activity and did the grounding did not change their models after it was completed.

Overall, the task of filling the empty model was found to be quite hard for junior high school students, indicating lack of capability to use the abstract template for constructing a meaningful representation by filling in relevant information - in this case facilitated to them as a resources list. In addition, when offered the possibility to change their models following its grounding, they did not consider that the newly acquired information should lead to changes in their models.

2.4. Concluding remarks

In this evaluation activity Junior High students attended a qualitative modelling workshop complementing their regular Science classes. The focus of the evaluation activity were questions concerning the contribution of LbQM to students' capability to model ecological systems, and to the contribution of grounding the models with relevant information to their learning. Following the detailed account of our observations presented in the previous sections, we will close this report with the following general remarks.

Concerning students' conceptual learning and ability to LbQM, our observations showed that along the classes an increase in mastery was reached. Although the models were in most cases of the hierarchical type, these became gradually more complex as well as more conceptually focused. Models 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.

Concerning the impact of the grounding activity, students' 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 serves 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.

At an additional level, technical difficulties affected the conduction of some activities and tasks. Because of the technical state of the software at the time the activities were conducted (release 0.8.8), students experienced different kinds of difficulties, e.g., wording of concepts that did not generate sound feedback (candidate definitions), or the need to try alternative syntactic forms or exclusion of characters in order to get feedback. Meanwhile these features were completed in current versions of the software.

3. Evaluation activity No. 4: High-School students' work with the TA

3.1. Introduction

During August 2011 a set of evaluation activities were conducted with a group of 15 High-School students, in a workshop focusing on learning by modelling systems. The aim of this activity was to assess the use and contribution of several, (as yet un-assessed) features of the software – specifically the use and contribution of the teachable agent (TA) to students' learning (Wißner et al., 2010).

Specifically, the questions addressed in the activity examined the contribution of students' interaction with a Virtual Character in its role as TA to:

- conceptual understanding
- motivation and appreciation of LbM using DynaLearn
- motivation for using the VC feature
- modelling capabilities

In the following sections the evaluation methods, results and conclusions are presented.

3.2. Method

3.2.1. Participants

Participants were 15 high school volunteers, aged 14-16. Only four of them had previous experience in working with DynaLearn - as a result of their participation in a previous-year's summer course. These youngsters were involved in a two-day DynaLearn workshop, with four ninety-minute sessions each day,. The students enrolled voluntarily in the workshop.

3.2.2. Variables

The variables for which data were collected and analyzed were defined in correspondence with the research questions:

- conceptual understanding related to the phenomenon to be modelled;
- motivation and appreciation of learning by modelling;
- motivation for interacting with the Virtual Characters while learning;
- modelling capability.

The detailed description of the variables scales and values, and corresponding scoring schemes, are presented in section 3.2.4.

3.2.3. Implementation instruments: The course and modelling workshop

The contents and tasks implemented in the workshop and the plan for the evaluation activities is shown in Table 3.1.

Table 3.1: Tasks and evaluation activities plan (Add another column – Evaluation Aim

Day/Sessions	Content	Evaluation Activity	Evaluation Aim
1/1	Introduction – learning by qualitative modelling, learning the software, and the modelling language. Demonstrations and explanations and "hands on" activities on how to build LS2 models.	NA?	Conceptual understanding
1/2	Students read a text on "the Jelly Fish invasion to the Mediterranean Sea" and answer eleven questions related to the text (see Appendix D).	Analysis of students' answers to the questionnaire.	Conceptual understanding
1/3,1/4,1/5	Students build a model at LS2 that represents the Jelly Fish invasion described in the text using given entities, quantities and relationships. Testing a given hypothesis related to the Jelly Fish phenomenon by using the Teachable Agent feature. The TA responses to the quiz are used by the students to revise their models.	"Learning by modelling" questionnaire (see Appendix E) administered at the end of the modelling experience. "Motivation in using VC" questionnaire (see Appendix F) administered after the TA-based task. Analysis of the post-TA-task revised models.	Modelling capabilities
2/1	Explanation and demonstration on how to build LS3 models.	Analyzing student models using a scoring guide (see bellow).	Modelling capabilities
2/2	Students are requested to build LS3 models using the text on the Jelly Fish invasion. Students answer once again the questionnaire (see Appendix D) related to the Jelly Fish invasion text.	Analysis of students' answers to the questionnaire.	Conceptual understanding

3.2.4. Study's design - intervention and data collection plan

The evaluation activity followed a "one-group pre- post test with repeated measures" design (Creswell, 2003). The design and allocation of treatment (X) and observation (O) activities is presented in Figure 3.1.

O	→	X	→	O	→	X	→	O
Text-related questionnaire		LS2 model and TA task		LbM questionnaire TA questionnaire Revised models analysis		LS3 model		models analyses Text-related questionnaire

Figure 3.1: the study design

3.2.5. Data collection instruments and analyses

Data analysis procedures were developed in correspondence with the evaluation goals, for the elaboration of the different types of materials collected in the study. The first stage in the process consisted in defining appropriate scoring guides for the students' responses, products, and observed performance. These are presented in this section.

Assessing students' conceptual understanding

After reading the text about "Jelly Fish invasion" the students were given a test, the items of which dealt with ecological concepts related to the phenomenon described. The test comprised of eleven items. Two were multiple-choice items - each item contributed one score point to the total score. The remaining items were of open response type with three scoring options: fully correct response-3 score points; partially correct response-2 points, and vague response-1 score point, 0 score for no response or incorrect response was not used as these types of answers do not appear in our data. The total possible score on this test was 29 score points.

The test was administered twice - the first time after introducing the text on the first day, and the second time after students' several modelling experiences at LS2 and LS3.

Assessing student motivation and appreciation of learning by modelling using DynaLearn:

Responses of students to a set of 12, Likert-type, items related to the modelling activities they experienced provided us with information on:

- Students' overall appreciation regarding the contribution of the modelling experience using DynaLearn for understanding (item 1).
- Their motivation to continue using the software for learning by modelling (item 2).
- Their appreciation of the contribution of learning by:
    - observing (Item 3).
    - Building their own models (Item 4).
    - Running a simulation (Item 5).
    - Using the feedback mechanism when comparing their models to their friends' model (Item 6); and to an expert model (Item 7).
- Two items assessed their appreciation of the "grounding" features (Items 8,9) -.

The questionnaire (see Appendix F) was administered at the end of the first day when the TA feature was implemented and students had the opportunity to build models in LS2, and again at the end of the second day after building models in LS3.

Assessing student's motivation about interacting with the VCs

A short questionnaire dedicated to the use and enjoyment of the VCs was administered at the end of the first day (See Appendix F, questions 10-12 in part I and 6 questions in part II). The questions focused on students' perceptions of features of the VCs and of their contribution to the modelling tasks.

Assessing students' model-building and model-revision capabilities

In this activity students were provided with a given set of entities and specified interrelationships between them that represented the Jelly Fish invasion phenomenon described in the text they had read. They were asked first to construct a LS1 model (concept map), then a LS2 model. They were also asked to raise hypotheses in regard to the phenomenon represented in their model and run a simulation to test these hypotheses. On the basis of their model, the TA responded to a quiz that tested the behaviour of the system under certain conditions. The success of the TA to correctly answer the quiz questions indicated a match between student's and expert's models. Students were informed that wrong answers appearing in red indicated a mismatch between the TA model (based on students' models) and an expert model (existing in the repository). This mismatch required a revision of the students' models. The students continued to revise their models in order to improve it, until all questions in the quiz were answered correctly. During this process, students were asked to save their models each time before a new revision. These models were then analyzed using the expert model as reference. It should be noted that the activity was conducted using release 0.8.8 of the software, before more recent modifications to the TA mode have been introduced.

The criteria used for analyzing the quality of students' models were:

a. Structure - whether all entities and given terms were integrated in the model and accurate relationships were defined.

b. Quantity space - whether the defined quantities and quantity spaces accurately fit each entity.

c. Causal relationships - appropriately defined between the quantities specified in the model.

A set of statements **deduced** from the expert model in relation to each of the above criteria served as standard, and was used as a scoring guide for analyzing statements based on students' models. These deduced statements were unknown to the students.

Table 3.2: Scoring guide for student models

Structure (maximum score = 6)
Marine ecosystem contained in Mediterranean Sea. (1)
Fishing Industry exploits marine ecosystem. (1)
Global warming affects marine ecosystem. (1)
Fish live in Mediterranean Sea. (1)
Crabs live in Mediterranean Sea. (1)
Jelly Fish live in Mediterranean Sea. (1)
Quantities (maximum score = 8)
Fishing industry – *Over-fishing*. (1)
Fish – Number of; *consumption*. (2)
Crab – *Number of* (1)
Jelly Fish – *Number of; consumption*.(2)
Global warming – *Heat*. (1)
Mediterranean Sea – *Temperature*. (1)
Causal Relationships (maximum score = 7)
Over fishing $\xrightarrow{-}$ number of fish. (1)
Number of fish $\xrightarrow{+}$ fish consumption. (1)
Jelly Fish consumption $\xrightarrow{-}$ number of crabs. (1)
Heat $\xrightarrow{+}$ temperature. (1)

Fish consumption $\xrightarrow{-}$ number of crabs. (1)
Temperature $\xrightarrow{+}$ number of Jelly Fish. (1)
Number of Jelly Fish $\xrightarrow{+}$ Jelly Fish consumption. (1)

3.3. Results

3.3.1. Students' gain in knowledge and conceptual understanding

Students were requested to answer a set of 11 questions related to the "Jelly Fish Invasion" phenomenon before and after the intervention. The average score for each question in the knowledge and understanding pre- and post-test is presented in Table 3.3.

Table 3.3: Pre- and post-test average scores in the knowledge and understanding test (N=15)

Item	Pre-tests Mean (*SD*)		Post-test Mean (*SD*)	
1	1.0	(0.00)	1.0	(0.00)
2	1.0	(0.00)	1.0	(0.00)
3	2.6	(0.65)	2.6	(0.51)
4	2.9	(0.27)	2.7	(0.49)
5	2.4	(0.65)	2.3	(0.65)
6	2.6	(0.51)	2.6	(0.49)
7	3.0	(0.00)	2.6	(0.51)
8	2.6	(0.50)	3.0	(0.00)
9	2.8	(0.37)	2.7	(0.47)
10	2.3	(0.50)	2.9	(0.26)
11	1.8	(0.42)	1.5	(0.42)
Total Raw	22.6	(3.1)	21.7	(4.6)
Total %	78%		75%	

The results show that students' conceptual understanding and knowledge in the pre-test was already high, close to 80%. Thus as could be expected, this score did not change following the intervention. This decrease can be attributed to the very short time interval between the two tests administration, and also probably, to the students' lack of motivation to respond again to the same questionnaire after the intervention.

Higher average scores were observed in questions related to processes and causal relationships (e.g., factors causing the invasive Jelly Fish population to increase - question 4), and even to hypotheses formation considering changing conditions (e.g., what will happen to the invasive population if the sea temperature increases by 1 C?). The lowest average scores were observed in questions demanding understanding of more complex relationships and accurate terminology (e.g., on the relationship between the local and invasive Jelly Fish: mutuality, competition, parasitism, prey-predator relationship -question 2- or on the meaning of "ecological equilibrium" - question 11). These results were consistent in both tests, and did not change as a result of the intervention.

3.3.2. Students' motivation to LbM and appreciation of DynaLearn features

At the end of each day of the workshop the students were requested to complete a 12 items questionnaire (5 points likert scale) focusing on their perception of the learning activities and several DynaLearn features. The results appear in Table 3.4.

The average scores at the end of each day were similar. In some items, scores were higher indicating an improvement in students' appreciation for learning by modelling, e.g., building their own models, or comparing their models to experts' ones. In general, responses to the items refer to the virtual characters indicated that the interaction with the VCs was enjoyable.

3.3.3. Students' perceptions of their interaction with VCs while learning

Data were collected with a seven items questionnaire (on a scale of 1-do not agree to 5-fully agree) administered after the completion of the tasks involving interaction with the Virtual Characters. The results are summarized in Table 3.5.

Table 3.4: Average scores in the appreciation-of-learning questionnaire (N=15)

Item		First day	Second day	Trend (Increase/ Decrease)
1	Using the software contributed to my understanding	2.9 (1.1)	3.3 (.90)	+
2	I like to learn by building models in other lessons	3.0 (1.0)	3.1 (.99)	+
3	Most of my learning occurs while observing the teacher	2.9 (.96)	2.6 (.89)	-
4	Most of my learning occurs while building my own model	2.9 (.86)	3.2 (.86)	+
5	Most of my learning occurs while running a simulation	2.5 (.99)	2.7 (.70)	+
6	Most of my learning occurred when comparing my model to my friend's one	2.2 (1.2)	2.2 (1.1)	-
7	Most of my learning occurred when comparing my model to an expert's one.	2.7 (1.0)	2.7 (.96)	-
8	The grounding activity contributed to my knowledge	3.0 (.96)	2.7 (1.0)	-
9	The grounding activity contributed to my understanding	2.9 (.92)	2.7 (.79)	-
10	I was helped al lot by the VC.	2.7 (1.2)	2.6 (1.2)	-
11	I loved the VC. They are fun and amusing	2.7 (.97)	2.9 (1.1)	+
12	I'd like to learn with the VC in other activities as well	2.9 (1.2)	3.0 (1.1)	+
Total		2.9 (1.1)	2.8 (.51)	+

Table 3.5: Average scores in the perception of VC features (N=15)

	Average	(*SD*)
The TA talk was understandable	3.7	(1.3)
The TA behavior was appropriate	4.2	(.67)
The TA facial expressions were appropriate	4.0	(.80)
I wanted my TA to succeed	4.3	(1.3)
I was glad when my TA succeeded	3.3	(1.2)
The TA helped me in building my model	4.0	(.75)
Total motivation	4.1	(.66)

Generally, students perceived positively their interactions with the VCs. However, in items related to their contribution to learning (3 last items in Table 3.4), students graded these in average at the midst of the scale. Concerning the VCs features, they found them most appropriate, with slightly lower mark for their voice. They also agreed upon the VCs motivational value.

3.3.4. Students' ability to construct and revise models of a system

For assessing students' evolving ability to construct models of ecological systems, and to revise and modify these following their interaction with the TA, we analysed the series of models constructed and saved by the students. The models were analyzed using the criteria presented in section 3.2.5 and Table 3.2, concerning the models' structure (max. score 6 structural relationships), quantities (max.score 8 quantities specified), and causal relationships represented (max. score 7 causal relationships indicated). The results are presented in Table 3.6 (see detailed results in Appendix H).

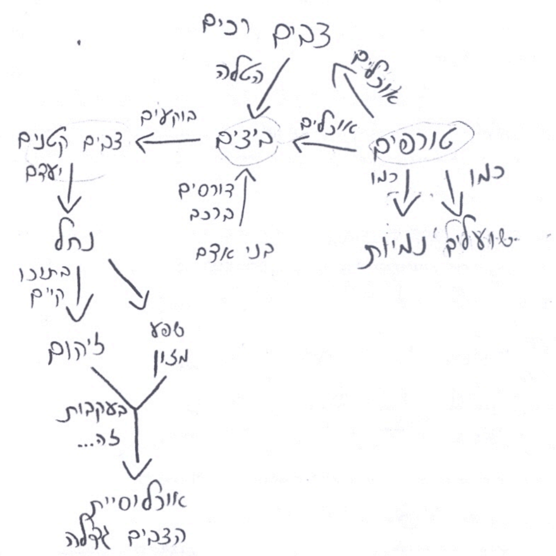
Due to technical difficulties, as well as to some students' decisions concerning the storage of intermediate products, we were only able to collect complete sets of models from 7 students. There was variation in the number of attempts made by the students to revise their models - from none to five revisions: two students revised their model five times; three only three times; and two students only provided one model – either because of lack of motivation or because they did not save the models produced.

All students obtained the maximum possible score points for the structural configuration criteria, specifying correctly the entities and the relationships between the entities involved. At lesser extent they succeeded in specifying quantities or quantity spaces for the entities (average score 78%), and in specifying appropriately causal relationships in the models (average score 49%).

Table 3.6: Students' evolving ability to construct and revise models

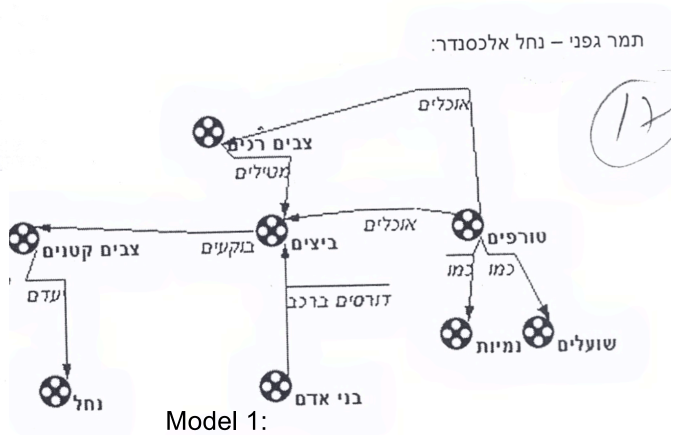
ID	Model #	Scores for Structural Configurations	Scores for Quantities	Scores for Causal Relationships
	1	6/6	2/8	0/7
1	2	6/6	2/8	4/7
	3	6/6	4/8	2/7
	4	6/6	6/8	2/7
	5	6/6	5/8	2/7
Total in %		100%	48%	29%
	1	6/6	6/6	1/7
2	2	6/6	6/6	2/7
	3	6/6	6/6	2/7
	4	6/6	5/6	1/7
	5	6/6	6/6	2/7
Total in %		100%	72%	23%
	1	6/6	6/6	4/7
3	2	6/6	6/6	4/7
	3	6/6	6/6	7/7
Total in %		100%	83%	71%
	1	6/6	8/8	7/7
4	2	6/6	8/8	7/7
	3	6/6	8/8	7/7
Total in %		100%	100%	100%
	1	6/6	5/8	3/7
5	2	6/6	6/8	3/7
	3	6/6	6/8	7/7
Total in %		100%	79%	62%
6	1	6/6	7/8	3/7
Total in %		100%	88%	43%
7	1	6/6	6/8	1/7
Total in %		100%	25%	14%

Looking at each student's sequence of models, we observe consistency in the models' features for all criteria either at the lower or the higher values of the scoring scale (e.g., ID4's models scored high for all criteria in all 3 models). Two students -ID3 and ID5- exemplify progression or improvement in capability by the different criteria from one model to the next in the sequence. ID3's sequence of models serves as illustration of this progressive improvement (see Figure 3.2). The drawing of the concept maps before are acquainted with the software, enable us to look into the 'brains' of the students (before they are forced to use the software). This drawing, thus gives us their intuitive perceptions.



Concept Map:

Structure: 6 (max. 6)
Quantities: 5 (max. 8)
Causal relationships: 3 (max. 7)



Model 1:

Structure: 6 (max. 6)
Quantities: 6 (max. 8)
Causal relationships: 3 (max. 7)

Model 2:

Structure: 6 (max. 6)
Quantities: 8 (max. 8)
Causal relationships: 7 (max. 7)

Figure 3.2: ID3's sequence of models. Scores taken from Table 3.6.

In the Concept Map and the first model, a fairly complete structural description is included (e.g., turtles population, river, predators). However, only few links indicate causal relationships. In Model 2 explicit reference to quantities as well as to processes and causal relationships appears. As the student gains more and better understanding of the systemic perspective, as well as better mastery of the modelling process and language, the model becomes more complex and comprehensive in relation to the modelled phenomenon.

3.4. Concluding remarks

In this evaluation activity Junior High students performed a series of tasks in which they were requested to interact with the Virtual Characters (VCs), specifically with the Teachable Agent.

The activity aimed to examine students' perceptions about the contribution of the interaction with the VCs to their understanding and modelling capabilities, and about the VCs motivational value for learning.

Generally, students perceived positively their interactions with the VCs, and found the activities enjoyable. They found most VCs features appropriate, except for their talk/voice. They also perceived that the task contributed to their learning.

However, concerning the actual contribution of the intervention to the modelling activity, the results were not conclusive. These 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 activity, no difference was found in student products at the lower and higher segment of the scoring scale of the variables considered, and only in 2 cases progression or improvement was observed. It seems that students subjectively perceived the value of the VC-based task higher than its actual effect on their work on the models (evident motivational effect).

4. Evaluation activity 5: Advanced modelling process - a case study

4.1. Introduction

Three of the fifteen high school volunteers who participated in the summer workshop conducted in July were invited to participate in a workshop that focused on building advanced models. The three were highly motivated to learn more with DynaLearn and they agreed to come for a full-day activity. This activity opened an opportunity for us to follow students' thinking closely, while building their models.

Students were asked to model and think aloud while modelling, Their thinking aloud was recorded and transcribed. Data for further analyses are these narratives as well as the sequence of models the students produced during the working session. The activity last 2 x 45 minute sessions.

In this report we will focus on one student, who provided us with the most detailed report of their work. The main goal of this evaluation activity is to examine in detail the way a student who has already had enough previous experience in modelling with DynaLearn (i.e., not a naïve student) approaches the modelling of a new complex phenomenon with minimal aid from the expert teacher. This objective is particularly important if we pursue the goal of integrating DynaLearn in regular teaching/learning processess in a continuous fashion, and not as one-time exceptional encounters of (mostly) naïve students with LbQM tasks and tools.

4.2. Method

The participant was a 16 years old student who had already participated in a previous workshop on Marine Biology using DynaLearn. He is very proficient with the technology. During the modelling session he was fully cooperative in describing his thoughts, reflections, decision-making and perceptions of success or failure. His performance was recorded and his models stored - the data was examined and interpreted by three members of our research team.

4.3. Results

4.3.1. Data

The data analysed were the transcripts of the taped student's think-aloud, and the models constructed by the participant. The transcription and its interpretation appears in Table 4.1. The models in Figures 4.1, 4.2, 4.3, 4.4 and 4.5.

The transcription was divided into segments indicative of the stages of I.'s work. Within the segments a distinction has been made between descriptive utterances indicating intention and actual modelling activities (e.g., "*I will add to the workers a measure called 'number-of'*"), and reflective utterances (in 'italics' in the text, e.g., "*one moment, I wonder if adding one quantity to workers will be inherited to all sub-quantities*"). A second column in the tables presents our interpretation of the actions made, and the student reflection on his actions (these appear as Rx.x, e.g. for the previous example - R5.1).

Table 4.1: Student's description of his work - transcription and interpretation

Student Text	Interpretation
1. Hello, my name is I. I am going to build a model at learning space 6 using DynaLearn. The topic I am choosing to model is "Airport" – its security issues and economic ones. Later on I will examine what will happen if suddenly a huge number of travellers will flood the airport, about one million a year. How will it affect the security aspects and the economic aspects?	The student defines the system and raises a problem related to the system phrasing it in an operational way: How will an increase in the independent variable (number of travellers) affect the dependent variables (security and economic aspects of an airport).
2. The first thing I am going to do is to build a concept map and then model it and "inherit" "characteristics" to each of the "concepts".	Early thinking using at this stage intuitive Concept Mapping rather than DynaLearn terminology, e.g., "concepts" or "characteristics" and not entities, quantities, etc. The system is conceived as having hierarchical structure, with levels of entities and inheritance of "characteristics" from class-level concepts to its instances.
3. We will start with the first entity called Airport. Out of this entity we can pull out many other entities, such as: plans, workers, travellers. We can attribute to the entities certain "characteristics" – for example for *planes,* we could specify number of travellers the plane can hold; amount of oil; level of security; number of pilots; frequency of security checks; number of staff needed for operating the plane. Under *workers*, we can specify number of; wages; skills. Under *travellers,* we can specify – number of; amount of free money for travelling; demand for vacations. We can classify workers to more entities, such as: stewards, pilots, security people, mechanics, engineers and a big list of others. All these entities "inherit" the characteristics workers have.	Preliminary thoughts about the static part of the model using a mixture of layman language and DynaLearn language. Thorough specification of the model's entities and "characteristics" (quantities). Depicting the hierarchical structure of the model: defining entities that include "sub-entities" that "inherit" the same "characteristics" of higher entities in the hierarchy Acknowledging the complexity of the system
4. Actually, we start now to model the system. A new DynaLearn project: Start new model at LS6. We will call it "Airport". Now we add to the master entity Airport, three entities – airplanes, workers and travellers. Then we will put under each of these entities, sub-entities like under workers: stewards, pilots, security people, mechanics, engineers. All inherit the same characteristics workers have. *(R) Now after adding all entities, I should move to the next stage. The problem is that I do not know what should be the next stage. Please wait for me I will be back soon (goes to seek help).*	Actually building the model using the software. I. seems to master knowledge and skills related to modelling as well as to the particular features of the software. R4 After completing the first stage of modelling -defining the static configuration of the model- the student seems to be lost and seeks help (the teacher).

5. I am back. What is really needed now is to add quantities to all entities. *(R) One moment, I wonder if adding one quantity to workers will be inherited to all sub quantities?*	R5 Questioning the logic behind the software. I.'s common sense and intuitive perception of the hierarchical model motivates this reflection: If there is a class with a given property (e.g., *number-of*), it seems only natural that members of this class inherit the property, and that an actual value for the inherited property is localized for each member. In practical terms: does the software support this?
6. I will add to the workers a "measure" called "number of". This measure describes the quantity of workers. The appropriate scale for this "measure" will be small, medium, large. This scale stays for – no workers, some workers, several workers, a specific point and large – from medium and above.	Thinking about a quantity space. First encounter with the problem of specifying an appropriate scale to describe the quantity space. Whether it should have discreet points on the scale or would be an interval scale.
7. Let's take another "measure" of workers – salary. This time I will specify a different scale: Zero, low, average, high. Zero – no salary; Low – a very low salary; Average – the average market salary and High – very high salary. *(R) A question can be asked – Is it realistic that there is no salary at all? I think it is realistic – for instance, if there are no workers, we need to "represent" it.*	Still thinking about an appropriate quantity space which is relevant for a specific quantity – What is the average salary? Specifying the relevant quantity space to each quantity. Not thinking in a routine mechanical way. R7 Being critical in checking the validity of his own decisions. I. looks at the model vis-à-vis the reality being modelled, but keeping in mind the logics of the representational structure - if there is a "no-salary" situation it should be appropriately represented.
8. Lets go now to another "measure" – skills. Here I think we need to use only two levels – Zero/Plus. Our works are either skilled or not. On second thoughts, maybe we should define our work as not skilled – Zero. A bit skilled – Low, and very skilled – High. I think this is better, so I will distinguish between Zero, Zero Plus, and Max.	Continue to debate regarding the best scale to represent a quantity space.
9. Now we will turn to the traveller. The measures needed are "number of travellers", the "free money they have for travelling" and their "demand for vacations". We can use the same measure "number of" we already specified. *(R) "Oops" I have just learnt that I made a mistake. I cannot add measures to each of my entities one by one. First I have to specify all needed measures and only then attribute them to the relevant entities.*	Continue to specify "measures" for another entity – travellers R9 Getting assistance and improving his modelling skills and working strategies with the software.

10. We will then add a new measure to the travellers "free money" – the amount of money travellers bring with them to be spent. We will use a scale: Zero; Small; Medium; Large; Maximum. *(R) Why do I specify 5 levels? It is not funny. This can lead to hundreds of options. But on the other hand, this measure is cardinal and has an effect on many things, that is why I chose 5 levels.*	Specifying another measure for travellers. R10 Exhibiting clear understanding of the systemic approach, and the consequences of constructing a complex model with very sensitive scales and complex relationships among components.
11. We will add now, another "measure" – the "demand for vacation" – we'll call it "Holiday Demand". We will give it a scale of Zero; Small; Medium' Large.	Completing the definition of the quantities for "travellers" initiated in stage 9.
12. Lets go now to the planes. We can use the same "number of" measure. We used previously to describe the number of airplanes. We can use it to describe the number of passengers that can enter the plane. At this moment I am not sure if it is important? I can just say that the plane is large or small and represent it as a Zero Plus scale. I can also represent it as a Low, Medium, High Scale and I will define between three sizes. So lets define a scale and call it S(small) M(medium) L(large). We will have three intervals. Maybe we don't need intervals, but we should fix points as there are only three kinds of planes – small, medium, or large. *(R) Now a suggestion for the programmers. Devise the opportunity to change intervals into points on a scale.*	Thought about the appropriate scale to represent the size of a plane: Two sizes, three sizes, discreet scale or an interval scale. R12 Thinking from the programmers' perspective – Another indication of I.'s understanding the logic of the software.
13. What is left is four more measures: The oil needed security check and staff. *(R) Maybe it is not wise to add these. It has no effect on the situation we want to check. What will happen if there will be a huge increase in the number of travellers? I think it is better to give it up, In any case, I will save what we have done up until now.*	R13 Signs of understanding the importance of reduction of the model and of keeping only relevant features that fit the problem under investigation.
14. The new fragment is Travellers. I will enter DynaLearn, Open model, Fragment, Editor, Statistics and we will create a new Fragment. We will call it Travellers. *(R) Here is a V sign on – Active. I do not know what it is, but I will leave it on.* We will now add a new entity and call it 'travellers". It already exists in the first file where we specified the hierarchy between the entities. We will add all measures created earlier: Number of; Free money for travelling; Holiday demand. *(R) As for now, I don't know whether I should specify all the quantities for this model. Let us ask Moshe.*	Evidence of I.'s mastering of functions and tools in the software. R14.1 Trying to learn on his own about functions of the software. R.14.2 Getting stuck – lets ask the expert.

15 *(R) Moshe's answer led me to deal with relationships with the fragment. We know that if there is more free money, there are more travellers. This relationship is an "I" type one. Wrong! This relationship is not a direct one – I. It is an indirect one – "P" type. The relationships have an effect not on the quantity, but rather on the direction of change. This means that if we will increase the amount of money in one unit on its scale, the number of travellers will grow, but not necessarily in one unit on its scale.* *All other relationships seem to be of this kind, e.g., between Holidays demand and number of travellers, etc. This is a P+ relationship.* 15 O.K., This is our first fragment. Lets save it on the Desktop as H. G.P. file calling it Airport.	R15 Clarifying the meaning of different types of relationships, i.e., I or P. This is done by considering actual examples while using previous knowledge. Evidence for generalization and understanding of generic types of relationships between model components.
16. We come back to our model. What is left for us to put into the models are the planes and the workers. The "measures" of planes are – "number of", "planes", "travellers", and workers. Workers are divided into sub-entities such as: "Stewards", "Pilots", "Security people", "Mechanics" and "Engineers". All of these sub-entities are included in the worker master-entity and are related to the master-entity in "Such as" links. I have to add the "measures" to all sub-entities. I will add three measures: "Number of", "Salary" and "Skills" should they be regarded as a "condition" or as a "consequence". I will add them as a condition to each worker. *(R) It will take me a minute.* Now will add relationship between the measures. The one I intend to add is a P link between "Skill" and "Salary". The more you are skilled, the more your salary increases. For example, your chance of earning more grows for each year of education you have. *(R) I finished adding all P relationships. Lets save the model.*	Deciding on "sub-entities", or ingredients grouped under a given category or "master-entity". Deciding on the relationships between a "master entity" and "sub-entities". Should quantities be regarded as a "condition" or as a "consequence" – evidence of being acquainted with, and able to elaborate upon, advanced modelling features. R16.1. Evidence of self-estimation of his own modelling capabilities. Thinking of causality. Identifying indirect and direct causal relationships, and their meaning vis-à-vis the modelled reality. R16.2. Evidence of having modelling capability.
17. What is left to model are the planes. The measures of this entity are: Quantity of planes, number of people needed to operate the plane, the size of the plane. These are related to the number of travellers the plane can carry. When we will finish this fragment, we will add a new one – Airport – which will include all entities we specified earlier. Going back to planes – the measure we need is the size of the plane. This affects all other measures. *(R) The scale that fits this measure is the one we already employed. S-small, M-medium and H-high.*	Once again, thinking about quantities for a new entity. After completing the top-down composition, I. reconstructs the hierarchy all the way to the top level conceptual umbrella of the whole model. R17 Evidence of advanced modelling skills, tailoring already existing ingredients into a new fragment.

18. Now we will treat another measure "Worker Requirement". *(R) Something is wrong. The problem is that worker requirement is substituting other already existing measures called "Skills". I need to delete both and create them from scratch.*	R18 The student traces mistakes in his modelling activity. He reviews his steps and plan repairing steps.
19. Now we go to add relationships. First we have to compare between the scales of measures - for instance between "Size" and "Worker Requirement"; whether there is small, medium and high correspondences in these. We have to identify the two scales and press C.	Understanding the meaning of correspondence between scales of quantities and knowing how to operate the software in such a case.
20. We can now start to build the real model with I type relationships, and check some hypotheses we have. Lets see what happens in the airport when there is a situation of overbooking. We will import all fragments we created until now. Airplanes, travellers, workers. Here the real task starts. How to tailor them all? The problem I am facing now is that I am short of space on the screen. I will try and arrange it. *(R) I am sorry there is not enough time to finish what I started with: building a scenario of "travellers-explosion". I will continue doing this at home*	Moving into composing the whole model to start examining and predicting the behaviour of the system in a certain scenario. Clear understanding of the main goal of the modelling process: to learn about a system and its behaviour under changing conditions - "here the real task starts" Facing a technical problem with too complex models. R20 Clear evidence of motivation to continue modelling and completing the task.

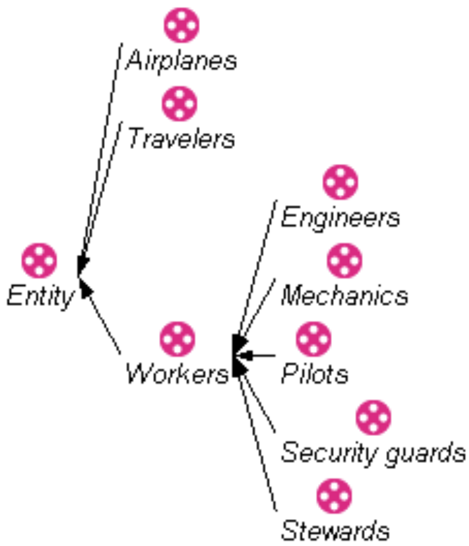


Figure 4.1: I.'s model of the Airport

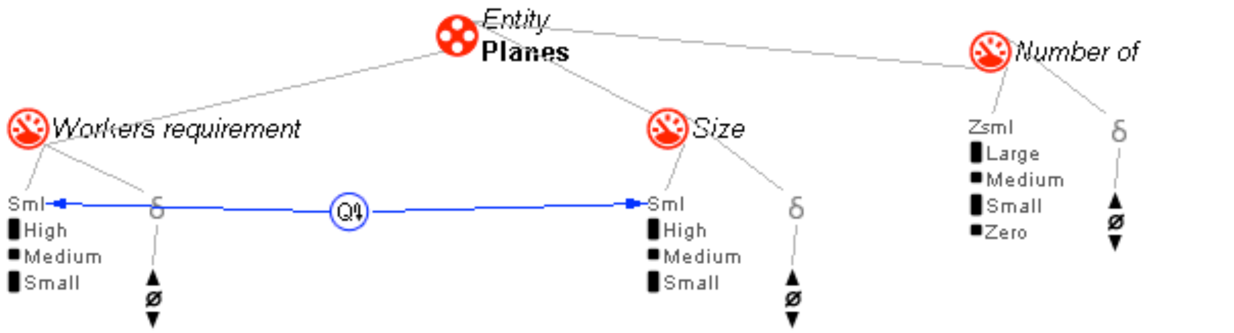


Figure 4.2: 'Planes' fragment in I.'s model (static)

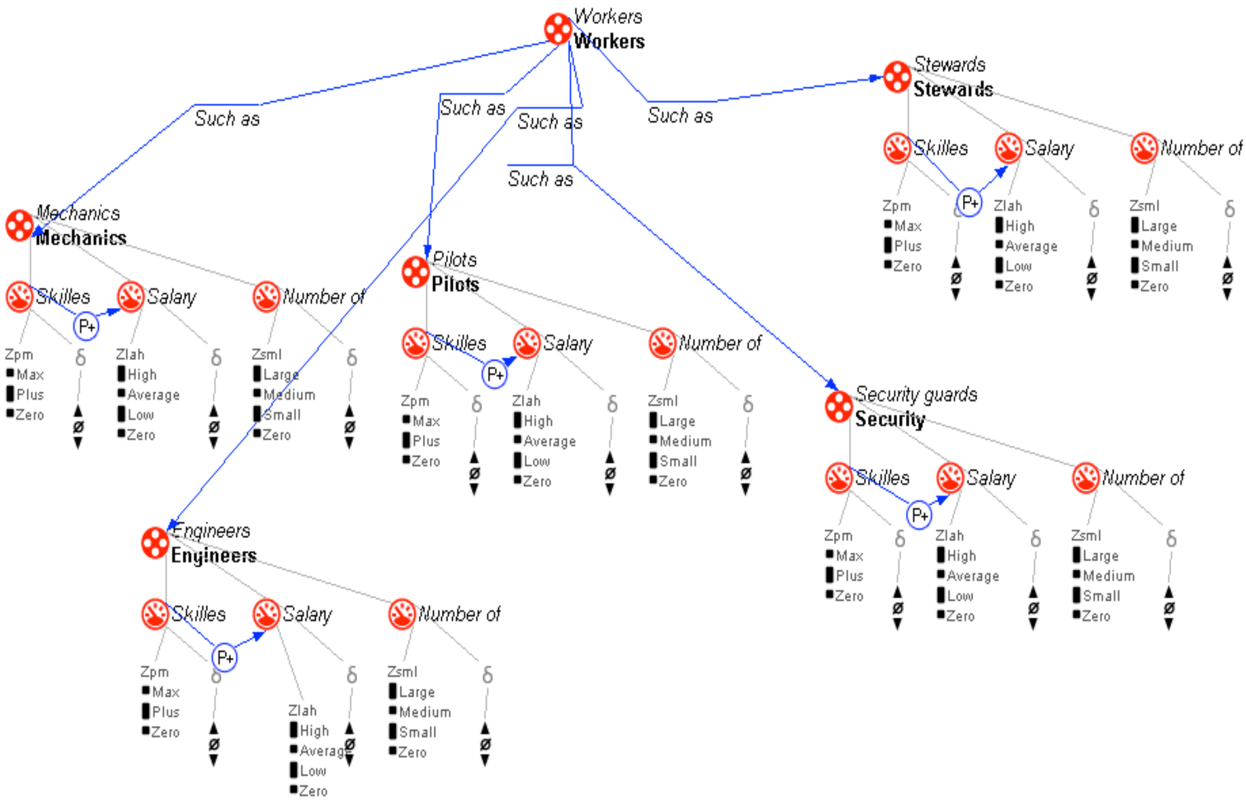


Figure 4.3: Workers fragment in I.'s model (static)

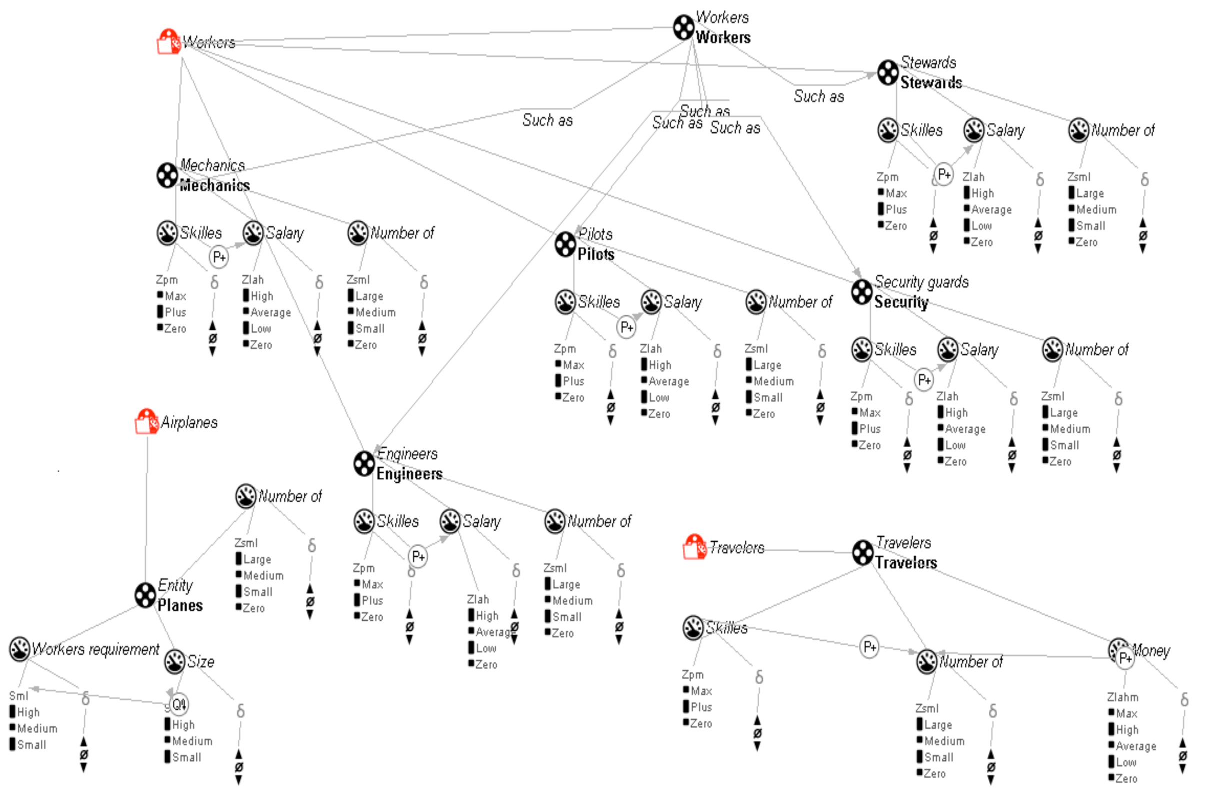


Figure 4.3: 'Overbooking' (process)

4.3.2. Analysis of the data

I.'s narrative comprises a total of 141 recorded statements. In general terms, these were classified as either descriptive or reflective statements. Within each category we distinguished between direct reference to actions made (e.g., a step in the modelling process) or reference to the meaning of actions (e.g., implications or significance of an action). The categorization appears in Table 4.2.

Table 4.2: Summary of student's recorded narrative

Statements categories	n (%)	Examples
Description of actions	79 (56)	"*The new fragment is Travellers. I will enter DynaLearn, Open model, Fragment, Editor, Statistics and we will create a new Fragmen*t"
Description of meaning of actions	24 (17)	*"The one I intend to add is a P link between "Skill" and "Salary". The more you are skilled, the more your salary increases. For example, your chance of earning more grows for each year of education you have*"
Reflection on actions	10 (7)	"*As for now, I don't know whether I should specify all the quantities for this model*"
Reflection on meaning of actions	28 (20)	"*Maybe it is not wise to add these. It has no effect on the situation we want to check*"
Total	141 (100)	

Above 70% of I.'s narrative are descriptive statements, and the remaining 30% reflective. This observation is of particular significance, indicating I.'s capability to intertwine between action and reflection during the model construction process. Equally significant is the observation that about 40% of the utterances -either descriptive or reflective- relate to the meanings of the actions taken. This denotes I.'s ability to think along the modelling process about "what to do" as well as "why to do that" - or thinking-in-action (Schon, 1983).

Another interesting observation relates to the weight of the different foci in the modelling process. As seen in Table 4.1., the first three stages -comprising about 12% of the statements- were thinking and planning stages, before any modelling with the software begun. The modelling spanned from stage 4 to 20. However, a rather large part of it was devoted by I. to the definition of quantities and quantity spaces -about 42% of the statements refer to these aspects- including revision of decisions made and modifications as needed. It seems that the appropriate definition of the quantities layer of the model was perceived by I. as crucial for the modelling of the system, as well as for the planned subsequent examination of the questions about the overcrowded-airport situation.

A detailed interpretation of I.'s thinking aloud protocol provides us with evidence on the modelling process he goes through. In the following we comment on I.'s narration referring to the specific stages (numbers in brackets) in the modelling process, as indicated in Table 4.1.

First I. defines the system he would like to model (Airport) and the problem this model might help to solve: "What will happen if suddenly a huge number of travellers will flood the airport, about one million a year? How will this affect the security and economic aspects of managing the airport"? He is able to phrase the problem in operational terms distinguishing between independent variables (e.g., number of travellers), and dependent variables (e.g., security and economic decisions). (1)

The next step is thinking about the static features of the model: concepts (entities) characteristics (quantities), configuration and hierarchy - entities and sub-entities that inherit the characteristics of the entities. These steps include a discussion regarding appropriate (relevant) characteristics -now called "measures"- and their scale, whether it should represent interval or discrete quantities, and how many levels should it include (2,3).

After completing the planning stage, I. starts building the model (stage 4). Once all entities are in place, I. feels that he does not know how to proceed to the next stage, moreover, does not know what the next stage should be. Then he turns for help (R4).

In the following stages -5, 6, 7, 8, 9, 10, 11, 12, 13- I. is busy in an intertwined process of thinking and doing while defining the "measures" (quantities) for the model's entities. Along the process, I.'s reflective thinking focuses not only on the technical aspects of the definition of quantity spaces. Rather, at some points (e.g., R10) it focuses on deeper understanding of the ways quantity scales and spaces might affect the behaviour of the various system's components.

As well, the reflective statements show I.'s ability to examine critically his own work and discard actions or decisions taken considered to be unnecessary (e.g., R13). Learning from his own failures, I. understands better the procedures he should follow. For example in R9 he concludes from his previous actions how to proceed to complete appropriately the definition of the quantities for the model's entities.

Another example of I.'s critical analyses of his own modelling actions is his reflective elaboration on the validity of his decisions concerning quantities - whether the scale he had defined corresponds to "reality", is relevant to the problem investigated, and is not too complex, what might result in too many mutual effects and difficulties for interpreting the system's behaviour. (R5, R7, R10, and R12).

All along the work, I. exhibits convincing mastery of the software functionalities and tools, and the skills required to build models and model fragments. However, when he faces problems and feels the need of additional knowledge he turns to seek aid from the teacher.

The teacher's helps I. mainly upon request, particularly in cases when the need for a transition in modelling phase is perceived (e.g., R4; R14.2). For example, at a given stage the Teacher directs I. to stop dealing with the "measures" of the entities and start thinking on the relationships between them (15; R15).

When moving to the new stage, I. faces the challenge to define different types of relationships, e.g., to deal with "I's and P's". It is clear that he understands the difference between these and applies this understanding in defining the set of relationships among the model's quantity spaces.

I. seems to move confidently towards higher modelling levels. He faces the construction of additional fragments, the definition of more complex relationships, the definition of conditions and consequences (16; 17). Finally, he envisions the hierarchical structure encompassing the different components and fragments of the model up to the systemic model of the airport (20).

At the end of the session he decides to move to what he considers "the real task": using the model to explore the system's behaviour, and in particular the question that triggered the whole model construction process - "Let's see what happens at the airport when there is a situation of overbooking".

Unfortunately, I.'s work comes to an end due to time limits. However, he shows motivation to continue the work at home.

Looking at the model-fragments and models in Figures 4.1, 4.2 and 4.3 offer clear evidence of I.'s capabilities as modeller, and the implementation of these capabilities along the modelling process. In the sample products in the figures, we observe the structural as well as conceptual complexity of the developed fragments. These partial products and revised versions also reflect I.'s learning process of new knowledge (e.g., features, functions, modelling strategies and tactics) required to proceed with the modelling task. As well they reflect I.'s learning of debugging procedures involving first and foremost the debugging of his own thought and decisions leading to the buggy representation.

First I. produced the depiction of the overall picture of the system - the airport (Figure 4.1.). Then the first fragment developed, which was quite complex (Figure 4.3) "workers". This complex chunk represents a rich arena for I. to face problems, understand underlying processes, make decisions, and debug his actions. It is in fact a rich thinking and learning opportunity already in at the beginning of the modelling task.

I.'s work progresses towards the completion of the complex total picture (Figure 4.3.). The whole model can be viewed actually as the external representation of I.'s inner (cognitive) understandings of the phenomenon modelled, its structure and hierarchical configuration, its ingredients, its corresponding quantity spaces, the complex set of influences and causal relationships among the ingredients, and the potential of its representation (the model) for exploring questions about the system's behaviour under different conditions.

4.4. Concluding remarks

The main goal of this case study was to follow closely the whole process of model-construction by a student with previous experience in modelling with DynaLearn - not a naïve modeller. In previous experiences in evaluation activities with DL we have concluded that recurrent modelling experience is crucial to ensure learning and performance at more advanced and complex levels. In the case of I. we had the opportunity to observe a student with such previous experience.

It should be noted that we did not look for the absolute accuracy or correctness of I.s' model - modelling at advanced levels requires ample expertise, and this is obviously not our goal with High-School students. Rather, the process is what is important here: the development of a systemic view, the gradual mastery of system thinking skills for analyzing and representing systemic phenomena, and the ability to create complex representations using the DL toolbox and language and exploring these to answer questions.

In this respect, an expert-modeller would probably raise many critical points about I.s' model. However from the learning perspective, we have collected important insights about I.s' thinking process, considerations, reflections and performance while building an advanced and fairly complex model. We can surely conclude that substantial learning and reasoning processes took place, and that real understanding of the systemic aspects of the phenomenon modelled and the process of modelling it is reflected in I.s' products.

Another important point relates to I.s' ability to mange and control his own learning as reflected in the process of model construction and the types of help requested and supplied (hen? by whom?) during the process. We have observed that most of the situations in which I. halted the modelling process to deal with a problem were, as reflected in the think aloud protocol, instances of self-regulated learning (Gibbons, 2002). In these cases the student poses questions, analyses the particular situation or stage, reflects and debates with himself, and devises his own solutions - for example:

1. Reflection stages related to the content/logics of the model

   (R) One moment, I wonder if adding one quantity to workers will be inherited to all sub quantities? [decision] I will add to the workers a "measure" called "number of". This measure describes…

   (R) A question can be asked – Is it realistic that there is no salary at all? I think it is realistic – for instance, if there are no workers, we need to "represent" it.

2. Reflection stage related to the software/modelling "language

   (R) "Oops" I have just learnt that I made a mistake. I cannot add measures to each of my entities one by one. First I have to specify all needed measures and only then attribute them to the relevant entities

Only in two situations I. requests help from the teacher - In both cases the problem faced are mainly of technical nature related to handling the software. For example:

(R) Now after adding all entities, I should move to the next stage. The problem is that I do not know what should be the next stage. Please wait for me I will be back soon (goes to seek help).

(R) As for now, I don't know whether I should specify all the quantities for this model. Let us ask [the teacher]

These observations suggest that future lines of work on help and support features (either by the software or as pedagogical model with the software) should focus on the logical aspects of the modelling process, in a modality that supports students' self-reflection about, and self-elaboration on, their own modelling process.

5. Evaluation activity 1: The contribution of grounding to learning

As mentioned in the introduction, an evaluation activity related to the main focus of phase II -i.e., semantic technologies and VC's- was conducted during the fall/winter academic course early in 2011. The detailed results appear in Mioduser et al., 2011. Here we present a brief summary of the results to complete the picture of phase II activities, and serve the overall discussion in section 7.

5.1. Summary of the acivity's plan

The context for the evaluation activity was the whole-semester undergraduate course on Marine Biology at Tel Aviv University. The lectures-based course was complemented every week with a modelling workshop using DL (14 weeks x 2 hours). One 2 hours class was dedicated to a task based on the use of the grounding feature. 14 students participated in the activity. Its main questions were:

Does students' involvement in grounding models contribute to:

- Their construction and revision of a model of a marine system?
- Their self-directed learning of concepts and causal relationships within the modelled system?

The activity conducted was similar to the "blind-model" activity described in section 2.2.3. Concepts and processes for the model were based on Chiavarini et al., 2003. Data collected were student models and explanations before and after the grounding task.

5.2. Summary of main results

Our main observations with the undergraduate students were:

- The students, performing independent work, reached success in the completion of all components of the task.
- Most candidate terms for grounding were appropriately handled by the students with their previous knowledge. However, there was a cluster of unknown terms which needed to be identified and defined. The grounding activity contributed to the understanding of these terms and concepts and to the successful incorporation of them into the models constructed.
- 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.
- Next on the extent of effect of grounding on students' performance relates to their ability to recognize the phenomena (to "tell the story of the system") and on the completeness of the explanations supplied by them.
- Additional observational data was obtained indicating the potential role of the grounding-based modelling activity to support self-directed learning - leading to students' success in completing a complex task in independent work (teacher's support was offered deliberately only about technical or procedural issues).

6. Evaluation activity 6: Expert-teachers pedagogical vision

This activity is being conducted as a cooperative endeavour of all WP7 partners. Its main goal is to compose a detailed picture of the pedagogical **rationale, goals and practices** for integrating DynaLearn in meaningful teaching/learning processes. The main sources of these data are expert-teachers - those who have had experience in teaching with DynaLearn. Essentially these are WP7 partners, other project partners and associates who actually had taught in the numerous activities conducted during the project.

6.1. Method and main instrument

Participants will be a group of expert-teachers with experience in using DynaLearn for teaching in the evaluation and additional activities conducted - about 10 teachers.

Data will be collected using a structured questionnaire during January 2012.

The questionnaire (see Appendix G) comprises 5 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' envision 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 DL.

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; the resources (besides DL) used during the teaching processes; pedagogical foci at each Learning Space; pedagogical added value of DL various features and learning tools; characteristics of various teaching modes implemented with DL.

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.

4. **Difficulties** and **problems** faced during the teaching processes, either pedagogical or technological.

5. Brief but illustrative accounts about **representative examples** and experiences registered during the implementation of the evaluation activities.

These data collected will be analyzed and processed during February 2012, and the report on the results will be integrated into a pedagogical chapter to be included in the final evaluation report, D7.4.

7. Discussion and concluding remarks

In this deliverable we report about phase II of evaluation activities conducted in Israel by TAU. The main goal of the second phase was to examine students work with the integrated DL software, with special emphasis on activities centred in features not evaluated in phase I, namely features related to the semantic technologies and Virtual Characters embedded in DL.

Activities were conducted with High School and undergraduate students, addressing questions related to the contribution of the three technologies integrated in DL to students' conceptual understanding, growth of system thinking skills, and mastery of modelling language and skills.

In this closing section of the report we will focus on the main insights we have gained from the activities, and their implications for the planning of future teaching with DL.

7.1. About conceptual understanding

Along the evaluation activities it became evident that the main learning gain takes place at the conceptual understanding level. Learning by modelling is a scarcely used approach in Science teaching. The teachers lack this kind of knowledge and training, and modelling activities are rarely encountered in the traditional curricula taught in schools.

With the advent of computer technology into the schools' landscape several decades ago, the lecture-like routine has been complemented with the use of simulation software of various degrees of sophistication (Honey & Hilton, 2011). At most, students are allowed to "run" a ready-made model and even manipulate its variables, but this is still far from the idea of affording the building of the model itself.

In recent years educational tools aiming to allow students to model have been developed and implemented in school settings (e.g., Jonassen & Strobel, 2006; Clariana & Strobel, 2007). The main rationale of these tools emphasizes their potential for supporting deep understanding of the structure and processes of the phenomena modelled, and the subsequent exploration of hypotheses and predictions concerning behaviour in changing conditions.

Indeed, a main insight obtained in our activities relates to the contribution of the modelling process with DL to students understanding of the structure and behaviour of the ecological/marine systems included in the course's curriculum. The need to "translate" data included in the scientific or descriptive texts into representations using DL language, and the actual manipulation of the systems' components and features as building blocks for composing these representations, supported students active (rather than receptive) understanding of the phenomena at hand. Undoubtedly, learning at the content level was an aspect highly benefited by the modelling activities.

7.2. About systems worldview and skills

Besides conceptual understanding, our main pedagogical goal was to support students’ development of a systems worldview and the acquisition of system thinking skills (Bar Yam, 1997; Assaraf & Orion, 2005). A system worldview implies a particular approach to look at phenomena in the world, to distil its essential structural, functional and behavioural features, and to be able to construct valid representations of these (Sterman, 1994). This worldview encompasses sets of skills -system thinking skills- allowing to operationalize the overall view into specific understandings and actions for generating the representations and exploring its behaviour under changing conditions, vis-à-vis the represented phenomenon.

In the activities conducted we had the opportunity to track closely the development of these capabilities by the students - and at the same time identify difficulties affecting their development. In most cases we have corroborated our previous observations indicating that recurrent experience and involvement in progressive modelling tasks is crucial for the construction of robust skills and capabilities. Our scoring and analysis schemes were applied once and again for analysing student products, from their very first concept maps to their advanced models (at least 2 by the High School students, several more by the undergraduate). These allowed us to identify changes (either improvement or stability) in the various aspects of the models, taken as external evidence of the acquired cognitive skills.

For example concerning the overall systemic approach, we had the opportunity to identify a progression path in High School student's thinking as reflected in the organizing principle used to depict the model of the system. Many students adopted a "storytelling" approach in their first concept maps: they represented the system as a linear chain of events much like the "story" told in the text (e.g., "a" happens, then "b", which results in "c"…). In the more advanced models the organizing principles were already conceptual or ecological oriented, leading to a systemic and web-like representation.

However, more fine grained analyses unveiled differential acquisition of skills, both within the set of skills and among students. For example, a detailed analysis of progressions of models by High School students during their work with the TA (most students generated 3 to 5 models) showed that: (a) **within** the set of skills, the ability to refer and represent the structural aspects of the system was achieved by all students and applied in all models; the lowest level of performance was related to the identification and representation of causal relationships among the system's ingredients; (b) **between** students, and for all sets of skills, consistent performance was observed either at the lower or higher levels, and only few showed progression and increasing mastery from model to model. It results evident that the robust acquisition of system thinking skills requires recurrent involvement in tasks over time, and that the understanding of the more profound aspects of the system's dynamics is even more cognitively demanding. In this, DL as model-construction playground allowing to "do-and-learn" and experiment with ideas has great potential for supporting the gradual development of the more complex skills, if only learning time is granted.

7.3. About the interaction with the Virtual Characters

The main insight resulting from the activities conducted was that the VC's added value is evident mainly at the motivational level. Overall, students enjoyed the interaction and valued positively the VCs features. In response to motivation questionnaires they even pointed out their (perceived) value in support for learning. The actual observations however indicated that interventions involving VCs had little effect on learning gain or student modelling performance (e.g., revision of previous models). This reinforces the observation that at least in the activities conducted (involving the TA and Quiz modalities) students valued the interaction modes themselves for being motivational and for affording a different mode of work.

7.4. About technical issues

Obviously there is inherent difficulty in conducting teaching/learning/evaluation activities while the tools being used are in continuous development. The activities are planned and conducted within considerable constraints (e.g., schedule of the academic year in schools and the University, availability of computer rooms and resources, availability of students due to the schools regular calendar of activities). In addition, vis-à-vis the development state of the software at the time of an activity. Another factor is that the features of the software are incorporated gradually according to the development plan, and in addition these are revised and improved along the versions. The result is that there are many factors affecting the implementation of the activities at any given time. For example, sometimes planned activities could not be conducted because it was still technically difficult . These difficulties were solved afterwards in subsequent versions.

At another level, the tool under development has several dimensions and features, and any attempt to cover all (or even many) of these in a sequence of evaluation activities is a challenging task. In our sequence of activities we chose to focus on a set of variables that we believe are the core reasons for introducing the LbQM approach into science education, namely to support the understanding of systems and the acquisition of system thinking skills, and to examine the achievement of these goals by using the different and continuously developing features of DynaLearn.

Many of our efforts are of the nature of exploratory studies, as demanded by the evolving character of the project and its accompanying evaluation activities. After the final completion of the integrated software, we believe that along the lines initiated in our and other partners' activities, a comprehensive study should be planned and conducted in larger scale implementation.

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.

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading, MA: Addison-Wesley.

Bredeweg, B., Liem, J., Beek, W., Salles, P., and Linnebank, F. (2010). Learning Spaces as Representational Scaffolds for Learning Conceptual Knowledge of System Behaviour. In Wolpers, M., Kirschner, P.A., Scheffel, M., Lindstaedt, S. and Dimitrova, V. (eds.), *5th European Conference on Technology Enhanced Learning* (EC-TEL 2010), LNCS 6383, 47-61.

Chiavarini, S., Massanisso, P., Nicolai, P., Nobili, C. and Morabito, R. (2003). Butyltins concentration levels and imposex occurrence in snails from the Sicilian coasts (Italy). *Chemophere, 50*, 311-319.

Clariana, R., and Strobel, J. (2007). Modeling technologies. In J. Spector, M. Merril, J. vanMerrienboer, and M. Driscoll (Eds.), Handbook of research on educational communications and technology. Mahwah, NJ: Erlbaum.

Conati, C., and Manske, M. (2009). Evaluating adaptive feedback in an educational computer game. In proceedings of the 9th International Conference on Intelligent Virtual Agents, pp. 146-158. Berlin: Springer-Verlag.

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

Honey, M., and Hilton, M. (Eds.) (2011). Learning Science through computer games and simulations. Washington, DC: Academic Press.

Jonassen, D., and Strobel, J. (2006). Modeling for meaningful learning. In D. Hung, and Khine, M., Engaged learning with emerging technologies. Berlin: Springer.

Lozano, E., Gracia, J., Gómez-Pérez, A., Liem, J., van Weelden, C. and Bredeweg, B. (2010). *Ontology-based feedback on model quality*. DynaLearn Deliverable D4.2.

Mioduser, D., Zuzovsky, R., Zurel, D., Leiba, M., Benayahu, Y., Nachmias, R., & Ram, J. (2011). Dynalearn Deliverable D7.2.4.

Novak, J., & Cañas, A. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report IHMC CmapTools, Florida Institute for Human and Machine Cognition. Available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf

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.

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

Sterman, J. (1994). Learning in and about complex systems. System Dynamics Review, 10(2-3), 291-330.

Wißner, M., Häring, M., Bühling, R., Beek, E., Linnebank, F., Liem, J., Bredeweg, B., André, E. (2010). *Basic Help and Teachable Agent*. DynaLearn Deliverable D5.3.

Yin, Y., Ruiz-Primo, M.A., Ayala, C.C., & Shavelson, R.J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use [Electronic version]. *Journal of Research in Science Teaching, 42,* 166-184.

Appendix A: Text and questionnaire on the topic "Soft Turtles"

Task

Read the story and describe it by drawing a concept map. Identify main entities and link them to each other. Write above each link title that describes the relationship number each link and explain the relationship.

The Sad Story about the Soft Turtle and its Happy Ending

Until the 1950s there were hundreds of soft turtles living in the rivers that led to the Mediterranean Sea. As the years went by, they disappeared mainly because of the polluted water and the destruction of their laid eggs. Ten years ago, scientists were surprised to find that, despite the pollution, the conditions of "Alexander River" were in favour of the turtle population.

The regional biologist working in the national authority of the natural reserves in Israel explains. The pollution had two different effects: first many of the natural predators of the turtles were affected by the pollution causing the turtle population to grow. Second, the pollution was mostly caused from remnants leftover meat from the poultry abattoir in the neighbourhood, on which young turtles feed, thus the population of the turtles increased both in number and size.

Every year in May, the female turtle leaves the water, digs a nest in the sand and lays her eggs. These eggs are considered a delicacy for predators. They are also destroyed by people driving their cars on the banks of the river. As a result of these hazards, only a few eggs are hatched and only a few young turtles survive and return to the water.

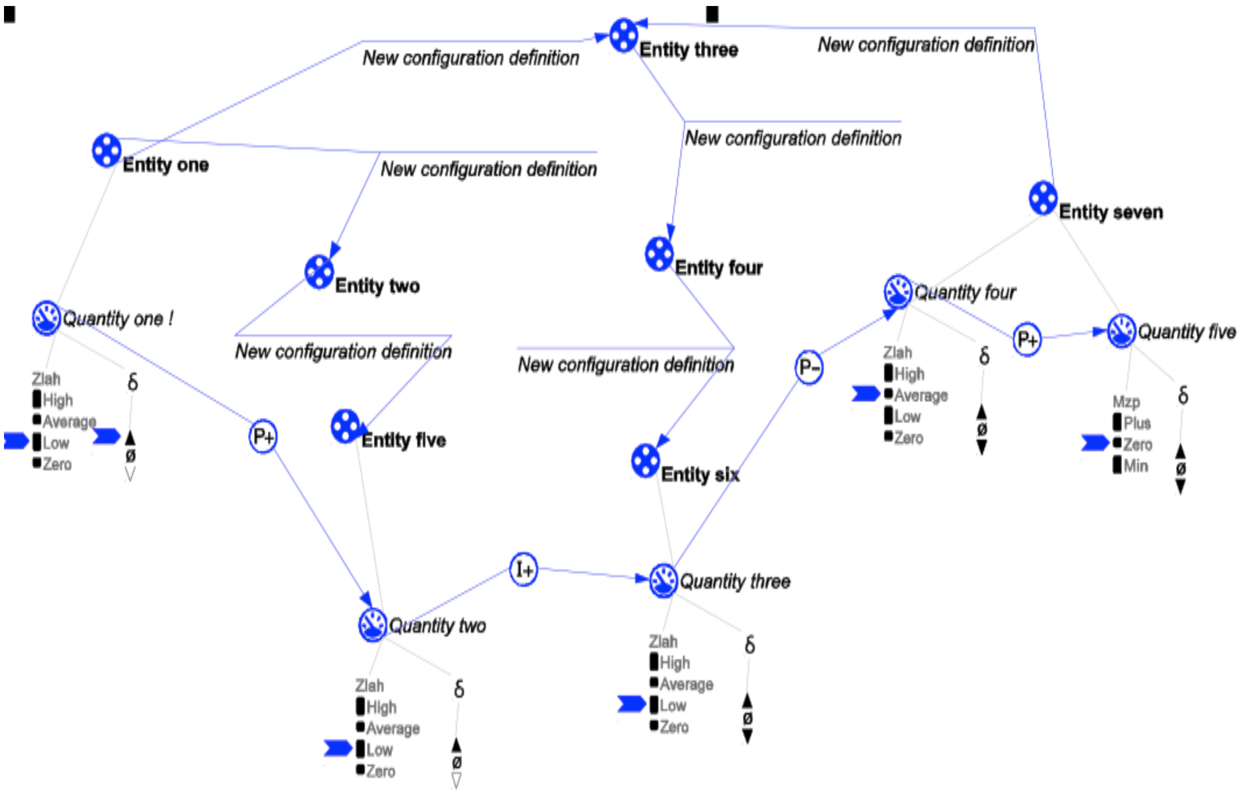
This was the way things occurred until the winter of 1992 when heavy rain caused flooding that resulted in the death of many turtles. Most of the turtle population that year, became extinct.

In May, the same year, scientists surveyed the river again and found 47 health turtles that had survived the flood – amongst them, females. The scientists hope that this small population will regenerate the turtle population in the coming years.

Answer the following questions

1. Explain the term "population".
2. Write in a table which factors caused an increase in the turtle population and which factors were responsible for its decrease.
3. One factor is mentioned in the story as causing both an increase and a decrease in the turtle population – what is the factor?
4. Explain.
5. The turtle story is an example of a disequilibrium state. Explain this term.
6. Use the following terms to explain what happened to the turtle population: Extinction; Natural growth; Survival rate; Predation.
7. Explain the following diagram that describes changes in the population of turtles in Israel from 1900 - 1990.

Appendix B: The "Blind Model" task - empty model and resources



Terms to be integrated in the "blind model", then grounded:

Ships

Antifouling Paint

Organotins

Gastropod

TBT

Concentration

Sediment

Mediterranean Sea

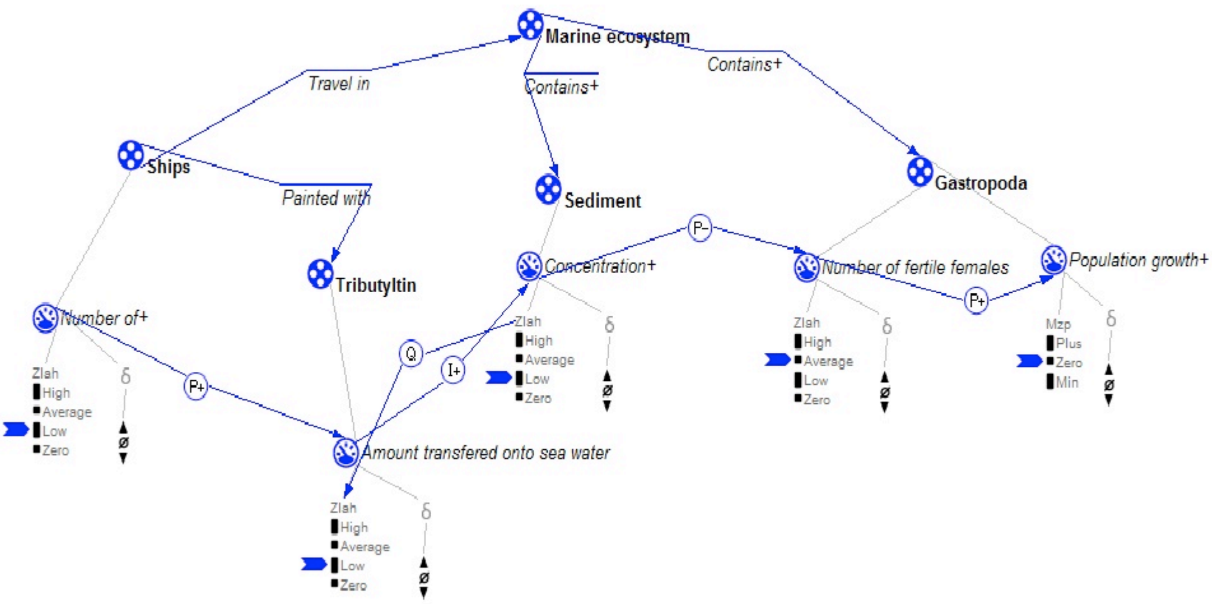
Number of

Population growth rate

Female fertility

Amount transferred to seawater

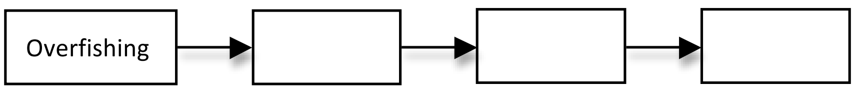
Appendix C: The "Blind Model" task - completed model



Appendix D: The "Jelly Fish" questionnaire

The Jelly Fish Invasion

1. What is the reason for the high number of Jelly Fish in the Mediterranean Sea? Which of the following is correct?
   a. Lack of competition?
   b. The higher temperature of water in the Mediterranean Sea?
   c. The lower temperature of the water in the Mediterranean Sea?
   d. Drivers destroying the nests of sea turtles on the Mediterranean shore?
2. Define the relationship between the local Jelly Fish and the invasive one. Which of the following is correct?
   a. Mutuality
   b. Competition
   c. Parasitism
   d. Pre-predator relationships
3. What are the factors that cause the fish population in the Mediterranean Sea to decrease?
4. What are the factors that cause the invasive Jelly Fish population in the Mediterranean Sea to increase?
5. What can man do in order to reduce the number of the invasive Jelly Fish in the Mediterranean Sea?
6. Specify the economic damage caused by the invasive Jelly Fish to Israel.
7. What will happen to the invasive Jelly Fish if the temperature of the Mediterranean Sea will increase by 1°C.
8. Explain what is an invasive species and how invasion of such species affects a local ecological system.
9. Complete the following boxes (you may add more if needed)



10. Explain what is "global warming"
11. Explain what is ecological equilibrium.

Appendix E: The modelling process questionnaire

Documenting the Process of Building a Model

1. While building a model, please answer the following questions (Please answer the same questions whenever you start a new model)
2. What is the phenomenon you wish to represent in the model?
3. Which entities have you chosen for this purpose? Explain.
4. What quantities are needed to be represented each entity in the model? Explain.
5. Which relationships are needed to be represented in the model? Explain.
6. State one hypothesis that predicts a change in the system behaviour if the quantity of one of the entities in your model changes. (Specify the direction of change).
7. What insight have you gained on the system behaviour after building your model and running a simulation?

Appendix F: Working with the VC questionnaire

To what extent you agree with the following statements (1-strongly disagree; 2-disagree; 3-neither agree nor disagree; 4-agree; 5-strongly agree)

1. Using the software contributed to my better understanding.
2. I would like to learn by building models in other lessons.
3. Most of my learning occurs while watching the teacher build a model.
4. Most of my learning occurred while I was building my own model.
5. Most of my learning occurred while I was running a simulation.
6. Most of my learning occurred while I compared my own model with my friend/s model.
7. Most of my learning occurred while I compared my own model with an expert model.
8. The grounding activity contributed to my knowledge.
9. The grounding activity contributed to my understanding.
10. I was helped a lot by the VC.
11. I loved the VC. They are fun and amusing.
12. I would like to learn with the VC in other activities as well.

To what extent do you agree (1-strongly disagree; 2-disagree; 3-neither agree nor disagree; 4-agree; 5-strongly agree).

1. The TA talk was understandable.
2. The TA behaviour was appropriate.
3. The TA expressions and gestures were appropriate.
4. I wanted the TA to succeed in the quiz.
5. I was glad when the TA succeeded in the quiz.
6. The activity with the TA helped me to build a model.

Appendix G: Expert-teachers questionnaire

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

		Structural configuration							Quantities							Causal Relationships								Av. total in % Structural config.	Av. total in % Quantities	Av. total in % Causal R/ship
ID	#	1	2	3	4	5	6	Total Av.	1	2-3	4	5-6	7	8	T otal Av.	1	2	3	4	5	6	7	Total Av.			
1	1	+	+	+	+	+	+	6	–	–+	–	–+	–	–	2	0	0	0	0	0	0	0	0			
	2	+	+	+	+	+	+	6	+	–+	–	–+	–	–	2	+	–	0	+	+	+	+	4			
	3	+	+	+	+	+	+	6	+	–+	+	–+	–	–	4	+	0	0	0	+	–	–	2			
	4	+	+	+	+	+	+	6	+	–+	+	++	+	–	6	+	0	0	0	0	0	+	2			
	5	+	+	+	+	+	+	6	+	–+	+	–+	–	+	5	+	–	–	–	+	–	–	2	100%	48%	29%
2	1	+	+	+	+	+	+	6	+	+–	+	++	–	+	6	+	–	–	–	–	–	–	1			
	2	+	+	+	+	+	+	6	+	+–	+	++	+	–	6	+	–	–	–	–	–	+	2			
	3	+	+	+	+	+	+	6	+	+–	+	++	+	–	6	+	–	–	–	–	–	+	2			
	4	+	+	+	+	+	+	6	–	+–	+	++	+	–	5	–	–	–	–	–	–	+	1			
	5	+	+	+	+	+	+	6	+	+–	+	++	+	–	6	–	–	–	–	–	–	+	2	100%	72%	23%
3	1	+	+	+	+	+	+	6	+	+–	+	++	+	0	6	+	0	0	+	+	0	+	4			
	2	+	+	+	+	+	+	6	+	+–	+	++	+	0	6	+	0	0	+	+	0	+	4			
	3	+	+	+	+	+	+	6	+	++	+	++	+	+	8	+	+	+	+	+	+	+	7	100%	83%	71%
4	1	+	+	+	+	+	+	6	+	++	+	++	+	+	8	+	+	+	+	+	+	0	7			
	2	+	+	+	+	+	+	6	+	++	+	++	+	+	8	+	+	+	+	+	+	7	7			
	3	+	+	+	+	+	+	6	+	++	+	++	+	+	8	+	+	+	+	+	+	7	7	100%	100%	100%
5	1	+	+	+	+	+	+	6	+	+–	+	+–	+	–	5	+	–	–	+	+	–	+	3			
	2	+	+	+	+	+	+	6	+	+–	+	++	+	–	6	+	–	–	–	+	0	+	3			
	3	+	+	+	+	+	+	6	+	++	+	++	+	+	8	+	+	+	+	+	+	+	7	100%	79%	62%
6	1	+	+	+	+	+	+	6	+	++	–	++	+	+	7	+	0	0	+	+	0	0	3	100%	88%	43%
7	1	+	+	+	+	+	+	6	+	++	+	+–	+	–	6	–	–	–	–	+	–	–	1	100%	75%	14%

Appendix H: Detailed account of results summarized in Table 3.6