Supporting Exploratory Learning by Offering Structured Overviews of Hypotheses

Melanie Njoo[1] and Ton de Jong[2]

[1]Department of Philosophy and Social Sciences, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
[2]Department of Education, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

Abstract: *Exploratory learning* with *computer simulations* is an approach that fits well within the current emphasis on viewing the learner as an active, constructive person. In previous studies we concluded that a valid performance of exploratory learning processes was a bottleneck and especially the process of *hypothesis generation* posed difficulties to learners. The major objective of the present study was to evaluate the effect of supporting hypothesis generation by offering structured overviews of predefined hypotheses. Subjects were 88 Mechanical Engineering students working in pairs, with a computer simulation program for control theory. Two experimental groups and one control group received an open-ended assignment for exploring a given modelled system. The major means of support that the experimental groups received was a *structured overview of hypotheses*. These overviews offered a list of, basically, the same set of eight predefined hypotheses from which subjects could choose. Two variations were designed: the *controller* structure followed types of controllers of increasing complexity and the *concept* structure organised the hypotheses according to fundamental domain concepts. The control group received the same assignment, but no support measures. Prior knowledge of all subjects was measured and at the end of the lab they were given a posttest that intended to measure 'deep' knowledge. Subjects worked on so-called 'fill-in forms' and their notes were used for analyzing their learning processes. Results showed that the Controller group scored higher on the posttest than the Concept group and subjects' level of prior knowledge influenced the posttest scores. Analysis of statements on the fill-in forms showed that among others the Controller group designed better (more complete) experiments than the Concept group.

Keywords: Computer simulations, Exploratory Learning, Instructional support

Introduction

Ever since cognitive psychology influences learning theory and research, learners are no longer seen as rather passive respondents to various environmental stimuli as was the case in the behaviourial tradition. On the contrary, cognitive approaches to learning stress that learning is active, constructive, generative, goal-oriented and dependent upon the mental activities of the learner. Therefore, the currently prevailing approach in learning and instruction, which is called *constructivism*, sees learners as actively constructing knowledge and strategies [3, 17, 22].

Computer simulations can offer a learning environment that is appropriate for active learning. A computer simulation is a program that incorporates a model of a process, phenomenon, system etc. The learner is able to control input values of the simulation model and examine the resulting changes in output. Computer simulations therefore may invite the learner to activity, both in manipulating the domain, i.e. the variables and parameters of the model, and in constructing a mental model of the system model or operational procedures offered in the simulation [4].

In the present study we examined the use of a computer simulation program for learning within the domain of control theory. Previous studies [12] in this domain showed that students have difficulties with generating hypotheses and showed some positive influence from access to predefined hypotheses. Students provided with hypotheses were more active and made less domain related mistakes in the formulation of their learning processes. Offering predefined hypotheses, however, is a rather directive measure [4] that easily interferes with the constructive nature of the exploratory process. In the present study we, therefore, increased the level of freedom for students and offered them an *overview of predefined hypotheses*. The next section provides background information for the most important aspects of our study: the role and support of hypothesis formation, sequencing hypotheses, and the role of prior knowledge.

Theoretical background

Learning with computer simulations is characterized as exploratory learning, which consists of active, constructive and goal-oriented processes. The learner has to discover general rules, procedures, or higher order skills. The general idea is that this active attitude of the learner encourages meaningful incorporation of information into the learner's cognitive structure. Njoo and De Jong [12] have specified in detail the learning processes in exploratory learning and have distinguished two main classes. The most important class is the one of *transformative* processes. These processes can be characterized as processes of scientific inquiry and consist of four main categories: analysis, hypothesis generation, testing and evaluation. These categories of processes were further subdivided into more detailed processes. For example, the category testing includes among others: designing an experiment, making predictions, and interpreting data (for similar processes [10, 14, 15, 18]).

In Njoo and De Jong [12] we described two studies in which we observed students working with a simulation program in the field of control theory. In the first study, we found that students made little use of some exploratory learning processes such as hypothesis generation. In the second study we offered learners fill-in forms which supported their exploratory learning processes. We analyzed the statements that students made on these fill-in forms. Results show that the *bottleneck* in exploratory learning is the *valid performance of exploratory learning processes*. Students were quite active, and once they made a statement that was valid in terms of a specific learning process they did not seem to have trouble with the domain itself or with relating different statements to each other. More specifically, we again saw that especially hypothesis generation caused difficulties. *In the present study we therefore have chosen to evaluate a way of supporting learners in this process of hypothesis formation.*

Several studies have tried to help students in creating hypotheses. 'Smithtown' [18, 19] offers the learner support for hypothesis generation by a hypothesis menu. This menu consists of four windows which present parts of a hypothesis e.g., variables, verbs to indicate change, and connectors. A similar means of support is a *hypothesis scratchpad* [6, 8]. Here, learners are offered different windows for selecting variables, relations and conditions. Hypotheses that are created can be placed on a 'hypotheses list' and be marked as 'under study', 'false', 'true' and 'unknown'.

These studies offer learners *elements* of hypotheses that they have to assemble themselves. A more directive support for creating hypotheses can be found in CIRCSIM-TUTOR [9], an ITS in the domain of medicine which treats problems associated with blood pressure. In CIRCSIM learners are posed with a perturbation of the cardio-vascular system. Subsequently they are asked to predict qualitatively what will happen to seven components of the cardio-vascular system. To be able to write down this prediction learners are offered a spreadsheet. In CIRCSIM-TUTOR the program takes the lead in choosing both input and output variables for the learner, and in determining the change in the input variable. From this it is only a small step to offering learners *complete hypotheses* to explore. And indeed, in 'Pathophysiology Tutor' (PPT) [11] learners can indicate a predefined hypothesis by using a list of nested menus that each give a more specific fixed list of hypotheses in the field of physiopathology.

The instructional support measures mentioned in this section differ in the level of being '*directive*'. The least directive measure is just prompting the learner to state a hypothesis [6, 12]. More direction is given when learners are offered hypothesis scratchpads [18, 19] and this has even more direction when the variables on the scratchpad are pre-selected and manipulated by the researcher [8]. Choosing from lists of predefined hypotheses is still more directive [11], finally followed by giving learners a specific hypothesis [12]. *In the present study we have chosen to support hypothesis formation by learners by giving them overviews of predefined hypotheses to select from.* The main reasons for making this choice, for a relatively directive measure, is that the domain we work in and the idea of exploratory learning are both difficult and novel for our students.

When hypotheses are offered to learners, a choice for a certain sequence should be made. This sequence is determined by the movement of the learner through the space of possible hypotheses. The sequence of hypotheses is frequently based on the *relations* or the *variables* in the hypotheses. For example, hypotheses could be offered in a qualitative to quantitative sequence [20, 21, 13]. This implies that relations are used to guide the sequencing. Another way to sequence hypotheses is to follow variables from the domain [6, 7, 8] e.g., ordered in a hierarchy, with global variables at the top and specific (instantiated) variables at the bottom.

Apart from this we may, within one domain, follow specific sequences by following different types of variables. An example of this can be found in Eylon and Reif [1], who presented a domain from physics, in two different structures. One structure followed a historical line of thought, the other structure followed essential concepts from the domain. In the domain that is involved in the present study, which is control theory, we may distinguish two types of structures. Basically, the two structures represent different methods for the solution of control problems. One structure concerns the different *types of controllers* that can be used for getting a system to perform in a specific way, the other concerns the *behaviour of the system*. Whereas students tend to start of from a specific controller and see if they can reach specific requirements for the behaviour of the system, teachers prefer that learners start from fundamental concepts (defined in terms of requirements for the behaviour of the system) and try to think of an adequate controller. *In the present study we have* *offered subjects basically the same set of hypotheses organized according to one of the two above mentioned structures.*

Offering hypotheses to learners only makes sense if they have some feeling for the content of these hypotheses. This implies that learners' prior knowledge partly will determine how they may profit from the hypotheses offered. Moreover, it is generally assumed that in order to profit from an exploratory learning environment prerequisite knowledge of the domain (and analogous domains) is necessary [10].

Schauble, Glaser, Raghavan, and Reiner [16] assessed the influence of prior knowledge on experimentation strategies and knowledge gains of students working with a computer simulation on basic electricity 'Voltaville'. They found that the quality of prior domain knowledge influenced the knowledge acquired. The more sophisticated models of the better learners also appeared to be related to a better performance of exploratory processes. Schauble et al. [16] write: "The unsophisticated models appear to encourage students to look for information in the wrong places, to develop misunderstandings about the functioning of components, and to provide misleading clues about how to interpret unexpected results" (p. 229). Also, students that started off with sophisticated models tried to explore the simulation or parts of it, as thoroughly as possible, whereas students with unsophisticated models worked unsystematically, made no generalisations of new findings, and were even unable to remember the discoveries they had been able to make. Similar findings were reported in a study that examined the relation between characteristics of domains and successful experimentation behaviour [2]. *In the present study the prior domain knowledge of our subjects was assessed and we examined its influence on exploratory processes and learning outcome.*

In the present section we subsequently introduced a number of aspects of our study. In sum, the goal of the study is to evaluate a type of off-line instructional support for exploratory learning with a simulation program for the domain of control theory. The main type of off-line support offered consisted of a structured overview of predefined hypotheses. For one experimental group these hypotheses were ordered according to the type of controllers, for the second experimental group hypotheses followed fundamental concepts from the domain. Additionally, all subjects in experimental groups received so-called 'fill-in forms' that presented learners separate cells for specific learning processes that could be filled with notes. Subjects also received information about the processes in each cell. A control group was provided with the same assignment and simulation program as the experimental groups, but no fill-in forms, and no general information on the learning processes and structured overviews of hypotheses were given to them. Students' prior knowledge was assessed and their learning outcome was measured through a posttest intended to appraise 'deep' knowledge of the domain. We made the following research predictions with regard to the scores on a posttest:

- Students who receive the off-line support will perform better on a posttest than students who have to work with the simulation without the off-line support.
- Students level of prior knowledge will influence posttest scores i.e., if the level of prior knowledge (as will be measured with a test for prior knowledge) increases then the posttest scores also increase.
- Students who receive hypotheses in a structure that follows the general concepts of the domain will score higher on a posttest than students who receive the hypotheses struc-

tured according to the controllers. The idea is that these general concepts address more fundamental aspects of the domain as compared to the types of controllers.

- The effect that students who receive the 'concept' structure of hypotheses perform better than students receiving the 'controller' structure will be more profound for the students with high prior knowledge as compared to the ones with low prior knowledge. The idea is that the structure according to fundamental concepts will be more in line with the knowledge of students with high prior knowledge.

For the two experimental groups we could assess the learning processes that they employed by analyzing the fill-in forms that the students used. These learning processes are seen as the intermediate between experimental condition and learning outcome. We expected that students with high prior knowledge who will use the structured overview following fundamental concepts will demonstrate, compared to the other groups of students, exploratory learning processes of a higher quality.

METHOD

Domain and simulation program

The domain involved in the present study is control theory, a subdomain of mechanical engineering. Within this domain the primary topic is to regulate systems so that they will behave in a required manner. Key concepts in the domain are: Laplace transform fundamentals, frequency domain analysis, and time domain characteristics. The domain is rather complex, has a number of specific dynamic key concepts, and provides clear experimentation possibilities. In this respect it presents a domain that is specifically appropriate for exploratory learning with a simulation.

The simulation program that is used is PCMatlab (© Mathworks). The program is originally intended for scientific and engineering numeric calculations and graphics. Input must be given by differential equations, which represent specifications of a mechanical system and the control law. Once the mechanical system is modelled the students can control the system by different types of controllers. The output consists of numeric data or two dimensional graphs.

Subjects

Participants were 88 students of the Mechanical Engineering Department taking part in a computer lab on control theory. Students in lab groups normally work in pairs and 43 pairs (two "pairs" consisted of three subjects) took part in our study. Three lab groups were assigned to two experimental groups (*n* is 15 and 14 pairs per group) and a control group (*n* is 14 pairs) on a random basis.

Support

All groups received an open-ended assignment to explore a given modelled system. Additionally, the experimental groups received support that consisted of fill-in forms, which was similar for both groups, and one of two variations of an overview of hypotheses.

Fill-in forms
The experimental groups received *fill-in forms* with additional information about the exploratory learning processes on an *information sheet* [12]. Subjects should work out the assignment with the simulation using the fill-in forms to note down their thoughts, actions, or results of the simulation for each of the exploratory learning processes that were explained on the information sheet.

The structure of the fill-in forms and information sheet was identical: they were both divided into six cells labelled VARIABLES & PARAMETERS, HYPOTHESIS, EXPERIMENT, PREDICTION OF THE EXPERIMENT, DATA INTERPRETATION, and CONCLUSION. The size was 16.5 x 11.8 inch (42 x 30 cm).

The cells of the information sheet contained information on the six exploratory learning processes. We offered general information about the six learning processes additionally with domain specific information or examples. For instance:

... Designing an experiment involves the following aspects:
- *Choice of input.*
  You can choose the variables (t, r(t)) or parameters (K_r, τ_d, τ_i) to vary.
- *Choice of output ...*

The fill-in forms had only blank cells with the six labels for processes. In the cell HYPOTHESIS they could write down one of the hypotheses that they had to choose from a structured overview of hypotheses. There was no compulsory order of cells on a form and subjects were urged to follow the order they preferred.

Structured overviews of hypotheses
We designed two variations of the *structured overviews of hypotheses*:

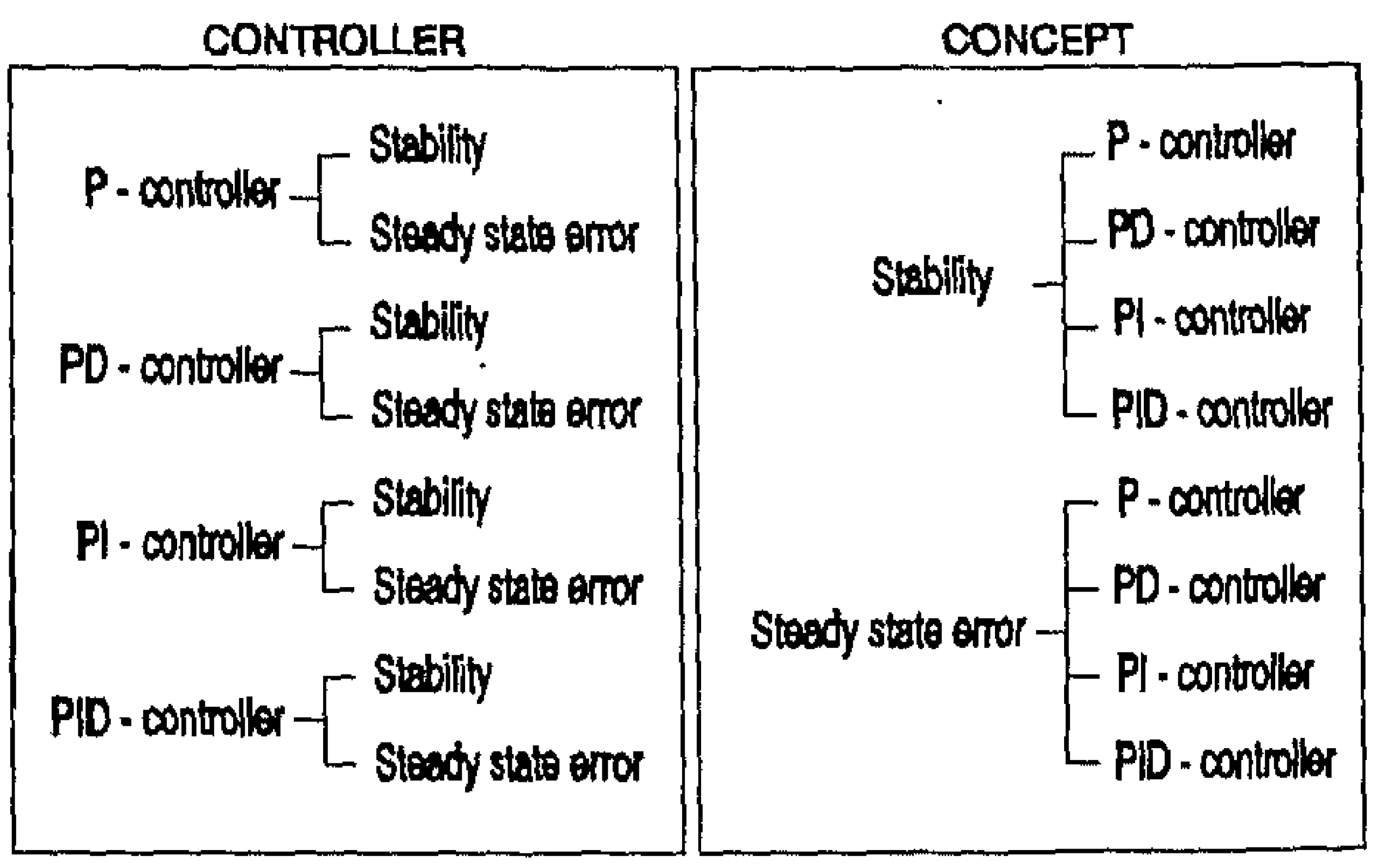

Figure 1: Structured overviews for the hypotheses

- The *controller* structure ordered the hypotheses by increasing complexity of the controllers, at the first level. The controllers are determined by three actions or a combination thereof. The actions are: a proportional (P), differential (D), or integral (I) action. These actions make up for four appropriate controllers: P-controller, PD-controller, PI-controller, and PID-controller. The controllers are presented in increasing complexity, with the exception of the PD- and PI-controller that are of the same complexity. At the second level of the structure the theoretical concepts were introduced.
- The *concept* structure ordered the hypotheses by theoretical concepts, at the first level. The two concepts are stability and steady state error. These concepts are two important characteristics of system's behaviour. The two concepts were presented in a hierarchical order. At the second level of the structure the four controllers were used. The structures of the overviews are presented in Figure 1.

Each of the two structures consisted of eight hypotheses, each expressing the relationship between the parameters of a controller and the behaviour of the system. Essentially the same eight hypotheses were used. The two structures only differed in the order of hypotheses and in the formulation of them, reflecting different points of view (controller v. behaviour of the system that had to be controlled). An example of a hypothesis formulated differently for both overviews is:

CONTROLLER: By adding a proportional controller the steady state error can be eliminated.[1]

CONCEPT: The steady state error can be eliminated by the implementation of a proportional controller.

Hypotheses were stated in an affirmative or a negative sense and five of the hypotheses presented were true and three hypotheses were false. We knew that subjects could fill in an average of five forms in one lab session [12], so they could not explore all eight hypotheses offered. Since we wanted subjects to follow the structure of the overviews, we gave them two rules for choosing hypotheses from the overviews. First, subjects could choose hypotheses form the overview but they had to hold on to the sequence of the overview. Thus, subjects could skip a hypothesis but they had to move through the overview from top to bottom and could not move upwards again. Secondly, they had to explore a minimum number of hypotheses per section. For the controller overview they had to explore one hypothesis per controller. For the concept structure they had to choose two hypotheses per concept.

Tests

Subjects' prior knowledge was assessed with a multiple choice test (18 items with four alternatives). Items deals with topics such as damping, feedback, and rise time.

At the end of the lab subjects completed a posttest, which consisted of 18 multiple-choice questions with four alternatives, that tested qualitative insight in the domain. Questions dealt with e.g., analysis of problems situations, relations between concepts and controllers, and specifying values for parameters of the controller.

---

[1] For the system in the assignment this hypothesis is false because the system is not stable. This implies that discussing the problem of eliminating steady state error is redundant.

Experimental procedure

The study was performed in the context of a second year course at our University of Technology. The course consisted of lectures and a computer lab. During the lab students worked in fixed pairs with self-selected partners. Usually about 10 to 20 pairs worked in a classroom at the same time. Two tutors per lab group were present to answer any domain- and simulation-related questions.

The lab took up 14 hours divided into four sessions (Session 1 through 4). Session 1 and 2 were the same for the experimental groups and the control group. The first two sessions were primarily intended as an introduction to PCMatlab. PCMatlab operations were explained and practised with simple assignments and gradually more emphasis on control theory was given. In Session 3 subjects from the experimental groups could practice with our support measures i.e., the fill-in forms with the additional information sheet and predefined hypotheses (not yet structured in an overview) while they explored systems of a moderate level of difficulty. The control group received exercises with study questions, dealing with the same problem situations as the experimental groups, but did not receive support or hypotheses. In Session 4 data for the study were gathered. All subjects received specifications of a modelled system[2] and an open-ended assignment to explore the modelled system. All subjects had the same amount of time to work on the assignment. The final goal was to construct an optimal controller for the system. The subjects in the experimental groups received the support measures that consisted of the fill-in forms, which they had practised to work with in the third lab session, and one of the two variations of the overviews of hypotheses. They were familiar with exploring with the help of hypotheses but the structured overview was new to them. They had to give their final choice for the type of controller with their motivation on a separate sheet. The control group received the same assignment but no additional support. In Session 3 and 4 the subjects were informed that the assignment(s) required free exploration and experimenting.

Table 1 summarises the experimental set-up as introduced in the preceding sections.

Table 1
Summary of the experimental set-up

Groups	Test for prior knowledge	Assignment	Fill-in forms & information sheet	Controller overview of hypotheses	Concept overview of hypotheses	Posttest
Controller	x	x	x	x		x
Concept	x	x	x		x	x
Control	x	x				x

[2] The system was an object with a certain mass that should be moved from an initial position to an end position. The influence of gravity is present but friction is not considered. Available was an imperfect actuator that could apply a force to the object. The actuator should be modeled as a first order system with a time constant and a static amplification.

Levels of analysis

For assessing the effects of the overview conditions on the exploratory learning processes we analyzed the statements that the subjects noted down on the fill-in forms with the assistance of a domain expert. The analysis was performed in a step wise order by introducing five levels of analysis [12]. The first level is the *global activity* level which is an assessment of the general activity level of the subjects. The second level is the *learning process validity* level which is an assessment of aspects of the statements given by the subjects in each cell. The aspects were related to general description of the cell as was given on the information sheet. At the third level we determined the *domain correctness* of the aspects of the statements that had proven to be learning process valid at the previous level. The fourth level was labelled the *consistency* level and was an assessment of the relations between contents of different cells at one fill-in form. The final and fifth level was called the *overall strategy* level and was an evaluation of the development of ideas through the stack of fill-in forms.

The general idea was that each level would work as a sieve; statements (or aspects of statements) that were not valid at a certain level would not be analyzed at a next level. At each level the qualitative assessment of the subjects' statements was summarized into a quantitative score. The scores on each level are related to the scores on the previous level. For example: scores on the learning process validity level are related to the maximum score the subjects could have achieved, given the number of cells they have used (and which was scored on the previous global activity level).

RESULTS

Test for prior knowledge

The overall mean score on the test for prior knowledge, which consisted of 18 items, was 8.3 (SD = 2.4). The group with the controller overview scored a mean of 8.0 (SD = 2.9), the group with the concept overview had a mean score of 8.8 (SD = 2.3), and the control group scored a mean of 8.2 (SD = 2.0). A one-way analysis of variance showed that this mean score did not differ significantly between groups. Thus, the entrance level of the groups did not differ.

On the basis of their level of prior knowledge we ordered subjects into three groups: subjects having a low level of prior knowledge ($< M - 1SD$); subjects having an average level of prior knowledge (within + 1 SD and - 1 SD around the mean); and subjects having a high level of prior knowledge ($> M + 1SD$). This subdivision resulted in 12 subjects with a low level of prior knowledge, 61 subjects with a middle level, and 15 subjects with a high level. The distribution of number of subjects from each level of prior knowledge over the Controller, Concept, and Control group is respectively:

- for low level of prior knowledge: 8, 2, and 2;
- for middle level of prior knowledge: 18, 23, and 20
- for high level of prior knowledge: 5, 4, and 6

Posttest

At the end of the lab, all subjects were given a posttest that consisted of 18 multiple choice questions. Table 2 presents the mean scores of the posttest.

A one-way analysis of variance showed that there was an effect for the experimental condition ($F_{(2,85)} = 8.03$, $p < 0.001$). The group with the controller overview of hypotheses had the highest mean score (10.8), the Concept group had a mean score of 9.3, and the control group had the lowest mean score (8.5). On the level of prior knowledge results also showed significant differences between groups (Kruskal-Wallis $KW_{(2, N=88)} = 6.31$, $p < 0.05$). The subjects with a high level of prior knowledge had the highest mean score (10.9), the subjects with the middle level of prior knowledge scored a mean of 9.5, and the subjects with a low level of prior knowledge had the lowest mean (8.5). Figure 2 depicts the scores of the posttest. To gain an impression of the interaction between experimental condition and level of prior knowledge a separate Kruskal-Wallis one-way ANOVAs were done for each experimental group. Results of these tests showed no differences and, therefore, interaction between experimental groups and level of prior knowledge is unlikely.

A comparison between the control group and both experimental groups combined gave a significant difference in favour of the groups with support (Mann-Whitney $U_{(1, N=88)} = 558.5$, $p < 0.05$). As was mentioned above, the control group had a mean score of 8.5 ($SD = 2.8$). The subjects that received support have a mean score of 10.1 ($SD = 2.2$) on the posttest.

Table 2
Mean and standard deviation of the posttest score for the groups and the levels of prior knowledge for subjects

	Level of prior knowledge							
	Low (n = 12)		Middle (n = 61)		High (n = 15)		Total (N = 88)	
Groups	M	SD	M	SD	M	SD	M	SD
Controller (n = 31)	9.4	2.0	11.1	1.9	12.2	2.7	10.8	2.2
Concept (n = 29)	6.5	0.7	9.4	1.9	10.5	1.3	9.3	1.9
Control (n = 28)	7.0	4.2	8.1	2.7	10.2	2.5	8.5	2.8
Total	8.5	2.4	9.5	2.4	10.9	2.3	9.6	2.5

Analysis of exploratory learning

Analyzing students' learning processes means looking at data from the fill-in forms and since students worked in pairs we had to combine individual prior knowledge scores to create a prior knowledge score for each of the pairs. When the combination of individual levels of prior knowledge in a pair was charted we found five combinations: low-low, low-middle, middle-middle, middle-high, and high-high[3]. This was arranged into three groups of

[3] Notable is that the combination low-high was not present.

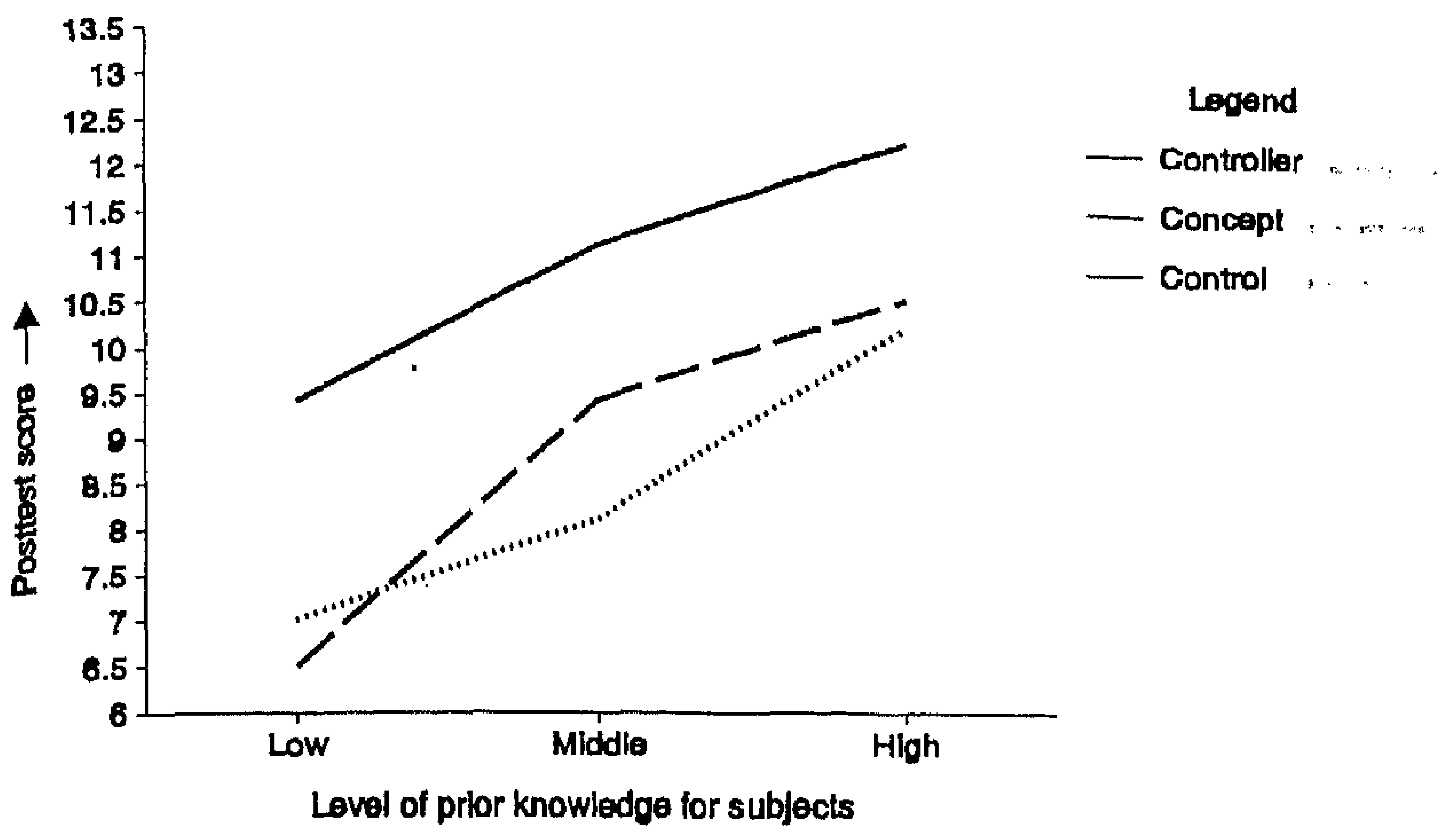

Figure 2: Mean posttest scores for the groups from the experiment and levels of prior knowledge for subjects

pairs: LOW pairs are pairs with low-low and low-middle combinations of individual scores, MIDDLE pairs are pairs where both subjects had a middle level of prior knowledge, and HIGH pairs are pairs with middle-high and high-high combinations of individual scores. There were 9 low pairs, 13 middle pairs, and 7 high pairs over both experimental groups.

Global activity level

The first level of analysis assessed the global activity of the subjects in terms of the number of forms, the number of hypotheses, and the percentage of cells that were filled in. The percentage of cells was a relative score; for each pair of subjects the number of cells filled in was divided by the maximum number of cells they could have scored given the number of forms this pair had used. Table 3 shows that overall the subjects used an average of 5 fill-in forms and that almost 87% of the cells on the forms were used. Furthermore, subjects explored an average of almost 5 hypotheses.

Table 3
Mean scores (M) and standard deviations (SD) of the general activity indicators for the overview conditions

	Overview of hypotheses					
	Controller (n = 15)		Concept (n = 14)		Total (n = 29)	
Indicators	*M*	*SD*	*M*	*SD*	*M*	*SD*
Number of forms	5.1	1.3	4.9	1.2	5.0	1.2
Number of hypotheses	4.7	0.8	4.6	0.7	4.6	0.7
Percentage cells	86.7	9.9	86.1	8.2	86.4	9.0

A t-test showed no significant differences between the Controller and Concept groups for each of the three general activity indicators. A Kruskal-Wallis one-way ANOVA showed no differences between the LOW, MIDDLE, and HIGH pairs for the indicators.

Learning process validity

At the second level of analysis the statements given by the subjects were assessed on their learning process validity i.e., it was evaluated whether the statement in each cell answered the general description of the cell as it was given on the information sheet. E.g., in the cell EXPERIMENT subjects should note down the input variables, the output variables, and the values of the input variables. For five cells together, a total of 19 aspects was stated.

The learning process validity scores were calculated first by counting the number of aspects from the assessment scheme that were included in the statements of the subjects and then relating this score to the maximum the subjects could have scored given the cells they used. Table 4 gives the learning process validity scores for the experimental condition.

Data in Table 4 show that overall the learning process validity score of the subjects was almost 37%. The highest score was in the cell EXPERIMENT where subjects scored 64.5% of the three aspects in this cell. In the cell PREDICTION OF THE EXPERIMENT a mean score of 11.7 % was achieved. Despite the heading, subjects did not predict the outcome of the experiment but usually predicted the validity of the hypothesis at hand. Furthermore, subjects sometimes noted down statements that could be scored in the cell VARIABLES & PARAMETERS or EXPERIMENT. The scores in the cells DATA INTERPRETATION and CONCLUSION were the result of positive scores on only a few of the aspects that could be scored. Aspects such as: noting down output and a remark about the validity of the hypothesis were frequently scored. However, aspects that required more profound analysis of output and generalization of findings were seldom or never found.

Table 4
Mean scores (M) and standard deviations (SD) of the learning process validity score for the overview conditions

	Overview of hypotheses					
	Controller (n = 15)		Concept (n = 14)		Total (n = 29)	
Cell	*M*	*SD*	*M*	*SD*	*M*	*SD*
VARIABLES & PARAMETERS	48.3	24.0	40.4	17.8	44.5	21.2
EXPERIMENT	75.1	21.3	53.1	20.7	64.5	23.5
PREDICTION OF THE EXPERIMENT	11.2	20.4	12.1	21.2	11.7	20.4
DATA INTERPRETATION	22.5	5.5	26.6	5.1	24.5	5.6
CONCLUSION	26.9	4.4	26.6	6.4	26.8	5.3
total[4]	38.9	9.3	34.4	6.8	36.7	8.4

[4] The total learning process validity is not an average of the scores of the five cells. For this score the total absolute score of all cells over all forms is taken and related to the maximum score (given the cells).

There are two significant differences between experimental groups on the learning process validity level. First, a difference was found for the cell EXPERIMENT ($t(28) = 2.83$, $p < 0.01$). Table 4 shows that the Controller group scored 75.1% of their maximum scores in the cell EXPERIMENT whereas the Concept group scored 53.1% of their maximum scores. Secondly, a difference was found for the cell DATA INTERPRETATION ($t\,(28)= 2.12$, $p < 0.05$). For the cell DATA INTERPRETATION Table 4 shows that the Concept group scored a higher score (26.6%) than the Controller group (22.5%). No significant differences between groups were found on the basis of their level of prior knowledge. Separate Kruskal-Wallis one-way ANOVA's for the experimental groups showed no differences for the total learning process validity score and, therefore, interaction between level of prior knowledge and overview structure is probably not present.

Domain correctness

For domain correctness we only analyzed aspects of the statements that were valid at the learning process validity level. If a subject made a mistake in one of the cells which caused mistakes in other cells, then the mistake was only scored once. Domain correctness scores were related to the scores of the previous level (learning process validity). For example, if a pair of subjects had two valid experiments, but one of them was incorrect, they scored 50% on domain correctness.

Results of the analysis at this level show that overall there is a high score on domain correctness with a mean of 89.7% (SD = 10.3). This means that almost 90% of the aspects that were scored learning process valid were also correct in domain-related sense. So, once the subjects made statements that were valid as an exploratory learning process, these statements were mostly correct at a domain level. The trend that we observed at the previous level, in which some of the cells had lower scores than others, is not seen at the domain correctness level. The percentages are between 80% and 99%. For example, overall the subjects had a mean score of 80.6% (SD = 25.1) in the cell DATA INTERPRETATION, and 99.1% (SD = 4.6) in the cell EXPERIMENT. There were no significant effects of the experimental conditions and/or the level of prior knowledge on this level of analysis.

Consistency

At the fourth level of analysis of the exploratory study process we assessed the consistency of statements in different cells on a fill-in form. So, for example, we estimated whether a statement given by students in the cell PREDICTION was consistent with their statement given in the cell EXPERIMENT at the same fill-in form. In our analysis we looked at four specific combinations and only included cells containing statements that were both learning process valid (now estimating a more limited number of what we called 'crucial' aspects) and domain correct. As a result not too many relations (25) were left for consistency assessment and of these relations only two were estimated as being inconsistent. Both regarded the relation between EXPERIMENT and DATA INTERPRETATION.

We made a separate and more detailed analysis of the relation between the cell HYPOTHESIS and EXPERIMENT since here differences between the two experimental groups existed (see section on learning process validity). Indicators for analyzing this relation were the number of appropriate experiments for the type of hypothesis at hand and the type of experiment that was used (analytic or simulation).

Analyzing the experiments that subjects designed in the cell EXPERIMENT showed that the mean number of experiments overall was 9.6 (SD = 4.9). Subjects hardly made mistakes with the appropriateness of the experiment for the hypothesis at hand. Just one pair of subjects made one inappropriate choice. A t-test showed ($t(29) = 2.22$, $p < 0.05$) that the Concept group designed significantly more analytic experiments (M = 2.3, SD = 1.7) that could be done without the simulation program than the Controller group (M = 1.1, SD = 1.1).

Overall strategy

As a final analysis level we assessed the development of subjects' ideas through the set of fill-in forms. We assumed that subjects' choices from the overview of hypotheses and the path through the hypotheses was an indication of their overall strategy.

In general, results showed that in both overview conditions the subjects had chosen a variety of paths. No single path was favoured and followed by more than three pairs of subjects. When we looked at the separate groups we saw that two pairs in the Controller group did choose the path which was preferred by the domain expert. For the Concept group none of the pairs had chosen the preferred path of the domain expert.

We also analyzed the number of pairs that had chosen a certain hypothesis. A chi-square test revealed that there is no overall difference in choice for hypotheses between the Controller and the Concept group ($\chi^2_{(7,\, N = 134)} = 7.9$; ns).

Conclusions

Our data bring us to a number of conclusions on the influence of the experimental conditions and prior knowledge on both exploratory learning and learning outcome.

Extra support for structuring the exploratory learning process and the specific structure of the overview of hypotheses has an effect on performance. First, our expectation that students who received the off-line support performed better on a posttest than students who had to work with the simulation without the off-line support is confirmed. Secondly, students who worked with the controller structure scored higher on the posttest than students working with the concept structure. This is, however, just the opposite of one of our expectations. We had expected that the concept structure would result in higher posttest scores because this overview addresses the fundamental aspects of the domain and supports the teacher's solution strategy to control problems. A possible explanation for our findings is that the controller structure could be more compatible with students needs and prior knowledge than the concept structure. In the textbooks, the controller structure is also employed and it is quite possible that students were more familiar with this point of view. Although teachers would prefer students to approach control problems with the concept structure, this structure does not recur in the textbooks. Likely, the concept approach is implicit, expert, knowledge and therefore not accessible, even not for students with high prior knowledge.

Students level of prior knowledge was important for the posttest scores i.e., the higher the level of prior knowledge the higher the posttest scores. This effect was present for both experimental groups and the control group. However, we did not find an interaction between level of prior knowledge and experimental condition. We had expected subjects with a high level of prior knowledge to gain a higher profit from the concept structure than the students with low prior knowledge, but our data show that both groups seem to prefer the controller structure.

Differences in quality of the exploratory learning processes matter; they can partly explain the results on posttest scores. One of the differences that we found is that students in the Controller group scored better in the cell EXPERIMENT on the learning process validity level than students in the Concept group. This means that the Controller group designed better experiments in the sense of more complete experiments. The Controller group more frequently stated the three required aspects of an experiment: their choice of input, the required output, and the value(s) of input variable(s) or parameter(s). It could be possible that this effect was strengthened by the fact that it is easier to design several experiments for one specific controller. This could have been an advantage for the Controller group since they had to explore the controllers one at a time. The Concept group, on the contrary, had to change between controllers all the time. This is in line with the findings on the consistency level that the Concept group designed more analytic experiments that could be done without the simulation than the Controller group. However, if we consider the number of experiments designed in the cell EXPERIMENT there are no differences between the two groups. Thus, the difference that we found in the cell EXPERIMENT is probably not caused by the above mentioned advantage.

The effects found in the quality of the learning processes are in line with the results of the posttest scores. They were generally positive for the Controller group (an exception was the cell DATA INTERPRETATION with a small difference in favour of the Concept group). However, differences in experimental conditions were reflected strongly in posttest scores but not profoundly in learning processes. Furthermore, the effect of the level of prior knowledge on the posttest scores was not found for the learning processes. It might be that our analysis measures for the exploratory processes do not sufficiently describe these processes and require more detailed and domain-related measurements. Additionally, we should keep in mind that the difference between experimental conditions is subtle.

Students do not perform exploratory processes too well, but practising the additional support may have been helpful. Although some aspects of the experimental set-up and the support measures are not exactly similar, we might compare results of the present study with results of Njoo and De Jong [12][5]. In the previous study we concluded that the valid performance of the learning processes was the bottleneck in exploratory learning. Results of the present study show that this remains a difficulty. Although in the current study students did not have to do the difficult process of hypothesis formation, they still did not do well on the other processes. In both studies interpreting data and drawing of conclusions was not done in depth and predicting the outcome of the experiment was seldom done in a correct way.

A significant finding in the present study is that experimental groups had higher learning outcomes. In the previous study we found no positive effects on learning outcome for the support measures. Students in this previous study had to perform two new tasks i.e., learn to work with the support measures and work out the assignment with the simulation, in the same session. An explanation could be found in related studies which show that additional support may draw away students' attention from the main task: learning about the domain

[5] It is possible to compare the (two) experimental groups of the present study with the group that was offered fill-in forms with predefined hypotheses in the previous study.

[5]. One of the advantages of the present study was that students had practised to work with the support measures prior to the evaluative session and although this did not result in a high quality exploratory learning process, this may have helped them in giving more attention to the main task. This underscores the general notion that experimental instructions should be worked with over a longer period of time before students get used to new ways of learning and improvements are to be expected.

With regard to the *design of simulation-based learning environments* we conclude that offering structures of hypotheses is a potential fruitful support measure. Our data show that students choose various paths through the hypotheses and no path was strongly favoured. From this we conclude that structured overviews of hypotheses still left sufficient exploratory freedom for learners. Secondly, we found that the form of the structure influences learning outcome (although not in the direction we expected). Studies on text structuring [1] show that the type of structuring of information is reflected in students' knowledge bases and consequently influences problem solving behaviour. In a similar vein, we assume that the particular structure of a set of hypotheses influences students' moves in hypothesis space [10] and, in this way, affects the resulting knowledge base. In summary we, therefore, conclude that offering structured overviews of predefined hypotheses as a mean of support is a promising method. We think, however, that structuring and supporting movements in hypothesis space needs to be studied in depth, before conclusive prescriptions for the design of simulation-based learning environments can be given.

Acknowledgements
The authors like to thank Christian Rademaker for his work as a domain expert in the studies reported here and Jos Banens from the Department of Mechanical Engineering (Eindhoven University of Technology) who, as a teacher of control theory, allowed us to perform the experiments and who supported us with his expert view on the domain throughout our work.

References

1. Eylon, B., & Reif, F.: Effects of knowledge organisation on task performance. Cognition and Instruction, 1, 5-44 (1984)
2. Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C.: Scientific reasoning across different domains. In E. de Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), Computer-based learning environments and problem solving (NATO ASI series F: Computer and Systems Series) (pp. 325-373). Berlin: Springer 1992
3. Jonassen, D.H.: Objectivism versus constructivism: Do we need a new philosophical paradigm? ETR&D, 39, 3, 5-14 (1991)
4. de Jong, T. (Ed.): Computer simulations in an instructional context. Education & Computing, 6 (1991)
5. de Jong, T., de Hoog, R., & de Vries, F.: Coping with complex environments: The effects of overviews and a transparent interface on learning with a computer simulation. International Journal of Man-Machine Studies, in press (1993)
6. van Joolingen, W.R., & de Jong, T.: Supporting hypothesis generation by learners exploring an interactive computer simulation. Instructional Science, 20, 389-404 (1991)
7. van Joolingen, W.R., & de Jong, T.: Modelling domain knowledge for Intelligent Simulation Learning Environments. Computers & Education, 18, 29-38 (1992)
8. van Joolingen, W.R., & de Jong, T.: Exploring a domain through a computer simulation: traversing variable and relation space with the help of a hypothesis scratchpad. Paper presented at the NATO ARW The use of computer models for explication, analysis and experiential learning. Bonas, France, October 1992
9. Kim, N., Evens, M., Micheal, J.A., & Rovick, A.A.: CIRCSIM-TUTOR: An intelligent tutoring system for circulatory physiology. In H. Maurer (Ed.), Computer Assisted Learning. Proceedings of the 2nd International Conference ICCAL (pp. 254-267). Berlin: Springer 1989

10. Klahr, D., & Dunbar, K.: Dual space search during scientific reasoning. Cognitive Science, 12, 1-48 (1988)
11. Michael, J.A., Haque, M.M., Rovick, A.A., & Evens, M.: The pathophysiology tutor: a first step towards a smart tutor. In H. Maurer (Ed.), Computer Assisted Learning. Proceedings of the 2nd International Conference ICCAL (pp. 390-400). Berlin: Springer 1989
12. Njoo, M., & de Jong, T.: Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. Journal of Research in Science Teaching (in press)
13. Plötzner, R., & Spada, H.: Analysis-based learning on multiple levels of mental domain representation. In: E. de Corte, M. Linn, H. Mandl & L. Verschaffel (Eds.), Computer-based learning environments and problem solving (pp. 103-129). Berlin: Springer 1992
14. Reimann, P.: Modelling scientific discovery learning processes with adaptive production systems. In D. Bierman, J. Breuker & J. Sandberg (Eds.), Artificial intelligence and education; synthesis and reflection. Proceedings of the 4th International Conference on AI and Education (pp. 218-227). Amsterdam: IOS 1989
15. Reimann, P.: Detecting functional relations in a computerized discovery environment. Learning and Instruction, 1, 45-65 (1991)
16. Schauble, L.S., Glaser, R., Raghavan, K., & Reiner, M.: Causal models and experimentation strategies in scientific reasoning. The Journal of the Learning Sciences, 1, 201-239 (1991)
17. Shuell, T.J.: Cognitive conceptions of learning. Review of Educational Research, 56(4), 411-436 (1986)
18. Shute, V.J., & Glaser, R.: A large-scale evaluation of an intelligent discovery world: Smithtown. Interactive Learning Environments, 1, 51-77 (1990)
19. Shute, V., Glaser, R., & Raghavan, K.: Inference and discovery in an exploratory laboratory, in P.L. Ackerman, R.J. Sternberg, and R. Glaser (eds.), Learning and Individual Differences, San Francisco: Freeman 1989.
20. White, B.Y. & Frederiksen, J.R.: Causal models as intelligent learning environments for science and engineering education. Applied Artificial Intelligence, 3(2-3), 83-106 (1989)
21. White, B.Y. & Frederiksen, J.R.: Causal model progressions as a foundation for intelligent learning environments. Artificial Intelligence, 42, 99-157 (1990)
22. Wildman, T.M.: Cognitive theory and the design of instruction. Educational Technology, July, 14-20 (1981)