

Promoting Self-Directed Learning in Simulation Based Discovery Learning Environments Through Intelligent Support

Koen Veermans, Wouter Van Joolingen, Ton De Jong

► **To cite this version:**

Koen Veermans, Wouter Van Joolingen, Ton De Jong. Promoting Self-Directed Learning in Simulation Based Discovery Learning Environments Through Intelligent Support. Interactive Learning Environments, Taylor

Francis (Routledge), 2000, 8 (3), pp.229 - 255. <10.1076/1049-4820(200012)8:3;1-D;FT229>. <hal-00197359>

HAL Id: hal-00197359

<https://telearn.archives-ouvertes.fr/hal-00197359>

Submitted on 14 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running head: PROMOTING SELF DIRECTED LEARNING

Promoting Self Directed Learning in Simulation Based Discovery Learning
Environments through Intelligent Support

Koen Veermans
Faculty of Educational Science and
Technology
University of Twente,
PO Box 217
7500 AE Enschede
The Netherlands
Phone: +31 53 489 46 59
Fax: +31 53 489 2849
e-mail: veermans@edte.utwente.nl

Wouter R. van Joolingen
Graduate school of teaching and learning
University of Amsterdam
Wibautstraat 2-4
1091 GM Amsterdam
The Netherlands
Phone: +31 20 525 15 95
Fax: +31 20 525 1290
wouter@ilo.uva.nl

Ton de Jong
Faculty of Educational Science and Technology
University of Twente,
PO Box 217
7500 AE Enschede
The Netherlands
Phone: +31 53 489 46 59
Fax: +31 53 489 2849
e-mail: jong@edte.utwente.nl

Abstract

Providing learners with computer generated feedback on their learning process in simulation based discovery environments cannot be based on a detailed model of the learning process due to the “open” character of discovery learning. This paper describes a method for generating adaptive feedback for discovery learning based on an “opportunistic” learning model that takes the current hypothesis of the learner and the experiments performed to test this hypothesis as input. The method was applied in a simulation based learning environment in the physics domain of collisions. Additionally, the method was compared to an environment in which subjects received pre-defined feedback on their hypotheses, not taking the experimentation behavior into account. Results showed that overall both groups did not differ on knowledge acquired. A further analysis indicated that in their learning processes the learners in the experimental condition built upon their intuitive knowledge base whereas, the learners in the control condition built upon their conceptual knowledge base. In addition, measures of the learning process showed that the subjects in the experimental condition adopted a more inquiry based learning strategy compared to the subjects in the control condition. We concluded, therefore, that the providing learners with adaptive feedback had a different and beneficial effect on the learning process compared to more traditional pre-defined feedback.

1 Introduction

Traditionally, Intelligent Tutoring Systems (ITSs) and environments for discovery learning were seen as incompatible. This is due to the origins of both types of learning environments which, when considered in their purest forms, use different paradigms of learning.

ITS systems tended to view as replacing the teacher or tutor. This means that their task is to maintain a tutoring dialogue with the learner, provide the learner with task-relevant information, set up exercises, and diagnose the learner's progress. Examples of such tutoring systems are the Lisp, and the geometry tutoring systems based on Act-R (Anderson, Farrell, & Sauers, 1984; Anderson, Boyle, & Jost, 1985). The main issue for an intelligent tutoring system is to decide what its next action will be in the interaction with the learner. Because the system is in control over the learning process, this decision is crucial for its success. ITSs base these decisions on a learner model, the system's representation of the learner's domain knowledge, which is adapted based on learner actions.

In discovery environments based on computer simulations (De Jong & Van Joolingen, 1998), learners can generate data and are expected to extract and construct knowledge from these data. The learners are in control and often have to set their own learning goals. At the same time, learners have to find the methods that help to achieve these goals. These activities would result in deeper rooting of the knowledge, enhanced transfer (Bruner, 1961), regulatory skill acquisition (Marchionni, 1988), and increased motivation. De Jong & Njoo (1992) describe discovery learning as comprising transformative processes that include analysis, hypothesis generation, testing and evaluation, and the regulative processes planning, verifying, and monitoring. Research shows that discovery learning is not easy for learners and that they may experience a large number of problems with these learning processes (de Jong & van Joolingen, 1998).

Due to the difference in locus of control and learning paradigm, ITSs and discovery learning seem to be incompatible. However, problems with the traditional ITS approach lead to systems that shift the control more to the learner (Holt, Dubs, Jones, & Greer, 1994; Shute & Psotka, 1993) and at the same time, problems with discovery learning (De Jong & Van Joolingen, 1998; Klahr & Dunbar, 1988) created a need for advanced support of discovery learning processes. In this paper, we will explore how we can use a method for learner modeling, usually associated with ITSs, in a discovery learning environment.

1.1 Discovery learning environments

One specific category of discovery learning environments is computer simulations. In computer simulations, learners change values of input variables and observe values of output variables, inducing characteristics of the underlying model. An example of such an environment is given in Figure 1. This is the interface of a simulation from the physics topic of collisions as it is used in the present study. Learners can do experiments by manipulating the variables' mass and velocity for both balls and then running the simulation by pressing the start button. The variables can be manipulated by clicking on the arrows next to the values or by just changing the values.

Visual information on the screen includes: graphs (displaying position, velocity, and kinetic energy of the balls), numerical output, and on the left side, an animation of the movement of the balls.

 INSERT FIGURE 1 ABOUT HERE

Support for discovery learning aims at providing context and tools for performing learning processes essential for discovery learning. The paradigm is that the environment provides the learner with cognitive tools (Lajoie, 1993; Van Joolingen, 1999). These tools help the learner perform learning processes by offering information, externalizing learning processes, or structuring the task. Examples of learning environments that follow this paradigm can be found in the environments based on the SimQuest authoring system for simulation-based discovery learning (Van Joolingen, King, & De Jong, 1997). Simulations are the core of the environments and learners are supported in various ways to help them in the learning process. SimQuest-based learning environments can structure the task using model progression (White & Frederiksen, 1990); provide support for regulation using assignments, small tasks that provide the learner with sub-goals that are within reach; provide the learner with just-in-time information in the form of explanations and allow learners to organize the experiments they have done with the simulation by offering a monitoring tool.

 INSERT FIGURE 2 ABOUT HERE

Typically, a learner will utilize the support in a SimQuest learning environment by opening an assignment and trying to reach the goal presented in the assignment. Figure 2 is an example of an assignment that goes together with the simulation presented in Figure 1. This particular type of assignment, an investigation assignment, asks the learner to investigate a certain relation in the simulation. The learner would conduct experiments with the simulation and the results would be displayed in the monitoring tool window (Figure 3). Next, the learner would analyze the results and choose one of the hypotheses in the bottom part of the assignment window. In the original SimQuest, the learner received feedback on this choice in the form of a pre-defined explanation containing information on the correctness of the hypothesis and optional extra information to support the learning process.

 INSERT FIGURE 3 ABOUT HERE

Using only standard techniques, the feedback that can be given on the learner's actions is limited. Given a list of hypotheses, and basic true/false feedback, the learner is easily entrapped in a trial and error-game. Moreover, giving direct feedback on an action (true or false with a pre-defined explanation) can be seen as conflicting with the ideas of discovery learning, because learners are supposed to find out for themselves what is right or wrong. Here is a point where intelligent support for discovery learning can prove its value.

1.2 Tutoring in Discovery Environments

There are two main arguments that explain why traditional ITS techniques such as learner modeling and instructional planning have not been used very often for discovery learning:

(a) the amount of learner freedom offered by discovery environments is so large that a full learner model is beyond the scope of practical application (i.e., the number of parameters is simply too large).

(b) often learner modeling is seen as incompatible with discovery learning and the related concept of constructivism, for which ‘measuring’ the learner's knowledge conflicts with the idea that the learner builds his or her own representation of the external world.

However, the techniques that have been developed for learner modeling can also be useful, in more open environments like discovery learning. The role of learner modeling in discovery learning is to generate support for the learner in the process of discovery, by inferring the inductive and deductive steps the learner takes in this process. The claim is not that the system maintains a full learner model of the discovery skills and steps taken by the learner, but that it infers just enough from the learner's behavior to support the discovery process.

The use of learner modeling in discovery learning also asks for a different attitude towards instructional planning, as directive teaching is not an appropriate teaching strategy for discovery learning. If the system is to give up control in favor of the learner this implies that tutoring should not be imperative, instead, the system should opt for a role of advising the learner. The implication is that the validity of the learner model is less critical. The aim is not to establish a complete model of the learners' knowledge, but to assist learners in their discovery processes enabling them to acquire the domain knowledge, using a representation of what the learner has actually done. Following Self (1990), we use a pragmatic approach in which we attempt to design interactions such that the information needed by the system is provided by the learner, link the content of the learner model to specific instructional actions, and assume a modest role for the model.

1.3 Intelligent Support for Discovery Learning

In this section, we outline a method for intelligently supporting the learner engaged in the discovery scenario that was sketched in the previous section. Using a learner model of the learner's discovery behavior and an expert model of discovery, it is possible to provide the learner with advice on experiment design, and data interpretation, while preserving the open nature of the discovery environment. The method is independent of domain knowledge.

Instead of providing the learner with direct predefined feedback on assignments as was explained in the previous section, the learner receives information to be used in evaluating the chosen hypotheses. This information is based on a comparison of a learner and an “expert” model; both abstracted from the experiments performed by the learner and the hypothesis that is investigated.

The algorithm for generating a learner model is based on the principles of induction and deduction. An induction process tries to infer a hypothesis from a given set of data, while a deduction process tries to predict experimental outcomes given a

hypothesis. In the algorithm we invert both processes. Instead of reasoning forward from given data to a candidate hypothesis, or from a given hypothesis to predicted data, we reason back from a candidate hypothesis to supportive data or from experimental outcomes to a hypothesis. In this way, we assess the steps the learner has taken and use this as a basis for generating tailored advice.

The starting point always is a set of experiments performed by the learner and the hypothesis that the learner is investigating. The hypothesis is obtained from an investigation assignment or from a source where the learner can enter the hypothesis in some format, like a hypothesis scratchpad (Van Joolingen & de Jong, 1993). In the assignment in [Figure 2](#), the learner has to investigate the relation between the variables v1 (initial velocity of ball 1) and Uk1_after (kinetic energy of ball 1 after the collision). The hypothesis the learner is working on is “If v1 becomes twice as great than the Uk1_after becomes four times as great”. Experiments are described as a set of values assigned to input and output variables.

In order to assess if a hypothesis can predict a given set of data, a stepwise process is applied to the set of data:

(1) First, a set of informative experiments about the relation is filtered from the complete set of performed experiments. In this case, the relation contains the variables v1 and Uk1_after. An experiment (or pair of experiments) is considered to be informative when the input variables that have been manipulated take part in the relation. If this set is empty, the process stops here.

(2) Then, a set of informative experiments about the hypothesis is filtered. This process uses the form of the hypothesis to divide the set of hypotheses resulting from the previous filter into sets, which can each generate a prediction using the hypothesis. For instance, for a hypothesis with the form: “If a doubles, b is divided by two”, experiments will be selected where a is doubled, quadrupled etc. and where all other input variables are kept constant.

(3) For each of the informative sets for the hypothesis, predictions are generated for the output variables. This can be done using the hypothesis. For instance, if the hypothesis is a quantitative relation, such as $y = 2x$. Then the output variable y can be computed directly from the input variable x . If the hypothesis is qualitative, such as: “When x increases, y increases”, it can be inferred from the (x, y) pairs: (1,5), (2,?) that the value on the question mark must be greater than 5.

(4) The generated predictions are compared to the values actually found in the experiments. On mismatches, advice can be generated.

The algorithm for analyzing experiments as described here results in two kinds of information. First, the presence of informative sets of experiments about a certain hypothesis contains information about what a learner could have inferred about a certain relation in the domain. Second, the results of the analysis can be used as a source to assess the learner’s discovery behavior as such. In this sense, the information constitutes a learner model of both the domain knowledge collected using the discovery environment and the discovery behavior of the learner. In both cases, this information can be used to generate advice, directed at improving the efficiency and effectiveness of the discovery process, without disrupting the self-directed nature of this process. A worked example on this method can be found in Veermans and Van Joolingen (1998).

The system can assess whether conclusions can be drawn about the hypothesis at all, and, if necessary the system can give advice on the kind of experiments needed to make an evaluation of a hypothesis possible. Also, the system can infer whether a conclusion drawn by the learner is correct or not, and draw a learner's attention to conflicting data in the case the learner makes incorrect inferences from the data.

 INSERT FIGURE 4 ABOUT HERE

In the example shown in Figure 4, the learner evaluated the hypothesis "If v_1 becomes twice as great than the Uk_1 after becomes four times as great" to be true. The analysis of the experiments resulted in two sets of two experiments that contain information that can be used for evaluation of the hypothesis. The other experiments that the learner did were not informative because variable v_1 was not changed, or did not fit the condition part of the hypothesis. The first experiment of each of the two sets is used to generate a prediction for the value of variable Uk_1 after and this value is then compared to the value that was actually observed in the simulation. It is then explained that a correct hypothesis should generate correct predictions in all cases. In this example, the predictions did not match the actual values. Thus, the learner is informed about this discrepancy and is advised to reject the hypothesis.

2 Evaluation Study

The method for intelligent support was evaluated in an experimental study in which a discovery environment with the new support method was compared with one that provided learners with the traditional pre-defined feedback. The main research question was whether intelligent feedback on the learners' experimentation behavior influenced learners' discovery behavior and the learning results. This was investigated using a pre-test – post-test design with different kinds of knowledge tests and by studying learners' behavior through logs of the interaction with the learning environment.

2.1 Experimental Conditions

In the study, two conditions corresponding with the aforementioned different environments were realized:

(1) A Control Condition: subjects interacted with a SimQuest simulation environment on Collisions (see Section 2.3). This simulation included assignments stimulating subjects to detect the principles behind moving and colliding particles by means of manipulating input variables and interpreting the outcomes of their experiments. The assignments covered relations in the domain and contained hypotheses about these relations. The learners were asked to select the correct hypothesis or hypotheses from this list. If they did so, they received feedback that was pre-defined and contained a statement about the truth-value of the hypothesis and additional information. For instance, if the hypothesis was correct, but there was another, more precise, correct hypothesis the additional information prompted the learner to look for this other hypothesis as well.

(2) An Experimental Condition: subjects interacted with basically the same SimQuest simulation environment on Collisions but there were two differences. First, in the experimental condition, subjects also had the option to state that a hypothesis was incorrect. Second, the feedback they received in the experimental condition depended on the experiments that the subjects did to support their statement about the hypothesis. The feedback contained an analysis of this evidence and, if needed, advice on the discovery processes along the lines described earlier in this article.

2.2 Subjects

Forty-six Dutch students, - fourth year of pre-scientific education, age 15-16 -, from two schools participated in this study. Subjects attended physics classes and had some computer experience. The subjects were transferred from their schools to the university to participate in the experiment. One school participated with twenty-three subjects from two classes in one experimental session. The other school participated with twenty-three subjects from one class. The subjects of these classes were distributed over the two experimental conditions. The subjects received no compensation or credit for their participation.

One subject was excluded from the analysis because the response times on most of the items in the post-test were close to the minimum time that is needed to respond, and the number of correct answers dropped dramatically from pre- to post-test. There is reason to believe that this subject did not try to answer the items correctly but merely tried to finish the test as fast as possible. Furthermore, for four subjects the result of one test was lost. The results of these subjects were excluded analysis by analysis.

2.3 The Learning Environment

The discovery learning environment used in this study was called on Collision and covered the physics domain of central collisions between two balls. Collision was developed in the SERVIVE project (Van Joolingen, King, & De Jong, 1997; De Jong et al., 1998; De Jong et al., 1999). The Collisions learning environment was designed for learners in the fourth year of secondary school. It included four levels of complexity: non-accelerated movement, collisions against a wall, completely elastic collisions between two balls in one dimension and, completely inelastic collisions between two balls in one dimension. Figure 1 displays a simulation interface for the third level. Apart from the four levels, the environment contained support in the form of assignments, a monitoring tool that registered the experiments with the simulation, background information about the simulations, and feedback explanations.

2.3.1 Assignments

A total of 41 investigation assignments guided the subjects in exploring the domain. Learners in both conditions were free to choose any assignment. The assignments offered four to eight hypotheses to the subjects. The subjects were advised to select one of the hypotheses from the list, to experiment with the simulation, to evaluate the evidence for the hypothesis and, if they felt this was necessary, to investigate other hypotheses as well. The content of the assignments was identical for the two conditions. In both conditions, assignments guided the subjects in investigating the relation between (a) mass, velocity, and momentum; (b) mass, velocity, and kinetic

energy; (c) mass, velocity, and resulting velocities after a collision. In addition, conservation of momentum was treated in assignments on elastic and inelastic collisions and contrasted with the loss of kinetic energy in inelastic collisions.

2.3.2 Monitoring Tool

Whenever a subject opened an assignment concerning a relation between variables, a monitoring tool (Figure 3) was activated which automatically registered the values of the important variables in that relation. This monitoring tool served as a kind of external memory for the subjects. After each experiment the values for the variables were listed in the monitoring tool. The subjects had the opportunity to re-run any of the experiments that were in this list and could re-arrange the experiments in the list to get a better overview of the experiments. The general idea behind an instructional measure like the monitoring tool is to allow the subjects to focus on discovering the relations in the domain. Without the presence of the monitoring tool, subjects would have to remember the results of their experiment and think of an appropriate next experiment at the same time and then interpret the results afterwards.

2.3.3 Tests

Three different tests were administered to assess the subjects' knowledge: the definitional knowledge test, the what-if knowledge test, and the what-if-why test. These tests were also used earlier in Swaak, De Jong, and Van Joolingen (submitted).

Definitional knowledge test. The definitional knowledge test consisted of three-answer items and aimed to measure conceptual knowledge of a declarative quality like definitions and equations. The same definitional test was given both as a pre- and as a post-test. Whenever learners selected an answer, the item disappeared from the screen and the next item popped-up. Learners were allowed to return to previously answered items. The definitional knowledge test consisted of 20 items.

Intuitive knowledge test (what-if test). To measure intuitive knowledge about the relationships between the variables of the domain, a test called the what-if test was created (Swaak & De Jong, 1996). In the what-if test, each test item contained three parts: conditions, actions, and predictions. The conditions and predictions were possible states of the system. The conditions were displayed in graphs. The action was presented in text. The predicted states were, like the conditions, presented in graphs. In the instructions of the what-if test the learners were asked to decide which state would follow from a given condition as a result of the action. The items of the task were kept as uncomplicated as the domain permitted. The items had a three-answer format. In order to prevent memorization effects, two parallel versions of the intuitive knowledge test were developed (however 9 of the 24 items were identical in both versions because no parallel item could be constructed). Whenever learners selected an answer, the item disappeared from the screen and the next item popped-up. Learners could not go back to previously answered items.

What-if-why test. The what-if-why test was essentially a paper version of the what-if test. It too required the learners to decide which of three situations followed from a given condition, given the action that was displayed. Additionally, the learners

had to state their answer in their own words. In this study, the learners were also asked to depict a situation in which the other answer alternatives would be true. A sub-test of 13 items from the what-if post-test was used for this purpose. Thus, the difficulty level and the level of detail of the two test formats were exactly identical. However, the what-if prediction task and the what-if-why prediction and explanation tasks contrasted, on the demand they placed on the conscious awareness of the learners with respect to the underlying principles and the implications of physics laws. This awareness was needed in the what-if-why test items, but was not necessary in the what-if test items.

2.3.4 Interaction Behavior

Many of the subjects' actions during the interaction with the environment were registered. This provided us with data on the use of the simulation, the assignments, and feedback to answers of assignments. In addition, time spent on different simulations was recorded. These data were used to compare the Experimental and Control condition in terms of the general interaction patterns of the subjects, and to associate the interaction within a condition with test outcomes.

2.3.5 Procedure

Each experimental session lasted approximately three hours. Each session was comprised of the following sequence of events:

(1) Introduction and pre-tests (40 minutes). Subjects were welcomed and given an overview of the activities in the session. After the introduction, the definitional knowledge and what-if pre-test were administered electronically.

(2) Introduction to the learning environment (10 minutes). Upon completion of the pre-tests, subjects read an introduction on the Collision environment. This was followed by a demonstration in which the experiment leader showed the function of the various elements of the learning environment and explained how they could be operated. It was explained to the subjects that both their performances on the tests and their interaction with the learning environment would be recorded.

(3) Interaction with Collision (set at 1 hour and 30 minutes). After the introduction, subjects worked with the Collision environment on their own. The experiment leader was present and could give assistance on questions concerning operating the environment, but not on questions concerning subject matter. Subjects were encouraged to use the full one and a half-hour available for the interaction. If they wanted finish earlier they were asked to explore more of the environment, but were not forced to do so.

(4) Post-tests (45 minutes). After the interaction with the learning environment the post-tests were administered. The definitional knowledge test was administered first, then the what-if test, and finally the what-if-why test followed. The what-if-why (with the prediction and the explanation part) test was administered using paper and pencil, the other two tests were presented electronically.

2.4 Predictions

The learning environments in the two conditions were identical except for the possibility to reject a hypothesis in the experimental condition and the feedback that the subject received upon evaluating a hypothesis. Both environments required the subjects

to investigate relations between the variables in the domain through experimenting with the simulation.

Our first prediction concerns the subjects' interaction behavior. Because subjects' experiments were used in the feedback the subjects in the experimental condition were invited to reflect more on their interaction with the learning environment. Consequently, we expected them to spend more time on designing and interpreting experiments resulting in more careful evaluation of the hypotheses.

Our second prediction was that both conditions would gain equally in definitional knowledge as measured by the definitional knowledge test, but that the experimental condition would also gain more knowledge as measured by the what-if and what-if-why tests. The definitional knowledge test assesses the formal principles of the domain, which are not explicitly dealt with in either of the two environments. We anticipated subjects in the experimental condition to perform better on the what-if test because, as a result of the more intense interaction, the subjects in this condition would have constructed more integrated knowledge of the relations of the domain. Finally, we expected the subjects in the experimental condition to perform better on the why part of the what-if-why because they would reflect more on their interaction with the environment, therefore being able to explain more adequately in their own words why a certain situation had occurred.

2.5 Results

The result section presents analyses of the learning outcomes and processes, and consists of five parts. The first three parts of this section present results on the knowledge tests. First, the overall results of these tests will be described, and the two conditions will be compared. Then, as a measure for the integration of knowledge, correlations between post-tests will be shown. We will go on by investigating the relationship between the pre-tests and the post-tests to see what role the definitional and intuitive knowledge play in the two conditions. The fourth part presents data on the interaction of learners with the learning environment, and in the fifth and final part the relation between the interaction and the knowledge measures is presented.

2.5.1 The Knowledge Measures

The definitional knowledge test was administered before and after the session. Reliability analysis (N=43; n=20 items) resulted in a reliability of .46 (Cronbach's α) for the pre-test and .38 for the post-test. The reliability of the pre-test was moderate, but the reliability of the post-test was relatively low. The what-if test was administered in two parallel versions as pre- and post-test. Reliability analysis of the pre-test (N=44; n=24 items) resulted in a reliability of .70 (Cronbach's α) and .56 for the post-test. The reliability of the pre-test was good and the reliability of the post-test moderate. The definitional knowledge test and the what-if test are assumed to measure different types of knowledge in learners. The low correlation between the pre-tests (.29) supported this assumption. The what-if-why test was scored on correctness of the prediction given and on the correctness of the explanation given for the prediction. The score for the explanation could be either 0 (incorrect), 0.5 (partly correct), or 1 (correct) and was rated by two independent domain experts. The inter-rater reliability yielded a Kappa of .65, which can be considered substantial. The reliability of the what-if-why correct test

(Cronbach's α) was somewhat low (.38) and the reliability of the what-if-why explanation test was good (.69).

The results of the knowledge tests are given in Table 1. As can be seen, learners gained on both the definitional and what-if tests in both conditions. Paired-samples T-Tests showed a significant within-subject effect for the definitional test across conditions ($t=-5.96$, $df=42$, $p=.001$) and in each of the conditions (for the experimental condition: $t=-3.35$, $df=19$, $p=.003$; for the control condition: $t=-5.14$, $df=22$, $p=.000$). The overall effect size was $d=.88$. Similarly, there was a significant within-subject effect for the what-if test (overall: $t=-7.81$, $df=42$, $p=.000$; experimental: $t=-5.11$, $df=20$, $p=.000$; control: $t=-6.04$, $df=21$, $p=.000$). Overall effect size again was $d=.88$.

There were no significant differences between conditions in the mean scores on the pre-tests, and on the post-tests, including the two measures on the what-if-why tests. T-tests on the measures show $p > 0.3$ in all cases.

Table 1
Mean Scores and Standard Deviations for the Different Knowledge Tests in the Different Conditions. Standard Deviations are given within Parentheses

| | Condition | | | | Total | |
|--------------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | Experimental | | Control | | Pre | Post |
| | Pre | Post | Pre | Post | | |
| DEFINITIONAL (MAX=20) | 11.6 (2.6) | 13.7 (2.4) | 10.9 (2.9) | 13.4 (2.3) | 11.3 (2.8) | 13.5 (2.4) |
| WHAT-IF (MAX. 24) | 14.7 (4.0) | 18.3 (3.2) | 14.1 (3.6) | 17.6 (2.7) | 14.4 (3.8) | 17.9 (3.0) |
| WHAT-IF-WHY CORRECTNESS (MAX. 13) | | 11.2 (1.4) | | 11.1 (1.6) | | 11.1 (1.5) |
| WHAT-IF-WHY EXPLANATION (MAX. 13) | | 7.0 (2.1) | | 7.2 (2.5) | | 7.1 (2.3) |

2.5.2 Relations Between the Different Knowledge Measures

In order to find out to what extent the different knowledge measures assess similar or different constructs, their correlations were computed. Table 2 displays the correlations between the post-tests. In the experimental group, correlations between the definitional post-test and the what-if and what-if-why explanation post-test were not significant. All other correlations were significant. In the control group all correlations were significant.

Table 2
Correlations Between the Scores on the Different Post-Tests, for Each of the Conditions.

| | | DEF. POST- TEST | WHAT-IF POST- TEST | WHAT-IF-WHY CORRECTNESS |
|----------------------------|--------------|--------------------|-----------------------|----------------------------|
| WHAT-IF POST-TEST | EXPERIMENTAL | .38 | - | - |
| | CONTROL | .51* | - | - |
| WHAT-IF-WHY CORRECTNESS | EXPERIMENTAL | .44* | .74** | - |
| | CONTROL | .75** | .54** | - |
| WHAT-IF-WHY EXPLANATION | EXPERIMENTAL | .13 | .60** | .58** |
| | CONTROL | .61** | .56** | .72** |

Note. * means $p < 0.05$, ** means $p < 0.01$.

These results were somewhat surprising. In the experimental condition, the definitional knowledge post-test seems to measure a different construct than the other tests, whereas in the control condition all post-test scores seem related, and therefore it can not be claimed that the post-tests measure different constructs. Given the low correlation between the pre-test scores on definitional knowledge and what-if knowledge, the correlations between the pre-tests and the post-test knowledge measures were computed to gain insight in the causes of the differences.

2.5.3 Relations Between Pre- and Post-Test Knowledge Measures

Table 3 displays the correlations between the pre- and post-tests scores, thus showing to which extent the two types of prior knowledge (definitional and intuitive knowledge) predicted the subjects' performance on the post-tests. A basic expectation is that a high score on a pre-test leads to a high score on the post-test of the same type. This expectation was met for the what-if tests in both conditions, for the definitional tests in the control condition, but not for the definitional tests in the experimental condition.

Comparing the experimental and the control condition we see remarkable differences in the pattern of correlations with post-tests for the definitional and what-if pre-test. For the definitional knowledge, in the experimental condition the pre-test does not correlate significantly with any of the post-tests. In the control condition, to the contrary, the definitional knowledge pre-test correlates significantly with all of the post-test measures. For the what-if knowledge pre-test, the pattern is reversed. In the experimental condition, the what-if pre-test correlates significantly with the post-tests. In the control condition, the what-if pre-test correlates significantly with the what-if post-test, but not with any of the other post-tests. In summary, it seems that in the experimental condition the what-if pre-test is the main predictor of all post-test scores whereas in the control condition the definitional pre-test is the main predictor.

Table 3
Correlations Between the Pre-Test Scores and the Post-Test Scores on the Knowledge Tests

| | | DEFINITIONAL POST-TEST | WHAT-IF POST-TEST | WHAT-IF- WHY CORRECTNESS | WHAT-IF- WHY EXPLANATION |
|--------------------------|--------------|---------------------------|----------------------|--------------------------------|--------------------------------|
| DEFINITIONAL PRE-TEST | EXPERIMENTAL | .30 | .39 | .33 | .15 |
| | CONTROL | .64** | .46* | .46* | .59** |
| WHAT-IF PRE-TEST | EXPERIMENTAL | .61** | .62** | .76** | .37 |
| | CONTROL | .20 | .68** | .21 | .32 |

Note. * means $p < 0.05$, ** means $p < 0.01$.

To further investigate these differences an extra analysis was performed. If the conditions differ with respect to the way their pre-test scores predicted their post-test scores, this should become apparent by comparing groups based on their pre-test scores. Therefore, both the experimental and control groups were divided into two groups, based on their score for the definitional pre-test (using median split, with the median left out). In Table 4, these groups are labeled LD and HD respectively (Low/High on Definitional test). These groups were compared for their scores on the four post-tests, using T-tests. The results are displayed in Table 4. The results from the correlational analysis reoccur on a more detailed level in this analysis. In the experimental condition,

the low definitional knowledge group can not be distinguished from the high definitional knowledge group on the basis of the post-test results, not even on the post-test version of the definitional test itself. In the control group, the difference between the groups becomes smaller but is still present on the definitional post-test, and differences are also found on the other post-tests, although the difference for the what-if post-test is not significant.

Table 4
Relations Between Pre-Test Score on the Definitional Knowledge Test and the Score on the Post-Tests.

| | EXPERIMENTAL | | | CONTROL | | |
|-------------------------|--------------|---------------|----------------|--------------|---------------|--------------------|
| | LD (9.33) | HD (14.11) | T (DF) P | LD (8.11) | HD (13.67) | T (DF) P |
| DEFINITIONAL POST TEST | 13.56 | 14.11 | -.48 (16) .635 | 11.56 | 14.33 | -3.04 (11.8) .010* |
| WHAT-IF POST TEST | 17.88 | 18.89 | -.67 (15) .535 | 16.44 | 18.67 | -1.64 (16) .121 |
| WHAT-IF-WHY CORRECTNESS | 11.11 | 11.44 | -.53 (16) .640 | 9.89 | 11.89 | -2.84 (16) .012* |
| WHAT-IF-WHY EXPLANATION | 7.00 | 7.22 | -.32 (16) .819 | 5.39 | 8.28 | -2.83 (16) .012* |

Note. * means $p < 0.05$, ** means $p < 0.01$. The p-values are computed using a T-test comparing the high scoring groups with the low scoring groups on the definitional pre-test. Between parentheses, the mean scores on the definitional pre-test are given for each group.

A similar analysis was done based on the scores for the what-if pre-test, yielding LW and HW groups. These results can be found in Table 5. Again, a pattern emerged in line with the correlational analysis. Now the low and high group in the experimental condition could still be distinguished in all the post-test scores, whereas in the control group only a difference was found on the what-if test post-test.

Table 5
Relations Between Pre-Test Score on the What-If Test and the Score on the Post-Tests.

| | EXPERIMENTAL | | | CONTROL | | |
|-------------------------|---------------|---------------|-------------------|---------------|---------------|------------------|
| | LW (11.27) | HW (18.18) | T (DF) P | LW (10.62) | HW (17.30) | T (DF) P |
| DEFINITIONAL POST TEST | 12.36 | 15.0 | -2.97 (20) .008** | 13.75 | 13.6 | .13 (11.4) .900 |
| WHAT-IF POST TEST | 16.72 | 20.1 | -2.79 (19) .012* | 16.25 | 19.2 | -2.47 (16) .025* |
| WHAT-IF-WHY CORRECTNESS | 10.27 | 12.09 | -3.81 (20) .001** | 11.0 | 11.2 | -.30 (16) .768 |
| WHAT-IF-WHY EXPLANATION | 5.95 | 8.0 | -2.61 (20) .017* | 6.5 | 7.45 | -.80 (16) .433 |

Note. * means $p < 0.05$, ** means $p < 0.01$. The p-values are computed using a T-test comparing the high scoring groups with the low scoring groups on the What-If pre-test. Between parentheses, the mean scores on the What-If pre-test are given for each group.

2.5.4 Interaction Measures

Actions that subjects performed while interacting with the learning environments were registered. This provided data on the use of the environments including time

distribution over the four levels of complexity, assignments, feedback on hypotheses, experiments, and the variability of experiments. Table 6 summarizes data on time spent in general and on the levels. Table 7 summarizes the data on the assignments, experimentation, and feedback.

Table 6
Mean and Standard Deviations of Time Spent on the Complexity Levels in Minutes

| TIME MEASURES (MINUTES) | CONDITION | | | |
|-----------------------------------|--------------|---------|---------|--------|
| | EXPERIMENTAL | | CONTROL | |
| LEVEL 1 (non-accelerated mov.) | 22:15 | (9:28) | 12:46 | (7:25) |
| LEVEL 2 (collisions against wall) | 20:17 | (6:09) | 21:40 | (7:36) |
| LEVEL 3 (elastic collisions) | 25:30 | (5:47) | 26:17 | (6:23) |
| LEVEL 4 (inelastic collisions) | 18:16 | (13:05) | 23:33 | (9:57) |
| TOTAL TIME GENERAL | 86:18 | (4:48) | 84:16 | (4:44) |

As shown in Table 6, the subjects in the experimental group spent considerably more time than the control group on the first level ($t= 3.73$, $df=39.79$, $p= .001$). Comments by subjects like “It does not say what the correct answer is” indicated that the extra time spent on this level was spent there to get acquainted with the feedback. Time spent on the second and third level was almost equal between the two groups ($t= -.673$, $df=43$, $p= .504$; $t= -.427$, $df=43$, $p= .671$). The experimental group spent less time on the last level although, due to the large variance, this difference is not significant ($t= -1.53$, $df=43$, $p= .159$).

Table 7
Mean and Standard Deviations of Interaction Measures within Conditions

| INTERACTION MEASURES | CONDITION | | | |
|-----------------------------|--------------|--------|---------|--------|
| | EXPERIMENTAL | | CONTROL | |
| ASSIGNMENTS | 33 | (8) | 40 | (16) |
| UNIQUE ASSIGNMENTS (MAX 41) | 30 | (8) | 34 | (9) |
| EXPERIMENTS | 111 | (50) | 84 | (22) |
| EXPERIMENTS DURING ASSIGNM. | 89 | (39) | 55 | (28) |
| FEEDBACK | 94 | (38) | 54 | (24) |
| AV. FEEDBACK /ASSIGNMENT | 3.2 | (1.3) | 1.6 | (0.5) |
| UNIQUE EXPERIMENTS | 65 | (31) | 46 | (12) |
| AV. TIME PER ASSIGNM. (MIN) | 2:32 | (1:44) | 1:44 | (1:43) |

Independent samples T-tests for the interaction measures showed a significant difference on the total number (multiple use allowed) of assignments used ($t= -2.04$, $df=33.44$, $p= .049$), with the control group using more assignments. Furthermore, significant differences were found on the overall number of experiments performed ($t= 2.34$, $df=28.65$, $p= .027$), the number of experiments performed during an assignment ($t= 3.32$, $df=43$, $p= .002$), the total amount of feedback ($t= 4.23$, $df=35.22$, $p= .000$), the average amount of feedback in an assignment ($t= 5.69$, $df=26.87$, $p= .000$), and the number of unique experiments performed ($t= 2.60$, $df=27.01$, $p= .015$), all with the larger numbers in the experimental group. On average, subjects in the experimental condition also spent more time on an assignment ($t= 3.63$, $df=43$, $p= .0001$). Only the number of unique assignments did not show a significant difference between the two conditions ($t= -1.52$, $df=43$, $p= .136$).

2.5.5 Relations Between Interaction and Knowledge Measures

We also investigated the relations between the learners' activities and the results on the various knowledge tests (i.e., whether the behavior in the learning environment could be related to the scores on the post-tests). A good measure from the learners' activities is the number of different assignments that were used as this provides an indication of the level coverage. The number of unique experiments learners conducted within each level is also an appropriate measure, because the number of unique experiments indicates the amount of evidence gathered that could be utilized for understanding the simulations' underlying principles. In Table 8, the correlations between the number of different assignments and the post-test scores are shown. Table 9 displays the correlations between the number of unique experiments and the post-test scores. In both tables, the correlations are computed for the overall number, but also for the number on each of the four levels that were present in the learning environment.

Table 8
Correlations Between the Number of Assignments used on the Different Levels and the Results on the Different Post Tests

| NUMBER OF ASSIGNMENTS | LEVEL 1 | LEVEL 2 | LEVEL 3 | LEVEL 4 | TOTAL |
|--------------------------------|--------------|--------------|--------------|---------|-------------|
| DEFINITIONAL POST TEST | | | | | |
| EXPERIMENTAL | -.21 | .21 | .34 | -.01 | .26 |
| CONTROL | .42* | .54** | .47* | .29 | .51* |
| WHAT-IF POST-TEST | | | | | |
| EXPERIMENTAL | -.04 | .22 | .22 | .06 | .24 |
| CONTROL | .38 | .50* | .53** | .33 | .52* |
| WHAT-IF-WHY CORRECTNESS | | | | | |
| EXPERIMENTAL | .16 | .38 | .22 | -.07 | .17 |
| CONTROL | .56** | .65** | .38 | -.04 | .40 |
| WHAT-IF-WHY EXPLANATION | | | | | |
| EXPERIMENTAL | -.02 | .32 | .21 | .04 | .27 |
| CONTROL | .40 | .42* | .45* | .17 | .42* |

Note. * means $p < 0.05$, ** means $p < 0.01$.

Table 8 shows the correlations between assignment use and post-test scores. As can be seen in this table there were quite a few significant correlations between assignment use and the post-test results in the control condition. The total number of different assignments in the control condition were significantly correlated with all post-tests scores with the exception of the what-if-why correctness post-test. On level one, the correlations between the use of assignments and scores on definitional knowledge and what-if-why correctness were significant. The number of assignments used on level 2 correlated with all post-test results. Correlations between assignment use on level 3 resembled these for the total number of assignments, and only the use of assignments on level 4 showed no significant relations with the post-test results. This contrasted sharply with the experimental condition, where no significant correlations were found between the use of assignments and the post-tests in the experimental condition.

Table 9
Correlations Between the Number of Unique Experiments Performed with the Simulations, and the Results on the Different Post Tests

| UNIQUE EXPERIMENTS | LEVEL 1 | LEVEL 2 | LEVEL 3 | LEVEL 4 | TOTAL |
|--------------------------------|---------|-------------|--------------|---------|-------------|
| DEFINITIONAL POST TEST | | | | | |
| EXPERIMENTAL | -.15 | .01 | .36 | -.11 | .08 |
| CONTROL | -.03 | -.28 | -.03 | .16 | -.06 |
| WHAT-IF POST-TEST | | | | | |
| EXPERIMENTAL | -.11 | .52* | .60** | .18 | .47* |
| CONTROL | -.06 | -.18 | .07 | -.06 | -.09 |
| WHAT-IF-WHY CORRECTNESS | | | | | |
| EXPERIMENTAL | .23 | .46* | .42 | -.08 | .29 |
| CONTROL | .29 | -.03 | -.13 | -.39 | -.20 |
| WHAT-IF-WHY EXPLANATION | | | | | |
| EXPERIMENTAL | .07 | .34 | .48* | -.05 | .28 |
| CONTROL | .14 | -.15 | .32 | -.01 | .16 |

Note. * means $p < 0.05$, ** means $p < 0.01$.

Table 9 shows a significant correlation between the number of unique experiments on level 3 and the results on the what-if post-test and the what-if-why explanation test for the experimental group. Significant correlations for the experimental group were also found between the number of unique experiments on level 2 and the what-if, and what-if-why correctness post-test. The total number of unique experiments correlates significantly with the what-if post-test. Only for the definitional knowledge test, there was no significant correlation between the number of experiments and the scores on the post-test. In the control condition, no significant correlations were found between unique experiments and post-tests.

3 Conclusions

We predicted that as a result of the type of feedback given (i.e., feedback that took the experimentation behavior of learners into account) the experimental group in our study would show a more reflective attitude while interacting with the learning environment. In the interaction data we found evidence that this was indeed the case. The interaction data revealed that on average subjects in the experimental group spent more time on an assignment, did more experiments when working with an assignment, did a larger percentage of their experiments during assignments, and did more unique experiments than subjects in the control group. The picture that emerges is that learners in the experimental group needed some time to get accustomed to the feedback, reflected in the time spent on the first complexity level. Afterwards, however, their behavior focussed on analyzing hypotheses rather than on solving assignments.

The effects of this different behavior did not directly show in the scores on the knowledge tests that were administered after the experimental session. Both conditions gained from pre- to post-test on both the definitional and what-if tests. Contrary to the predictions, no significant differences in favor of the experimental condition were found

for the what-if test or the what-if-why test. Actually, when looking at the average scores, there is no difference at all between groups, so it seems that there is no influence of the experimental treatment on the definitional and intuitive domain knowledge that learners gained during the interaction with the learning environment.

However, close examination of the results reveals effects in the way the overall means were constituted by the individual subjects' scores. This becomes clear when looking at the relation between the pre-tests and post-tests, as shown in Table 3 and as it is elaborated in Table 4 and Table 5. The results display a completely different picture for the subjects in the two groups. In the control group, the major predictor of a post-test result was the result on the same test as the pre-test. Of course, this is nothing special, starting higher on a test means that the same person will probably score higher when the same test is taken again. In the control group, this was true for both the definitional test and the what-if test. In the experimental group, however, the what-if pre-test was a good predictor for the results on both the definitional post-test and the what-if post-test. In this group, the result of the definitional pre-test did not seem to have any relation to any of the post-tests, not even on the definitional post-test. Table 4 showed that the difference between the high scoring group and the low scoring group on the definitional pre-test had completely vanished at the time the subjects took the post-tests.

Another difference between the experimental and control group is that only in the experimental group a relation was found between the experimentation as reflected in interaction data and post-test scores. This is an indication that experimenting is a factor that contributed to learning in this condition. A similar relation was absent for the control group. In the control group, we found that the use of assignments correlated with the post-tests scores. Such a correlation was not found in the experimental condition. These results indicate that in the experimental condition it does not matter that much how many assignments the learner used, but more how they used them, whereas in the control condition it was merely using the assignments that contributed to the post-test scores.

Although on the surface both groups of subjects did not appear to show any difference, differences were revealed when a more fine-grained analysis of the data was conducted. It appears that the overall means of subjects in the two experimental are equal, but that they are the result of a completely different learning process.

To understand what happened in the two experimental groups, it could be illuminating to follow two 'typical' subjects in the group. However, it must be emphasized that these two scenarios require a great deal of data interpretation. Further research should be done to investigate the extent to which these scenarios are true.

A subject in the control group would perform the pre-tests, and study the material in the learning environment. The results on the post-test are best predicted from the definitional pre-test scores. There is no relation between the experimenting behavior and the results on the post-tests, but there is a relation between the number of assignments used and the results on the post-tests. The knowledge gained in the learning environment therefore appears not to be related to the fact that the subject was engaged in a discovery environment, but to the fact that for 90 minutes the learner processed new information on the topic. During this time, the subject merely builds on the existing definitional knowledge and utilizes this knowledge on the post-tests.

A typical subject in the experimental group engaged more in experiments. The results on the post-test are best predicted based on the what-if pre-test scores. Subjects in this group activated their intuitive knowledge (as measured by the what-if test) in order to generate and interpret experiments. The feedback given by the learning environment seems to have triggered this more in-depth processing of the information. The more intuitive knowledge the subject had in the beginning, the more he or she could benefit from the information generated in the experiments and present in the feedback. This knowledge gain extended to definitional knowledge and even to the extent that the result on this test is no longer dependent on the subject's definitional pre-knowledge. Therefore, it seems that definitional knowledge is not activated in the experimental learning environment. Also, the more experiments the learner did, the more intuitive knowledge he or she gained.

The above are stereotypes to contrast the learning in the two conditions. Even if these were true, a number of questions remain unanswered after this study. The main question is why we did not find any differences between the two conditions on the knowledge tests. Three explanations come to mind.

One is that we based these expectations on the idea that learners in both conditions had to show discovery learning behavior, but that learners in the experimental condition were better supported by the feedback on the discovery learning processes. The results indicate that the learners in the control condition compensated for the absence of feedback on the discovery learning processes by adopting a more traditional learning style in which they made extensive use of the assignments and drew heavily on their definitional knowledge. This way of learning was not anticipated when we made our predictions about the learning outcomes. Removing the domain-oriented feedback from the control environment to force discovery learning behavior upon the learners would be a way to look at the influence of the feedback on the discovery learning processes in more detail.

The second explanation is that each of the two environments is better suited for learners with a specific learning style. Closer examination of learners working with the environments should reveal whether this is the case and if so, we would like to be able to distinguish between the categories. For the moment, it seems that definitional and intuitive knowledge are possible selection criteria.

A third explanation for not finding the expected differences lies in the time that the learners in the experimental condition needed to get acquainted with the new feedback. Perhaps, if they had been able to work with the environment longer we would have found differences. The feedback given in the experimental condition differs from the feedback learners usually receive in that the content of the feedback is based on the learner's own experiments. Thus, learners were required to think about experiment design if they wanted to receive feedback on the correctness of a hypothesis. That is, they had to think about the interpretation of the experiments in relation to the hypothesis. This required the learners to take a more active role than the learners in the control condition and it took them more time to adjust. This automatically leads us to the following issue. Our idea was that feedback in the experimental condition would help learners in learning the domain knowledge through supporting the discovery process. We believe that extra time spent on the first level and the other differences in behavior are indicators that the learners in the experimental condition, apart from learning domain knowledge, learned discovery skills as well. As this is indirect

evidence, we would like to substantiate this claim in a follow-up study explicitly testing for discovery skills. This might also provide an answer to the issue of the relation between intuitive knowledge and discovery skills. In this study, we found that in the experimental condition intuitive knowledge was the main predictor for the post-test results. This raises the question of the exact role of the intuitive knowledge in the discovery processes. Is domain specific intuitive knowledge a necessary condition for discovery learning unrelated to the discovery skills, or is there interdependency between the discovery skills and the domain specific intuitive knowledge? By testing for the discovery skills in our next study, along with investigating the learning processes of a few learners in more detail, we hope to be able to answer this question in the future.

In this study, we have introduced a potentially effective way of supporting learners in a simulation based discovery environment. The approach is general and can be used in domains in which hypotheses can be interpreted by the system, because they are specified beforehand, as was the case in this study, or because they are created by tools like a hypothesis scratchpad yielding well-formed hypotheses. The results show that providing learners with specific information on the relationship between their experiments and the hypotheses they study changed the overall behavior of the learners in the learning environment and lead to more discovery oriented behavior. In our next study, we hope to show that learners working with this kind of learning environment not only gain domain knowledge, as we found in this study, but also gain discovery skills.

4 References

Anderson, J.R., Boyle, C.F., & Yost, G. (1985). The Geometry Tutor. In *Proceedings of the international joint conference on artificial intelligence*, pp. 1-7. Los Altos CA: International Joint Conference on Artificial Intelligence.

Anderson, J.R., Farrell, R., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science*, 8, 87-129.

Bruner, J.S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21-32.

Holt, P., Dubs, S., Jones, M., & Greer J.E. (1994). The State of Student Modelling. In J.E. Greer & G. I. McCalla (Eds.), *Student Modelling: The key to Individualized Knowledge-Based Instruction* (NATO ASI series F: Computer and Systems Series, Vol 125) (pp. 3-35). Berlin, Germany: Springer-Verlag.

Jong, T. de, Joolingen, W.R. van, Swaak, J., Veermans K., Limbach R., King S., & Gureghian D. (1998). Self-directed learning in simulation-based discovery environments. *Journal of Computer Assisted Learning*, 14, 235-246.

Jong, T. de, Martin, E., Zamarro J-M., Esquembre, F., Swaak, J., & Joolingen, W.R. van (1999). The integration of computer simulation and learning support; an example from the physics domain of collisions. *Journal of Research in Science Teaching*, 5, 597-615

Jong, T. de & Joolingen, W.R. van (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68, 179-201.

Jong, T. de & Njoo, M. (1992). Learning and instruction with computer simulations: learning processes involved. In E. de Corte, M. Linn, H. Mandl & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving* (pp. 411-429). Berlin: Springer-Verlag.

Joolingen, W.R. van (1999). Cognitive tools for discovery learning. *International Journal of Artificial Intelligence and Education*, 10, 385-397.

Joolingen, W.R. van, & Jong, T. de (1993). Exploring a domain through a computer simulation: traversing variable and relation space with the help of a hypothesis scratchpad. In D. Towne, T. de Jong & H. Spada (Eds.), *Simulation-based experiential learning* (pp. 191-206). Berlin, Germany: Springer-Verlag.

Joolingen, W.R. van, King, S., & Jong, T. de (1997). The SimQuest authoring system for simulation-based discovery environments. In B. du Boulay & R. Mizoguchi (Eds.), *Knowledge and media in learning systems* (pp. 79-87). Amsterdam: IOS.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.

Lajoie, S.P. (1993). Computer Environments as Cognitive Tools for Enhancing Learning. In S. Lajoie & S. Derry (Eds.), *Computers as Cognitive Tools* (pp. 261-288). Lawrence Erlbaum Associates, NJ.

Marchionni, G. (1988). Hypermedia and learning: Freedom and Chaos. *Educational Technology*, 28, 8-12.

Self J. A. (1990) Bypassing the intractable problem of student modeling. In C. Frasson & G. Gauthier (Eds.). *Intelligent Tutoring Systems: At the Crossroadsof Artificioal Intelligence and Education* (pp. 107-123). Norwood, NJ: Ablex.

Shute, V.J., & Psotka, J. (1996). Intelligent tutoring systems: Past, Present and Future. In D. Jonassen (Ed.), *Handbook of Research on Educational Communi-cations and Technology* (pp. 570-601). London: Prentice Hall International.

Swaak, J., & Jong, T. de (1996). Measuring intuitive knowledge in science: the development of the what-if test. *Studies in Educational Evaluation*, 22, 341-362.

Swaak, J., Jong, T. de, & Joolingen, W.R. van (submitted). Do different learning environments make a difference? The effects of discovery learning and expository instruction on the acquisition of intuitive and explicit knowledge

Veermans, K., & Joolingen, W.R. van (1998). Using induction to generate feedback in simulation-based discovery learning environments. In: B.P. Goetl, H. M. Half, C.L. Redfield, & V.J. Shute (Eds.) *Intelligent Tutoring Systems, 4th International Conference, San Antonio, TX USA*. (pp. 196-205). Berlin: Springer.

White, B. Y., & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 42, 99-157.

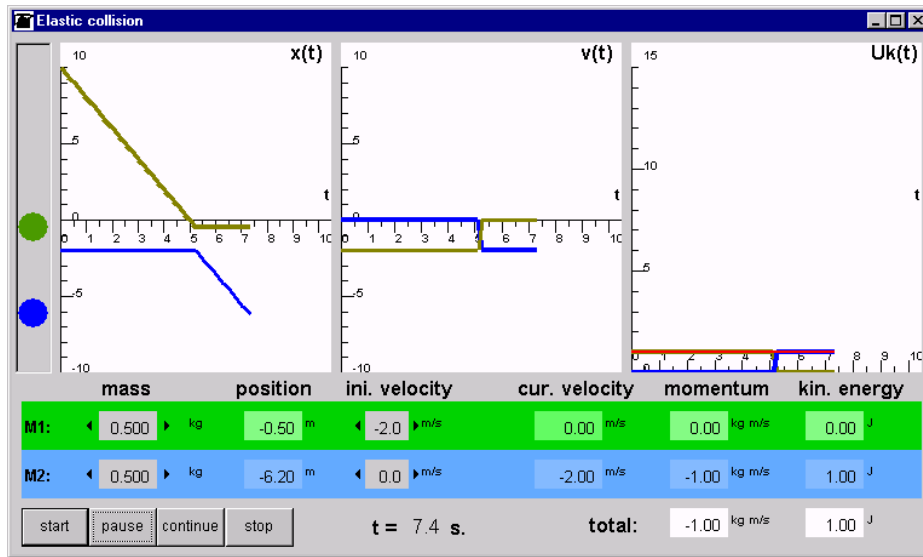
Figure Captions

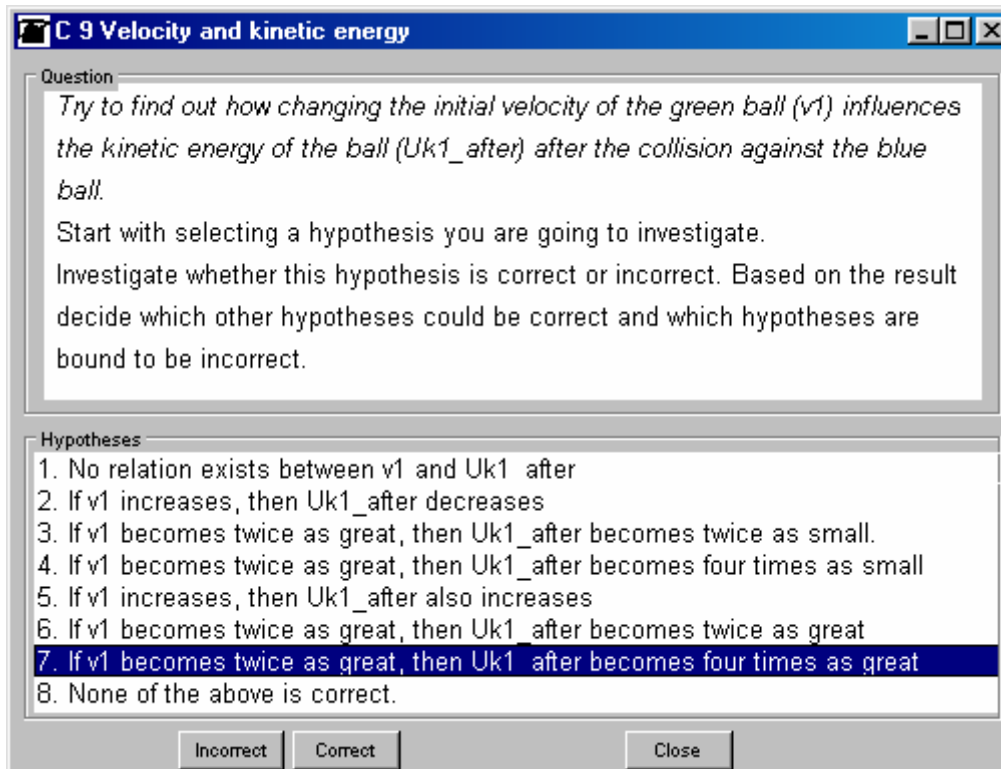
Figure 1. Elastic collision simulation window

Figure 2. An investigation assignment, the seventh alternative is in focus

Figure 3. The monitoring tool, with experiments.

Figure 4. Feedback explanation with analysis of the experiments in figure 3 and the hypothesis from figure 2.





C 9 Velocity and kinetic energy

Question

Try to find out how changing the initial velocity of the green ball (v_1) influences the kinetic energy of the ball (U_{k1_after}) after the collision against the blue ball.

Start with selecting a hypothesis you are going to investigate.
Investigate whether this hypothesis is correct or incorrect. Based on the result decide which other hypotheses could be correct and which hypotheses are bound to be incorrect.

Hypotheses

1. No relation exists between v_1 and U_{k1_after}
2. If v_1 increases, then U_{k1_after} decreases
3. If v_1 becomes twice as great, then U_{k1_after} becomes twice as small.
4. If v_1 becomes twice as great, then U_{k1_after} becomes four times as small
5. If v_1 increases, then U_{k1_after} also increases
6. If v_1 becomes twice as great, then U_{k1_after} becomes twice as great
7. If v_1 becomes twice as great, then U_{k1_after} becomes four times as great
8. None of the above is correct.

Incorrect Correct Close

| Exp | m1 | m2 | v1 | v2 | Uk1_na |
|-----|-----|-----|------|-----|--------|
| 1 | 0.5 | 0.5 | -2.0 | 1.0 | 0.25 |
| 2 | 0.5 | 1.0 | -3.0 | 1.0 | 1.361 |
| 3 | 0.5 | 1.0 | -2.0 | 2.0 | 2.778 |
| 4 | 0.5 | 1.0 | -2.0 | 0.0 | 0.111 |
| 5 | 0.5 | 1.0 | -3.0 | 0.0 | 0.25 |
| 6 | 0.5 | 1.0 | -4.0 | 0.0 | 0.444 |
| 7 | 0.5 | 1.0 | -4.0 | 2.0 | 4.0 |

Maximum number of experiments

Feedback on statement and experiments

You claim that the hypothesis *7. If v_1 becomes twice as great than the U_{k1_after} becomes four times as great* is true.

Below you will find your experiments that are relevant for testing this hypothesis.

| v_1 | v_2 | m_1 | m_2 | prediction |
|-------|-------|-------|-------|------------|
| -2.0 | 2.0 | 0.5 | 1.0 | 2.78 |
| -4.0 | 2.0 | 0.5 | 1.0 | 4.0 11.11 |

| v_1 | v_2 | m_1 | m_2 | prediction |
|-------|-------|-------|-------|------------|
| -2.0 | 0.0 | 0.5 | 1.0 | 0.11 |
| -4.0 | 0.0 | 0.5 | 1.0 | 0.44 0.44 |

If this hypothesis were correct, then the values you observed for U_{k1_after} in these experiments should match the values that you would expect based on the hypothesis. This is however not the case. *This means there is reason to believe that this hypothesis is not correct.*

Close